Associations Between Long- and Short-Term Exposure to Neighborhood Social Context and Pregnancy-Related Weight Gain

By

Irene Elizabeth Headen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mahasin S. Mujahid, Chair
Professor Nicholas P. Jewell
Professor Barbara F. Abrams
Professor Julianna Deardorff

Summer 2015

Abstract

Associations Between Long- and Short-Term Exposure to Neighborhood Social Context and Pregnancy-Related Weight Gain

by

Irene Elizabeth Headen
Doctor of Philosophy in Epidemiology
University of California, Berkeley

Professor Mahasin Mujahid, Chair

Background: Weight gained during childbearing has significant implications for maternal and child health. Both too much and too little weight gained during pregnancy can result in adverse outcomes. Recommendations for ideal gestational weight gain (GWG) have been developed by the Institute of Medicine, but achieving these standards remains a challenge. Better understanding of the wider context in which women experience pregnancy may aid in the development of novel interventions to improve trends in healthy GWG. Neighborhoods define one such dimension of women's wider context that is emerging as a promising factor in this area of research. However, limited work has considered long-term exposure to neighborhood environments or the role of women's perceptions of their neighborhood environments in relation to either inadequate or excessive GWG.

Methods: This dissertation explores the associations between long- and short-term exposure to neighborhood social and socioeconomic context and GWG using data from the 1979 National Longitudinal Survey of Youth. It additionally investigates associations between objective and perceived measures of neighborhood social context in relation to GWG. The first paper investigates associations between long-term, cumulative neighborhood socioeconomic deprivation and GWG. The second paper investigates associations between objectively measured and perceptions of point-in-time neighborhood social environment and GWG. Objective neighborhood social environment is measured using neighborhood socioeconomic deprivation. Perceived neighborhood social environment is assessed from women's self-report of problems within their neighborhood environment. The final paper in this dissertation conducts a systematic review of the literature to characterize the reporting error associated with use of self-reported, pregnancy-related weight in efforts to move the field toward developing bias correction techniques to address methodological limitations of this measure. While

not directly related to understanding neighborhoods and GWG, this issue is relevant to future studies in this area that rely on self-reported weight.

Conclusion: The papers included in this dissertation illustrate the importance of considering both long-term and short-term measures of neighborhood social context in order to fully understand how neighborhood environments impact inadequate and excessive GWG. In particular, long-term measures of exposure to neighborhood environments should be more fully considered in order to better understand how neighborhoods can support healthy GWG. Interventions based on this improved knowledge of the environment in which women experience weight gain during pregnancy may improve GWG outcomes and health trajectories of both mother and child. Future studies in this area may also benefit from more rigorous study of variation of reporting error in self-reported pregnancy-related weight by maternal characteristics, which will aid in the development of bias correction techniques for these widely used measures.

For my parents, Dr. Alvin Eugene Headen, Jr. and Dr. Sandra Louise Wilson Headen, whose wisdom, support, and relentless encouragement brought me through this dissertation. I am thrilled to be carrying on the "family business" and excited to be the newest addition to the "Dr. Headen" line.

TABLE OF CONTENTS

Acknowledgements……………………………......……………………………………………..iii

Chapter 1. Background………………………….……………………………………………..1

Chapter 2. Associations between cumulative neighborhood deprivation and gestational weight gain………………………………..…………………………………………………..9

Chapter 3. Perceived and objective neighborhood social environment and gestational weight gain……............................. ……………………………………………37

Chapter 4. The accuracy of self-reported, pregnancy-related weight: A systematic review……………………………………………66

Chapter 5. Conclusion…………………….… ……………………………………………98

Appendix Sections

Appendix 1…………………………...…… …………………………………………..101

Appendix 2……………………………..… …………………………………………..105

Appendix 3…………………………...…… …………………………………………..110

Acknowledgements

I first and foremost would like to thank my dissertation chair, Professor Mahasin Mujahid, for being an amazing adviser every step of the way. Her immense patience and support have helped this dissertation evolve from the thoughts of a very “green” young student into the mature product that it is today. Thank you for holding me to the highest standards and teaching me what it means to be a rigorous scholar.

I also owe a special debt of gratitude to Professor Barbara Abrams, who has been an active co-adviser in every capacity. Through working with her and the NLSY research team over the past five years, I have had limitless opportunities for academic development and collaboration with an outstanding set of colleagues. This experience has been key to my development as an epidemiologist. Thank you for your foresight and commitment to my development; it has brought me further than I thought possible.

I am also thankful for the support of my other committee members. Professor Nicholas Jewell has provided rigorous statistical feedback throughout the development of this dissertation. He has also made biostatics a fun and accessible topic for me, which has driven me to be more engaged with the statistical aspects of my scholarship. Professor Julie Deardorff has provided insightful feedback on the role of cultural contexts in shaping health and development that has been instrumental to the interpretation of findings in this dissertation.

I would also like to extend a special thank you to the colleagues and classmates that I have worked with during my time in this doctoral program. They are an incredibly bright and talented bunch and I appreciate the very high standards that they have held me to. Specifically, I would like to thank Dr. Ben Chaffee, Dr. Divya Vohra, Alison Cohen, and Paul Wesson for not only being the most rigorous (and constructive) critics of my work, but also being most supportive friends throughout this process.

Finally, I would like to thank my parents, Drs. Alvin and Sandra Headen, and grandmother, Mrs. Irene Wilson, for being amazing role models. You all have believed in me and kept me going through the highs and lows of this journey. I would also like to thank Dr. Delbert André Green II for being an amazing friend throughout this process. You all have provided me with an invaluable example of what rigorous Black scholarship looks like. Thank you for listening to me and providing guidance at every step of the way.

CHAPTER 1: BACKGROUND

The childbearing experience has been identified in a wide body of literature as key to shaping the short and long-term health trajectories of women and their children. (1-5) In particular, pregnancy is one of the only periods in a woman's life during which she intentionally gains a substantial amount of weight over such a short amount of time. However, gaining too much weight during this time can contribute to obesity development in both mother (6) and her child. (7) Too little weight gain during this time, alternatively, is associated with adverse pregnancy outcomes including increased risk of delivering a low birth weight or small-for-gestational age infant and preterm birth. (1,3,8) Not only do both of these outcomes increase risk of infant mortality at time of birth, (1) but research on fetal programming has linked these pregnancy outcomes in the infant to adverse health outcomes that occur in adulthood, such as adult metabolic risk and risk for cardiovascular heart disease. (9-13) Thus, associations between gestational weight gain (GWG) and adverse maternal and child health outcomes make it a key modifiable factor of interest in improving both current and future health of mother and child.

In light of the potential importance of GWG as a predictor of positive maternal and child health outcomes, the Institute of Medicine (IOM) developed guidelines to aid women in achieving adequate GWG. However, the majority of women still do not gain within this ideal range. Most women (41-51%) gain in excess of the guidelines and a number still gain inadequately, or below these recommended amounts (17-28%). (14) A wide body of research has identified individual risk factors driving both inadequate and excessive GWG in hopes of using this knowledge to develop interventions to reduce non-ideal GWG. (14-16) Some success has been made, (17) but the majority of studies have been conducted in clinical or controlled medical settings. (17) To continue to make gains in increasing adequate GWG, focus on factors beyond the individual woman may be needed to better contextualize the environment in which she is engaging in weight management during pregnancy. (4,18) Neighborhoods are a potentially key dimension of this external environment that may be relevant for understanding, and intervening to improve, trends in GWG.

The Neighborhood Context

The relationship between health and place, which is defined by the geographically located physical and social spaces in which individuals and populations live, work, and play, has gained increasing attention over the last few decades. (19,20) The spatial distributions of resources define the risk and protective factors that individuals encounter on a daily basis. (9) Additionally, because social interactions are structured within these spatial environments, aspects of place impact the social context in which health outcomes occur. (21) Neighborhoods have been a main focus of this burgeoning area of research on place because they define a geographically and socially relevant level at which such risk and protective factors act to effect health. (19,22,23)

Theoretical motivations underlying the exploration of neighborhoods and their effects on health suggest joint biological and behavioral pathways through which daily and cumulative exposures to the residential environment produce adverse health outcomes. The concept of embodiment from ecosocial theory (24) proposes that chronic exposure to a lack of resources and other environmental risk factors negatively affects biological functioning across multiple regulatory systems and thereby increases susceptibility to disease. (24) Specifically, chronic activation of the stress response, which involves inputs from multiple biological regulatory systems, has been identified as a key mechanism through which the functioning of these systems is disregulated over time. The accumulation of multi-system disregulation interferes with the individual's ability to avoid disease, resulting in adverse health outcomes. (24,25)

These biological mechanisms are accompanied by behavioral actions that shape adverse health outcomes as well. First, the psychological consequences of stress can lead to increased participation in adverse coping behaviors, such as overeating, drinking, and smoking. (4) Furthermore, structural barriers in the immediate neighborhood can constrain participation in health-promoting behaviors. (26) For example, lack of available and affordable food resources can limit residents' ability to consume fruits and vegetables, even if they would like to do so. These situations create a double burden on individuals. Those who are already more susceptible to disease due to the biological consequences of chronic stress are also more likely to adopt health behaviors that increase the risk of adverse health outcomes. (4,27)

Neighborhoods are also a key dimension across which social stratification occurs. (28,29) In particular, racial residential segregation systematically exposes racial/ethnic minorities to more disadvantaged neighborhoods over time. (28,29) Furthermore, even across these disadvantaged neighborhood environments, there may be variation in how different racial/ethnic groups experience these neighborhoods due to the presence of other social and physical characteristics that support or inhibit access to opportunity. (29,30) For example, Blacks may experience deprived neighborhoods that are more likely to exist in areas of concentrated poverty and experience higher levels of economic divestment. (29,30) Alternatively, Latinos may be more likely to reside in ethnic enclaves, which may experience a richer resource and social environment due to the presence of culturally oriented businesses and tight social networks. (31) Such qualitative differences in the neighborhood environments that different racial/ethnic groups are exposed to may influence the way that neighborhood deprivation is associated with health outcomes as well.

Neighborhoods and Pregnancy Outcomes

Neighborhoods may be particularly relevant in shaping the health of pregnant women. Pregnant women experience physiological and psychosocial changes associated with pregnancy (4,32) that may cause them to perceive (33) and respond differentially to the neighborhood environment. Research investigating the association between neighborhood environment and various pregnancy outcomes, such as low birth weight, preterm birth, and infant mortality, suggest that neighborhood factors (e.g. deprivation (34,35), physical incivilities (36), and crime (37)) are associated with increased risk of these outcomes. However, little work has been done specifically looking at the relationship between neighborhood environment and weight gain during pregnancy.

Only four studies to date have investigated neighborhood context and GWG, (33,36,38,39) and three were conducted in the same study population or a direct offshoot of this study population. (33,36,39) Findings from these studies, generally suggest that elements of the neighborhood physical environment (e.g. presence of parks and supermarkets, walkability; 36,39,40) and social environment (e.g. presence of abandoned buildings, litter, and graffiti indicating social disorder; 33,36,38,39) are associated with increased risk of both inadequate and excessive GWG. However, more research is needed to better understand links between the neighborhood environment and GWG.

In particular, while existing studies on GWG have explored different dimensions of the neighborhood environment, they have not focused on the time scale across which these elements may act to impact GWG. Considering lengthier time scales across which neighborhood environments work to impact health (4,24,41) more accurately captures the longer-acting biological mechanisms that shape the underlying health risk that a woman brings to her pregnancy. Another gap in the current literature is that analyses assessing residents' perceptions of, rather than researchers' reports of, their neighborhood environments in relation to these outcomes do not exist. Using measures of women's perceptions of their neighborhood environment may more accurately reflect their appraisal of whether factors in the neighborhood amount to actual barriers to daily life functioning. (42) Thus, this appraisal may be more proximal to elements of stress activation and behavioral responses that lead to adverse health outcomes. (4,42) Both of these elements of residents' experience in neighborhood environments may be key to fully understanding how neighborhoods are linked to GWG.

To address the gaps in the literature discussed above, this dissertation uses data from a national sample of women participating in the 1979 National Longitudinal Survey of Youth (NLSY79) to investigate the association between neighborhood social context and GWG. Chapter two assesses the relationship between a long-term, cumulative measure of neighborhood socioeconomic deprivation experienced by women from baseline to pregnancy and GWG. Chapter three investigates two point-in-time measures of the neighborhood social environment, assessed both through perceptions and objectively measured, in relation to GWG. Women in the NLSY79 were asked to report their perceptions of problems within the neighborhood, including crime and violence, transportation, and lack of trust among neighbors; this constituted the perceived measure of neighborhood social environment. Objective neighborhood social environment was assessed using a census-based measure of neighborhood socioeconomic deprivation. In both analyses, racial/ethnic differences in associations between neighborhood social environment and GWG were assessed.

Chapter four shifts focus to consider an important measurement-related methodological challenge faced by many observational studies assessing GWG, including our own. Self-report of pregnancy-related weight is a key way to measure weight in these studies. This chapter contains a systematic review assessing the accuracy of such measures with the goal of more completely characterizing reporting error in self-reported weights at different times during pregnancy. The challenge of using self-reported pregnancy-related weight is likely to remain an issue in future studies geared toward investigating neighborhoods and GWG. For this reason, chapter four concludes with suggestions for ways that the literature can support the development of bias correction techniques that can be applied to self-reported pregnancy-related weight measures.

The overall goal of this dissertation is to determine whether consideration of different time-scales and inclusion of perceived, as well as objective, assessments of neighborhood environment provides a more nuanced understanding of how neighborhoods shape trends in GWG. Such knowledge can better contextualize the external environment in which women become pregnant and lead to the development of improved interventions to support healthy GWG.

REFERENCES

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

2. Bodnar LM, Hutcheon JA, Platt RW, et al. Should gestational weight gain recommendations be tailored by maternal characteristics? *Am J Epidemiol*. 2011;174(2):136–146.

3. 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.

4. Davis EM, Stange KC, Horwitz RI. Childbearing, stress and obesity disparities in women: A public health perspective. *Matern Child Health J*. 2012;16(1):109–118.

5. Siega-Riz AM, Viswanathan M, Moos MK, et al. A systematic review of outcomes of maternal weight gain according to the Institute of Medicine recommendations: birthweight, fetal growth, and postpartum weight retention. *Am J Obstet Gynecol*. 2009;201(4):339.e1–339.e14.

6. Davis EM, Zyzanski SJ, Olson CM, et al. Racial, Ethnic, and Socioeconomic Differences in the Incidence of Obesity Related to Childbirth. *Am J Public Health*. 2009;99:294–299.

7. Diesel JC, Eckhardt CL, Day NL, et al. Gestational weight gain and the risk of offspring obesity at 10 and 16 years: A prospective cohort study in low-income women. [Epub ahead of print May 29, 2015]. *BJOG: Int J Obstet Gy*. (doi: 10.1111/1471-0528.13448).

8. Bodnar LM, Siega-Riz AM, Simhan HN, et al. Severe obesity, gestational weight gain, and adverse birth outcomes. *Am J Public Health*. 2010;91:1642–1648.

9. Macintyre S, Ellaway A, Cummins S. Place effects on health: How can we conceptualise, operationalise and measure them? *Soc Sci Med*. 2002;55:125–139.

10. Freeman DJ. Effects of maternal obesity on fetal growth and body composition: Implications for programming and future health. *Semin Fetal Neonatal Med*. 2010;15(2):113–118.

11. Lu MC, Halfon N. Racial and ethnic disparities in birth outcomes: A life-course perspective. *Matern Child Health J*. 2003;7(1):13–30.

12. Sandman CA, Davis EP, Buss C, et al. Exposure to prenatal psychobiological stress exerts programming influences on the mother and her fetus. *Neuroendocrinology*. 2012;95(1):8–21.

13. Godfrey KM, Barker DJ. Fetal programming and adult health. *Public Health Nutr*. 2001;4(2b):611–624.

14. Deputy NP, Sharma AJ, Kim SY, et al. Prevalence and characteristics associated with gestational weight gain adequacy. *Obstet Gynecol*. 2015;125(4):773–781.

15. Hickey CA. Sociocultural and behavioral influences on weight gain during pregnancy. *Am J Public Health*. 2000;71(suppl):1364S–70S.

16. Webb JB, Siega-Riz AM, Dole N. Psychosocial determinants of adequacy of gestational weight gain. *Obesity (Silver Spring)*. 2009;17(2):300-9.

17. Muktabhant B, Lawrie TA, Lumbiganon P, et al. Diet or exercise, or both, for preventing excessive weight gain in pregnancy (Review). *Cochrane Database Syst Rev*. 2015;(6):1–260.

18. Culhane JF, Elo IT. Neighborhood context and reproductive health. *American Journal of Obstetrics and Gynecology*. 2005;192(5):S22–S29.

19. Diez Roux AV. Investigating Neighborhood and area effects on health. *Am J Public Health*. 2001;91(11):1783–1789.

20. Macintyre S, Ellaway A. Neighborhoods and health: An overview. In: Kawachi I, Berkman LF, eds. *Neighborhoods and Health*. New York: Oxford University Press, USA; 2003:20–42.

21. Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing "neighborhood effects": Social processes and new directions in research. 2002;28:443-78.

22. Diez Roux AV. Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *Am J Public Health*. 1998;88:216–222.

23. Shankardass K, Dunn JR. How goes the neighbourhood? Rethinking neighbourhoods and health research in social epidemiology. In: O'Campo P, Dunn JR, eds. *Rethinking social epidemiology: Towards a science of change*. Dordrecht: Springer Netherlands; 2012:137–156.

24. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol*. 2001;30:668–677.

25. McEwen B. Protective and damaging effects of stress mediators. *N Engl J Med*. 2000;338(3):171–179.

26. Bird CE, Rieker PP. Gender and health: The effects of constrained choices and social policies. New York: Cambridge University Press; 2008.

27. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci*. 2010;1186:125–145.

28. Williams DR, Collins C. Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Rep*. 2001;116(5):404–416.

29. Massey DS. Segregation and stratification: A biosocial perspective. *The DuBois Review: Social Science Research on Race*. 2004;1:1–19.

30. Kramer MS. Race, place, and space: Ecosocial theory and spatiotemporal patterns of pregnancy outcomes. In: Howell FM, Porter JR, Mathews S, eds. *Recapturing Space: New midle-Range theory in Spatial Demography*. Dordrecht: Springer, Inc; (in press)

31. Eschbach K, Ostir GV, Patel KV, et al. Neighborhood context and mortality among older Mexican Americans: Is there a barrio advantage? *Am J Public Health*. 2004;94:1807–1812.

32. Halbreich U. The association between pregnancy processes, preterm delivery, low birth weight, and postpartum depressions—The need for interdisciplinary integration. *Am J Obstet Gynecol*. 2005;193(4):1312–1322.

33. Laraia B, Messer L, Evenson K, et al. Neighborhood factors associated with physical activity and adequacy of weight gain during pregnancy. *J Urban Health*. 2007;84(6):793–806.

34. Janevic T, Stein CR, Savitz DA, et al. Neighborhood deprivation and adverse birth outcomes among diverse ethnic groups. *Ann Epidemiol*. 2010;20(6):445–451.

35. Schempf A, Strobino D, OŁCampo P. Neighborhood effects on birthweight: An exploration of psychosocial and behavioral pathways in Baltimore, 1995-1996. *Soc Sci Med*. 2009;68(1):100–110.

36. Vinikoor-Imler LC, Messer LC, Evenson KR, et al. Neighborhood conditions are associated with maternal health behaviors and pregnancy outcomes. *Soc Sci Med*. 2011;73(9):1302–1311.

37. Messer LC, Kaufman JS, Dole N, et al. Neighborhood crime, deprivation, and preterm birth. *Ann Epidemiol*. 2006;16(6):455–462.

38. Mendez DD, Doebler DA, Kim KH, et al. Neighborhood socioeconomic disadvantage and gestational weight gain and loss. *Matern Child Health J*. 2014;18(5):1095-103.

39. Messer LC, Vinikoor-Imler LC, Laraia BA. Conceptualizing neighborhood space: Consistency and variation of associations for neighborhood factors and pregnancy health across multiple neighborhood units. *Health Place*. 2012;18(4):805–813.

40. Laraia BA, Messer L, Kaufman JS, et al. Direct observation of neighborhood atributes in an urban area of the US south: Characterizing the social context of

pregnancy. *Int J Health Geogr*. 2006;5:11.

41. McEwen B. Protective and damaging effects of stress mediators: Central role of the brain. *Dialogues in Clinical Neuroscience*. 2006;8(4):367–381.

42. Bailey EJ, Malecki KC, Englelman CD, et al. Predictors of discordance between perceived and objective neighborhood data. *Annals of Epidemiology*. 2014;24(3):214–221.

CHAPTER 2. ASSOCIATIONS BETWEEN CUMULATIVE NEIGHBORHOOD DEPRIVATION AND GESTATIONAL WEIGHT GAIN

INTRODUCTION

Weight gain during pregnancy impacts both short- and long-term health trajectories of women and their children. (1-3) Too little gestational weight gain (GWG) has been linked to low birth weight and preterm birth, (3,4) both of which are driving factors in infant mortality. Too much GWG has been linked to high maternal postpartum weight retention, (3,4) which is also associated with increased risk of long-term maternal obesity, (5,6) and childhood obesity. (7) As such, GWG remains a potentially important modifiable pregnancy factor through which to potentially improve infant outcomes and address the obesity epidemic in the U.S.

The Institute of Medicine (IOM) developed recommendations for ideal amounts of gain (3) in order to aid women in avoiding these adverse outcomes. However, data from recent studies suggests that only a minority of women gain adequately (23-34%). (8) Excessive GWG is far more prevalent (41-51%) and inadequate GWG also impacts a substantial number of women (17-28%). (8)

While a large body of literature identifying individual risk factors for both inadequate and excessive GWG exists, (8-10) the role of neighborhood environments in shaping GWG remains largely unexplored. In general, interest in neighborhoods as relevant determinants of health is linked to wider recognition that individuals are rooted in many external environments and contexts that are equally important for overall health as their individual attributes. (11) Neighborhoods, in particular, are key dimensions that define the day-to-day physical and social environments that individuals must engage with in order to achieve positive health behaviors. (12) Especially for pregnant women, elements of the social, service, and built environment may be key barriers or facilitators for accessing the support and resources needed to achieve positive pregnancy outcomes. (13)

The few existing studies investigating neighborhood factors and GWG suggest that such factors, including neighborhood deprivation, (14) presence of parks, porches, and other social spaces, (15-17) and walkability, (16,17) are associated with inadequate and excessive GWG. However, all of these studies use cross-sectional or point-in-time measures of neighborhood context. No studies that we are aware of have investigated the impact of prolonged exposure to adverse neighborhood environments on GWG.

Theoretical pathways through which neighborhoods impact health are largely dependent on the length of exposure or time frame over which neighborhoods “act” to result in poor health. Specifically, while some mechanisms may be mediated through more immediate time frames, such as lack of material or service resources in the neighborhood serving as point-in-time barriers to women achieving healthy pregnancy behaviors, (13,18) others are mediated through long-term or *chronic* exposure to adverse environments. (1,13) For example, stress-based neuroendocrine mechanisms propose that it is not acute or transient exposure to stressors that is problematic for health. (19) Rather, it is the continued, chronic exposure to stressors that prevents the stress response from resolving and leads to biological cascades that disregulate multiple health-related systems. (19) Thus, understanding both the impact of short- and long-term mechanisms on any particular health outcome is necessary to fully reveal the extent to which the neighborhood environment impacts health.

Understanding the impact of chronic exposure to adverse neighborhood environments may be particularly important for pregnancy outcomes. Pregnancy is a critical period, both psychosocially (20) and physiologically. (21-23) During this time prolonged exposure to structural barriers preventing engagement in positive health behaviors as well as underlying biological health risk may manifest as adverse outcomes. Increased demand for services and other resources needed to achieve a healthy pregnancy may exacerbate the adverse impact of chronic lack of resources in neighborhoods that experience chronic deprivation. (13,24) Furthermore, the additional biological demand on the woman's system during pregnancy may trigger the latent risk of disease. (23) Weight gain mechanisms, in particular, may be sensitive to both disregulation of the stress response that results from prolonged exposure to neighborhood stressors (23,25,26) and mal-adaptive coping that may have developed over time. (25) Thus, GWG is a key outcome to investigate in understanding impacts of chronic neighborhood deprivation on pregnancy.

Histories of discrimination and social stratification in the U.S. also mean that the association of chronic neighborhood deprivation on GWG may vary by race/ethnicity. Such histories of social sorting have resulted in long-term spatial segregation of racial/ethnic minorities into poor neighborhood environments. (27) These populations are more likely to reside in deprived environments at any point in time, and are more likely to have resided in these environments for prolonged periods of time. (27,28) However, the same level of neighborhood disadvantage may look and be experienced in different ways by different racial/ethnic groups. (29) For example, Blacks may be more likely to live in disadvantaged neighborhoods that also are subject to long-term economic divestment. They also exist in proximity to other neighborhoods that lack beneficial opportunity structures as well. (29) The "same" disadvantaged neighborhoods for Latinos, however, may not be paired with similar economic divestment due to the presence of businesses catering to cultural demand. Ethnic enclaves that often characterize disadvantaged Latino neighborhoods may provide social support and social networks to buffer negative impacts of these neighborhoods. (30) Extended residence in either of these two deprived neighborhoods could potentially have quite different implications for health. (29)

We investigated the association between cumulative neighborhood deprivation and GWG in a national cohort of women who were followed for over 30 years. We hypothesized that higher cumulative neighborhood deprivation would increase the risk of inadequate as well as excessive GWG. Additionally, we investigated whether associations varied by race/ethnicity.

METHODS

Subjects

Our sample included women giving birth to singleton, live infants in the 1979 National Longitudinal Survey of Youth (NLSY79). The NLSY79 is a cohort study conducted by the National Bureau of Labor Statistics composed of a nationally-representative sample of 12686 men and women, who were 14-21 years old at the time of enrollment. (31) The cohort was constructed from three, multi-stage probability samples: a nationally representative sample, a military sample, and a supplemental sample that overrepresented Blacks, Hispanic/Latino/Spanish (referred to as Latino/a going forward), and economically disadvantaged non-Black, non-Latinas. The military and economically disadvantaged samples were discontinued after 1984 and 1990, respectively. Subjects were followed from 1979 to 2012. (31)

Participants were interviewed annually from 1979 to 1994 and biennially 1994 to 2012. For women who had children over the course of follow up (n = 4931), information on their pregnancies and children was collected starting in 1986. (31) Information about pregnancies occurring before 1986 was retrospectively collected. (32) Pregnancies occurring between interview periods were documented in closest subsequent interview of the mother. Of the 11504 total pregnancies, we restricted our sample to singleton gestations ranging from 22 to 44 weeks, ((33); n = 8860) and births that had non-missing information on both exposure and outcome, which resulted in a final sample size of 6772 pregnancies to 3689 women. (Figure 1) While these pregnancies spanned the full length of follow up, the majority (66.1%) occurred prior to 1990 and another 31.7% occurring between 1990 and 2000. Our analysis was approved by the University of California, Berkeley Center for the Protection of Human Subjects. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

Analytic Variables

Exposure:

Census tracts were used as proxies for neighborhood boundaries. Census tracts are subdivisions of counties that contain an average of 4000 individuals per tract and are selected to be homogeneous on sociodemographic characteristics. (34) To maintain consistency, all census tract boundaries were standardized to year 2010 boundaries. Standardized census tract data was obtained through the GeoLytics, Inc Neighborhood Change Database. (35)

Socioeconomic data from the 1980, 1990, 2000, and 2010 U.S. census were used to construct a measure of neighborhood deprivation. We used geometric interpolation (35) to predict values of census variables in inter-censal years. From these data, we constructed a deprivation index for each census tract for each year based on measures previously discussed in the literature. (36) This index included eight items spanning five domains: occupation (% adults in management and professional occupations), employment (% unemployed), housing (% crowded households), poverty (% families in poverty, % female headed households w/ dependents, % households on public assistance, % of families earning < $30,000), and education (% adults who earned less than a high school degree). Using factor analysis, we reduced these factors into a weighted index score (weighted by factor loadings) ranging from 0-100. Factor loadings for each item were moderate to high, ranging from 0.57 to 0.91, and consistent across years. (Appendix 1 Tables A1.1-A1.1b) Year-specific scores for each census tract were linked to each woman using her census tract of residence for each interview. Scores from 1979 to the closest prior year to the year in which the pregnancy occurred were then averaged to create a cumulative measure of deprivation for each pregnancy. For women giving birth in an interview year, this constituted a 1-year lag from baseline to 1994 and a 2-year lag from 1994 to 2012. For women giving birth between interview years, this coincided with a 1-year lag.

We analyzed cumulative deprivation as a continuous increase and as “patterns” of lifetime deprivation. For the continuous measure, a 1-unit increase corresponded to a one standard deviation change (8.3 points) in deprivation score. For patterns of lifetime deprivation we created four groups of residential deprivation trajectories: staying affluent, staying deprived, moving on an upward residential affluence trajectory, or moving on a downward residential poverty trajectory. We chose these categories *a priori* to maintain consistency with existing literature that has investigated cumulative deprivation and pregnancy outcomes (e.g. (37,38)). To create these categories, we used tertiles of neighborhood deprivation to define deprivation categories. Women were categorized as staying affluent if they stayed in the lowest tertile of deprivation over the course of follow up, as staying deprived if they stayed in the highest tertile of deprivation over follow up. For upward mobility trajectories, women were grouped into this category if they continually moved into lower deprivation tertiles over the course of follow up. Conversely, downward mobility was categorized as women who continually moved into higher tertiles of deprivation over the course of follow up. For this sub-analysis, we chose to restrict our population to women who fell within these pattern (n = 3715), thus excluding women who had more erratic patterns over the course of follow up.

Outcome

Women self-reported their weight prior to pregnancy and immediately before delivery when they were interviewed about the index pregnancy. Self-reported pre-pregnancy weight is highly correlated with medically reported pre-pregnancy weight, (39-41) GWG was calculated as the difference between self-reported pre-pregnancy weight and delivery weight. We then calculated a measure of GWG adequacy standardized by length of gestation to address the non-linear association of GWG with pregnancy duration, as described previously in the literature. (42,43) GWG adequacy was the ratio of a woman’s observed and expected GWG for her particular length of gestation. Expected GWG was calculated using the 2009 IOM recommendations (3) for amount of weight gain during the first trimester and rate of weight gain during the second and third trimesters, using the following equation: expected GWG= recommended first trimester weight gain + (gestational age-13)*(rate of weight gain during the second and third trimesters). Trimester (1^{st}, 2^{nd}, 3^{rd}) recommendations were specific to the category of pre-pregnancy BMI, with lower recommended amounts of weight gain among women who had a higher pre-pregnancy BMI. (Table 1) GWG adequacy was further categorized into inadequate, adequate, and excessive weight gain based on the 2009 IOM recommendations. (Table 1)

We conducted several sensitivity analyses using secondary classifications of our outcome. First, we used 1990 IOM recommendations (44) to classify GWG adequacy. (Table 1) The pregnancies in our data occur before 2009; however, in main analyses we report results using the 2009 recommendations because they use a more widely accepted pre-pregnancy BMI classification criterion and provide an upper limit for GWG for obese women. Additionally, 93% of women (100% for Inadequate GWG; 87% for Adequate GWG; 93% for excessive GWG) were categorized in the same GWG adequacy category between these two different classification criteria. Results from analyses using 1990 GWG recommendations are reported in Appendix 1. We also used alternative classifications of GWG to determine whether findings were robust to approaches to decouple the dependency of GWG on gestational age. Recent literature (45) has suggested that GWG adequacy does not completely reduce the bias due to this dependency when investigating associations with gestation-based birth outcomes such as preterm birth. Thus, we estimated models that used GWG categories based on non-gestation standardized GWG, controlling for gestational age in the model, and ran the same model among full term births only (37 to 44 weeks of gestation; n = 5952). Point estimates from these analyses were largely similar to those observed for GWG adequacy, so we report estimates from models using GWG adequacy going forward (data not shown). While substantial progress toward an improved measure of standardized GWG, (46) GWG z-score, is in the process of being developed, it does not yet have clinically relevant cut-offs similar to those presented by the IOM. Thus we do not present sensitivity analyses using this z-score in the current paper.

Covariates

We measured a number of additional maternal characteristics considered to be potential confounders based on the literature (3,9,10): marital status, immigrant status, maternal age, birth year of the child, parity, race/ethnicity, education, employment, income, and home ownership. Race/ethnicity was categorized as Black, Latino, Asian, and non-Black, non-Latino individuals. The non-Black, non-Latino racial/ethnic group was 88% white and so are referred to subsequently as “white.” Additionally, there were only 32 Asian women in our sample, so they were combined into the “white” group. Immigrant status (yes/no) was self-reported at baseline. Marital status was classified as currently married (yes) or never being married/divorced/separated (no). Maternal age was measured continuously based on the age at which the woman gave birth. Parity was measured as the number of live births that a woman had prior to the index pregnancy. Birth year of child was self-reported by the mother at the time of interview for the index pregnancy. Employment was based on the average hours that women reported working per week in the year prior to interview. Women were classified as unemployed (≤ 10 hrs/week), part-time employed (10-34 hrs/week), and full time employed (≥ 35 hrs/week). Educational attainment at time of birth was classified as follows: less than high school education (<12 yrs), high school education and/or some college (12-15yrs), and college graduates (≥ 16yrs). Household income was measured continuously and adjusted for family size (47) and log transformed to adjust for non-normal distribution.

Home ownership by participant or her spouse in year of birth (yes/no) was also measured.

We additionally controlled for a number of residential characteristics that capture aspects of geographical and residential location. Region of residence at the time of birth of the child was based on Census-defined regional definitions (48) and were classified as northeast, north central, south, and west. Rural residence at time of birth (yes/no) was categorized based on census definitions of urbanized areas. (31) Moving in the year of birth (yes/no) was determined based on changes in reported census tract of residence between consecutive interview years. Length of residence for women who did not move was determined based on number of years that a woman had lived in her census tract of residence at the time of the index pregnancy. Cumulative number of times moved was assessed as a count of different values for census tract of residence that a woman had over the course of follow up.

Statistical Analysis

Survey weighted means, standard deviations, and frequencies were calculated for all analytic variables. Bivariable associations used Chi squared tests or ANOVAs, where appropriate, to assess whether covariates varied by GWG adequacy. We also assessed whether average cumulative deprivation varied across covariates using ANOVAs. Multivariable regression was used to estimate both crude and adjusted associations between our cumulative neighborhood deprivation and GWG. Adjusted analyses included all covariates described above. Generalized linear models with log link functions were used to estimate relative risks. (49) Survey weights were used to account for the sample design and also reweighted the population to be representative of the national population of women in 1979. Because clustering by census tract was small[1], we did not explicitly account for this correlation. However, we did use robust standard errors to account for any correlation resulting from clustering of pregnancies within mothers. We additionally assessed whether associations between cumulative neighborhood deprivation and GWG varied by race/ethnicity. We created cross-product terms for race/ethnicity (white/Black/Latina) and cumulative neighborhood deprivation to be included in our models. We used overall Wald tests to assess whether effect measure modification was significant, using a $p \le 0.10$ threshold for significance. (50) Race/ethnicity stratified models were presented if effect measure modification was significant. All continuous variables were median centered and adjusted for the covariates described above. All analyses were conducted in Stata 12.0.

Our analytic sample was complete case for exposure and outcome. Prevalence of missing data in covariates ranged from 0 to 40% (Appendix 1 Table A1.2). Those excluded from our sample had slightly lower scores on cumulative neighborhood

[1] Clustering for a subsample of the population giving birth from 1990 to 2000 was on average 1.2 women per census tract with a standard deviation ranging from 0.6-0.7 over all years measured within this time frame. Specific estimates for the clustering for the full 1979-2010 births were not available for the current analysis. However, other analyses (e.g. Chase-Lansdale PL and Gordon RA. Economic Hardship and the development of five- and six-year-olds: Neighborhood and regional perspectives. *Child Development*. 1996; 67: 3338-3367.) that have looked at clustering by neighborhood in the NLSY have similarly found that it is minimal.

deprivation and were more likely to gain inadequately. They were also younger, more likely to be multiparous, less likely to be employed, more likely to lack a high school education, had a lower household income and slightly larger family size. They were more likely to reside in the south or the north central region at time of their pregnancy, more likely to be residentially stable and living in an urban area. We used regression-based multiple imputation using chained equations in Stata (51) to impute missing values for covariates. Separate models were fit for inadequate and excessive GWG, such that women who gained excessively were excluded from analyses looking at inadequate gain and vice versa. Sensitivity analyses using multinomial logistic regression to pool all women were conducted and findings were similar to stratified models. Thus, results from outcome-stratified models are reported going forward.

RESULTS

Among the 6772 pregnancies in our final sample, the mean age was 27.2 years (SD = 5.0) and 81.8% of women were white, 12.9% were Black, and 5.4% were Latina. (Table 2) Women were most likely to gain excessively (45.3%) with a substantial minority still gaining inadequately (22%). Cumulative neighborhood deprivation was relatively low, on average, women scored 18.1 (SD = 8.3) on a scale from 0 to 100. For patterns of lifetime neighborhood deprivation, 45.1% of women continuously resided in an affluent neighborhood. More women experienced downward mobility (21.6%) than upward mobility (17.0%), and 16.3% of women continuously resided in a deprived neighborhood. The majority of women in our sample were giving birth to their first (41.8%) or second (34.8%) child at time of participation in our study and were married (67.3%) at the time of the index pregnancy. Most women were employed at the time of their pregnancy, with 40.1% being employed full-time and 28.2% being employed part-time. The majority of women did not own their home (54.3%) and had been living in their current census tract an average of 1.5 years (SD = 2.2).

In bivariable analyses, most covariates varied across GWG adequacy categories. (Table 2) Adequate gainers were more likely to be white (85.3%), married (71.7%), home owners (50.5%), and had higher prevalence of college graduates (25.9%) as well as higher income households. However, women who gained excessively were more likely to be employed full-time (43.1%). Only two residential characteristics, cumulative times moved and rural residence, varied across GWG adequacy categories. On average, excessive gainers were more likely to have moved more than five times (8.3%), while the prevalence of rural residence was higher among inadequate gainers (17.5%).

Cumulative neighborhood deprivation also varied across most covariates. (Table 3) Women who gained inadequately or excessively both had higher mean cumulative deprivation scores compared to women who gained adequately. Across racial/ethnic groups, white women had the lowest mean cumulative deprivation scores (15.8; SD = 4.8) with Black (28.7; SD = 13.3) and Latina (26.1; SD =14.8) women both having significantly higher cumulative deprivation scores. Higher mean cumulative deprivation scores were observed among younger, unmarried, unemployed women with less than a high school degree. Women with higher parity also had higher mean cumulative deprivation scores. Patterns of lifetime residential deprivation also varied by race (data not shown). Among women experiencing continuous residence in an affluent neighborhood, 96% were white. Women experiencing continuous residence in deprived neighborhoods were much more likely to be Black (57.2%) or Latina (19.7%). Among women experiencing upward mobility, 76.0% were white, 14.8% were Black, and 7.9% were Latina. Among women experiencing downward mobility, 82% were white, 11.5% were Black, and 5.9% were Latina.

Average Cumulative Neighborhood Deprivation: Cumulative neighborhood deprivation (Table 4) was associated with inadequate GWG (RR: 1.19; 95% CI: 1.13, 1.24) in crude analyses, but after controlling for covariates, this association was no longer significant (RR: 1.06; 95% CI: 0.99, 1.14). Models did not detect significant effect modification of the association between neighborhood deprivation and inadequate GWG by race/ethnicity (Wald p = 0.18).

For excessive GWG (Table 4), cumulative neighborhood deprivation increased risk by 1.05 times (95% CI: 1.02, 1.09) in adjusted models. After adjusting for covariates, this association persisted (RR: 1.05; 95% CI: 1.00, 1.11). Furthermore, models assessing effect modification by race/ethnicity indicated that race/ethnicity did significantly modify the association between cumulative neighborhood deprivation and excessive GWG (Wald p = 0.07). In race/ethnicity-stratified analyses adjusting for covariates, higher cumulative neighborhood deprivation only significantly increased the risk of excessive GWG for white women (RR: 1.11; 95% CI: 1.02, 1.20), with null associations observed among Black (RR: 1.00; 95% CI: 0.94, 1.06) and Latina (RR: 0.97; 95% CI: 0.89, 1.06) women.

Patterns of Neighborhood Deprivation: In analyses assessing the role of lifetime patterns of deprivation on inadequate GWG (Table 5), crude analyses indicated that stable residence in neighborhoods with high deprivation increased risk of inadequate GWG compared to stable residence in neighborhoods with low deprivation (RR: 1.63; 95% CI: 1.38, 1.93). Similar associations were observed between stable-high deprivation and upward mobility (RR: 1.50; 95% CI: 1.22, 1.83; data not shown) as well as downward mobility (RR: 1.38; 95% CI: 1.15, 1.66; data not shown). However, in all cases, after adjusting for covariates, point estimates were attenuated and borderline significant.

Analyses investigating the relationship between lifetime patterns of deprivation and excessive GWG found that stable residence in a high deprivation neighborhood (RR: 1.23; 95% CI: 1.09, 1.39), upward residential mobility (RR: 1.15; 95% CI: 1.01, 1.31), and downward mobility (RR: 1.18; 95% CI: 1.04, 1.33) all increased risk of excessive GWG compared to stable residence in low-deprivation neighborhoods (Table 6). After adjusting for relevant covariates, associations were still significant and of the same magnitude across all of the lifetime patterns of residential mobility: all conveyed an approximately 20% increased risk of excessive weight gain compared to stable residence in a low-deprivation neighborhood

Sensitivity Analysis
Results from models using the 1990 IOM GWG guidelines (rather than the 2009 guidelines) were similar in direction, magnitude, and statistical significance for both inadequate and excessive GWG (Appendix 1, Table A1.3). Results using the 1990 GWG guidelines, however must be interpreted with the caveat that they use outdated classification criteria to define pre-pregnancy weight class (44) and do not suggest an upper limit of weight gain for obese women. Thus, an upper limit for the overweight women was used to identify excessive gainers in this sub-analysis. This threshold was 2.5kg higher than the threshold eventually implemented for obese women in 2009. (3)

Findings from multinomial logistic regression models, as opposed to log-linear regression models that stratified the outcome (GWG adequacy) into two dichotomous outcomes (inadequate GWG and excessive GWG), were largely similar to those discussed above. (Appendix 1, Table A1.4) Point estimates were further from the null, as expected. An overall Wald test for effect modification by race/ethnicity of the association between cumulative neighborhood deprivation and GWG indicated that effect modification was not significant ($p = 0.27$), but this model may have been underpowered to detect such an association. In particular, when comparing point estimates from race stratified models, (Appendix 1, Table A1.4) findings are the same as those we observed in our main analyses. White women seem to largely be driving the association between cumulative neighborhood deprivation and excessive GWG.

DISCUSSION
Given the limited research on exposure to long-term neighborhood deprivation and GWG, we investigated associations between a measure of cumulative neighborhood deprivation and lifetime patterns of neighborhood deprivation in relation to GWG in a national sample of women. We found that higher cumulative deprivation was associated with increased risk of excessive GWG. Lifetime patterns of deprivation also seemed to matter. Continuous residence in an affluent neighborhood was protective of excessive GWG. Interestingly, all other lifetime patterns of neighborhood deprivation increased risk of excessive GWG by about the same amount compared to lifetime residence in an affluent neighborhood. For inadequate GWG, all findings for both cumulative neighborhood deprivation and lifetime patterns of deprivation were marginally or not significant after adjusting for covariates.

This is the first study, to our knowledge, to investigate cumulative deprivation and GWG. There are other studies that investigate the link between cumulative and/or life course deprivation and birth outcomes. Results from these studies suggest that women with higher levels of long-term neighborhood deprivation have increased risk of low birth weight and small for gestational age infants. (52,53) There is also evidence to support the link between lifetime neighborhood deprivation and increased risk of preterm birth, (37,38,54) especially among women who are low birth weight themselves. (37,38) Our findings provide further support for the importance of cumulative neighborhood deprivation for maternal and child health outcomes. Specifically, while many of the outcomes that have been studied to date are more strongly associated with inadequate

GWG, (3,4) our finding indicated that cumulative neighborhood deprivation mattered for excessive GWG and associated adverse outcomes as well.

Compared to existing studies that have investigated point-in-time measures of neighborhood deprivation, our findings suggest that cumulative neighborhood deprivation has implications for excessive GWG as opposed to inadequate GWG. For example, the only other analysis investigating point-in-time measures of neighborhood socioeconomic deprivation and GWG, (14) finds that neighborhood socioeconomic deprivation is associated with inadequate GWG or weight loss during pregnancy. (14) Three other studies investigating associations between measures of neighborhood social environment, such as physical incivilities and social spaces, (15-17) and GWG also find more consistent associations across measures for inadequate GWG. When these studies do find associations between neighborhood factors and excessive GWG, they are tenuous: among two studies conducted in the same general population (the Wake County area in North Carolina), one finds that physical incivilities is associated with excessive GWG (16) and the other finds that social spaces is associated with excessive GWG. (15) Neither are able to replicate the other's findings. When considered with this literature, our findings suggest that investigating both cumulative and contemporary deprivation is critical to fully understand how neighborhood environments are associated with GWG.

We found evidence to support effect modification of the association between cumulative neighborhood deprivation and excessive GWG by race/ethnicity. Specifically, the association with excessive GWG was driven by white women; cumulative deprivation was not associated with excessive GWG in Black or Latina women. This highlights an important nuance in how cumulative neighborhood deprivation relates to excessive GWG. Other studies assessing effect modification by race/ethnicity of associations between either point-in-time neighborhood deprivation and GWG (14,16,17) or cumulative deprivation and other birth outcomes (52,53) find similar results: associations are stronger in or only significant in whites. (14,16,17,24,52,53) These stronger associations among whites are potentially unintuitive given that social stratification and residential segregation are more likely to concentrate Black and Latino populations into low-income, chronically deprived areas. However, interpretation of findings should consider the fact that many of the factors driving social stratification by race (i.e. discrimination and institutionalized racism (25,27,55)) transcend the economic gradient for Black women. (24,56,57) Thus, Black women may benefit less from upward economic mobility afforded by living in neighborhoods with less socioeconomic deprivation. (28) For Latina women, the lack of association may be due to ethnic enclaves and cultural ties providing support for achieving healthy pregnancy behaviors across the socioeconomic gradient. Long-term, low-income Latina women may be more likely to live in ethnic enclaves that provide support in terms of social connections and resources to buffer the expected adverse impacts of deprivation on health. (30) Additionally, a recent study (58) found that the social connections cultivated within ethnic enclaves remain intact for U.S. born Latinos and are present even after controlling for socioeconomic status. For both of these reasons, more work is needed to understand the contextual environment in which minority women gain weight during pregnancy, even as cumulative neighborhood deprivation should continue to be investigated as a factor to improve GWG adequacy among whites.

Few studies in the literature have investigated how trajectories of neighborhood deprivation are associated with pregnancy outcomes. The majority of studies have been conducted in an inter-generationally linked birth cohort in Chicago (1989-1991). (37,38,53) Overall, these studies find that mobility patterns matter for both low birth weight (53) and preterm birth. (37,38) However, due to high degrees of social and economic stratification, the authors are only able to study downward mobility among whites and upward mobility among Blacks. Another study, conducted using birth data from women in California, (59) also finds that neighborhoods that are consistently impoverished, compared to those that were consistently affluent, increase the risk of preterm birth. A third study in a population of women giving birth in Georgia between 1994 and 2004, (54) examines trajectories of neighborhood deprivation. Findings are mainly descriptive, and these authors find that continued residence in deprived neighborhoods is much more likely among women who gave birth to their first child in a high deprivation neighborhood compared to a low or average deprivation neighborhood. They also observe that teen child bearing interferes with upward mobility across the reproductive life course. (54) Our findings support this existing evidence, suggesting that patterns of continued neighborhood affluence and continued neighborhood deprivation matter for both excessive GWG and inadequate GWG, although confidence intervals for inadequate GWG included the null. Overall, this work indicates that lifetime mobility patterns matter for pregnancy outcomes, and while we were not able to look at this by race/ethnicity in our population, future studies should aim to investigate racial/ethnic differences in lifetime residential mobility patterns and GWG.

Our findings should be considered in light of a few limitations. First, we could not rule out the impact of selection. (60) Selection factors driving women to live in high-deprivation rather than low-deprivation neighborhoods may also be linked to how much a woman gains during pregnancy. If this is the case, the presence of these factors would bias our observed associations. However, this is of less concern in the current study because we controlled for an extensive set of maternal socioeconomic characteristics that have been identified as key predictors of neighborhood selection. (61)

Next, we chose to analyze associations using census tracts as the proxy for neighborhood environments. However, this may not be the ideal geographic level of resolution. If smaller or larger scales of geography are more relevant for this outcome, then our findings will be biased due to differential misclassification of the exposure. This is known as the modifiable area unit problem (MAUP). (60) However, in an analysis looking at the association between a number of pregnancy-related outcomes, including smoking during pregnancy, maternal weight gain, low birth weight, and neighborhood deprivation defined at a number of different spatial levels including census block and tract, Messer et al. (17) found that spatial scale does not largely change the observed association with pregnancy outcomes.

In this analysis we used an administrative index of general socioeconomic deprivation. While this is a relevant reflection of the fact that adverse social and economic exposures often cluster within neighborhoods with poor socioeconomic indicators as well, (61) it does not allow us sufficient resolution to investigate specific characteristics of neighborhood over time in relation to GWG. (62) Future studies should prioritize this avenue of investigation. For example, determining whether it is long-term exposure to depleted service environments or chronic exposure to crime and violence that are responsible for trends in excessive GWG would greatly strengthen our understanding of mechanisms underlying cumulative deprivation and weight gain during pregnancy. Currently, there are not similar data sets, that we are aware of, that could address this limitation. While data linkage is a promising solution to this gap, (54) more work is needed in this area to overcome feasibility barriers for national neighborhood studies, such as cost, adequate data storage, and computing capacity to analyze such large data sets.

Finally, our study relies on self-reported data, which is of particular concern for our outcome, GWG. Reporting bias in self-reported weight in non-pregnant populations has been consistently reported and bias correction techniques have arisen to address these limitations. (63) However, the magnitude of bias for self-reported weights during pregnancy differs from those observed in non-pregnant populations and varies by type of weight outcome being recalled. (64) Pre-pregnancy weight has a higher reporting error (-2.94k -1.09kg) than delivery weight (-0.91-3.01kg). Furthermore, reporting error for each varies by maternal and demographic characteristics. (64) More work is needed to develop appropriate bias correction techniques for self-reported pregnancy-related weight measures (65) so that this limitation can be addressed in future studies. However, in the current analysis, we cannot rule out the impact of such reporting error in biasing our findings, especially for excessive GWG.

There are a number of key strengths of the current study. First, it is one of the first studies to have multiple repeated observations of women's residential census tract that can be linked to administrative data to assess their life-time deprivation over a 30-year follow up for GWG. While other studies have been able to look at "cumulative" deprivation in relation to pregnancy outcomes, many of these measures of cumulative deprivation are based on exiting birth records (53,54,59) which are limited to information on socioeconomic factors at the mother's time of birth and factors at the time that she subsequently gives birth to her offspring, with no resolution on the intervening years. Additionally, we were able to assess associations in a national sample. Prior studies on cumulative or lifetime deprivation have been limited to very specific geographic locations (Chicago, IL (53), Atlanta, GA (54), California (59)). An additional strength of the current study is the racial/ethnic diversity of our population. Additionally, we were among the first to additionally be able to look at associations of cumulative neighborhood deprivation and GWG among Latina women. Finally, in the NLSY79, we were able to control for a wide range of socioeconomic indicators over time. Socioeconomic status is a multidimensional construct that includes multiple domains. (66) Exclusion of any of these domains may result in incomplete measurement of socioeconomic status, and thus, incomplete control of the associated confounding effects. This is particularly of concern in neighborhood studies as the resulting bias from excluded individual SES measures may be picked up in the neighborhood-level point estimate of interest. (62)

CONCLUSION

Overall, our study highlights the association between cumulative neighborhood deprivation and excessive GWG and illustrates the importance of investigating patterns of lifetime deprivation, as well as the overall level of deprivation, when studying GWG. When considered in conjunction with existing literature on neighborhoods and GWG, women's point-in-time and chronic exposure to neighborhood factors will need to be considered in order to better comprehend the neighborhood context in which they are gaining weight during pregnancy. Understanding the dynamic roles of neighborhood deprivation across extended time frames will be key to designing and implementing successful interventions to improve GWG and, thus, the health trajectories of both mother and child.

REFERENCES

1. Messer LC, Boone-Heinonen J, Mponwane L, et al. Developmental programming: Priming disease susceptibility for subsequent generations. *Curr Epidemiol Rep*. 2015;2(1):37–51.

2. Herring SJ, Oken E. Weight gain during pregnancy: Importance for maternal and child health. *Ann Nestlé [Engl]*. 2010;68(1):17–28.

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

4. 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.

5. Amorim AR, Rossner S, Neovius M, et al. Does excess pregnancy weight gain constitute a major risk for increasing long-term BMI? *Obesity*. 2007;15:1278–1286.

6. Cohen AK, Chaffee BW, Rehkopf DH, et al. Excessive gestational weight gain over multiple pregnancies and the prevalence of obesity at age 40. *International Journal of Obesity*. 2014;38(5):714–718.

7. Diesel JC, Eckhardt CL, Day NL, et al. Gestational weight gain and the risk of offspring obesity at 10 and 16 years: a prospective cohort study in low-income women. [Epub ahead of print May 29, 2015]. *BJOG: Int J Obstet Gy*. (doi: 10.1111/1471-0528.13448).

8. Deputy NP, Sharma AJ, Kim SY, et al. Prevalence and characteristics associated with gestational weight gain adequacy. *Obstet Gynecol*. 2015;125(4):773–781.

9. Webb JB, Siega-Riz AM, Dole N. Psychosocial determinants of adequacy of gestational weight gain. *Obesity (Silver Spring)*. 2009;17(2):300-9.

10. Hickey CA. Sociocultural and behavioral influences on weight gain during pregnancy. *Am J Public Health*. 2000;71(suppl):1364S–70S.

11. Krieger N. Theories for social epidemiology in the 21st century: An ecosocial perspective. *Int J Epidemiol*. 2001;30:668–677.

12. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci*. 2010;1186:125–145.

13. Culhane JF, Elo IT. Neighborhood context and reproductive health. *American Journal of Obstetrics and Gynecology*. 2005;192(5):S22–S29.

14. Mendez DD, Doebler DA, Kim KH, et al. Neighborhood Socioeconomic

Disadvantage and Gestational Weight Gain and Loss. *Matern Child Health J*. 2014;18(5):1095-103.

15. Laraia B, Messer L, Evenson K, et al. Neighborhood factors associated with physical activity and adequacy of weight gain during pregnancy. *J Urban Health*. 2007;84(6):793–806.

16. Vinikoor-Imler LC, Messer LC, Evenson KR, et al. Neighborhood conditions are associated with maternal health behaviors and pregnancy outcomes. *Soc Sci Med*. 2011;73(9):1302–1311.

17. Messer LC, Vinikoor-Imler LC, Laraia BA. Conceptualizing neighborhood space: Consistency and variation of associations for neighborhood factors and pregnancy health across multiple neighborhood units. *Health Place*. 2012;18(4):805–813.

18. Morenoff JD. Neighborhood mechanisms and the spatial dynamics of birth weight. *Am J Sociol*. 2003;108(5):976–1017.

19. McEwen B. Protective and damaging effects of stress mediators: Central role of the brain. *Dialogues in Clinical Neuroscience*. 2006;8(4):367–381.

20. 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.

21. Ladyman SR, Augustine RA, Grattan DR. Hormone interactions regulating energy balance during pregnancy. *Journal of Neuroendocrinology*. 2010;22(7):805-17.

22. Schetter CD. Stress Processes in pregnancy and preterm birth. *Curr Dir Psychol*. 2009;18(4):205–209.

23. Nelson SM, Matthews P, Poston L. Maternal metabolism and obesity: Modifiable determinants of pregnancy outcome. *Human Reproduction Update*. 2010;16(3):255–275.

24. Kramer MS. Race, place, and space: Ecosocial theory and spatiotemporal patterns of pregnancy outcomes. In: Howell FM, Porter JR, Mathews S, eds. *Recapturing Space: New midle-Range theory in Spatial Demography*. Dordrecht: Springer, Inc; (in press).

25. Davis EM, Stange KC, Horwitz RI. Childbearing, Stress and Obesity Disparities in Women: A Public Health Perspective. *Matern Child Health J*. 2012;16(1):109–118.

26. Dallman MF. Stress-induced obesity and the emotional nervous system. *Trends in Endocrinology & Metabolism*. 2010;21(3):159–165.

27. Williams DR, Collins C. Racial residential segregation: A Fundamental cause of racial disparities in health. *Public Health Rep*. 2001;116(5):404–416.

28. Collins JW, Schulte NF. Infant health: Race, risk, and residence. In: Kawachi I, Berkman LF, eds. *Neighborhoods and Health*. Oxford University Press, USA; 2003:223–232.

29. Kramer MR, Cooper HL, Drews-Botsch CD, et al. Do measures matter? Comparing surface-density-derived and census-tract-derived measures of racial residential segregation. *Int J Health Geogr*. 2010;9(1):29.

30. Eschbach K, Ostir GV, Patel KV, et al. Neighborhood context and mortality among older Mexican Americans: Is there a barrio advantage? *Am J Public Health*. 2004;94:1807–1812.

31. Bureau of Labor Statistics. *National Longitudinal Survey of Youth*. Washington, DC: US Department of Labor.1979.

32. Bureau of Labor Statistics. *National Longitudinal Survey of Youth, 1979*. Washington, DC: US Department of Labor. 1986.

33. Oken E, Kleinman KP, Rich-Edwards J, et al. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatr*. 2003;3:6.

34. US Census Bureau. Geographic Terms and Concepts-Census Tract. *www.census.gov*. (https://www.census.gov/geo/reference/gtc/gtc_ct.html). (Accessed August 26, 2014)

35. Tatian PA. Appendix J: Description of tract remapping methodology. *www.geolytics.com/pdf/NCDB-LF-Data-Users-Guide.pdf*. (http://www.geolytics.com/pdf/NCDB-LF-Data-Users-Guide.pdf). (Accessed August 26, 2014)

36. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health*. 2006;83(6):1041–1062.

37. Collins JW Jr, Rankin KM, David RJ. African American women's lifetime upward economic mobility and preterm birth: The effect of fetal programming. *Am J Public Health*. 2011;101(4):714–719.

38. Collins JW, Rankin KM, David RJ. Downward economic mobility and preterm birth: An exploratory study of Chicago-born upper class white mothers. *Matern Child Health J*. 2015;19(7):1601–1607.

39. 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.

40. Yu SM, Nagey DA. Validity of self-reported pregravid weight. *Ann Epidemiol*. 1992;2(5):715–721.

41. 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.

42. Siega-Riz AM, Hobel C. Predictors of poor maternal weight gain from baseline anthropometric, psychosocial, and demographic information in a hispanic population. *J am Diet Assoc*. 1997;97(11):1264–1268.

43. Bodnar LM, Hutcheon JA, Platt RW, et al. Should gestational weight gain recommendations be tailored by maternal characteristics? *Am J Epidemiol*. 2011;174(2):136–146.

44. Institute of Medicine, ed. Nutrition during pregnancy: Part I-Weight gain. Washington, DC: National Academies Press (US); 1990.

45. Hutcheon JA, Bodnar LM, Joseph KS, et al. The bias in current measures of gestational weight gain. *Paediatric and Perinatal Epidemiology*. 2012;26(2):109–116.

46. Hutcheon JA, Platt RW, Abrams B, et al. A weight-gain-for-gestational-age z score chart for the assessment of maternal weight gain in pregnancy. *Am J Public Health*. 2013;97(5):1062-7.

47. Rehkopf DH, Krieger N, Coull B, et al. Biologic risk markers for coronary heart disease. *Epidemiology*. 2010;21:38–46.

48. Bureau UC. Geographic Terms and Concepts- Census Divisions and Census Regions. *www.census.gov*. (https://www.census.gov/geo/reference/gtc/gtc_census_divreg.html). (Accessed August 26, 2014)

49. Zou G. A Modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*. 2004;159(7):702–706.

50. Greenland S. Interactions in Epidemiology: Relevance, identification, and estimation. *Epidemiology*. 2009;20(1):14–17.

51. StataCorp. Stata multiple-imputation reference manual. College Station, TX: StataCorp LP; 2013.

52. Collins JW, David RJ, Rankin KM, et al. Transgenerational effect of neighborhood poverty on low birth weight among African Americans in Cook County, Illinois. *American Journal of Epidemiology*. 2009;169(6):712–717.

53. Collins JW, Wambach J, David RJ, et al. Women's lifelong exposure to neighborhood poverty and low birth weight: A population-based study. *Matern*

Child Health J. 2008;13(3):326–333.

54. Kramer MR, Dunlop AL, Hogue CJR. Measuring women's cumulative neighborhood deprivation exposure using longitudinally linked vital records: A method for life course MCH Research. *Matern Child Health J*. 2013;18(2):478–487.

55. Braveman P. Health inequalities by class and race in the US: What can we learn from the patterns? *Social Science & Medicine*. 2012;74(5):665–667.

56. Colen CG. Addressing racial disparities in health using life course perspectives: Toward a constructive criticism. *Du Bois Review*. 2011;8(1):79–94.

57. Braveman PA, Heck K, Egerter S, et al. The role of socioeconomic factors in Black–White disparities in preterm birth. *Am J Public Health*. 2015;105(4):694–702.

58. Viruell-Fuentes EA, Morenoff JD, Williams DR, et al. Contextualizing nativity status, Latino social ties, and ethnic enclaves: an examination of the "immigrant social ties hypothesis." *Ethnicity & Health*. 2013;18(6):586–609.

59. Margerison-Zilko CE, Cubbin C, Marchi KS, et al. Beyond the cross-sectional: Neighborhood poverty histories and preterm birth. *Am J Public Health*. 2015;105:1174–1180.

60. Mujahid MS, Roux AVD. Neighborhood Factors in Health. In: Steptoe A (ed.). *Handbook of Behavioral Medicine: Methods and Applications*. New York, NY: Springer New York; 2010:341–354.

61. Sampson RJ. Neighborhood-level context and health: Lessons from Sociology. In: Kawachi I, Berkman LF, eds. *Neighborhoods and Health*. New York: Oxford Univ Press; 2003:132–178.

62. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *J Epidemiol Community Health*. 2001;55:111–122.

63. 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.

64. Headen I, Cohen AK, Abrams B. The accuracy of self-reported pregnancy-related weight: A systematic review. (Unpublished Manuscript).

65. Bodnar LM, Siega-Riz AM, Simhan HN, et al. The impact of exposure misclassification on associations between prepregnancy BMI and adverse pregnancy outcomes. *Obesity (Silver Spring)*. 2010;18(11):2184–2190.

66. Williams DR, Collins C. US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology*. 1995;21:349–386.

TABLES AND FIGURES

Figure 1. Observations Included in Final Analytic Sample from the 1979 National Longitudinal Survey of Youth Based on Inclusion and Exclusion Criteria

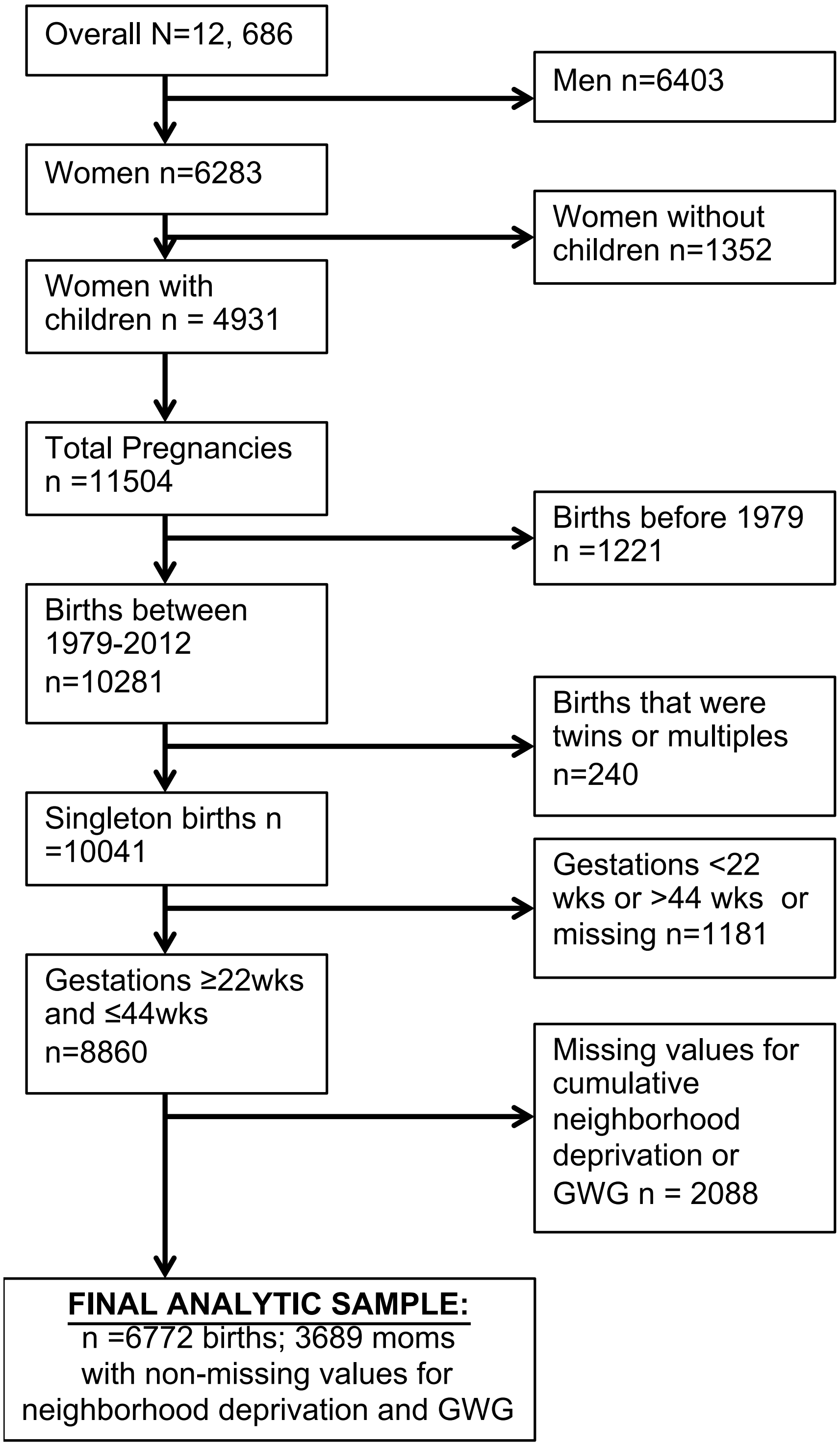


Table 1. Institute of Medicine Gestational Weight Gain Recommendations

	Overall Weight Gain (kg)	Weight Gain by Trimester	
		1st Trimester Total Gain	2nd and 3rd Trimester Rate of Gain (kg/week)
2009 Recommendations			
Underweight (<18.5kg/m^2)	12.5-18	2	0.51
Normal Weight (18.5-24.9 kg/m^2)	11.5-16	2	0.42
Overweight (25.0-29.9 kg/m^2)	7-11.5	1	0.28
Obese (≥30.0kg/m^2)	5-9	0.5	0.22
1990 Recommendations			
Underweight (<19.8kg/m^2)	12.5-18	n.s.*	0.50
Normal Weight (19.8-26.0 kg/m2)	11.5-16	n.s.*	0.40
Overweight (>26.0-29.0kg/m2)	7-11.5	n.s.*	0.30
Obese (>29.0kg/m2)	≥7	n.s.*	n.s.

*Estimates for first trimester were not specified (n.s.), the IOM report mainly relied on overall growth charts spanning the duration of pregnancy and thus focused on rates per week.

Table 2. Survey Weighted Descriptive Statistics for Analytic Sample by Gestational Weight Gain (GWG) Adequacy

	Un-weighted N	Total Sample	Inadequate GWG (n=1663; 22.3%)[¥]	Adequate GWG (n=2044; 12.9%)[¥]	Excessive GWG (n=3095; 45.3%)[¥]	P-value
Cumulative Neighborhood Deprivation (mean (SD))	6772	18.1 (8.31)[§]	19.5 (8.9)[§]	17.2 (7.4)[§]	18.1 (8.0)[§]	**<0.0001**
Neighborhood Deprivation Patterns[†]						
Stable-Affluent	1,118[†]	45.1%	41.1%	51.6%	42.0%	**<0.0001**
Upward Mobility	672[†]	17.0%	15.0%	16.4%	18.4%	
Downward Mobility	729[†]	21.6%	20.6%	19.8%	23.5%	
Stable Deprived	1,196[†]	16.3%	23.4%	12.1%	16.1%	
Race/Ethnicity*						
White	2194	81.8%	76.3%	85.3%	81.9%	**<0.001**
Black	912	12.9%	17.5%	10.3%	12.4%	
Latina	583	5.4%	6.2%	4.4%	5.6%	
Foreign born*						
No	3415	95.8%	95.3%	96.2%	95.8%	0.71
Yes	274	4.2%	4.7%	3.9%	4.2%	
Marital Status						
No	2,730	32.7%	36.2%	28.3%	34.1%	**<0.0001**
Yes	3,805	67.3%	63.8%	71.7%	65.9%	
Employment						
Unemployed	2,485	31.7%	37.7%	32.0%	28.6%	**<0.001**
Part-Time Employed	1,803	28.2%	26.0%	29.5%	28.4%	
Full-time Employed	2,314	40.1%	36.3%	38.4%	43.1%	
Region of Residence						
Northeast	1,217	20.2%	20.0%	19.5%	20.8%	0.45
North Central	1,559	29.2%	29.3%	30.0%	28.5%	
South	2,190	31.1%	31.4%	32.9%	29.7%	
West	1,533	19.6%	19.3%	17.7%	21.1%	

Table 2. Survey Weighted Descriptive Statistics for Analytic Sample by Gestational Weight Gain (GWG) Adequacy--CONTINUED

	Un-weighted N	Total Sample	Inadequate GWG (n=1663; 22.3%)[¥]	Adequate GWG (n=2044; 12.9%)[¥]	Excessive GWG (n=3095; 45.3%)[¥]	P-value
Home Ownership						
No	4,079	54.3%	59.1%	49.5%	55.4%	**<0.0001**
Yes	2,160	45.7%	40.9%	50.5%	44.6%	
Child's Birth year (mean (SD))	6772	1988 (5.1)[§]	1988 (5.3)[§]	1988 (4.8)[§]	1989 (5.1)[§]	**<0.0001**
Education						
<12 yrs	1,505	16.1%	20.4%	13.6%	15.7%	**<0.0001**
12-15 yrs	3,996	63.0%	63.2%	60.5%	64.8%	
≥16 yrs	1,024	20.9%	16.5%	25.9%	19.6%	
Equivalized Income (mean (SD))	5841	9.9 (1.2)[§]	9.7 (1.2)[§]	10.0 (1.04)[§]	9.9 (1.1)[§]	**<0.0001**
Quartile 1	1,387	16.4%	19.8%	14.5%	16.1%	**<0.01**
Quartile 2	1,422	19.1%	21.7%	17.0%	19.4%	
Quartile 3	1,460	28.1%	27.3%	29.7%	27.4%	
Quartile 4	1,572	36.4%	31.2%	38.9%	37.0%	
Household Income (mean (SD))	6772	$39,729 (39,999)	$35193 (37,857)	$42762 (40,855)	$39,801 (39,300)	**<0.001**
Family Size (mean (SD))	6536	3.1 (1.52)[§]	3.3 (1.7)[§]	3.1 (1.4)[§]	3.0 (1.5)[§]	**<0.001**
Maternal Age (mean (SD))	6772	27.3 (5.0)[§]	26.7 (5.3)[§]	27.4 (4.7)[§]	27.5 (5.0)[§]	**<0.001**
<20 yrs	500	5.5%	6.7%	5.2%	5.1%	**0.01**
20-24 yrs	2,091	26.9%	31.0%	25.6%	25.8%	
25-29 yrs	2,303	35.0%	33.4%	36.6%	34.6%	
30-34 yrs	1,331	23.3%	18.9%	23.8%	25.2%	
35-39 yrs	489	8.4%	9.1%	7.8%	8.5%	
≥40 yrs	58	1.0%	0.9%	1.1%	1.0%	

Table 2. Survey Weighted Descriptive Statistics for Analytic Sample by Gestational Weight Gain (GWG) Adequacy--CONTINUED

	Un-weighted N	Total Sample	Inadequate GWG (n=1663; 22.3%)¥	Adequate GWG (n=2044; 12.9%)¥	Excessive GWG (n=3095; 45.3%)¥	P-value
Parity (mean (SD))	6772	0.9 (1.0)§	1.0 (1.1)§	1.0 (1.0)§	0.9 (1.0)§	**<0.001**
0	2,654	41.8%	35.9%	40.3%	45.8%	**<0.0001**
1	2,337	34.8%	38.1%	35.1%	33.0%	
2	1,149	15.6%	16.0%	17.1%	14.3%	
3	408	5.4%	6.6%	5.1%	5.0%	
≥4	224	2.4%	3.6%	2.3%	1.9%	
Moved in Birth Year						
No	3,196	70.5%	69.4%	72.2%	69.8%	0.33
Yes	1,433	29.5%	30.6%	27.8%	30.3%	
Length of residence (mean (SD))	6189	1.5 (2.22)§	1.5 (2.3)§	1.6 (2.2)§	1.5 (2.1)§	0.20
Long term resident (≥ 5years)						
No	5,640	90.2%	90.3%	89.7%	90.5%	0.78
Yes	549	9.8%	9.7%	10.3%	9.5%	
Cumulative time moved (mean (SD))	6772	1.67 (1.64)§	1.5 (1.6)§	1.6 (1.5)§	1.8 (1.7)§	**<0.001**
Moved a lot (≥5 times)						
No	6,351	93.1%	94.7%	93.9%	91.7%	**0.01**
Yes	421	6.9%	5.3%	6.1%	8.3%	
Rural Residence						
No	5,437	84.6%	82.5%	84.1%	86.1%	**0.04**
Yes	855	15.4%	17.5%	15.9%	13.9%	

*Numbers refer to counts of mothers, not pregnancies

¥Reported n's are for the un-weighted sample; percentages are survey weighted

§ Means and standard deviations are reported as descriptive statistics in this cell

†Sample size for this overall analysis differs from the full analytic sample; 3715 observations had non-missing values for patterns in neighborhood deprivation and fell into one of the a prior defined categories listed in the table

Table 3. Survey Weighted Mean Cumulative Neighborhood Deprivation by Study Covariates for Analytic Sample

	Cumulative Neighborhood Deprivation Mean (SD)	P-value
Gestational Weight Gain		
Inadequate	19.5 (8.9)	**<0.0001**
Adequate	17.2 (7.4)	
Excessive	18.1 (8.0)	
Race/Ethnicity		
White	15.8 (4.8)	**<0.0001**
Black	28.7 (13.3)	
Latina	26.1 (14.8)	
Foreign born		
No	17.9 (7.8)	**<0.001**
Yes	21.5 (11.6)	
Marital Status		
No	21.8 (10.8)	**<0.0001**
Yes	16.3 (6.0)	
Employment		
Unemployed	20.6 (10.1)	**<0.0001**
Part-Time Employed	18.1 (7.6)	
Full-time Employed	16.3 (6.3)	
Region of Residence		
Northeast	18.1 (8.7)	0.09
North Central	17.2 (7.3)	
South	19.2 (8.2)	
West	17.8 (7.5)	
Home Ownership		
No	20.3 (10.0)	**<0.0001**
Yes	15.1 (4.9)	
Education		
<12 yrs	23.4 (11.0)	**<0.0001**
12-15 yrs	18.3 (7.5)	
≥16 yrs	13.6 (4.7)	

Table 3. Survey Weighted Mean Cumulative Neighborhood Deprivation by Study Covariates for Analytic Sample--CONT'D

	Cumulative Neighborhood Deprivation	
	Mean (SD)	P-value
Equivalized Income		
Quartile 1	22.9 (10.9)	**<0.0001**
Quartile 2	21.3 (9.7)	
Quartile 3	17.0 (5.9)	
Quartile 4	14.2 (4.6)	
Maternal Age		
<20 yrs	23.4 (10.7)	**<0.0001**
20-24 yrs	21.0 (9.0)	
25-29 yrs	17.7 (7.4)	
30-34 yrs	15.7 (6.4)	
35-39 yrs	14.5 (6.3)	
≥40 yrs	14.5 (8.5)	
Parity		
0	17.3 (7.4)	**<0.0001**
1	18.1 (7.8)	
2	19.1 (8.9)	
3	20.0 (9.4)	
≥4	22.0 (12.2)	
Moved in birth year		
No	17.6 (8.0)	0.08
Yes	18.1 (7.6)	
Long term resident (≥ 5years)		
No	18.3 (8.1)	**0.02**
Yes	17.2 (8.1)	
Moved a lot (≥5 times)		
No	18.3 (8.2)	**<0.0001**
Yes	15.6 (5.5)	
Rural Residence		
No	18.1 (8.4)	0.65
Yes	17.8 (5.9)	

Table 4. Main Associations Between Cumulative Neighborhood Exposures and Gestational Weight Gain (GWG), Overall and Stratified by Race For Significant Interaction

	Crude		Adjusted†		Wald test for Interaction§
	RR	95% CI	RR	95% CI	P-value
Inadequate GWG	**1.19**	**1.13, 1.24**	1.06	0.99, 1.14	0.18
Excessive GWG	**1.05**	**1.02, 1.09**	**1.05**	**1.00, 1.11**	**0.07**
White	1.06	0.99, 1.13	**1.11**	**1.02, 1.20**	
Black	1.00	0.95, 1.05	1.00	0.94, 1.06	
Latina	0.99	0.91, 1.07	0.97	0.89, 1.06	

†Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.
§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

Table 5. Associations Between Patterns of Neighborhood Deprivation and Gestational Weight Gain (GWG)[§]

	Inadequate GWG				Excessive GWG			
	Crude		Adjusted[†]		Crude		Adjusted[†]	
	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Stable-Affluent	-	-	-	-	-	-	-	-
Upward Residential Mobility	1.09	0.86, 1.38	0.99	0.77, 1.26	**1.15**	**1.01, 1.31**	**1.19**	**1.02, 1.38**
Downward Residential Mobility	1.18	0.94, 1.48	1.08	0.86, 1.37	**1.18**	**1.04, 1.33**	**1.21**	**1.06, 1.38**
Stable-Deprived	**1.63**	**1.38, 1.93**	1.22	0.98, 1.53	**1.23**	**1.09, 1.39**	**1.21**	**1.03, 1.42**

[§]Sample size for this analysis is restricted to 3715 observations that fell into one of the *a priori* specified categories in the table: stable-affluent—n = 1118; upward residential mobility—n = 672; downward residential mobility—n = 729; stable-deprivation—n = 1196.

[†]Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.

CHAPTER 3. PERCEIVED AND OBJECTIVE NEIGHBORHOOD SOCIAL ENVIRONMENT AND GESTATIONAL WEIGHT GAIN

INTRODUCTION

Healthy weight gain during pregnancy is key to supporting both short and long-term positive health trajectories for mother and child. (1,2) Too little weight gain has been linked to increased risk of preterm birth and small for gestational age (SGA), both of which contribute to infant mortality rates in the U.S. (3) However, too much weight gain during pregnancy has been liked to adverse outcomes as well, including increased risk of maternal postpartum weight retention, later life obesity development in the mother, (3-5) and increased risk of childhood obesity in her offspring. (6) Thus, gestational weight gain (GWG) is an important modifiable factor that can contribute to both improving infant outcomes and addressing the growing obesity epidemic in the U.S.

Ideal weight gain recommendations, developed by the Institute of Medicine (IOM), to aid women in avoiding these adverse outcomes have failed to increase the prevalence of adequate GWG among childbearing women. (3) Recent data from the 2011 Pregnancy Nutrition Surveillance System (PNSS, (7)) indicates that 48% of women still gain in excess of the guidelines and 21% gain below the guidelines. Although this sample only tracks weight gain among low income women, similar trends in other studies across different populations have been observed. (8-11) A wide body of literature exists identifying individual risk factors associated with both low and high weight gain, (8,12-14) but few studies have focused on neighborhood-level risk factors. Factors outside of the individual critically define the context in which they must attempt to engage in health-promoting behaviors. Such contextual determinants of health may be just as important for understanding health outcomes as the individual's personal attributes. (15) Neighborhoods, in particular, have arisen as a relevant dimension of this external milieu. (16) Pregnant women may rely on contextual factors, such as goods and services or social support within their neighborhood, in order to maintain a healthy pregnancy.

Studies that have investigated neighborhood factors in relation to GWG indicate that both poor physical and social neighborhood environments increase the risk of excessive (17-19) and inadequate GWG. (17-20) Measures of the physical environment included in these studies contain items such as presence of parks, sidewalks and distance to supermarkets. The social environment is assessed using indicators such as presence of vacant property, graffiti, and litter because these factors usually denote the presence of social disorder in an area. However, only two of these studies (18,20) use relevant, IOM-based criteria to classify inadequate and excessive GWG and they all rely on data from women in only two geographically localized areas (five counties in the Wake County area of North Carolina and Allegheny County, Pennsylvania). Thus, more population-based studies using appropriate outcome classifications are needed to understand these associations in geographically diverse populations.

Furthermore, research is needed to better understand the impact of both perceived and objective neighborhood social factors on GWG. Previous studies have only considered objective measures of neighborhood disadvantage, (20) physical incivilities, lack of social spaces, (17-19) and poor walkability. (17,19) No studies to date have additionally assessed residents' perceptions of such factors. While perceived and objective measures of neighborhood environment are correlated, existing research (e.g. (21-23)) indicates that there is still a fair amount of discordance between the two. While the focus on objective measures is warranted in order to avoid concerns about reporting bias associated with use of perceived measures, (21,24) these may omit key elements that are relevant to residents' health. Residents are able to report on social contextual factors that may be imperceptible to administrative sources or neighborhood "outsiders." (21,24) Additionally, objective measures may incompletely reflect residents' appraisal of neighborhood factors as problematic. (16,25) Perceptions of neighborhood environments may better capture residents' appraisal of stressors and thus the extent to which these elements are functional barriers to engaging in health-promoting behaviors. (23,26) Given that embodiment pathways are mediated through stress activation, (25) risk appraisal of neighborhood factors, as reported by residents, may be a better indicator of whether the requisite stress activation needed for stress-based pathways to operate is present. Thus, investigation of perceived measures of neighborhood environment may capture key elements of the lived experience of residence in the neighborhood that are relevant to GWG outcomes.

Investigating both perceived and objective neighborhood factors may be especially relevant for pregnancy-related outcomes, such as GWG. Pregnancy is a time during which physical and psychological change often increase a woman's material demands on her environment, as well as shift her risk assessment or environmental appraisal in light of a new or growing family. (27-30) For both of these reasons, elements within the environment that were previously inconvenient but innocuous to health may become barriers to women engaging in health promoting behaviors during their pregnancy, especially as they relate to weight gain processes.

Due to persistent residential segregation racial/ethnic minorities are more likely to reside in neighborhoods with higher levels of deprivation. (31,32) Furthermore, because such segregation is rooted in institutional policies of housing discrimination and social stratification, disadvantaged neighborhoods that minorities may reside in may be characteristically different from deprived neighborhoods that whites reside in. (32) For example, they may have experienced more chronic levels of deprivation over time and reflect social and structural deterioration as a result of continued economic and structural divestment. Thus, associations between neighborhood factors and GWG likely vary by race/ethnicity.

Thus, to address these gaps in the literature, we examined the relationship between perceptions of neighborhood social deprivation, objectively measured neighborhood socioeconomic deprivation, and GWG in a national sample of women. The three objectives of our analyses were 1) to estimate the association between perceived neighborhood social deprivation, as assessed by neighborhood problems, and GWG, 2) to assess whether associations between neighborhood social deprivation and GWG were different for objective and perceived measures and 3) to determine if associations between neighborhood social deprivation, both objective and perceived, and GWG varied by race/ethnicity. We hypothesized that both perceived and objective measures of neighborhood social deprivation would be associated with GWG, and that perceptions of neighborhood social deprivation may have a stronger (i.e. higher magnitude) association with GWG as it is more proximal to stress appraisal for women. Furthermore, we hypothesized that associations would vary across racial/ethnic groups.

METHODS

Study Sample

Our sample included singleton births occurring from 1992-2000 to women in the 1979 National Longitudinal Survey of Youth. The NLSY79 is a cohort study conducted by the National Bureau of Labor Statistics composed of a nationally-representative sample of 12686 men and women, who were 14-21 years old at the time of enrollment. (33) The cohort was constructed from three, multi-stage probability samples: a nationally representative sample, a military sample, and a supplemental sample that overrepresented Blacks, Hispanic/Latino/Spanish (referred to as Latino/a going forward), and economically disadvantaged non-Black, non-Latinos. The military and economically disadvantaged samples were discontinued after 1984 and 1990, respectively. Subjects were followed from 1979 to 2012.

Participants were interviewed annually from 1979 to 1994 and biennially to 2012. Information on women's pregnancies over the course of follow up was collected starting in 1986. (34) Pregnancies occurring between interview dates were documented at the closest subsequent interview. We restricted our sample to births from 1992-2000 (n = 2031) based on the timing of data collection for our perceived measure of neighborhood problems. We further restricted our sample to gestations of 22-44 weeks ((35); n = 1553) that had non-missing information on both exposures and outcomes. Thus, our sample size for investigating perceived neighborhood problems was 1310 pregnancies to 1026 mothers and our sample size for objective neighborhood deprivation was 1243 pregnancies to 984 mothers (Figure 1). Our analysis was approved by the University of California, Berkeley Center for the Protection of Human Subjects. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

Analytic Variables

Exposures

Perceived Neighborhood Problems (PNP): From 1992 to 2000 women with children in the NLSY79 cohort were asked to report on a number of neighborhood characteristics relevant to raising children in their neighborhood. Women rated the following seven items as either (1) a big problem, (2) somewhat of a problem, or (3) not a problem: crime and violence in the neighborhood, abandoned and rundown buildings, police protection, public transportation, parent supervision of their children, neighbors keeping to themselves, and individuals respecting rules and laws. Items were highly correlated (Cronbach's alpha = 0.81). We used factor analysis with oblique rotation to create a score for PNP for each woman in the year that she gave birth or in the closest prior year. From this factor analysis we identified one factor. Item loadings onto this factor were relatively high for most items (0.58-0.85), although loading for public transportation was somewhat low (0.24). However, excluding this factor did not greatly change internal consistency across items (data not shown), so we included it in the final score. We created an overall perceived neighborhood score as an un-weighted sum of the item scores. Items were reverse coded so that higher scores indicated more problematic neighborhoods and the index was re-scaled so that women reporting no problems received a score of 0. Scores ranged from 0 to 14. PNP was analyzed as a categorical variable divided into the following categories: no problems, 1-2 problems, 3-4 problems, and 5-14 problems.

Neighborhood Deprivation (NDI): We used socioeconomic data from the 1990 and 2000 U.S. Census to create a socioeconomic deprivation index. Census tracts were used as proxies for the neighborhood. Census tracts are administrative subdivisions of counties that contain on average 4000 individuals. (36) Boundaries for these geographic elements are based on population density, socioeconomic homogeneity, and any relevant physical boundaries (e.g. highways, rivers). All census tract boundaries were standardized to year 2000. (37)

Census data was obtained from the Neighborhood Change Database (Geolytics, Inc., 2003). Eight items covering five socioeconomic domains (occupation, education, housing, employment, and poverty) were used to construct a deprivation index. Items were selected based on existing literature (38) and included the following items: percent individuals in management and professional occupations, percent unemployed, percent crowded households, percent families in poverty, percent female-headed households w/ dependents, percent households on public assistance, percent of families earning < $30,000, and percent adults who earned less than a high school degree. Geometric interpolation was used to assign values to inter-censal years. (39) Items were moderately correlated across years (r = 0.54-0.68) and factor analysis with oblique rotation was used to reduce these items into a weighted score ranging from 0-100. Year-specific scores were linked to each birth using the woman's census tract of residence in the closest prior interview year to ensure temporality. For pregnancies occurring in interviews, this amounted to a 2-year maximum lag, and for pregnancies occurring between interview years, this amounted to a 1 year maximum lag. Neighborhood deprivation was moderately correlated with perceived neighborhood problems (r = 0.43). Neighborhood deprivation was categorized into quartiles for analysis.

Outcome

Observed GWG was calculated as the difference between self-reported pre-pregnancy and delivery weight. We then calculated a measure of GWG adequacy standardized by length of gestation to account for the non-linear association between GWG and pregnancy duration, as described previously in the literature. (40-42) GWG adequacy was the ratio of a woman's observed to her expected GWG for her length of gestation.

Expected GWG was based on the 2009 IOM recommendations for amount of weight gain during the first trimester and rate of weight gain during the second and third trimesters (Table 1) using the following equation: expected GWG= recommended first trimester weight gain + (gestational age-13)*(rate of weight gain during the second and third trimesters). Trimester (1^{st}, 2^{nd}, 3^{rd}) recommendations were specific to the category of pre-pregnancy BMI, with lower recommended amounts of weight gain among women who had a higher pre-pregnancy BMI (Table 1). The GWG adequacy ratio used in our main analysis was further categorized into inadequate, adequate, and excessive weight gain based on the 2009 IOM recommendations (Table 1). (3)

We conducted a number of sensitivity analyses using secondary classifications of GWG. First, we conducted a sensitivity analyses using the 1990 IOM recommendations (43) rather than 2009 recommendations. The majority of births in our sample occurred prior to the creation of the 2009 recommendations. The 1990 recommendations were in place at the time that women in our cohort were giving birth. However, the 1990 guidelines use an outdated classification criterion to group women by pre-pregnancy BMI and do not have an upper limit for obese women, preventing the accurate classification of women gaining excessively. For our sensitivity analysis, we applied the upper weight gain limit for obese women found in the 2009 guidelines in order to identify obese women who may have gained excessively. Using this modification, 93% (Inadequate: 99%; Adequate: 86%; Excessive: 94%) of women were classified in the same GWG adequacy group regardless of which set of guidelines were used. Thus while we report findings using 2009 recommendations in main analyses, we report findings using the1990 classifications in Appendix 2. Additionally, we conducted sensitivity analyses to determine robustness of findings to different approaches to decouple the dependency of GWG on gestational age. Recent literature (44,45) indicated that the GWG adequacy may not completely eliminate the correlation between length of gestation and GWG leading to residual bias in associations with key gestation-dependent birth outcomes. Thus, we assessed robustness of associations across multiple approaches that address this dependency. We controlled for gestational age both linearly in the model and restricted to the subset of full term births (n = 985) among which gestation-based differences in GWG are not as much of a concern.

Covariates

Covariates identified in the literature as potential confounders (12,13,46) included the following: race/ethnicity, immigrant status, marital status, maternal age, parity, birth year of the child, region of residence at the time of birth of the child, employment, education, income, and home ownership. Race/ethnicity was categorized as Black, Latina, and non-Black, non-Latina. Our non-Black, non-Latina group was 88% white, so they are referred to as “white” going forward. Immigrant status (yes/no) was self-reported at baseline. Marital status was classified as currently married (yes) or never being married/divorced/separated (no). Maternal age was measured continuously based on the age at which the woman gave birth. Parity was measured as the number of live births that a woman had prior to the index pregnancy. Birth year of child was self-reported by the mother at the time of interview for the index pregnancy. Region of residence at the time of birth of the child was based on Census defined regional definitions (47) and were classified as northeast, north central, south, and west. Employment was based on the average hours per week that women reported working in the year prior to interview. Women were classified as unemployed (≤ 10 hrs/week), part-time employed (10-34 hrs/week), and full time employed (≥ 35 hrs/week). Educational attainment was classified as women who had less than high school education (< 12 yrs), high school education and/or some college (12-15 yrs), and a college education (≥ 16 yrs). Household income was measured continuously, adjusted for family size, (48) and log transformed to account for non-normal distribution. Home ownership of the woman or her spouse (yes/no) was also measured.

Finally, we measured three residential selection characteristics to capture residential stability. All measures were taken from the closest prior interview. Moving (yes/no) was determined based on changes in reported census tract of residence between consecutive interview years. Length of residence for women who did not move was determined based on number of years that a woman had lived in her census tract of residence at the time of the index pregnancy. Cumulative number of times moved was assessed as a count of different values for census tract of residence that a woman had over the course of follow up. We also assessed whether the woman lived in a rural area (yes/no) based on census definitions of urbanized areas. (33)

Analytic Approach

Survey weighted descriptive statistics were calculated to determine the mean, standard deviation, and frequency of all analytic variables. Bivariable analysis used ANOVA's and chi-squared tests when appropriate to assess whether covariates varied by GWG as well as neighborhood deprivation and perceived problems. Multivariable regression was used to estimate crude and adjusted associations between our exposures of interest and GWG. Adjusted analyses included the covariates described previously; continuous covariates were median-centered to improve interpretability of estimates. Generalized linear models with log link functions and Poisson distributions were used to estimate risk ratios. (49) Survey weights were used to account for the sampling design as well as national representativeness of the sample. Clustering of women within census tracts was negligible (1.2 women per tract; SD = 0.6-0.7 across years of follow up), but we still specified robust standard errors to account for any remaining non-independence between observations due to clustering of pregnancies within the mother.

We assessed effect measure modification of our associations of interest between neighborhood deprivation and PNP and GWG by race/ethnicity. Cross-product terms were calculated between race/ethnicity (White/Black/Latina) and exposures of interest (neighborhood deprivation or PNP). Wald tests were used to test for significant effect measure modification based on the set of cross-product terms in each model. The significance threshold was set *a priori* at $p \leq 0.10$ to account for the fact that models may be less powered to detect interaction. (50) Race/ethnicity stratified models were reported if significant effect measure modification was detected. All analyses were conducted in Stata 12.0 (StataCorp LP College Station, TX, 2011-2013)

Prevalence of missing data in our covariates ranged from 0% to 13%. (Appendix 2 Table A2.1) We used regression-based multiple imputation using chained equations to impute missing values for covariates. (51) Our sample remained complete case for each exposure and outcome. However, women who were excluded from our sample were more likely to gain adequately, be married, be employed part-time or unemployed, live in the South, and have lower income, slightly larger family sizes, be multiparous, and have moved fewer times over follow up. Models assessing inadequate GWG (compared to adequate GWG) were estimated separately from models assessing excessive GWG. As a sensitivity analysis, we additionally fit multinomial logistic regressions to pool all women across GWG adequacy categories (rather than models run separately for inadequate compared to adequate and excessive compared to adequate GWG). Findings were further from the null, as expected for odds ratios, but comparable in both direction and significance to those in stratified models, so we only report estimates from stratified models.

RESULTS

Based on our final analytic sample for analyses examining PNP and GWG (n = 1310 pregnancies; 1026 mothers), the mean age was 33.0 (SD = 2.9) and 83.4%, 11.0%, and 5.6% of women were white, Black, or Latina, respectively. (Table 2) The distribution of GWG adequacy closely reflected those previously observed in the literature, with the majority of women gaining excessively (49.1%) with 17.4% still gaining inadequately. On average, women did not perceive their neighborhoods to be very problematic, with a third of women (33.9%) perceiving no problems in their neighborhood. Similarly, mean neighborhood deprivation score was low, on average 16.5 (SD = 10.5) on a scale ranging from 0 to 100, but scores up to 76 were observed for neighborhoods in the highest quartile of deprivation (data not shown). The majority of mothers in our sample were giving birth to their first (32.4%) or second (33.0%) child, were married (78.1%), employed full time (54.2%), and had received at least a high school education (95.0%). Most women also lived in urban areas, in homes owned by themselves or a spouse, and had lived in their census tract of residence an average of two years (SD = 2.6).

Adequate gainers varied from non-adequate gainers across a number of our covariates. Women who gained adequately were more likely to be married (84.5%), college graduates (47.3%), in the highest quartile of income (40.0%), and move less over the course of follow up (2.8 times; SD = 2.0). Alternatively, inadequate gainers were more likely to be minority women (23.4%) and had a higher neighborhood deprivation score (18.2; SD = 11.7).

Racial and socioeconomic composition varied across categories of PNP and neighborhood deprivation (Table 3) as expected. Neighborhoods in the higher quartiles of deprivation and experiencing the most problems had more minorities and poorer socioeconomic indicators. Women in the most deprived neighborhoods were also less likely to gain adequately compared to women in the other quartiles of deprivation, but GWG did not vary across PNP categories. Employment did not vary across categories of PNP. Approximately half of women worked full time across all categories of perceived problems, with the remaining women almost evenly split between being unemployed and working part-time. None of the residential characteristics varied across categories of neighborhood deprivation or PNP, with the exception of rural residence.

Perceived Neighborhood Problems: In crude analysis (Table 4), women reporting the highest number of problems in their neighborhood, compared to women reporting no problems, had higher risk of inadequate GWG (relative risk (RR): 1.28; 95% confidence interval (CI): 0.94, 1.75). However, this difference was not significant. The association was attenuated and remained non-significant after adjusting for covariates (RR: 1.09; 95% CI: 0.78, 1.51). For excessive GWG, women who perceived their neighborhood to be most problematic, compared to women who did not think that their neighborhood was problematic at al, had an increased risk of excessive GWG in unadjusted models (RR: 1.21; 95% CI: 1.02, 1.44). After adjusting for covariates, this association was no longer significant (RR: 1.16; 95% CI: 0.95, 1.40).

Neighborhood Deprivation Index: In crude analysis (Table 4), risk of inadequate GWG increased across quartiles of high deprivation (Table 4). For the two highest quartiles of neighborhood deprivation, unadjusted risk of inadequate GWG was significantly higher (Quartile 3: RR: 1.52; 95% CI: 1.11, 2.10; Quartile 4: RR: 1.80; 95% CI: 1.33, 2.44) compared to women in the lowest quartile of neighborhood deprivation. Associations were attenuated and no longer statistically significant after adjusting for covariates (Quartile 3: RR: 1.32; 95% CI: 0.93, 1.87; Quartile 4: RR: 1.36; 95% CI: 0.93, 2.01).

Crude analyses for excessive GWG suggested that those in the highest quartile of neighborhood deprivation had a significantly increased risk of excessive GWG (RR: 1.20; 95% CI: 1.02, 1.39) compared to the lowest quartile. This association was attenuated and not statistically significant after adjusting for covariates (RR: 1.16; 95% CI: 0.95, 1.40).

Variation by Race/Ethnicity: Interaction by race/ethnicity was not significant for associations between PNP or neighborhood deprivation and GWG in models adjusted for all covariates. (Table 4)

Sensitivity Analyses

Perceived Neighborhood Problems: For PNP, different approaches to address dependency of GWG on gestational age produced point estimates that were generally further from the null (Appendix 2, Table A2.2). Overall similar trends were observed, but associations with excessive GWG remained significant after controlling for covariates for the group of women experiencing the highest number of perceived problems. Thus, our use of GWG adequacy presented more conservative estimates of the extent to which perceived neighborhood social environment is a risk factor for either inadequate or excessive GWG. Differences observed across these models re-emphasize the importance of using methods that best reduce the bias arising from dependency of GWG on gestational age. (45)

Sensitivity analyses using the 1990 guidelines to classify GWG also produced point estimates generally further from the null (Appendix 2 Table A2.3). This only impacted significance of findings for women experiencing the highest level of perceived problems. However, such findings must be interpreted in context: the 1990 guidelines (43) implemented a less widely used classification category for pre-pregnancy weight and did not have an upper limit for weight gain for obese women so the upper limit for overweight women was used. This limit is 2.5kg higher than the upper limit that was eventually introduced for obese women in 2009.

Neighborhood Deprivation Index: Different approaches to address dependency between GWG and gestational age did not largely impact associations between neighborhood deprivation and either inadequate or excessive GWG (Appendix 2, Table A2.4). However, sensitivity analyses using 1990 guidelines to classify GWG produced point estimates that were further from the null (Appendix 2, Table A2.5). In this analysis, increased risk of both inadequate and excessive GWG associated with higher levels of neighborhood deprivation remained significant after adjusting for covariates. These findings should be interpreted with the same caveats for the 1990 guidelines mentioned above.

DISCUSSION

In a national sample of women giving birth to singleton infants between 1992-2000, we found that neighborhood social environment, measured objectively or through perceptions, was not significantly associated with inadequate or excessive GWG after controlling for relevant covariates. These findings did not support our hypothesis that neighborhood social environment would be associated with non-adequate GWG and that perceived neighborhood environment would have a stronger association with non-adequate GWG as it is more conceptually proximal to stress appraisal along the embodiment pathway. (25)

To our knowledge, this analysis is the first to investigate both perceived and objective measures of neighborhood social environment and GWG. While none of the existing studies have measures of perceptions of neighborhood environment with which to compare our findings, many of the objectively measured neighborhood factors in these studies correlate with the constructs assessed in our measure of perceived neighborhood problems. For example, some of the specific factors included in objective measures of physical incivilities assessed by Laraia et al. (18), Vinikoor-Imler et al. (17) and Messer et al. (19), such as “vacant or burned property” and “condition of housing, yards and public spaces” are similar to the “abandoned or run-down buildings” and “people keep to themselves, don’t care about neighborhood” items in our PNP scale. Compared to these studies, our findings for PNP are similar to those of Laraia et al. (18), who did not find that physical incivilities were associated with GWG after controlling for relevant covariates. However, they contrast with findings from Vinikoor-Imler et al. (17) and Messer et al., (19) who, using the same study population, found that increased physical incivilities in the environment increased the risk of both inadequate and excessive GWG. It is difficult to further contextualize our findings in relation to these studies because although they contrast to each other in their findings, they were sampled from the same underlying target population. Specifically, both studies were conducted as a part of or as an ancillary study to the Pregnancy, Infection, and Nutrition study in North Carolina. However, the studies differed in their actual study sample (one obtained recruits from a clinic-based sample (18) and the other used birth certificate data (17,19)) and in how GWG was measured (one was unable to use IOM classifications due to lack of data on pre-pregnancy weight. (17,19)) Further explanation for the divergence of these findings was not evident. In the future, such inconsistencies among studies with just objective measures of neighborhood context will need to be resolved in order to better understand differences in perceived and objective measures of neighborhood environment in relation to GWG.

Our findings for objective neighborhood deprivation and GWG contrast with the small body of literature that has investigated this association. Mendez et al. (20), who used a similar neighborhood deprivation index to study GWG in population of 55608 Pennsylvania mothers, found that neighborhood deprivation was associated with inadequate GWG in both Black and white women, as well as with weight loss during pregnancy. In addition to differences in age and geographic location, difference in sample size is likely the main reason for our disparate findings. Mendez et al. (20) had access to over 50,000 births whereas our cohort included around 1200 births. In comparing the magnitude of association between the two studies, estimates were not qualitatively different: the adjusted relative risk for quartile 4 of deprivation and inadequate GWG in Mendez et al. was 1.2 (95% CI: 1.1, 1.2) compared to 1.36 (95% CI: 0.93, 2.01) in the current study. In fact, our point estimates actually suggested a slightly stronger association between neighborhood deprivation and inadequate GWG. Our confidence intervals, however, were wider and did not exclude the null.

We propose several explanations for the lack of association between neighborhood social environment, measured objectively or through perceptions, and GWG in our population. First, our measure of neighborhood deprivation may be too broad to capture the most relevant neighborhood characteristics. (52) Using an index, such as neighborhood deprivation, approximates the aggregate exposure to “stressors” in the neighborhood environments as many toxic neighborhood characteristics across different domains cluster in low-SES neighborhoods. (53-55) It may mask potential heterogeneity in the impact of stressors from different sources aggregated into this summary measure. (52) For example, stress from lack of a supportive services/retail environment may differ from stress resulting from lack of structural elements, such as sidewalks and parks, both of which may be limited to varying degrees in deprived neighborhoods. In fact, Vinikoor-Imler et al. (17) directly illustrate that neighborhood deprivation is not always highly correlated with other, more specific measures of neighborhood environment, such as walkability or presence of sidewalks and parks and porches. Thus, going forward, future research should continue to focus on specific domains of the neighborhood environment that shape weight gain behaviors during pregnancy.

While we theorized that perceived neighborhood problems would be more proximal to stress activation by capturing stress appraisal, we may not have observed an association because stress activation relies on both appraisal and risk assessment as well as coping resources. (25) We did not have information on individuals' direct assessment of how stressful reported neighborhood problems were or information on additional coping strategies or resources that may buffer the impact of neighborhood problems. Especially during pregnancy, women may be motivated to seek out formal and informal sources of aid that would attenuate the detrimental impact of perceived neighborhood problems on health. (53) A recent review of coping styles during pregnancy further suggests that coping style matters: avoidant or disengaged coping is associated with increased risk of adverse pregnancy outcomes. (56) Latendresse et al. (57) also found that type of coping style was associated with CRH levels in early pregnancy, further suggesting that coping style, in addition to stress appraisal, is necessary to understand the pregnancy outcomes of any external stressors occurring during this time. Furthermore, self-report of problems still may not completely capture the true biological stress activation occurring when individuals report not perceiving certain factors to be stressors. Especially in the context of long-term exposure to highly stressful environments, women may have resting stress profiles that are always "on" as a result of long-term programming, (58) even when they do not consciously perceive the environment to be problematic. This may be particularly relevant in our study population as women tended to rate their neighborhoods low on the problems scale. Even in the highest quartile of socioeconomic deprivation, almost one third of women ranked their neighborhood as having 2 or fewer problems on a scale of 0 to 14. Thus, future research should aim to include both measures of coping and biomarkers of stress to further understand neighborhoods and GWG.

Finally, neighborhoods may not be the most proximal or relevant factor shaping trends in weight gain during pregnancy. A complex constellation of macro- and meso-level environments contributes to whether a woman gains adequately, (24,25,59,60) including work, family, and medical environments. Thus, one of these other areas may be more relevant to women's weight gain during pregnancy, and truly understanding GWG may depend on understanding how these dimensions interrelate. (55,61) The way in which neighborhood environments interact with the other spatial geographies that women move through should be a future direction of investigation in order to better understand how context supports or deters optimal weight gain during pregnancy.

This study has a number of key limitations. First, our findings may not correspond to the most important geographical "unit of impact" for GWG. (52,62) Misspecification of the unit of influence may misclassify people into the wrong neighborhood 'exposure,' and thus lead to a false conclusion of lack of association (or presence of association) depending on how the neighborhood factors being investigated operate with respect to the health outcome of interest. (52,62) However, evidence from Messer et al. (19) suggests that this may not be a large concern for pregnancy outcomes. They investigated the impact of different spatial definitions of neighborhoods and found that scale did not greatly bias observed associations across a number of birth outcomes, including GWG. (19) The magnitude of the point estimates were fairly consistent across nested spatial scales, although the precision varied.

Second, our study population was restricted to women giving birth relatively late in the their reproductive life course. These women may be characteristically different from women at the beginning of their reproductive trajectory on our key exposure and outcome of interest. Older women may be more likely to live in less deprived neighborhoods and have been shown to be more likely to gain inadequately compared to their younger counterparts. (3) Furthermore, by not having access to younger women, we are not able to observe the impacts of early exposure to neighborhood environments on GWG. This may be particularly important because neighborhood deprivation experiences influencing weight gain in these early pregnancies may set women on a trajectory of weight gain patterns for their subsequent pregnancies. (5) Thus, beyond the fact that our restricted sample is likely not generalizable to women of all reproductive ages, we also may not be capturing the total impact of neighborhood deprivation at different points in the life course. The smaller sample size available within this restricted population may have increased the imbalance across exposure and covariates and limited our power to detect an effect. (52) This is a persistent concern in research on neighborhoods and health due to the strong social patterning of health and class within American society. (63) Studies including larger sample sizes across wider ranges of women's reproductive life course areß necessary to overcome, at least in part, these methodological challenges.

Use of self-reported data for the study outcome is another limitation of our study. Reporting bias in self-reported pregnancy-related weight suggests that, on average, error is small (-2.94-1.09kg for pre-pregnancy weight and -0.91-3.01kg for GWG; (64)), but varies widely by maternal characteristics and other factors. Currently, bias correction approaches for pregnancy-related weight measures do not exist and differences in magnitude of bias between weights reported during pregnancy and weights reported during non-pregnant periods (65) likely make exiting bias correction techniques created for non-pregnant populations inapplicable. More work is needed to develop appropriate bias correction techniques (e.g. regression calibration) for use in observational studies on self-reported pregnancy-related weight. (66)

Despite these limitations, our study has a number of key strengths. First, it is one of the few studies to investigate neighborhood environment and GWG in a national sample. Prior studies have been limited to only two geographic locations (a five county area around Wake County, North Carolina (17-19) and Allegheny County, Pennsylvania (20)), thus restricting the ability to determine whether observed associations are region specific. Our analysis addresses this limitation. Second, due to the longitudinal nature of the NLSY79, we were able to establish temporality between our exposure and our outcome. Many studies conducted to date are limited by their cross-sectional design. Our data indicated that while a large majority of women (61.3%) did, in fact, have a stable ranking of their neighborhood before and after pregnancy, a substantial minority (38.7%) changed the ranking of their neighborhood by at least two points before and after pregnancy. Furthermore, this is one of the only cohorts in which information on objective and perceived neighborhood environments is available along with measures of GWG and detailed sociodemographic data. Finally, we were able to control for a wide range of socioeconomic factors. Socioeconomic status is a complex construct representing multiple domains. Exclusion of any of these domains may result in incomplete control of confounding due to socioeconomic status. This concern is particularly relevant in studies assessing the role of neighborhood social and socioeconomic environment on health outcomes because observed point estimates may reflect these excluded individual level socioeconomic factors rather than an association with the neighborhood. (52)

CONCLUSION

Overall, the current study did not observe statistically significant associations between perceived or objectively measured neighborhood social environment and GWG in a national sample of women. However, the magnitude of association observed in point estimates across quartiles of neighborhood deprivation and GWG suggest that better powered analyses are warranted to further investigate this association in current, nationally representative samples. Neighborhood environments remain an important factor shaping pregnancy outcomes at large, (60,67) but interventions to improve GWG may need to additionally focus on other contextual factors in order to optimally improve maternal and child health trajectories.

REFERENCES

1. Messer LC, Boone-Heinonen J, Mponwane L, et al. Developmental programming: Priming disease susceptibility for subsequent generations. *Curr Epidemiol Rep*. 2015;2(1):37–51.

2. Herring SJ, Oken E. Weight gain during pregnancy: Importance for maternal and child health. *Ann Nestlé [Engl]*. 2010;68(1):17–28.

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

4. Amorim AR, Rossner S, Neovius M, et al. Does excess pregnancy weight gain constitute a major risk for increasing long-term BMI? *Obesity*. 2007;15:1278–1286.

5. Cohen AK, Chaffee BW, Rehkopf DH, et al. Excessive gestational weight gain over multiple pregnancies and the prevalence of obesity at age 40. *International Journal of Obesity*. 2014;38(5):714–718.

6. Diesel JC, Eckhardt CL, Day NL, et al. Gestational weight gain and the risk of offspring obesity at 10 and 16 years: A prospective cohort study in low-income women. [Epub ahead of print May 29, 2015]. *BJOG: Int J Obstet Gy*. (doi: 10.1111/1471-0528.13448).

7. Center for Disease Control and Prevention (CDC). Table 16D: 2011 Pregnancy Nutrition Surveillance. 2012; Summary of trends in maternal health indicators:1–2.

8. Deputy NP, Sharma AJ, Kim SY, et al. Prevalence and characteristics associated with gestational weight gain adequacy. *Obstet Gynecol*. 2015;125(4):773–781.

9. Pawlak MT, Alvarez BT, Jones DM, et al. The effect of race/ethnicity on gestational weight gain. *J Immigr Minor Health*. 2015;17(2):325-32.

10. Stuebe AM, Oken E, Gillman MW. Associations of diet and physical activity during pregnancy with risk for excessive gestational weight gain. *Am J Obstet Gynecol*. 2009;201:58.e1–58.e8.

11. 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.

12. Hickey CA. Sociocultural and behavioral influences on weight gain during pregnancy. *Am J Public Health*. 2000;71((suppl)):1364S–70S.

13. Webb JB, Siega-Riz AM, Dole N. Psychosocial determinants of adequacy of

gestational weight gain. *Obesity (Silver Spring)*. 2009;17(2):300-9.

14. Weisman CS, Hillemeier MM, Downs DS, et al. Preconception predictors of Weight Gain During Pregnancy. *Women's Health Issues*. 2010;20(2):126–132.

15. Sallis JF, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz K, Rimer BK, Viswanath K, eds. *Health Behavior and Health Education: Theory, Research, and Practice*. Jossey-Bass; 2008:465–486.

16. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci*. 2010;1186:125–145.

17. Vinikoor-Imler LC, Messer LC, Evenson KR, et al. Neighborhood conditions are associated with maternal health behaviors and pregnancy outcomes. *Soc Sci Med*. 2011;73(9):1302–1311.

18. Laraia B, Messer L, Evenson K, et al. Neighborhood factors associated with physical activity and adequacy of weight gain during pregnancy. *J Urban Health*. 2007;84(6):793–806.

19. Messer LC, Vinikoor-Imler LC, Laraia BA. Conceptualizing neighborhood space: Consistency and variation of associations for neighborhood factors and pregnancy health across multiple neighborhood units. *Health Place*. 2012;18(4):805–813.

20. Mendez DD, Doebler DA, Kim KH, et al. Neighborhood Socioeconomic Disadvantage and Gestational Weight Gain and Loss. *Matern Child Health J*. 2014;18(5):1095-103.

21. Elo I, Mykyta L, Margolis R, et al. Perceptions of neighborhood disorder: The role of individual and neighobrhood characteristics. *Social Science Quarterly*. 2009;90(5):1298–1320.

22. Boehmer TK, Hoehner CM, Deshpande AD, et al. Perceived and observed neighborhood indicators of obesity among urban adults. *International Journal of Obesity*. 2007;31(6):968–977.

23. Bailey EJ, Malecki KC, Englelman CD, et al. Predictors of discordance between perceived and objective neighborhood data. *Annals of Epidemiology*. 2014;24(3):214–221.

24. Culhane JF, Elo IT. Neighborhood context and reproductive health. *American Journal of Obstetrics and Gynecology*. 2005;192(5):S22–S29.

25. Davis EM, Stange KC, Horwitz RI. Childbearing, stress and obesity disparities in women: A Public Health Perspective. *Matern Child Health J*. 2012;16(1):109–118.

26. Powell-Wiley TM, Ayers CR, de Lemos JA, et al. Relationship between perceptions about neighborhood environment and prevalent obesity: data from the dallas heart study. *Obesity*. 2013;21(1):E14–E21.

27. Mendez DD, Burke JG, Jones JR, Salter C. Perspectives From Community-Based Doulas and Mothers: Neighborhood Context and Pregnancy. *Journal of Health Disparities Research and Practice.* 2014;7(4):91–106.

28. Schetter CD. Stress Processes in Pregnancy and Preterm Birth. *Curr Dir Psychol*. 2009;18(4):205–209.

29. McDonell JR. Neighborhoods and Families. In: *Children's Well-Being: Indicators and Research*. Dordrecht: Springer Netherlands; 2010:55–73.

30. Rosenthal SR, Vivier PM, Rogers ML, et al. Measures of neighborhood quality: Self-reports of mothers of infant children. *J Child Fam Stud*. 2015;24:1256-61.

31. Williams DR, Collins C. Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Rep*. 2001;116(5):404–416.

32. Kramer MR, Cooper HL, Drews-Botsch CD, et al. Do measures matter? Comparing surface-density-derived and census-tract-derived measures of racial residential segregation. *Int J Health Geogr*. 2010;9(1):29.

33. Bureau of Labor Statistics. *National Longitudinal Survey of Youth*. Washington, DC: US Department of Labor.1979.

34. Bureau of Labor Statistics. *National Longitudinal Survey of Youth, 1979*. Washington, DC: US Department of Labor. 1986.

35. Oken E, Kleinman KP, Rich-Edwards J, et al. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatr*. 2003;3:6.

36. US Census Bureau. Geographic Terms and Concepts-Census Tract. *www.census.gov*. (https://www.census.gov/geo/reference/gtc/gtc_ct.html). (Accessed August 26, 2014)

37. Tatian PA. Appendix J: Description of tract remapping methodology. *www.geolytics.com/pdf/NCDB-LF-Data-Users-Guide.pdf*. (http://www.geolytics.com/pdf/NCDB-LF-Data-Users-Guide.pdf). (Accessed August 26, 2014)

38. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health*. 2006;83(6):1041–1062.

39. Escarse JJ, Lurie N, Jewell A. RAND Center for Population Health and Health Disparities (CPHHD) Data Core Series: Decennial Census Abridged, 1990-2010 [United States].

40. Bodnar LM, Hutcheon JA, Platt RW, et al. Should gestational weight gain recommendations be tailored by maternal characteristics? *Am J Epidemiol*. 2011;174(2):136–146.

41. Siega-Riz AM, Hobel C. Predictors of poor maternal weight gain from baseline anthropometric, psychosocial, and demographic information in a Hispanic Population. *J am Diet Assoc*. 1997;97(11):1264–1268.

42. Siega-Riz AM, Adair LS, Hobel C. Maternal underweight status and inadequate rate of weight gain during the third trimester of pregnancy increases the risk of preterm delivery. *J. Nutr.* 1996;126:146–153.

43. Institute of Medicine, ed. Nutrition During Pregnancy: Part I-Weight Gain. Washington, DC: National Academies Press (US); 1990.

44. Hutcheon JA, Bodnar LM, Joseph KS, et al. The bias in current measures of gestational weight gain. *Paediatric and Perinatal Epidemiology*. 2012;26(2):109–116.

45. Hutcheon JA, Platt RW, Abrams B, et al. A weight-gain-for-gestational-age z score chart for the assessment of maternal weight gain in pregnancy. *Am J Public Health*. 2013;97(5):1062-7.

46. Fleischer NL, Roux AVD. Using directed acyclic graphs to guide analyses of neighbourhood health effects: An introduction. *Journal of Epidemiology & Community Health*. 2008;62(9):842–846.

47. US Census Bureau. Geographic Terms and Concepts- Census Divisions and Census Regions. *www.census.gov*. (https://www.census.gov/geo/reference/gtc/gtc_census_divreg.html). (Accessed August 26, 2014)

48. Rehkopf DH, Krieger N, Coull B, et al. Biologic risk markers for coronary heart disease. *Epidemiology*. 2010;21:38–46.

49. Zou G. A Modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*. 2004;159(7):702–706.

50. Greenland S. Interactions in epidemiology: Relevance, identification, and estimation. *Epidemiology*. 2009;20(1):14–17.

51. StataCorp. Stata multiple-imputation reference manual. College Station, TX: StataCorp LP; 2013.

52. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health*. 2001;55:111–122.

53. Morenoff JD. Neighborhood mechanisms and the spatial dynamics of birth weight. *Am J Sociol*. 2003;108(5):976–1017.

54. Sampson RJ. Neighborhood-level context and health: Lessons from Sociology. In: Kawachi I, Berkman LF, eds. *Neighborhoods and Health*. New York: Oxford Univ Press; 2003:132–178.

55. Macintyre S, Ellaway A, Cummins S. Place effects on health: How can we conceptualise, operationalise and measure them? *Soc Sci Med*. 2002;55:125–139.

56. Guardino CM, Dunkel-Schetter C. Coping during pregnancy: A systematic review and recommendations. *Health Psychology Review*. 2014;8(1):70–94.

57. Latendresse G, Ruiz RJ. Maternal coping style and perceived adequacy of income predict CRH levels at 14-20 weeks of gestation. *Biological Research For Nursing*. 2010;12(2):125–136.

58. McEwen B. Protective and damaging effects of stress mediators: Central role of the brain. *Dialogues in Clinical Neuroscience*. 2006;8(4):367–381.

59. Devine CM, Bove CF, Olson CM. Continuity and change in women's weight orientations and lifestyle practices through pregnancy and the postpartum period: The influence of life course trajectories and transitional events. *Soc Sci Med*. 2000;50(4):567–582.

60. Vos AA, Posthumus AG, Bonsel GJ, et al. Deprived neighborhoods and adverse perinatal outcome: a systematic review and meta-analysis. *Acta Obstet Gynecol Scand*. 2014;93(8):727–740.

61. Cummins S, Curtis S, Diez Roux AV, et al. Understanding and representing "place" in health research: A relational approach. *Soc Sci Med*. 2007;65(9):1825–1838.

62. Mujahid MS, Roux AVD. Neighborhood Factors in Health. In: Steptoe A (ed.). *Handbook of Behavioral Medicine: Methods and Applications*. New York, NY: Springer New York; 2010:341–354.

63. Williams DR, Collins C. US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology*. 1995;21:349–386.

64. Headen I, Cohen AK, Abrams B. The accuracy of self-reported pregnancy-related weight: A systematic review. (Unpublished Manuscript).

65. 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.

66. 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.

67. Metcalfe A, Lail P, Ghali WA, et al. The association between neighbourhoods and adverse birth outcomes: A systematic review and meta-analysis of multi-level studies. *Paediatric and Perinatal Epidemiology*. 2011;25(3):236–245.

FIGURES AND TABLES

Figure 1. Observations Included in Final Analytic Sample from the 1979 National Longitudinal Survey of Youth Based on Inclusion and Exclusion Criteria

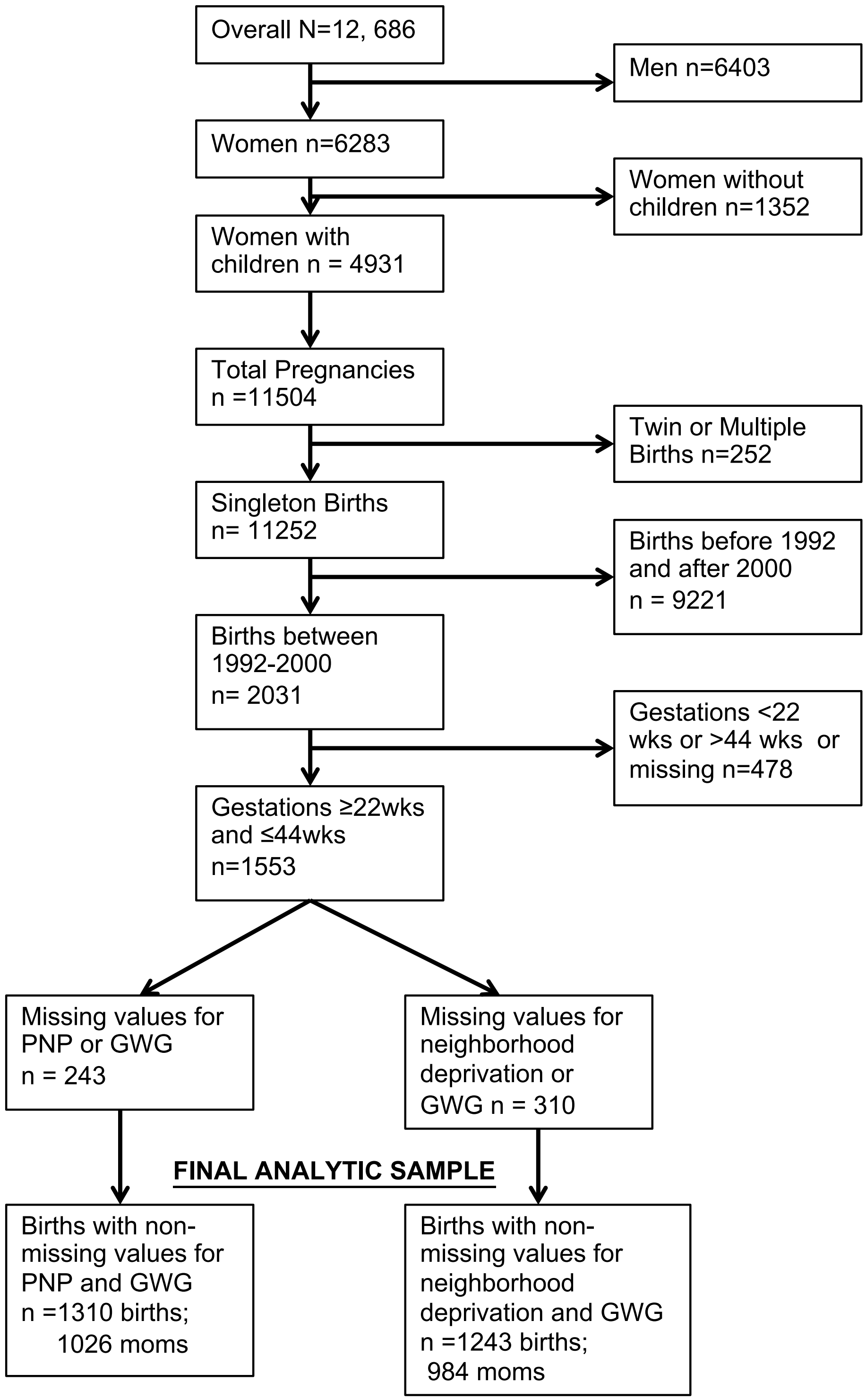


Table 1. Institute of Medicine Gestational Weight Gain Recommendations

	Overall Weight Gain (kg)	Weight Gain by Trimester	
		1st Trimester Total Gain	2nd and 3rd Trimester Rate of Gain (kg/week)
2009 Recommendations			
Underweight (<18.5kg/m^2)	12.5-18	2	0.51
Normal Weight (18.5-24.9 kg/m^2)	11.5-16	2	0.42
Overweight (25.0-29.9 kg/m^2)	7-11.5	1	0.28
Obese (≥30.0kg/m^2)	5-9	0.5	0.22
1990 Recommendations			
Underweight (<19.8kg/m^2)	12.5-18	n.s.*	0.50
Normal Weight (19.8-26.0 kg/m2)	11.5-16	n.s.*	0.40
Overweight (>26.0-29.0kg/m2)	7-11.5	n.s.*	0.30
Obese (>29.0kg/m2)	≥7	n.s.*	n.s.

*Estimates for first trimester were not specified (n.s.), the IOM report mainly relied on overall growth charts spanning the duration of pregnancy and thus focused on rates per week.

Table 2. Survey Weighted Descriptive Statistics by Gestational Weight Gain Adequacy (GWG)[†]

	Un-weighted N	Survey weighted descriptive statistics	Inadequate GWG (n=245; 17.4%[¶])	Adequate GWG (n=402; 33.5%[¶])	Excessive GWG (n= 663; 49.1%[¶])	P-value
Perceived Neighborhood Problems (Mean (SD))	1310	2.2 (2.6)[¥]	2.3 (2.8)[¥]	2.0 (2.3)[¥]	2.4 (2.8)[¥]	0.14
No Problems	383	33.9%	33.3%	34.3%	32.3%	0.21
Low Problems (1-2)	440	34.9%	35.1%	35.6%	33.2%	
Medium Problems (3-4)	203	14.8%	13.9%	18.0%	14.9%	
High Problems (5-14)	284	16.4%	17.7%	12.2%	19.6%	
Neighborhood Deprivation Index (Mean (SD))	1126	16.5 (10.5)[¥]	18.2 (11.7)[¥]	15.2 (8.8)[¥]	16.47 (11.0)[¥]	**<0.01**
Quartile 1 (0.9-11.4)	321	37.2%	30.7%	41.0%	37.1%	0.16
Quartile 2 (11.5-18.0)	297	29.7%	28.7%	30.4%	29.6%	
Quartile 3 (18.1-28.8)	273	21.8%	25.1%	21.0%	21.3%	
Quartile 4 (28.9-76.2)	235	11.3%	15.6%	7.7%	12.1%	
Race/Ethnicity*						
White	614	83.4%	76.6%	88.5%	82.6%	**<0.0001**
Black	236	11.0%	16.8%	7.4%	11.3%	
Latina	176	5.6%	6.6%	4.2%	6.1%	
Foreign born*						
No	951	96.1%	96.3%	96.5%	95.7%	0.87
Yes	75	3.9%	3.7%	3.5%	4.3%	
Marital Status						**<0.01**
No	377	21.9%	24.9%	15.5%	25.2%	
Yes	924	78.1%	75.1%	84.5%	74.8%	
Employment						
Unemployed	333	24.0%	28.0%	24.6%	22.1%	0.37
Part-Time Employed	274	21.8%	16.8%	22.6%	23.1%	

Table 2. Survey Weighted Descriptive Statistics by Gestational Weight Gain Adequacy (GWG)[‡]-CONT'D

		Survey weighted descriptive statistics				
	Un-weighted N		Inadequate GWG (n=245; 17.4%[¶])	Adequate GWG (n=402; 33.5%[¶])	Excessive GWG (n= 663; 49.1%[¶])	P-value
Full-time Employed	682	54.2%	55.2%	52.8%	54.8%	
Region of Residence						
Northeast	239	21.7%	23.1%	19.6%	22.7%	0.63
North Central	311	28.8%	28.2%	29.4%	28.6%	
South	435	30.5%	30.8%	34.2%	27.8%	
West	302	19.0%	17.9%	16.8%	21.0%	
Home Ownership						
No	594	37.3%	34.4%	36.9%	38.6%	0.65
Yes	704	62.7%	65.5%	63.1%	61.4%	
Child's Birth year (Mean (SD))	1310	1995 (2.4)[¥]	1995 (2.4)[¥]	1995 (2.2)[¥]	1995 (2.4)[¥]	0.21
Education						
<12 yrs	113	5.0%	4.9%	3.8%	5.8%	**0.03**
12-15 yrs	771	55.3%	59.4%	48.9%	58.3%	
≥16 yrs	417	39.7%	35.7%	47.3%	35.9%	
Equivalized Income (Mean (SD))	1227	10.4 (0.9)[¥]	10.3 (1.0)[¥]	10.5 (0.9)[¥]	10.3 (0.9)[¥]	**0.01**
Quartile 1	261	14.1%	20.3%	10.9%	14.1%	0.06
Quartile 2	311	23.4%	20.2%	22.2%	25.4%	
Quartile 3	311	28.0%	26.4%	26.8%	29.3%	
Quartile 4	344	34.5%	33.1%	40.0%	31.2%	
Household Income	1310	$62,205 (52,880)[¥]	$57,450 (49203)[¥]	$71,094 (58246)[¥]	$57,830 (48592)[¥]	**<0.01**
Family Size	1301	3.1 (1.3)[¥]	3.2 (1.4)[¥]	3.2 (1.3)[¥]	2.9 (1.3)[¥]	**0.04**
Maternal Age	1310	33.0 (2.9)[¥]	33.5 (3.0)[¥]	33.0 (2.9)[¥]	32.9 (2.8)[¥]	0.10
25-29 yrs	165	12.4%	11.0%	14.1%	11.7%	0.13
30-34 yrs	707	55.2%	48.5%	55.5%	57.3%	
35-39 yrs	418	31.0%	39.3%	28.1%	30.0%	
≥40 yrs	20	1.5%	1.3%	2.3%	0.9%	

Table 2. Survey Weighted Descriptive Statistics by Gestational Weight Gain Adequacy (GWG)‡-CONT'D

		Survey weighted descriptive statistics				
	Un-weighted N		Inadequate GWG (n=245; 17.4%¶)	Adequate GWG (n=402; 33.5%¶)	Excessive GWG (n= 663; 49.1%¶)	P-value
Parity		1.2 (1.2)¥	1.4 (1.3)¥	1.3 (1.2)¥	1.2 (1.2)¥	0.08
0	389	32.4%	24.7%	31.1%	36.1%	0.14
1	414	33.0%	35.8%	33.2%	31.8%	
2	273	19.9%	19.7%	21.6%	18.8%	
3	143	10.1%	14.7%	8.9%	9.3%	
≥4	91	4.6%	5.0%	5.2%	3.9%	
Moved in birth year						
No	853	73.9%	75.1%	76.0%	71.9%	0.48
Yes	328	26.1%	24.9%	24.0%	28.1%	
Length of residence (Mean (SD))	1226	2.3 (2.7)¥	2.4 (3.1)¥	2.4 (2.6)¥	2.1 (2.5)¥	0.28
Long term resident (≥ 5years)						
No	1,034	84.9%	83.2%	84.5%	85.9%	0.66
Yes	192	15.1%	16.8%	15.6%	14.1%	
Cumulative time moved (Mean (SD))	1309	3.1 (2.1)¥	3.0 (2.1)¥	2.8 (2.0)¥	3.2 (2.2)¥	**0.04**
Moved a lot (≥5 times)						
No	1,004	76.3%	78.3%	79.8%	73.2%	0.12
Yes	305	23.7%	21.7%	20.2%	26.8%	
Rural Residence						
No	1,063	81.2%	77.6%	81.8%	82.2%	0.44
Yes	217	18.8%	22.4%	18.2%	17.9%	

†N's for this table taken from the complete case sample for perceived neighborhood problems (n= 1310). The complete case sample for neighborhood deprivation was smaller (n-1243)
¥ Cell reports mean and standard deviation for variable reported
* N's based on mothers in the analytic sample, not births

Table 3. Survey Weighted Descriptive Statistics by Perceived Neighborhood Problems and Neighborhood Deprivation

	No Problems (score= 0; n=383)	Low Problems (score 1-2; n=440	Medium Problems (score 3-4; n=203)	High Problems (score 5-14; n=284)	P-value	Quartile 1 (score 0.9-11; n=360)	Quartile 2 (score 12-18; n=324)	Quartile 3 (score 18.1-29; n=300)	Quartile 4 (score 29.1-76; n=261)	P-value
Perceived Neighborhood Problems (Mean (SD))	-	-	-	-		1.6 (2.0)[¥]	2.1 (2.3)[¥]	2.8 (2.9)[¥]	4.6 (4.5)[¥]	**<0.001**
No Problems	-	-	-	-		42.2%	30.0%	27.0%	11.4%	**<0.001**
Low Problems	-	-	-	-		36.7%	39.0%	29.4%	21.7%	
Medium Problems	-	-	-	-		12.8%	16.8%	20.5%	16.3%	
High Problems	-	-	-	-		8.4%	14.2%	23.2%	50.6%	
Neighborhood Deprivation Index (Mean (SD))	13.3 (7.1)[¥]	15.2 (9.2)[¥]	17.7 (9.5)[¥]	25.4 (16.8)[¥]	**<0.001**	-	-	-	-	
Quartile 1	49.4%	40.0%	29.7%	17.2%	**<0.001**	-	-	-	-	
Quartile 2	28.0%	33.9%	31.1%	23.3%		-	-	-	-	
Quartile 3	18.6%	18.8%	27.8%	28.0%		-	-	-	-	
Quartile 4	4.1%	7.2%	11.5%	31.6%		-	-	-	-	
Gestational Weight Gain										
Inadequate	17.5%	17.8%	15.4%	18.3%	0.21	14.1%	17.6%	22.6%	23.7%	**0.05**
Adequate	34.6%	34.7%	38.2%	24.3%		34.7%	33.2%	28.7%	22.0%	
Excessive	47.9%	47.5%	46.5%	57.4%		51.2%	49.3%	48.7%	54.4%	
Race/Ethnicity										
White	49.4%	54.7%	46.1%	61.7%	**<0.001**	95.3%	88.9%	81.2%	34.8%	**<0.001**
Black	35.2%	26.1%	41.4%	15.0%		2.1%	6.5%	12.5%	49.1%	

Table 3. Survey Weighted Descriptive Statistics by Perceived Neighborhood Problems and Neighborhood Deprivation--CONTINUED

	No Problems (score= 0; n=383)	Low Problems (score 1-2; n=440	Medium Problems (score 3-4; n=203)	High Problems (score 5-14; n=284)	P-value	Quartile 1 (score 0.9-11; n=360)	Quartile 2 (score 12-18; n=324)	Quartile 3 (score 18.1-29; n=300)	Quartile 4 (score 29.1-76; n=261)	P-value
Latina	15.4%	19.2%	12.5%	23.3%		2.6%	4.5%	6.3%	16.0%	
Foreign born										
No	96.7%	96.1%	97.7%	94.4%	0.26	95.9%	96.9%	96.9%	92.5%	0.20
Yes	3.3%	3.9%	2.3%	5.6%		4.1%	3.1%	3.1%	7.5%	
Marital Status										
No	13.4%	17.6%	26.8%	42.6%	**<0.001**	10.4%	22.0%	30.4%	56.9%	**<0.001**
Yes	86.6%	82.4%	73.3%	57.4%		89.6%	78.0%	69.6%	43.1%	
Employment										
Unemployed	27.7%	18.9%	24.0%	26.8%	0.16	20.9%	21.3%	30.8%	33.6%	**<0.001**
Part-Time Employed	18.1%	24.7%	22.6%	22.5%		23.5%	21.4%	12.0%	31.1%	
Full-time Employed	54.2%	56.3%	53.4%	50.7%		55.5%	57.3%	57.2%	35.2%	
Region of Residence										
Northeast	20.1%	22.3%	23.0%	22.4%	0.14	26.5%	19.1%	15.6%	21.4%	0.27
North Central	30.7%	32.8%	27.0%	18.8%		30.2%	32.7%	26.0%	21.5%	
South	30.8%	27.9%	33.0%	32.5%		24.8%	27.1%	35.2%	34.0%	
West	18.4%	16.9%	17.1%	26.4%		18.4%	21.1%	23.2%	23.2%	
Home Ownership										
No	25.3%	35.6%	43.3%	58.7%	**<0.001**	28.5%	38.2%	43.1%	68.0%	**<0.001**
Yes	74.7%	64.4%	56.7%	41.3%		71.5%	61.9%	56.9%	32.0%	

Table 3. Survey Weighted Descriptive Statistics by Perceived Neighborhood Problems and Neighborhood Deprivation--CONTINUED

	No Problems (score= 0; n=383)	Low Problems (score 1-2; n=440	Medium Problems (score 3-4; n=203)	High Problems (score 5-14; n=284)	P-value	Quartile 1 (score 0.9-11; n=360)	Quartile 2 (score 12-18; n=324)	Quartile 3 (score 18.1-29; n=300)	Quartile 4 (score 29.1-76; n=261)	P-value
Child's Birth year (Mean (SD))	1995 (2.3)[¥]	1995 (2.2)[¥]	1995 (2.3)[¥]	1995 (2.7)[¥]	0.90	1995 (2.1)[¥]	1995 (2.3)[¥]	1995 (2.4)[¥]	1995 (3.4)[¥]	0.29
Education										
<12 yrs	4.2%	3.2%	5.1%	9.9%	**<0.001**	0.8%	4.7%	9.3%	16.2%	**<0.001**
12-15 yrs	54.3%	49.7%	60.1%	64.3%		41.8%	60.2%	67.6%	71.0%	
≥16 yrs	41.5%	47.1%	34.8%	25.8%		57.5%	35.0%	23.2%	12.8%	
Equivalized Income (Mean (SD))	10.5 (0.8)[¥]	10.5 (0.9)[¥]	10.2 (1.0)[¥]	10.0 (1.1)[¥]	**<0.001**	10.8 (0.7)[¥]	10.3 (0.7)[¥]	9.9 (1.0)[¥]	9.5 (1.2)[¥]	**<0.001**
Quartile 1	8.8%	10.8%	16.8%	29.3%	**<0.001**	04.1%	10.4%	26.9%	45.4%	**<0.001**
Quartile 2	21.1%	20.1%	28.0%	31.0%		14.9%	29.0%	31.1%	33.7%	
Quartile 3	29.2%	29.8%	32.0%	17.7%		23.8%	41.3%	27.8%	13.3%	
Quartile 4	40.9%	39.3%	23.3%	22.1%		57.1%	19.3%	14.2%	07.7%	
Maternal Age										
25-29 yrs	12.8%	11.8%	12.8%	12.3%	0.46	9.6%	11.8%	18.0%	19.8%	**0.02**
30-34 yrs	53.3%	58.9%	53.8%	52.6%		56.0%	60.4%	53.3%	45.1%	
35-39 yrs	32.5%	29.0%	30.8%	32.3%		33.0%	26.0%	28.4%	32.9%	
≥40 yrs	1.4%	0.3%	2.6%	2.8%		1.4%	1.8%	0.3%	2.3%	
Parity										
0	29.6%	37.2%	31.7%	29.0%	**0.01**	37.5%	32.9%	27.2%	21.8%	**<0.001**
1	38.6%	33.1%	24.3%	29.9%		34.4%	36.0%	28.3%	20.9%	
2	18.3%	18.6%	26.6%	19.6%		20.2%	14.9%	27.1%	21.4%	
3	9.9%	7.8%	11.7%	14.0%		5.7%	10.4%	13.4%	21.4%	
≥4	3.6%	3.4%	5.8%	7.5%		2.2%	5.8%	4.0%	14.5%	

Table 3. Survey Weighted Descriptive Statistics by Perceived Neighborhood Problems and Neighborhood Deprivation--CONTINUED

	No Problems (score= 0; n=383)	Low Problems (score 1-2; n=440	Medium Problems (score 3-4; n=203)	High Problems (score 5-14; n=284)	P-value	Quartile 1 (score 0.9-11; n=360)	Quartile 2 (score 12-18; n=324)	Quartile 3 (score 18.1-29; n=300)	Quartile 4 (score 29.1-76; n=261)	P-value
Moved										
No	74.7%	76.6%	71.2%	69.4%	0.39	75.8%	70.7%	77.0%	68.4%	0.27
Yes	25.3%	23.4%	28.8%	30.6%		24.2%	29.3%	23.0%	31.6%	
Length of residence (Mean (SD))	2.3 (2.4)[¥]	2.1 (2.3)[¥]	2.2 (2.6)[¥]	2.5 (3.8)[¥]	0.54	2.1 (2.0)[¥]	2.0 (2.2)[¥]	2.2 (2.8)[¥]	2.7 (4.9)[¥]	0.13
Long term resident (≥ 5years)										
No	82.7%	88.3%	86.0%	81.4%	0.19	86.3%	87.9%	86.8%	80.1%	0.21
Yes	17.3%	11.7%	14.0%	18.6%		13.7%	12.1%	13.2%	19.9%	
Cumulative time moved (Mean (SD))	3.0 (2.0)[¥]	3.0 (1.9)[¥]	3.1 (2.2)[¥]	3.2 (2.7)[¥]	0.92	3.2 (1.9)[¥]	3.2 (1.9)[¥]	3.1 (2.4)[¥]	2.8 (3.1)[¥]	0.36
Moved a lot (≥5 times)										
No	78.4%	78.3%	73.8%	70.3%	0.21	76.5%	75.4%	70.9%	83.3%	0.17
Yes	21.6%	21.7%	26.2%	29.7%		23.5%	24.6%	29.1%	16.8%	
Rural Residence										
No	82.2%	82.4%	78.6%	79.5%	0.76	91.8%	78.0%	67.8%	84.4%	**<0.001**
Yes	17.8%	17.6%	21.4%	20.5%		8.2%	22.1%	32.2%	15.6%	

[†]N's for this table taken from the complete case sample for perceived neighborhood problems (n= 1310). The complete case sample for neighborhood deprivation was smaller (n=1243)

[¥] Cell reports mean and standard deviation for variable reported

Table 4. Crude and Adjusted Associations Between Perceived Neighborhood Problems, Neighborhood Deprivation and Gestational Weight Gain (GWG)

	Perceived Neighborhood Problems					Neighborhood Deprivation Index				
	Crude		Adjusted†		Wald test for interaction§	Crude		Adjusted†		Wald test for interaction§
	RR	95% CI	RR	95% CI	p-value	RR	95% CI	RR	95% CI	p-value
Inadequate GWG					0.22					0.36
Low Problems/ Quartile 2	1.01	0.75, 1.36	1.03	0.76, 1.38		1.20	0.86, 1.67	1.12	0.81, 1.54	
Medium Problems/ Quartile 3	0.86	0.57, 1.28	0.74	0.50, 1.11		**1.52**	**1.11, 2.10**	1.32	0.93, 1.87	
High Problems/ Quartile 4	1.28	0.94, 1.75	1.09	0.78, 1.51		**1.80**	**1.33, 2.44**	1.36	0.93, 2.01	
Excessive GWG					0.55					0.69
Low Problems/ Quartile 2	1.00	0.86, 1.16	1.02	0.87, 1.21		1.00	0.85, 1.18	0.97	0.82, 1.15	
Medium Problems/ Quartile 3	0.95	0.78, 1.15	0.89	0.72, 1.10		1.06	0.90, 1.24	1.01	0.84, 1.21	
High Problems/ Quartile 4	**1.21**	**1.02, 1.44**	1.16	0.95, 1.40		**1.20**	**1.02, 1.39**	1.15	0.93, 1.42	

†Adjusted for race/ethnicity, immigrant status, marital status, maternal age, parity, birth year of the child, region of residence, employment, equivalized income, education, home ownership, moving in the birth year, length of residence, cumulative times moved, and rural residence

§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

CHAPTER 4. THE ACCURACY OF SELF-REPORTED, PREGNANCY-RELATED WEIGHT: A SYSTEMATIC REVIEW

INTRODUCTION

While anthropometrically-measured weight is ideal when investigating the impact of weight on health outcomes, many studies must still rely on self-reported weight, especially in populations of pregnant women. (1) A thorough investigation of maternal weight during the childbearing year requires data on weight and height at conception (pre-pregnancy), weight gain during pregnancy (usually studied as the total amount gained between the delivery weight and pre-pregnancy weight) and weight at some point after women have recovered from the pregnancy (postpartum weight, often measured at 6-12 months after delivery). (2) Pre-pregnancy weight is almost universally based on maternal recall, since there are few opportunities to measure women's weight before they conceive. (2) Weight during pregnancy, at delivery and postpartum is more easily measured. Ideal weight measurement assesses weight using a standardized procedure, performed by trained personnel, who use a calibrated scale to weigh subjects wearing light-weight clothing and no shoes. (3) However, lack of resources in some studies may preclude ideal weight measurement, thus, many studies rely on maternal self-reported weight obtained through survey or interview techniques. Additionally, since the 2003 revision of the birth certificate to include pre-pregnancy weight, (4) there has been a renewed call (2,5) to use birth certificates, which often rely on women's self-report for studies of maternal weight and birth outcomes. However, as with most self-reported measures, self-reported weight may be subject to reporting error, which can introduce bias if it is non-randomly distributed within the population and/or differential by outcome. (6) Thus, it is important to understand the magnitude, direction and impacts of such error on observed associations between pregnancy-related weight and pregnancy outcomes.

In non-pregnant populations of women of childbearing age, self-reported and measured weights are highly correlated, but women tend to underreport their weight. (7-9) The magnitude of the reporting error ranges from 0.2kg to 3.54kg. (7-9) Studies aimed at understanding this error, and how it varies by different demographic characteristics, have contributed to the development of correction techniques such as regression calibration (10-12) and probabilistic bias correction. (6) However, the characteristics of reporting error may be different for women recalling weights pertaining to their pregnancy. Pregnancy is one of the few periods in a woman's life defined by rapid, intentional weight gain, significant life changes across a short period of time, and increased interaction with the medical context, all of which may impact the accuracy with which a woman recalls her weight during this period, even after they are no longer pregnant. (13) Furthermore, the magnitude of error may differ depending on the time during pregnancy that the weight is reported or recalled: during the beginning of pregnancy, at the end of pregnancy, or during the postpartum period.

While the magnitude and impact of reporting error has been much less investigated in pregnant women, enough literature exists to inform two non-systematic reviews on reporting error in relation to pregnancy-related weight measures. (10,14) The first focuses on error surrounding different techniques used to measure weight during pregnancy, (10) while the second focuses only on pre-pregnancy weight, not extending their review to understand error around gestational weight gain (GWG), delivery weight, or postpartum weight. (14) However, a number of key studies have been published since these reviews were conducted, indicating the need for an updated, systematic review.

In this study, we systematically review observational studies, validation studies, and clinical interventions and trials in order to better characterize accuracy of self-reported pregnancy-related weight through the following four objectives. First, we aim to document the size and direction of reporting error due to self-report of weight by pregnant women or women who are not currently pregnant, but are reporting on weights occurring during pregnancy. We focus on self-reported weight at four key time points during pregnancy: pre-pregnancy, delivery, weight gain over gestation, and weight retention after delivery. Second, we will assess variation in such error across maternal demographic and health characteristics as well as the potential impact that reporting error has on misclassification into clinically relevant weight categories. Third, we will assess the impact of observed reporting error on estimated associations between pregnancy-related weight and birth or pregnancy outcomes. Fourth, we will conclude with a discussion of methodological limitations facing this body of literature and recommendations for future researchers in assessing validity of these self-reported measures.

METHODS

We identified relevant articles through the MEDLINE database and the Google Scholar search engines. In particular, we found Google Scholar a useful complement to the MEDLINE database because Google Scholar searches the full text of articles for key words and scans a wider array of web content to identify relevant articles. (15) Google Scholar also automatically includes variants of search terms in the search results (i.e. “recalled” also searches for “recall,” “recalling,” ect.). (16) This was useful in identifying some articles that reported relevant information related to accuracy of self-reported pregnancy-related weight in their methods, results, or discussion sections, but did not highlight this information as a main finding or mention it in the abstract. These papers would have been excluded in the traditional PubMed abstract review approach often used when conducting literature reviews.

To be included in this systematic review, a study had to compare a maternal-reported measure of pregnancy-related weight (pre-pregnancy weight, GWG, delivery weight, and/or postpartum weight) with an anthropometric or medically reported measure of that weight. Medical or clinical reports are often used as reference points for assessing accuracy of self-reported health because they are thought to reflect the medical procedures conducted by trained personnel during an appointment. (17) However, for weight, they often instead rely on self-reports by patients to the medical provider. We wanted to capture reporting error based on both sources of reference in order to assess whether differences in findings would emerge. All study designs were included in our review. We also did not restrict the inclusion of studies based on method of height measurement, although this is also a potential source of bias in assessing the validity of body mass index (BMI) for pregnant women. Self-reported height, similar to self-reported weight, has been show in non-pregnant populations to have some error associated with it, (7-9) and, ideally, height in women during pregnancy would be measured for inclusion in BMI calculations. However, height is not likely to vary much over pregnancy, suggesting that errors found in non-pregnant populations likely apply to pregnant populations as well. Thus, assessing error introduced into height measurement by self-report was considered to be beyond the scope of the present review. Finally, we further limited our included studies to English language only.

In MEDLINE, we executed the following queries using MeSH terms: “pregnancy”[MeSH] AND “body weight”[MeSH] AND “self report”[MeSH], “pregnancy”[MeSH] AND “body weight”[MeSH] AND “mental recall”[MeSH], and “pregnancy”[MeSH] AND “body weight”[MeSH] AND “reproducibility of results”[MeSH]. We also did MEDLINE searches without MeSH terms, using the following phrases: gestational weight gain accurate, pre-pregnancy weight accurate, postpartum weight accurate, and BMI pregnancy accurate. We completed the MEDLINE searches on September 18, 2013 -March 31, 2014. (Figure 1) In Google Scholar, we used the following queries to identify articles: “self-reported postpartum weight accuracy”, “self-reported pregnancy delivery weight accuracy”, “self-reported pregnancy-related weight accuracy”, “self-reported pre-pregnancy weight bias”, “self-reported postpartum weight bias”, “self-reported pregnancy delivery weight bias”, “self-reported pregnancy-related weight bias”, “recalled postpartum weight bias”, “recalled pregnancy-related weight bias”, “self-reported gestational weight gain accuracy”, “self-reported gestational weight gain bias”, and “recalled gestational weight gain bias”. These searches were completed August 29-31, 2013. We searched results from Google Scholar queries in the order provided (sorted by citation frequency; (15)) until 50 results in a row failed to provide any relevant article based on title and abstract review. We used this technique to mitigate the fact that the number of articles returned for each query ranged from 2,780- 28,800. (Figure 1) In both searches, we additionally reviewed reference lists of selected articles to further identify relevant papers that may have been missed in our search technique. Five additional studies were identified through this method, two of which met all inclusion criteria. Title review was used to assess initial inclusion of studies based on inclusion/exclusion criteria. This was followed by abstract review for studies included based on title content. Studies that were selected as relevant through both of these mechanisms were then included for full text review.

During full text review, we developed a data extraction form to extract information on self-reported weight and medically recorded or anthropometrically measured weight at the above specified times during pregnancy. All available measures of agreement were collected, including magnitude of reporting error (i.e. mean difference between self-reported measure and “gold-standard” measure), direction of reporting error, variability of reporting error (standard deviation and/or range), correlations between self-reported and “gold standard” weight, percent agreement, misclassification based on relevant categories (if available), and sensitivity and specificity. We further abstracted this information by any demographic or health status characteristic in order to determine how reporting error varied across these subgroups. Information on length of recall and timing of "gold standard" measurement were also collected. We abstracted any information included in studies on the impact of reporting error or misclassification on estimated associations between pregnancy-related weight and birth outcomes. Information on source of "gold standard" weight reference was collected in order to assess bias in included studies. A true gold standard weight reference was defined as the use of measured weight at all times during pregnancy. As discussed by the IOM, (2,18) for pre-pregnancy weight, this should be based on measured weight at the preconceptional visit. For delivery weight this should be based on measured weight at the last clinical visit prior to delivery or at the time of delivery. For postpartum weight this should be based on measured weight at some point after delivery of the infant. For GWG, a true gold standard reference weight should rely on both measured pre-pregnancy weight prior to conception and measured weight at time of admission for delivery or last clinical visit prior to delivery. (2,18) All weight measures should be taken with light clothing on, without shoes on, and using appropriately calibrated equipment. All other ways of assessing reference weight measurements for pregnancy-related weight measures (e.g. measured weight at the wrong time or medical records) were considered "alloyed" gold standards. We also abstracted information on sampling procedures, response and/or participation rates, and location of study population in order to assess methodological strength of studies.

RESULTS

Our search of MEDLINE identified 426 studies total; our Google Scholar searches identified a large number of total studies, ranging from 2,780-28,800, which were reviewed as described above, until 50 consecutive titles provided no relevant studies. Across all of our searches, 37 studies met all inclusion and exclusion criteria. Six papers reported on the same study populations: both Herring et al. (19) and Oken et al. (20) reported on the Project Viva cohort, Ferrara et al. (21) and Hedderson et al. (18) both reported on the Kaiser Permanente Northern California population, and Tomeo et al. (22) and Buka et al. (23) used retrospective data from the National Collaborative Perinatal Project. (22,23) Additionally, multiple studies were conducted on women from the McGee Women's Hospital, (5,24,25) although they did not include exactly the same study populations. Four additional studies (26-29) focused exclusively on reliability rather than validity, and thus are not discussed here. Two of the studies identified were the previously discussed reviews, which focus on variation in methods of measuring pregnancy-related weight gain or validity of pre-pregnancy weight only. (10,14) Thus, overall, we abstracted data on reporting error from 28 articles identified through our search. Our findings were organized by type of pregnancy-related weight: pre-pregnancy weight, delivery weight, GWG, and postpartum weight. We also organized studies based on whether they used a true gold or alloyed gold standard. Some studies reported on multiple pregnancy-related weight outcomes, resulting in a total of 20 studies on pre-pregnancy weight, 4 studies on delivery weight, 13 studies on GWG, and 1 study on postpartum weight. (Figure 1)

The majority of studies (89%) were conducted in the United States, (Appendix 3, Table A3.1) with three being conducted outside of the U.S. Of the studies conducted outside of the U.S., one was conducted in Canada, (30) one in Denmark, (31) and one in Portugal. (32) Cohorts represented births from a wide range of time periods spanning from 1959 to 2011. Many study populations were racially and ethnically homogenous, but a substantial minority of studies (43%) included women from more than one racial or ethnic background. (5,17,21,23,24,33-38) Within studies that were racially/ethnically diverse, white, Black, and Latina women were most widely represented. All studies except five (22,23,33,39,40) were based on convenience samples of women recruited from medical or hospital institutions either at the time of their first prenatal visit or at time of delivery. Five studies were based in US national samples. Two studies used a subset of women from the National Collaborative Perinatal Project, (22,23) one used a subset of women from the National Maternal and Infant Health Survey, (40) and two assessed validity of reported pre-pregnancy weight using women who were pregnant in the National Health and Nutrition Examination Study (NHANES) data sets. (33,39) While a number of studies explicitly stated that assessing validity of self-reported pregnancy-related weight was their main objective, other studies included in this review were designed to answer research questions unrelated to assessing validity. (18-21,24,31,34,41-43) Thus, while they did provide some relevant information on reporting error, key metrics, such as mean difference and correlation, were not consistently reported and there was often limited information about variation of error across demographic sub-groups.

Measures of weight at one time point

Pre-pregnancy weight
Across several populations, self-reported pre-pregnancy weight appears to be well correlated with measured or medically reported pre-pregnancy weight. (Table 1) Correlation coefficients range from 0.86 (22) to 0.99. (19,20,43) Mean differences range from -2.94kg (36) to -0.60kg. (32) While average differences in self-reported and measured weight are not large, reporting error varies widely. Studies that reported ranges of reporting error (30,35,44,45) suggest that values range from underreporting by 12.58kg (45) to over-reporting by 22.04kg. (30) Additionally, there are trends of regression toward the mean: underweight women over-report their weight and overweight women tend to underreport their weight.

<u>True Gold Standard Studies:</u> Only four of the studies assessing accuracy of pre-pregnancy weight use true gold standard measures to assess reporting error. (18-21,36,46) In a study of young, unmarried, primiparous Black women in Rochester, NY giving birth between 1986 and1989, there is no statistically significant difference in self-reported (62.5kg) and measured pre-pregnancy weight (63.8kg). (46) However, there is a trend toward regression to the mean: underweight women over-report their weight by 1.4kg (SD = 2.5kg) and overweight women underreport their weight by 5.0kg (SD = 6.0kg). In a more diverse sample of women attending a safety-net, academic clinic in Ohio between 1990 and 2009, the authors find a statistically significant difference in self-reported and measured pre-pregnancy weight: women underreport their pre-

pregnancy weight by about 2.94kg. (36) Additionally, these authors find that this reporting error impacts misclassification across pre-pregnancy BMI classes. The lowest amount of misclassification occurs among normal weight women and the highest amount occurs among overweight women, with the latter being more likely to underreport their weight. A study in women giving birth to singleton infants within the Northern California Kaiser Permanente Network between 1996 and 1998 also finds that women underreport their pre-pregnancy weight. (18,21) The authors observe a mean difference of -1.6kg, although correlation between self-reported and measured pre-pregnancy weight remains high (r = 0.97). Data from women participating in Boston's Project Viva cohort between 1999 and 2000 also report a high correlation between self-reported and measured pre-pregnancy weight (r = 0.99) and a mean difference of only 1kg. (19,20)

Alloyed Gold Standard Studies: The remaining studies that assess accuracy of pre-pregnancy weight use alloyed gold standards. Nine of these studies use a measured weight that reflects their target measure (i.e. weight measured before conception), (17,30,33,35,37,39,41,45,47) but is measured at either the first prenatal visit or during the first trimester, so does not qualify as a true gold standard. Findings are similar to those observed among studies using true gold standards. Correlations between self-reported and measured early pregnancy weight are high, ranging from 0.92-0.99. (17) Women still underreport their weight anywhere from 1.52kg (30) to 2.3kg. (39) In a study of older predominantly black, unmarried women giving birth between 1986 and 1989, the authors report wide variability in reporting error with some women underreporting by as much as 12.56kg, while others over-report by as much as 8.64kg. (45) A Canadian study of women giving birth in 2011 confirms this high level of variation in reporting error: some women underreport their pre-pregnancy weight by 8.18kg while others over-report by 22.04kg. (30) The studies using measured alloyed gold standards do, however, less consistently observe regression toward the mean for reporting error across pre-pregnancy BMI groups. Three of these studies find that women across all pre-pregnancy weight classes underreport their weight, (17,37,45) and one finds that underweight women tend to over-report their weight while all other weight classes tend to underreport their weight. (39)

A number of these studies also look at percent agreement between self-reported pre-pregnancy weight and measured weight at the first prenatal visit. In a sample of predominantly white and rural pregnant women seeking obstetrical care within a primary care clinic in Upstate New York between 1995 and 1996, agreement between BMI using self-reported and measured weight at first prenatal visit is high (86%). (41) A study among women receiving prenatal care in Massachusetts between 2007 and 2008 finds similar rates of agreement when women are grouped into pre-pregnancy BMI weight classes. (35) Eighty-seven percent of women are classified into the same category regardless of weight measure. However, data from the 1999-2004 NHANES finds much lower rates of agreement for women being classified into the same BMI category based on self-reported and measured weight within the first trimester: only 60% of women remain in the same group. (33) Another study in the 2003-2006 NHANES assessing agreement between self-reported and weight measured in the first trimester also finds that both sensitivity and specificity of correct classification varies by pre-pregnancy weight class. (39) Obese women have the highest sensitivity and overweight women have the lowest sensitivity. Normal weight women have the lowest specificity, while specificity remains relatively high for all other pre-pregnancy weight groups. The only study assessing agreement that is not conducted in the U.S. finds that while there is variation in reporting error by pre-pregnancy BMI, the impact of this error on classification into pre-pregnancy BMI groups is not significant. (30)

The remaining seven studies that assess accuracy of self-reported pre-pregnancy weight and use alloyed gold standards to assess reporting error rely on medical or clinical records rather than measured weight. (5,22,24,31,32,43,44) These studies similarly find high correlation between self-reported and medically recorded pre-pregnancy weight, ranging from 0.95 (24) to 0.99, (43) although one study (22) with a much longer recall period observes a correlation of 0.86. However, most of these studies find that the magnitude of mean difference between self-reported and medically recorded pregnancy weight is either not significant (5,44) or slightly smaller, around -0.6 kg, (22,31,32) than that observed in studies using measured weight for their gold or alloyed gold standards. The only study that reports ranges of reporting error also finds that women underreport by as much as 6.8 kg and over-report by as much as 4.08 kg. (44)

A number of these studies also assess misclassification of women by pre-pregnancy weight class. In a large study within the Danish National Birth cohort between 1996 and 2002, 91% of women are classified within the same BMI group using self-reported and medically recorded pre-pregnancy weight. (31) Another study in a Portuguese birth cohort recruited from public maternity clinics in Porto, Portugal between 2005 and 2006, also finds that only 10.3% of women are misclassified by BMI category when using self-reported rather than medically recorded pre-pregnancy weight. (32) A study conducted in the U.S., however, among a predominantly white population of women giving birth between 2003 and 2010 at Magee-Women's hospital in Pittsburgh, PA, finds a much wider range of agreement across pre-pregnancy BMI groups (51.7%-100%) and also finds that this agreement varies by maternal characteristics. (5)

Variation by Maternal Characteristics: Overall, results from different cohorts and across studies using both gold and alloyed gold standards indicate that reporting error between self-reported pre-pregnancy weight and medically recorded or measured pre-pregnancy weight is not large (Table 2). However, there is some evidence to indicate that the magnitude of this error or the subsequent impact of this error on misclassification varies by a woman's pre-pregnancy weight class. Most notably, underweight women are more likely to over-report their weight (37,46) and misclassification tends to be highest among overweight or obese women. (36,41) Bodnar et al. (5) further find that misclassification varies by obesity subtypes, with obese class I women having higher misclassification than obese class 2 and 3 women. Conversely, Russell et al. (30) find that obese class I women have the highest agreement while obese classes 2 and 3 women have lower agreement. However, Russell et al. (30) use an agreement criterion of within 1kg rather than the BMI class based criteria that Bodnar et al. (5) use.

A subset of studies that we identified also investigate whether reporting error varies across different maternal demographic or socioeconomic characteristics. (5,17,20,37,39,44-46) (Table 2) These studies most consistently report on maternal age, socioeconomic status, race/ethnicity, and timing of measurement of gold or alloyed gold standard weight. Three (5,37,39) of the six (5,17,20,37,39,45) studies that investigate race/ethnicity find that reporting error varies by race/ethnicity. In particular, the magnitude of reporting error is greater for Black women compared to white women. (5,37,39) However, all of the studies that observe variation in reporting error by race/ethnicity use alloyed gold standards, rather than true gold standards.

None of the studies that investigate variation in reporting error by SES find significant associations. (37,39,45) Two (37,39) of the five studies (17,37,39,40,44-46) that investigate variation in reporting error by maternal age find that it does vary, but conflict on the age groups that have the greatest reporting error. One suggests that error is greatest among older women, (37) and the other suggests that error is greatest among younger women. (39) However, the magnitude of this difference in reporting error in the latter paper is small (1.5kg vs 1.4kg) and likely not clinically significant. (39) Only one (37) of the four studies (37,39,44,46) that investigate whether timing of gold or alloyed gold standard measurement mattered finds that it does, suggesting that reporting error is greater when medically recorded weights are measured later in pregnancy (e.g. if the first prenatal visit occurs later in the first trimester). The only gold standard study to assess variation in reporting error by timing of weight measurement does not find an association. (46)

Delivery weight

We identified only four studies (24,32,36,40) that report on the magnitude of reporting error between self-reported weight and medically recorded delivery weight. (Table 1) These studies suggest that reporting error is small and generally not significant. Mean difference in delivery weight ranges from -1.28kg (40) to 0.07kg. (36) All studies use a true gold standard of measured weight at the time of admission for delivery or at the closest prior prenatal visit. In a subset of women in the National Maternal and Infant Health Survey (NMHIS) giving birth in 1988, self-reported delivery weight among women recalling weight 6 to 31 months after pregnancy is 1.28kg less than measured weight at the time of delivery. (40) Sixty-one percent of women report weights within 2.27kg (5 pounds) of their measured weight, 28% underestimate by more than 2.27kg (5 pounds), and 11% overestimate by more than 2.27kg (5 pounds). Alternatively, in a subset of women from a diverse cohort recruited from a safety net medical center in Ohio between 1990 and 2009, women tend to over-report their weight by a non-significant 0.07 kg. (36) A study among women giving birth in Allegheny County, PA between 2003 and 2010 finds that self-reported delivery weight is highly correlated with medically recorded delivery weight (r = 0.96). (24) However, the authors do not further report on mean difference or variability of individual reporting error. The only study conducted in a non-U.S. population finds that women tend to underreport their delivery weight. (32) Women's self-reported delivery weight is 0.3kg lower than medically recorded delivery weight. Additionally, even though this magnitude of reporting error is small, only 50.6% of women are correctly classified within 1kg of their medically recorded delivery weight.

Variation by Maternal Characteristic: Only one (40) of these four studies further investigates whether delivery weight varies by maternal characteristics. Schieve et al. (40) find that reporting error varies across many of the maternal characteristics that they look at, including the following: higher pre-pregnancy and current BMI, non-adequate weight gain during pregnancy (either low or high weight gains), positive or negative weight change from delivery to time of reporting, non-white race/ethnicity, being unmarried, lower education, unintended pregnancy, late or no prenatal visits, longer length of recall, subsequent pregnancy between measured and reported weight, and reporting a delivery weight ending in zero. (40) While reporting error is of higher magnitude in all of these subgroups, it is not necessarily in the same direction. The majority of these subgroups are more likely to underreport their delivery weight compared with their medical records, but women who gain less than 15 pounds during pregnancy and women who gain weight between delivery and recall are more likely to over-report their delivery weight.

Measures of weight Change

Gestational Weight Gain

Correlation between self-reported and either measured or medically recorded GWG is lower than that seen for pre-pregnancy weight or delivery weight as a result of differences in accuracy of reporting for these two weight metrics. Overall, the reporting error associated with self-report of pre-pregnancy weight is of greater magnitude than the reporting error associated with delivery weight. This results in greater error in GWG estimated from self-reported measures of these two weight metrics. (5,14) Correlation between self-reported GWG and medically recorded GWG range from 0.42 (22) to 0.99. (48) (Table 3) Mean differences in self-reported and measured or medically recorded GWG range from -0.91kg (38) to 3.01kg. (36) There is a fair amount of variability around these estimates. One study (44) reports individual reporting errors ranging from -2.3kg to 20.5kg, and another reports errors ranging from -19kg to 32kg. (25) However, on average, women overestimate their GWG.

True Gold Standard Studies: Only two of the studies that we identified use true gold standard weight measurements to assess validity of self-reported GWG. (36,46) In a population of adolescents from upstate New York giving birth between 1986 and 1989, mothers over-report their weight gain during pregnancy by 1.2 kg. (46) Correlation between self-reported and documented GWG is moderate (r = 0.6). Furthermore, when self-reported weight gain is used, it appears that underweight women gain about 6kg less than overweight women, whereas this difference is not present when measured GWG is used. These findings are similar to those observed in a diverse population of women attending a safety net academic medical clinic in Ohio between 1990 and 2009. (36) Women over-report their weight gain during pregnancy by 3.01kg. The authors also find that this results in more women being classified as gaining excessively (63.2% vs 47.4%) when self-reported versus measured GWG is used.

Alloyed Gold Standard Studies: The remaining eleven (5,22,23,25,30,31,38,42,48-50) studies assessing validity of GWG rely on alloyed gold standards. One uses measured weight from the first prenatal visit, (30) while the other ten use weight from medical or clinical records. Still, most studies (22,23,25,30,48,51,52) find that women tended to over-report their GWG by 0.33kg (52) to 1.61kg. (30) Correlation between self-reported and medically recorded GWG is only moderate, ranging from 0.42 (22) to 0.63 (25). However, one study does find high correlation (r = 0.99) between self-reported and medically recorded GWG, although this study is conducted among adolescents. (48) One study alternatively reports that women underreport, rather than over-report, their GWG. (38) In a population of low-income women participating in the New Jersey HealthStart program from 1989 to 1992, mothers' self-reported weight gain during pregnancy on their birth certificates is 0.91kg lower than GWG calculated from clinical records. (38) However, this study uses a unique form of clinical record that is collected by HealthStart staff rather than a doctor or other medical personnel. The authors still find that self-reported and clinically recorded weight are only moderately correlated (r = 0.57).

A subset of these studies assesses agreement, rather than mean difference, between self-reports and medical records of GWG. Agreement ranges from moderate to high. A secondary analysis conducted among women participating in the Women and Infants Study of Healthy Hearts between 1997 and 2002 finds that 45% of the population self-reported GWG within 5 pounds of the documented weight; 33% of women overestimate their weight by more than 2.27kg and 22% of women underreport their weight by more than 2.27kg. (25) Another study comparing GWG recorded on birth certificates to weights recorded in obstetric electronic medical records of women giving birth in 2007 finds similar agreement rates. Just under half (48.2%) of women accurately report their GWG within 4.5kg, with 36.6% over-reporting and 15.2% underreporting. (50) However, a study using birth records from a sample of women from North Carolina giving birth in 1989, finds a somewhat higher level of agreement: 82.8% of self-reports of GWG agree exactly with the medically recorded amount for the women. (49) A study conducted among adolescent females receiving care from a large urban hospital prior to 1999 supports these findings. (48) Most (85%) adolescents 1-15 months postpartum accurately report their GWG within 0.91kg compared with their medical records. All of the prior studies are conducted in the U.S., however, two are conducted in non-U.S. cohorts. A large study conducted in the Danish National Birth Cohort between 1996 and 2002, finds only moderate (64%) agreement between self-reported and medically recorded GWG. (31) In a cohort of Canadian women giving birth in 2011, accuracy of reporting is fairly high (70.7%). (30) However, obese women are more inaccurate in self-reporting their GWG (50% agreement within 1kg) than other pre-pregnancy BMI groups. Overweight women are the most accurate reporters (81.3% agreement within 1kg); normal weight women fall in between (68.0% agreement within 1kg).

A few of the alloyed gold standard studies additionally assess agreement between GWG adequacy categories. Similar to the one gold standard study that assesses misclassification by GWG adequacy groups, (36) these studies suggest that reporting error impacts excessive GWG most, although this varies by pre-pregnancy BMI. In one study among women in New England giving birth between 1959 and 1966 recalling pregnancy-related events an average of 30 years after delivery, women over-report excessive GWG (as defined as weight gain of more than 25 pounds): only 28% of women have excessive weight gain as recorded in their charts, but 43% of women report excessive GWG thirty years later. (23) This corresponds to a sensitivity of 56% and a specificity of 62% (kappa = 0.15). McClure et al. (25) find that the use of recalled GWG misclassifies 36% of women overall. Among women gaining excessively, 20% are misclassified, while 45% of women gaining inadequately and 53% of women gaining adequately are misclassified. While this does not result in large differences in marginal prevalence of each GWG adequacy group, prevalence of excessive gain is slightly higher when self-reported GWG is used (52% vs 48%). A validation study conducted in a cohort of women giving birth at McGee Women's hospital in Pittsburgh, PA between 2003 and 2010 finds 73% agreement between self-reported and medically recorded measures, but also reports wide variability in this agreement (40.9-83.3%) depending on maternal characteristics. (5) Using data from the paper, we found that sensitivity was highest for women gaining adequately (81%) and was lower for inadequate and excessive gainers (67.6% for both groups). Wright et al. (50) also find that women who experience excessive GWG are more likely to under-report their gain, while women who experience inadequate gain are more likely to over-report. GWG adequacy and pre-pregnancy weight class interact to bias reporting of GWG: among women who gain adequately according to medical records, overweight women are more likely to overestimate their GWG compared to normal-weight women. Among normal-weight women, women with excessive GWG according to medical records are more likely to underreport their GWG compared to women who gain adequately. The authors do not further report differences in overall prevalence of inadequate, adequate, and excessive GWG by birth certificate compared to medically recorded weights.

<u>Variation by Maternal Characteristics:</u> Reporting error varies across some characteristics in studies that investigate this issue. Pre-pregnancy BMI is the most commonly investigated factor. (5,30,36,42,46,50) Most of the studies (using both gold and alloyed gold standards) find that normal weight women have the smallest error and the highest accuracy. (5,36,46,50) Studies are inconsistent in identifying groups with the poorest accuracy in recalling their GWG. Some find that obese women have the highest reporting error, (36) while others find that overweight women have the highest reporting error. (46,50) Still others find that the reporting error varies even within categories of obesity. (5,30)

Other maternal characteristics investigated include education, race, parity, and age. (Table 4) (5,22,23,25,46,48,50) Specifically, two (22,23) of three studies (22,23,25) that look at variation by maternal education find that reporting error does not vary by this socioeconomic indicator. The third study alternatively finds that women with lower education are more likely to over-report their GWG. (25) Of the three studies (5,25,50) that look at variation by race/ethnicity, two (25,50) find that minority women are less likely to accurately report their GWG, with McClure et al. (25) additionally indicating that Black women are more likely to underreport (rather than over-report) their weight by

2.27kg. Two studies, both conducted in the same study sample, (22,23) investigate whether reporting error varies by parity. One finds that there is a higher correlation between self-reported and medically recorded GWG for the first pregnancy, (23) but the other study does not support this finding. (22) Finally, most of these maternal characteristics are assessed in studies that rely on alloyed gold standards. The one gold standard study that looks at variation in reporting error by any maternal characteristics besides pre-pregnancy BMI, looks at age. (46) They find that adolescents appear to gain less, compared to their adult counterparts, when self-reported measures are used to assess GWG, but this difference is not observed when measured GWG is used.

An additional study characteristic important to explore in understanding variation in reporting error is timing of self-reported GWG. Across the studies that we identified, timing of GWG report varies from being taken at time of delivery (5,30,36,38,46,49,50) to being recalled 30 years in the past. (22,23) Correlations between self-reported and medically recorded weight are somewhat lower for studies with lengths of recall longer than 30 years. (22,23) However, there is not a clear trend in increasing or decreasing correlations between recalled and medically recorded GWG for lengths of recall between 10 and 20 years. Two of the studies that specifically assess whether there is variation in reporting error by length of recall have conflicting results. Biro et al.(48) do not find that length of recall is associated with differences between self-reported and medically recorded weight. McClure et al. (25) find that women who have given birth within 8 years of when they are asked to recall GWG are more likely to over-report their weight gain by 2.27kg. All of these studies rely on alloyed gold standards.

Post-partum weight

Little work explicitly investigates reporting error associated with postpartum weight, but one study finds that self-reported weights at 12 months postpartum are, on average, 1.23kg lower than medical records (alloyed gold standard) among women in New Haven giving birth between 2001-2004. (43) The corresponding correlation coefficient or variation in such error by demographic or socioeconomic characteristics is not reported.

One limitation we encountered with identifying studies that assess reporting error in postpartum weight retention was distinguishing it from regular weight reported in a non-pregnant woman who happens to be post-delivery. Studies that use self-reported postpartum weight may assume that errors associated with it are similar to those seen in non-pregnant populations of women of reproductive age (e.g. (53-55)), which is why literature on this topic is not as widely available. However, especially because weight gain and loss patterns during the postpartum period may be characteristically different from such patterns outside of this time period, due to lifestyle changes associated with adding a new or additional child to the family, stress of caring for a newborn, and breastfeeding, (13) work should be conducted to confirm that weight reporting errors for postpartum weight mirror those for weight in women of childbearing age.

Implications for Pregnancy and Birth Outcomes

A key concern about error in self-reported pregnancy-related weight is the introduction of bias in observed associations between pregnancy-related weight measures and birth outcomes. Quantifying the overall impact of such reporting error is important. We identified only four studies that assess the impact of reporting error on associations between pregnancy-related weight and birth outcomes. (5,25,35,56) Overall, these studies do not indicate that associations between pregnancy-related weight measures and birth outcomes are greatly biased even when misclassification resulting from reporting error is moderate.

Two studies investigate the impact of reporting error in pre-pregnancy weight on birth and pregnancy outcomes. One study specifically focuses on effects of misclassification of pre-pregnancy BMI on GWG adequacy classifications and finds that such error does not impact the prevalence of inadequate, adequate, and excessive gainers. (35) For both self-reported and measured weight, around 16% of women gain inadequately, around 27% gain adequately, and around 55% of women gain excessively. Another study assesses whether agreement of pre-pregnancy BMI measures varies by preterm birth status (i.e differential misclassification by outcome). (5) They find some evidence of differences in agreement for term, compared to preterm births, but do not further follow up to determine the impact that this variability has on estimates of association between pre-pregnancy BMI, GWG, and preterm birth.

Although not included in our review due to the fact that they do not specifically report on magnitude and direction of reporting error for any pregnancy-related weight metric, a 2010 paper by Bodnar et al. (6) investigates the impact of probabilistic bias adjustment to correct for bias resulting from reporting error in the associations between pre-pregnancy BMI and adverse pregnancy outcomes. They find that after accounting for misclassification of women into the wrong pre-pregnancy weight class, point estimates for the association between pre-pregnancy BMI and pregnancy outcomes including preterm birth, small for gestational age (SGA), large for gestational age (LGA), and gestational diabetes are generally in the same direction, but slightly attenuated. The extent to which the point estimates are attenuated varies by the outcome, but the only associations that are no longer significant are those between underweight pre-pregnancy BMI and SGA and spontaneous preterm birth. (6) A similar study that more extensively assesses the impact of using probabilistic bias correction to account for bias due to error in pre-pregnancy weight similarly finds that associations between pre-pregnancy weight and preterm birth are attenuated for underweight women and slightly more exaggerated for overweight and obese women. (57) Across all weight classes, however, pre-pregnancy weight increases the risk of preterm birth regardless of whether self-reported or bias corrected measures are used. (57)

Two studies assess whether reporting error in GWG impacts associations between GWG and pregnancy or birth outcomes. One study finds that such error does not impact associations between inadequate GWG and SGA, excessive GWG and postpartum weight retention, or GWG and cesarean delivery (null for both inadequate and excessive GWG). (25) However, relationships between inadequate GWG and postpartum weight retention and between excessive GWG and preterm birth are attenuated with self-reported weight. Another study also looks at whether agreement of self-reported and medically recorded GWG varies by preterm birth status. They find some support for the fact that agreement is actually higher for women who have preterm deliveries if they also have a low GWG. (5) However, the authors do not extrapolate on overall implications of such differences in agreement for bias in associations between GWG and preterm birth.

Finally, one study assesses the impact of reporting error in delivery weight on associations with of pregnancy outcomes, specifically GWG. (56) The authors report that 60-70% of women are classified in the same GWG category when self-reported delivery weight is used compared to measured delivery weight. In general, disagreement is due to underreporting delivery weight; reported delivery weight is more likely to place women in a lower weight gain category than her measured weight. Given this mediocre level of correct classification, associations between GWG and birth weight still persist when self-reported measures are used, although the relationship is attenuated.

Methodological Considerations

Two major methodological limitations impact the studies that we identified. First, as has been mentioned throughout, many studies rely on “alloyed gold standard” against which to compare women’s self-report. (5,19-25,31,32,38,48,49,51) This is especially problematic for studies relying on medical or clinical records as “alloyed gold standards” because weight measurements during clinical visits are often obtained through asking the mother her weight (e.g. (5,17,31)). In fact, one study assesses how many weights recorded in clinical records are likely to be actual measured weights and finds that only 6 cases in their sample are measured. (17) The rest are self-reported by the mother at one of her prenatal visits. Thus, studies using medical records as an alloyed gold standard may actually be capturing reliability of women’s report rather than validity. Alloyed gold standards relying on measured weight at the first clinic visit may still limit true assessment of validity, although not as much as medical records. (17,21,30,33-35,37,39,41,43-45,56) The timing of the first prenatal visit in the studies described here ranges from conception to 22 weeks of gestation. (Appendix 3 Table A3.1) Estimates of reporting error using weight measured during pregnancy may be capturing actual differences in a woman’s weight gain during her first trimester rather than error in her ability to recall her pre-pregnancy weight. (17,45) While we did not observe great variation in estimated correlations or mean differences for self-reported pregnancy-related weight measures when alloyed gold or true gold standards are used, there are not enough true gold standard studies to make a reliable comparison, especially for further assessing impacts of reporting error on misclassification by either pre-pregnancy BMI or GWG adequacy.

Another methodological challenge faced in this body of literature is sampling design and sample size used to evaluate reporting error. Many studies were not intentionally designed to assess reporting error, resulting in sample sizes that are limited, post hoc, to the women who happen to have existing values for their medical record as well as self-reports. Thus, while many studies still have reasonable sample sizes (range: 40

(48) to approximately 35,000 (38); mean = 2970) estimates are based on convenience samples of women who happen to have non-missing information for both self-report and measured or medically recorded weight. If reporting error, in fact, varies across some of the same predictors that also determine whether a woman has a complete medical record, (58) the estimates of bias in self-reporting error may not be accurate, especially in specific subgroups of women. One recent study intentionally sampled the population to have variation in factors they considered to be important for characterizing reporting error. (5) This approach directly addresses the possible issue of lack of variability in some subgroups due to convenience sampling. The study finds that while, on average, correlation between recorded and self-reported weight may be high, wide variation across demographic and maternal health characteristics still remains. (5) They specifically focus on maternal race/ethnicity, actual maternal pre-pregnancy BMI category, and preterm birth. Going forward, more studies could follow this approach in order to make sure that appropriate sample sizes are available for assessing validity of pregnancy related weight gain across additional maternal characteristics.

DISCUSSION

Overall we found that correlation is high between self-reported and objectively measured pre-pregnancy and delivery weight and moderate for GWG. Women tend to underestimate their pre-pregnancy weight and delivery weight while overestimating their GWG. The mean difference, however, between self-reported and medically recorded weight is relatively small, ranging from -2.94kg to 1.09kg for pre-pregnancy weight, -0.91kg to 3.01kg for GWG, and -1.28kg to 0.07kg for delivery weight. While reporting error is small in magnitude for the various types of pregnancy-related weight, it still results in moderate misclassification when grouping women into pre-pregnancy weight groups or GWG adequacy groups. Misclassification is most problematic for pre-pregnancy weight and the resulting errors carry over to measurements of GWG. This increases the magnitude of error and misclassification of women into GWG adequacy groups. Furthermore, there is some indication that pre-pregnancy BMI, race/ethnicity, maternal age and maternal education (for GWG only) influences the magnitude of reporting error for pregnancy-related weight measures, although these findings were based on a small number of studies. Nonetheless, reporting error and misclassification do not greatly bias associations between pre-pregnancy BMI or GWG and birth outcomes.

This is the first systematic review, to our knowledge, to assess reporting error in multiple measures of weight over the course of pregnancy. However, our findings build on two previous non-systematic reviews on this topic. (10,14) These reviews explore practical techniques and estimation methods for obtaining accurate measurements of weight prior to, during, and after pregnancy. They also focus on addressing many of the methodological issues inherent in obtaining pregnancy-related weight measures from women (e.g. participation of women in prenatal care, reliable adherence to ideal weight measurement protocols by medical staff, availability of optimal weight measurement equipment). As such, both reviews highlight the fact that while correlation is high and mean difference is relatively small, there is a good deal of individual variation in the error of self-reported weight measures. Errors also vary across sociodemographic characteristics. Furthermore a review by Gunderson et al. (59), although not systematic, finds that on average, the correlation between self-reported and measured (or medically recorded) weight during pregnancy is high. Mean differences for self-reported pregravid weight range from -0.5kg to -1.95kg and vary by age and race/ethnicity. These findings are consistent with the findings of the current review.

In the current review, however, we also reported on misclassification of women by pre-pregnancy BMI weight classes or GWG adequacy categories. Misclassification seems to lead to potentially important differences in estimates of the marginal prevalence of different measures of pregnancy-related weight, particularly GWG. However, based on the studies that we identified, misclassification based on self-reported weight does not substantially impact estimates of associations between pregnancy-related weight and birth outcomes. There is evidence to suggest that use of self-reported weight underestimates the strength of association for some outcomes, (25,60) but not to a large degree in most cases, and associations using more error-prone measures of pregnancy-related weight are still significant. However, less work has been done assessing the impact of misclassification of pregnancy-related weight measures when they are the outcome, rather than the exposure. This may be of particular concern given that self-reported estimates of GWG seem to overestimate the marginal prevalence of excessive GWG, thus skewing the distribution of this outcome. This is an important future direction for research as it impacts whether and to what degree associations of risk factors for these outcomes are potentially misestimated. Once this is more clearly understood, several bias adjustment techniques may be useful in addressing this concern in futures studies. (12,61,62)

Reporting error in self-reported weight measures has been widely studied in non-pregnant women. (3,62,63) Correction techniques have been developed to address such error, (12,62,63) although debate over their usefulness remains. Overall, data from non-pregnant women suggest that, on average, women underreport their weight by 0.1kg to 6.5kg with substantial variation in individual reporting error. (7,63) The range of mean reporting error in these reviews is much larger than the range that we found in our review. However, estimates from studies in non-pregnant populations, which include women of all ages, indicate that self-reported weight error may increase with age. (8,33,63,64) Studies that specifically investigate women of childbearing age find that error ranges from -2.09 kg (7) to -0.25kg (64). The range of average reporting errors that we observed for pregnancy-related weight measures is more similar to those observed in non-pregnant women of reproductive age. This suggests that correction algorithms and techniques designed for self-reported weight correction in non-pregnant women of reproductive age may, in fact, be potentially applicable to pregnant women. However, it will be necessary to confirm that correction methods apply to all pregnancy-related weight measures including self-reported pre-pregnancy weight, GWG, and delivery weight.

Based on our review, we propose three main recommendations for improving the current body of literature on reporting error in self-reported weight. First, studies should more consistently report the same measures of agreement (i.e. mean error, standard deviation, misclassification, sensitivity, and specificity). Second, more work should be done investigating variation of reporting error by maternal characteristics to assess their contribution to the variability that currently exists in reporting error in self-reported weight metrics. And finally, studies should pursue better gold standard measures. Measured weight should be used more often when assessing validity.

On this first recommendation, while we were able to draw conclusions across different metrics of agreement in the studies that we identified, developing a more cohesive understanding of reporting error across measures of pregnancy-related weight will require more consistent reporting of the same validity measures across studies. (63) Authors report a range of metrics including correlation coefficients, means, standard deviations, ranges of differences, percent agreement, and misclassification based on categorization by either BMI weight class or GWG adequacy categories. However, all of these pieces of information are needed to give a complete picture of the impact of reporting error, especially in order to consider the impact of such error on bias in particular associations of interest. (63) For example, just using correlation between self-reported and medically recorded (or measured) weight does not give a clear picture of the magnitude of the reporting error and the individual variability associated with it. Additionally, mean difference does not give a clear picture of how this difference is distributed within the population and how it impacts classification into relevant groups. Since there are widely used clinical classifications for both pre-pregnancy weight and GWG that are relevant to health outcomes, it is important to report such metrics of agreement. Consistent reporting on all such measures would contribute to the development of bias correction techniques to be applied across populations. (61)

Our second recommendation is based on the finding that, across all pregnancy-related weight measures, studies found that individual variability in self-reported weight and measured or recorded weight could be rather large. Understanding whether there are characteristics that systematically identify women prone to larger reporting errors is key to developing appropriate correction methods. Although approximately half of the studies identified investigated variation across subgroups in some way, (5,17,20,22,23,25,33,37,39,44-46,48,50,56) there was a lack of consistency regarding which maternal characteristics were investigated. For example, nine studies, overall, (5,17,20,33,37,39,44-46) investigate subgroup differences in reporting error by pre-pregnancy weight class, but only three investigate this variation in relation to education. (37,39,45) Many studies were not sampled in a way that would allow for enough subgroup variation to achieve the power needed to investigate such differences. This issue is based on the fact that many studies rely on existing data. In these data sets, the subset of women for whom both information on medical records and self-report are available may be limited, especially for particular socio-demographic characteristics. (58) As discussed previously, a recent study by Bodnar et al. (5) highlights a useful methodology for designing studies that are appropriately powered to understand variation in reporting error by subgroups. In particular, they use a balanced design to make sure that each subgroup of interest has enough women with both self-reported and medically recorded weight to be able to estimate reporting error with better precision. A key aspect of being able to conduct such a design is having large numbers of records available to sample from, which the authors achieved using birth certificate data and medical records for women at a large urban medical center. (5) Going forward, electronic medical records may provide a way to more consistently access health information for large portions of the population. (65) However, implementing ideal weight measurement techniques and ensuring complete data collection is necessary to create a high quality clinical database that can be used to assess validity of self-reported weight in pregnancy.

The implementation of studies using better gold standard weight references is critically needed to strengthen the ability to draw conclusions from this body of literature. As we discussed previously, studies relying on medical records, which served as the alloyed gold standard in many of the studies we identified, are likely assessing reliability rather than validity. Ideal gold standard measures may be difficult to obtain, especially for pre-pregnancy weight, due to the fact that women may not engage as frequently with the medical system prior to pregnancy. Nonetheless, future studies that use these gold standard weight measurements are needed. One thing that may help is development of preconception care programs. (66-68) More universal collection of measured weight within a few months prior to conception or at a first prenatal visit very early in pregnancy before substantial weight is gained is needed. (2)

One of the strengths of our review is that it includes a consideration of multiple and different measures of weight that are important over the course of pregnancy. We also used Google Scholar to include studies that would have been less likely to be identified in PubMed when validity was not the main aim of the study. This allowed us to obtain a more complete picture of reporting error based on existing studies. One main limitation of the current review was that studies were limited to English language only, therefore our conclusions about the validity of self-reported pregnancy-related weight was largely limited to studies conducted in the U.S.

Overall, this is the first review to our knowledge that assesses multiple aspects of accuracy in pregnancy related weight, including the validity of multiple self-reported pregnancy-related weight measures, the variability of validity across maternal subgroups, and impacts on misclassification. Our findings suggest that researchers using self-reported pregnancy weight measures should be cautious, especially in using self-reported pre-pregnancy weight and GWG. However, because self-reported weight remains an easy and efficient way of collecting weight information on women, especially in longitudinal, non-clinical settings, it will be important to address current limitations of the literature in order to develop correction techniques for such measures. Studies should routinely incorporate validation analyses into their initial study design when self-reported pregnancy-related weight is a main exposure or outcome. These actions will limit the potential impacts of bias related to self-reported pregnancy-related weight metrics and allow researchers more confidence when using such measures.

REFERENCES

1. Powell-Young YM. The validity of self-report weight and height as a surrogate method for direct measurement. *Applied Nursing Research*. 2012;25(1):25–30.

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

3. Willett W, Hu F. Anthropometric measures and body composition. In: Willett W, ed. *Nutritional Epidemiology*. New York: Oxford University Press, USA; 2013:213–240.

4. Osterman MJ, Martin JA, Mathews TJ, Hamilton BE. Expanded data from the new birth certificate, 2008. *Natl Vital Stat Rep*. 2011;59(7):1–28.

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. Bodnar LM, Siega-Riz AM, Simhan HN, et al. The impact of exposure misclassification on associations between prepregnancy BMI and adverse pregnancy outcomes. *Obesity (Silver Spring)*. 2010;18(11):2184–2190.

7. Brunner Huber LR. Validity of self-reported height and weight in women of reproductive age. *Matern Child Health J*. 2006;11(2):137–144.

8. 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.

9. 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(1):28–34.

10. 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.

11. Nyholm M, Gullberg B, Merlo J, et al. The validity of obesity based on self-reported weight and height: Implications for population studies. *Obesity*. 2007;15:197–208.

12. Willett W. Correction for the effects of measurement error. In: Willett W, ed. *Nutritional Epidemiology*. New York: Oxford University Press, USA; 2013:287–304.

13. 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.

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

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

16. Hartman KA, Mullen LB. Google Scholar and academic libraries: An update. *New library world*. 2008;109(5/6):211–222.

17. Ledeman 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.

18. Hedderson MM, Gunderson EP, Ferrara A. Gestational weight gain and risk of gestational diabetes mellitus. *Obstet Gynecol*. 2010;115(3):597–604.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. Mendez DD, Doebler DA, Kim KH, et al. Neighborhood socioeconomic disadvantage and gestational weight gain and Loss. *Matern Child Health J*. 2014;18(5):1095-103.

25. McClure CK, Bodnar LM, Ness R, et al. Accuracy of maternal recall of gestational weight gain 4 to 12 years after delivery. *Obesity (Silver Spring)*. 2011;19(5):1047–1063.

26. Kramer MS, Coates AL, Michoud MC, et al. Maternal anthropometry and idiopathic preterm labor. *Obstet Gynecol*. 1995;86(5):744–748.

27. Hinkle SN, Sharma AJ, Schieve LA, et al. Reliability of gestational weight gain reported postpartum: A comparison to the birth certificate. *Matern Child Health J*. 2012;17(4):756–765.

28. Herrmann D, Suling M, Reisch L, et al. Repeatability of maternal report on prenatal, perinatal and early postnatal factors: Findings from the IDEFICS parental questionnaire. *International Journal of Obesity*. 2011;35(S1):S52–S60.

29. Schlaff RA, Holzman C, Maier KS, et al. Associations among gestational weight gain, physical activity, and pre-pregnancy body size with varying estimates of pre-pregnancy weight. *Midwifery*. 2014;30(11):1124–31.

30. Russell A, Gillespie S, Satya S, et al. Assessing the accuracy of pregnant women in recalling pre-pregnancy weight and gestational weight gain. *J Obstet Gynaecol Can*. 2013;35(9):802–809.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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*. 2013;17(10):1872-8.

36. 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.

37. 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.

38. 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.

39. Shin D, Chung H, Weatherspoon L, et al. Validity of prepregnancy weight status estimated from self-reported height and weight. *Matern Child Health J*.

2014;18(7):1667-74.

40. 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. NMIHS Collaborative Working Group. *Am J of Epidemiol*. 1999;150(9):947–956.

41. Olson CM, Strawderman MS. Modifiable behavioral factors in a biopsychosocial model predict inadequate and excessive gestational weight gain. *Journal of the American Dietetic Association*. 2003;103(1):48–54.

42. Hunt SC, Daines MM, Adams TD, et al. Pregnancy weight retention in morbid obesity. *Obes Res*. 1995;3(2):121–130.

43. 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.

44. 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.

45. Yu SM, Nagey DA. Validity of self-reported pregravid weight. *Ann Epidemiol*. 1992;2(5):715–721.

46. 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.

47. Hickey CA, McNeal SF, Menefee L, et al. Prenatal weight gain within upper and lower recommended ranges: Effect on birth weight of Black and White infants. *Obstet Gynecol*. 1997;90:489–494.

48. Biro FM, Wiley-Kroner B, Whitsett D. Perceived and measured weight changes during adolescent pregnancy. *J Pediatr Adolesc Gynecol*. 1999;12:31–32.

49. 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.

50. Wright CS, Weiner M, Localio R, et al. Misreport of gestational weight gain (GWG) in birth certificate bata. *Matern Child Health J*. 2010;16(1):197–202.

51. Hunt KJ, Alanis MC, Johnson ER, et al. Maternal pre-pregnancy weight and gestational weight gain and their association with birthweight with a focus on racial differences. *Matern Child Health J*. 2013;17(1):85–94.

52. Nohr EA. *Obesity in pregnancy* [dissertation]. Denmark: University of Aarhus;

2005.

53. Keppel KG, Taffel SM. Pregnancy-related weight gain and retention: Implications of the 1990 Institute of Medicine guidelines. *Am J Public Health*. 1993;83:1100–1103.

54. 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.

55. Gunderson EP, Rifas-Shiman SL, Oken E, et al. Association of fewer hours of sleep at 6 months postpartum with substantial weight retention at 1 year postpartum. *Am J of Epidemiol*. 2007;167(2):178–187.

56. Schieve LA, Cogswell ME, Scanlon KS. An empiric evaluation of the Institute of Medicine's pregnancy weight gain guidelines by race. *Obstet Gynecol*. 1998;91:878–884.

57. Lash TL, Abrams B, Bodnar LM. Comparison of bias analysis strategies applied to a large data set. *Epidemiology*. 2014;25(4):576–582.

58. Smith PC, Araya-Guerra R, Bublitz C, et al. Missing clinical information during primary care visits. *JAMA*. 2005;293:565–571.

59. Gunderson EP, Abrams B. Epidemiology of gestational weight gain and body weight changes after pregnancy. *Epidemiologic Reviews*. 2000;22(2):261–274.

60. 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.

61. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. New York: Springer-+Business Media. New York; 2009.

62. Shields M, Gorber SC, Janssen I, et al. Bias in self-reported estimates of obesity in Canadian health surveys: An update on correction equations for adults. *Health Rep*. 2011;22(2):35–45.

63. 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.

64. 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.

65. Dean BB, Lam J, Natoli JL, et al. Review: Use of Electronic medical records for health outcomes research: A literature review. *Medical Care Research and Review*. 2009;66(6):611–638.

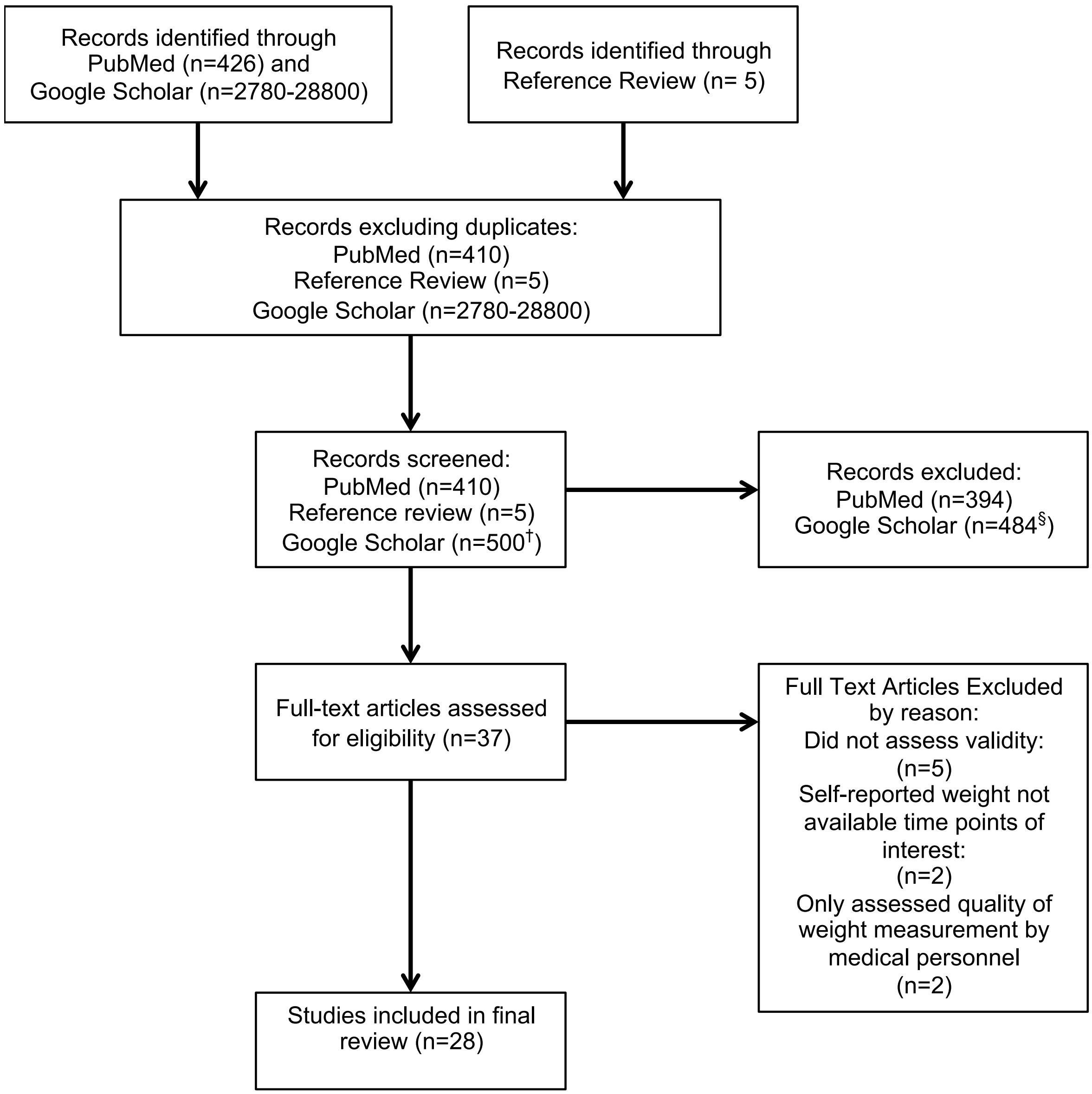
66. Korenbrot CC, Steinberg A, Bender C, et al. Preconception care: A systematic review. *Matern Child Health J*. 2002;6(2):75–88.

67. Paden MM, Avery DM. Preconception counseling to prevent the complications of obesity during pregnancy. *American Journal of Clinical Medicine*. 2012;9(1):30–35.

68. 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.

FIGURES AND TABLES

Figure 1. Flow Diagram for Study Selection into Systematic Review



†Based on our systematic search strategy, titles from all google scholar searches were reviewed until 50 consecutive titles did not produce a relevant article. The number of studies that were reviewed to reach this point varied by search query, but did not exceed approximately 500 studies.
§Sixteen total relevant articles were identified through initial screening as being eligible for full text review. Exact number of excluded studies varied by query; this number is an approximation based on maximum number of studies included for screening

Table 1. Summary of magnitude of reporting error for point in time measures of pregnancy-related weight (pre-pregnancy weight and delivery weight)

Study	Year of Birth Cohort	Gold Standard Rating	Mean difference (SD)	Correlation	Correct Classification	Misclassification by BMI Weight Class			
						Into higher weight class	Into lower weight class	Sensitivity	Specificity
Pre-pregnancy Weight									
Stevens-simon et al. (1992)	1986-1989	***	-1.3kg (5kg)	0.96	-	-	-	-	-
Mandujano et al. (2012)	1990-2009	***	-2.94kg	-	87%	NW: 1% OW: 2% OB: n/a	NW: 1% OW: 25%; OB: 14%	NW: 98%; OW: 73%; OB: 86%	NW: 91% OW: 92% OB: 99%
Ferrera et al. (2006)	1996-1998	***	-1.6kg	0.97	-	-	-	-	-
Oken et al. (2007), Herring et al. (2008)	1999-2002	***	-1kg	0.99	-	-	-	-	-
Hickey et al. (1993)	1985-1988	**	-	0.96	-	-	-	-	-
Yu et al. (1992)	1986-1989	**	-1.95kg	0.94	-	-	-	-	-
Ledeman et al. (1998)	1991-1993	**	-2.00kg (4.08kg)	0.92-0.99 (by pre-pregnancy BMI class)	85%	UW: 67% NW: 2% OW: 28% OB: n/a	UW: n/a NW: 2% OW: 0% OB: 11%	UW: 33% NW: 96% OW: 72% OB: 89%	UW: 98% NW: 82% OW: 98% OB: 94%
Olson et al. (2003)	1995-1996	**	-	-	86%	-	-	-	-
Craig et al. (2008)	1999-2004	**	-	-	60%	-	-	-	-

Table 1. Summary of magnitude of reporting error for point in time measures of pregnancy-related weight (pre-pregnancy weight and delivery weight)--CONTINUED

Study	Year of Birth Cohort	Gold Standard Rating	Mean difference (SD)	Correlation	Correct Classification	Misclassification by BMI Weight Class			
						Into higher weight class	Into lower weight class	Sensitivity	Specificity
Shin et al. (2013)	2003-2006	**	-2.3kg (0.7kg)	0.98	72%	UW: 50% NW: 9% OW: 8% OB: n/a	UW: n/a NW: 9% OW: 44% OB: 16%	UW: 50.0% NW: 82.4% OW: 48.0% OB: 83.3%	UW: 96.6% NW: 75.4% OW: 91.4% OB: 95.4%
Park et al. (2009)	2005	**	-1.93kg (0.04kg)	0.95	76%	UW: 23% NW: 7% OW: 6% OB: n/a	UW: n/a NW: 7% OW: 33% OB: 23%	UW: 77.3%; NW: 86.0%; OW: 61.1%; OB: 76.45%	UW: 96.8%; NW: 82.4%; OW: 88.4%; OB: 97.5%
Holland et al. (2012)	2007-2008	**	-1.81kg (3.27kg)	-	87%	UW: 46% NW: 10% OW: 15% OB: n/a	UW: n/a NW: 3% OW: 1% OB: 5%	UW: 54% NW: 88% OW: 83% OB: 95%	UW: 99% NW: 96% OW: 93% OB: 95%
Russell et al. (2013)	2011	**	-1.52kg	-	75%	-	-	-	-
Tomeo et al. (1999)	1959-1965	*	-0.66kg (3.7kg)	0.86	-	-	-	-	-
Stevens-simon et al. (1986)	1981-1983	*	-1kg	0.98	-	-	-	-	-
Nohr et al. (2008)	1996-2002	*	-0.66kg (2.42kg)	-	91%	UW: 14% NW: 1% OW: 2% OB: n/a	UW: n/a NW: 2% OW: 14% OB: 16%	UW: 86% NW: 96% OW: 84% OB: 84%	UW: 99% NW: 90% OW: 96% OB: 99%

Table 1. Summary of magnitude of reporting error for point in time measures of pregnancy-related weight (pre-pregnancy weight and delivery weight)--CONTINUED

Study	Year of Birth Cohort	Gold Standard Rating	Mean difference (SD)	Correlation	Correct Classification	Misclassification by BMI Weight Class			
						Into higher weight class	Into lower weight class	Sensitivity	Specificity
Gould-Rothberg et al. (2011)	2001-2004	*	-	0.96-0.99	-	-	-	-	-
Bodnar et al. (2014)	2003-2010	*	-	-	83%	UW: 30% NW/OW: 3% OB I: 14% OB II-III: n/a	UW: n/a NW/OW: 0% OB I: 18% OB II-III: 8%	UW: 70% NW/OW: 97% OB I: 68% OB II-III: 92%	UW: 100% NW/OW: 84% OB I: 97% OB II-III: 95%
Mendez et al. (2013)	2003-2010	*	-	0.95	-	-	-	-	-
Alves et al. (2011)	2005-2006	*	-0.6kg	-	90%	-	-	-	-
Delivery Weight									
Schieve et al. (1999)	1988	***	-1.28kg	-	61%¶	11%†	28%†	§	§
Mandujano et al. (2012)	1990-2009	***	0.07kg	-	-	-	-	§	§
Mendez et al. (2013)	2003-2010	***	-	0.96	-	-	-	§	§
Alves et al. (2011)	2005-2006	***	-0.3kg	-	51%¶	-	-	§	§

***Measured weight, at the right time, for all weight measurements were used
** Measured weight, for all weight measurements were used (but not at the ideal time)
* Medically or clinically recorded weight was used for any of the weight measures
¶ Heterogeneity in definitions of agreement: Schieve--within 2.27kg (5lbs); Alves--within 1kg
†Percentages refer to women reporting outside of 2.27kg of their measured, gold standard weight. BMI-based weight classes do not apply to these classifications.
§Sensitivity and specificity not reported because of lack of clinically relevant categories for delivery weight

Table 2. Variation in magnitude of reporting error for pre-pregnancy weight by maternal characteristics

	Stevens-simon et al. (1992)	Oken et al. (2007)	Yu et al. (1992)	Ledeman et al. (1998)	Shin et al. (2013)	Park et al. (2009)	Stevens-simon et al. (1986)	Bodnar et al. (2014)
Height			o			↑		
Age			o	o	↓	↑	o	
Race (Non-white race/ethnicity)		o	o	o	↑	↑		↑
Education			o		o	o		
Insurance Status			o					
Marital Status			o			↓		
Gestational Age		o						↓*
Gravidity	o							
Parity	o							
Socioeconomic Status	o				o			
Time between last menstrual period and recall	o				o	↑	o	
Site of prenatal care							o	
Number of prenatal visits prior to last menstrual period							o	
Tobacco use during pregnancy						↑		
Country of birth					o			

*They report that variability in reporting error was greater for preterm births, but they also had lower numbers for preterm births

↑ Increases magnitude of error/decreases accuracy

↓ Decreases magnitude of error/increases accuracy

o No Association

Table 3. Summary of magnitude of reporting error for gestational weight gain (GWG)

	Year of Birth Cohort	Gold Standard Rating	Mean Difference (SD)	Correlation	Agreement[§]	Sensitivity	Specificity	Over-report or Into higher GWG class	Underreport or Into lower GWG class
Stevens-simon et al. (1992)	1986-1989	***	1.2 kg (5.0)	0.6	n/a	-	-	-	-
Mandujano et al. (2012)	1990-2009	***	3.01kg	-	Self-reported: Below--13.2% Within--23.5% Above--63.2% Documented: Below--19.7% Within--33.3% Above--47.4%	-	-	-	-
Russell et al. (2013)	2011	**	1.61 kg	-	71%	-	-	-	-
Buka et al. (2004)	1959-1966	*	-	0.15[†]	-	56%	62%	-	-
Tomeo et al. (1999)	1959-1966	*	-	0.42	n/a	-	-		-
Reichman et al. (2001)	1989-1992	*	-0.91kg	0.57	-	-	-	-	-
Buescher et al. (1993)	1989	*	-	-	83%	-	-	-	-
Hunt et al. (1995)	1991-1992	*	Control: 1.3kg Morbidly obese: 0.6kg	-	-	-	-	-	-
Nohr et al. (2008)	1996-2002	*	0.33kg (3.7)	-	64%	64%	-	-	-
McClure et al. (2011)	1997-2002	*	1.0kg	0.63	45%	Inadequate: 55% Adequate: 47% Excessive: 80%	Inadequate: 88% Adequate: 64% Excessive: 73%	Inadequate: 45% Adequate: 32% Excessive: n/a	Inadequate: n/a Adequate: 21% Excessive: 20%

Table 3. Summary of magnitude of reporting error for gestational weight gain (GWG)--CONTINUED

	Year of Birth Cohort	Gold Standard Rating	Mean Difference (SD)	Correlation	Agreement§	Sensitivity	Specificity	Over-report or Into higher GWG class	Underreport or Into lower GWG class
Biro et al. (1999)	1999	*	0.54 kg	0.99	85%	-	-	-	-
Bodnar et al. (2014)	2003-2010	*	-	-	Overall: 72.5% Range: 40.9%-83.3%	Inadequate: 67.6% Adequate: 80.7% Excessive: 67.6%	Inadequate: 92.2% Adequate: 71.0% Excessive: 94.2%	Inadequate: 32.4% Adequate: 6.8% Excessive: n/a	Inadequate: n/a Adequate: 12.4% Excessive: 32.3%
Wright et al. (2012)	2007	*	-	-	48.20%	-	-	36.60%	15.20%

*** Measured weight, at the right time, for all weight measurements were used

** Measured weight, for all weight measurements were used (but not at the ideal time)

* Medically or clinically recorded weight was used to obtain either of the weight measurements needed to calculate GWG. Most one star scores, however, were driven by the use of medically recorded pre-pregnancy weight. Most delivery weights were measured

† Correlation statistic is a kappa statistic

§ Heterogeneity of definition: Buescher--exact; Nohr-- IOM categories; Mandujano--IOM categories; McClure--within 5lbs; Biro--within 2lbs; Bodnar--IOM categories; Wright--IOM categories; Wright--within 10lbs; Russell-- within 1kg

Table 4. Variation in accuracy of reporting in gestational weight gain (GWG) by maternal characteristics

	Stevens-simon et al. (1992)	Buka et al. (2004)	Tomeo et al. (1999)	McClure et al. (2011)	Biro et al. (1999)	Bodnar et al. (2014)	Wright et al. (2012)
Age	↓						
Non-white Race/ethnicity				↑		o	↑
Education		o	o	↓			
Insurance Status							↓
Gestational Age						↑	↓
Parity		o	↑				
Length of recall				↓	o		
Prenatal care							o
Tobacco use during pregnancy				↑			

↑ Increases magnitude of error/decreases accuracy

↓ Decreases magnitude of error/increases accuracy

o No Association

CHAPTER 5. CONCLUSION

Achieving healthy weight gain during pregnancy remains a challenge for many women. Continued progress in helping women achieve ideal GWG may require better understanding of the contexts in which they are managing their weight during pregnancy. This dissertation explores associations between one such contextual factor, the neighborhood social environment, in relation to GWG. Findings demonstrate that both length of exposure and pattern of exposure are important to consider in understanding how to engage within these environments to promote optimal GWG.

In chapter two, I find that cumulative neighborhood deprivation increases risk of excessive GWG, most strongly for white women, and that residential patterns of upward mobility, downward mobility, and continuous residence in deprived neighborhoods all increase risk of excessive GWG, compared to continued residence in affluent neighborhoods. Prior studies have not investigated associations between long-term neighborhood deprivation and our findings suggest that this will be an important factor to consider going forward. Understanding such associations between chronic neighborhood deprivation and GWG can better inform how to direct women enmeshed in these environments towards weight management resources or other support mechanisms to aid in achieving an ideal weight gain during pregnancy. (1,2) Furthermore, while it is not possible to change a woman's lived experience at the time that she enters into a clinic or care for pregnancy, the role of chronic deprivation during this critical time of pregnancy contributes additional supporting evidence for the importance of developing community or institutional policies targeted toward reducing long-term impoverishment in neighborhoods. (3,4) Such reductions in community poverty will have impacts that extend beyond improving the woman's health. By decreasing her lifetime experience of neighborhood deprivation, her offspring will be able to avoid the adverse health outcomes associated with maternal excessive GWG as well. Finally, given that white women were most strongly influenced by cumulative neighborhood deprivation in relation to excessive GWG, the differential impact of any interventions and policies by race/ethnicity should be considered. In minority communities, ancillary actions may be needed to address more widespread factors that place these women at risk for poor GWG. (2)

Chapter three in this dissertation continues on to investigate associations between point-in-time measures of objective and perceived neighborhood social environment and GWG and finds that neither are associated with inadequate or excessive GWG. However, when considered in conjunction with findings from paper one, these analyses highlight the importance of considering multiple approaches to assessing the contextual environments in which women experience pregnancy. Just focusing on these point-in-time measures, rather than also being able to explore long-term neighborhood deprivation, would have given an incomplete understanding of how neighborhoods matter for GWG. Existing studies have investigated numerous point-in-time measures across social and physical dimensions of the neighborhood environment. (5-8) Future studies that continue with exploration of these neighborhood domains should consider how different approaches to measuring or operationalizing these domains may bring depth to our understanding of how these factors influence GWG.

As investigators continue to embark on the research needed to better understand the role of neighborhood environments on GWG, they will need to consider how to address an important methodological limitation explored in the fourth chapter in this dissertation. Reporting error in self-reported pregnancy-related weight measures, while small on average, varies greatly across individuals. More work, using better gold standards, is needed to understand the degree to which this variability may be rooted in maternal characteristics. Improved understanding of reporting error across self-reported pregnancy-related measures can aid in the development of bias correction techniques that can be applied to studies using self-reported weight measures. (9) This has implications for future work on neighborhoods and GWG because, while some studies will be able to rely on better weight measurements and obtain ideal weight measurements among pregnant women they recruit, studies that follow women over long-periods of time and across multiple pregnancies in their reproductive life course may still need to rely on self-reported pregnancy-related weight measures. Such studies are needed in order to better understand dynamic patterns of exposure to neighborhood environments over time on GWG. Availability of bias correction techniques will provide options for addressing this methodological limitation.

Taken together, results from this dissertation highlight the complexities that must be considered in order to fully understand how neighborhood environments shape GWG. Length of exposure as well as type of exposure will be relevant for successfully identifying which neighborhood factors are most salient to helping women achieve healthy weight gain during pregnancy. Once such dynamic relationships between women and their neighborhood environment are understood, better interventions can be developed to improve GWG, and, consequently, the health trajectories of both mother and child.

REFERENCES

1. Culhane JF, Elo IT. Neighborhood context and reproductive health. *American Journal of Obstetrics and Gynecology*. 2005;192(5):S22–S29.

2. Davis EM, Stange KC, Horwitz RI. Childbearing, stress and obesity disparities in women: A public health perspective. *Matern Child Health J*. 2012;16(1):109–118.

3. Mendez DD, Hogan VK, Culhane JF. Stress during pregnancy: The role of institutional racism. *Stress and Health*. 2013;29(4):266-74.

4. Kramer MS. Race, place, and space: Ecosocial theory and spatiotemporal patterns of pregnancy outcomes. In: Howell FM, Porter JR, Mathews S, eds. *Recapturing Space: New midle-Range theory in Spatial Demography*. Dordrecht: Springer, Inc; (in press).

5. Mendez DD, Doebler DA, Kim KH, et al. Neighborhood socioeconomic disadvantage and gestational weight gain and loss. *Matern Child Health J*. 2014;18(5):1095-103.

6. Vinikoor-Imler LC, Messer LC, Evenson KR, et al. Neighborhood conditions are associated with maternal health behaviors and pregnancy outcomes. *Soc Sci Med*. 2011;73(9):1302–1311.

7. Laraia B, Messer L, Evenson K, et al. Neighborhood factors associated with physical activity and adequacy of weight gain during pregnancy. *J Urban Health*. 2007;84(6):793–806.

8. Messer LC, Vinikoor-Imler LC, Laraia BA. Conceptualizing neighborhood space: Consistency and variation of associations for neighborhood factors and pregnancy health across multiple neighborhood units. *Health Place*. 2012;18(4):805–813.

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

APPENDIX 1. ADDITIONAL TABLES AND FIGURES FOR CHAPTER 2

Table A1.1a Factor Loadings for Neighborhood Deprivation Index by Year, All US Census Tracts Included

	1980	1990	2000	2010
% Professionals in Management Occupations	0.68	0.65	0.66	0.71
% Crowded Households	0.64	0.57	0.61	0.52
% Poverty	0.84	0.89	0.90	0.84
% Less Than High School Education	0.83	0.85	0.87	0.81
% Unemployed	0.70	0.81	0.73	0.62
% Households on Public Assistance	0.88	0.89	0.86	0.61
% Female Headed Households	0.75	0.76	0.75	0.70
% Families Earning less than $30,000	0.73	0.84	0.91	0.89

Table A1.1b Correlation Between Neighborhood Deprivation Indices Across Years, Census Tracts in Analytic Sample

	1980	1990	2000	2010
1980	1.00			
1990	0.56	1.00		
2000	0.53	0.66	1.00	
2010	0.46	0.57	0.68	1.00

Table A1.2. Percent Missing in Analytic Variables

	Percent Missing
Cumulative Neighborhood Deprivation	21%
GWG	14%
Race/Ethnicity	0%
Foreign born	0%
Marital Status	10%
Employment	8%
Region of Residence	10%
Home Ownership	15%
Child's Birth year	0%
Education	10%
Equivalized Income	22%
Household Income	5%
Family size	9%
Maternal Age	0%
Parity	0%
Moved	40%
Length of residence	18%
Cumulative times moved	5%
Rural	15%

Table A1.3. Associations Between Cumulative Neighborhood Deprivation and Gestational Weight Gain (GWG), Stratified by Race For Significant Interaction, 2009 Institute of Medicine Recommendations Compared to 1990 Institute of Medicine Recommendations

	GWG Using 2009 Guidelines					**GWG Using 1990 Guidelines**				
	Crude		Adjusted[†]		Wald test for interaction[§]	Crude		Adjusted[†]		Wald test for interaction[§]
	RR	**95% CI**	**RR**	**95% CI**	**P-value**	**RR**	**95% CI**	**RR**	**95% CI**	**P-value**
Inadequate	**1.19**	**1.13, 1.24**	1.06	0.99, 1.14	**0.18**	**1.18**	**1.13, 1.23**	**1.07**	**1.00, 1.14**	**0.18**
Excessive	**1.05**	**1.02, 1.09**	**1.05**	**1.00, 1.11**	**0.07**	**1.07**	**1.04, 1.11**	**1.08**	**1.03, 1.13**	**0.02**
White	1.06	0.99, 1.13	**1.11**	**1.02, 1.20**		**1.10**	**1.03, 1.17**	**1.14**	**1.06, 1.23**	
Black	1.00	0.95, 1.05	1.00	0.94, 1.06		1.01	0.97, 1.06	1.01	0.95, 1.07	
Latina	0.99	0.91, 1.07	0.97	0.89, 1.06		1.01	0.94, 1.09	1.02	0.93, 1.11	

[†]Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.

[§]Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

Table A1.4. Associations Between Cumulative Neighborhood Exposures and Gestational Weight Gain (GWG) Using Outcome Stratified Regression Compared to Multinomial Logistic Regression; Stratified by Race for Significant Interaction

	Outcome Stratified Log-Linear Regression					**Multinomial Logistic Regression**				
	Crude		**Adjusted**†		Wald test for interaction§	**Crude**		**Adjusted**†		Wald test for interaction§
	RR	**95% CI**	**RR**	**95% CI**	**p-value**	**OR**	**95% CI**	**OR**	**95% CI**	**p-value**
Inadequate	**1.19**	**1.13, 1.24**	1.06	0.99, 1.14	0.18	**1.38**	**1.25, 1.51**	1.13	0.99, 1.30	
Excessive	**1.05**	**1.02, 1.09**	**1.05**	**1.00, 1.11**	**0.07**	**1.14**	**1.04, 1.25**	**1.15**	**1.01, 1.31**	0.27
White	1.06	0.99, 1.13	**1.11**	**1.02, 1.20**		1.15	0.96, 1.36	**1.29**	**1.05, 1.63**	
Black	1.00	0.95, 1.05	1.00	0.94, 1.06		1.00	0.87, 1.16	1.01	0.86, 1.20	
Latina	0.99	0.91, 1.07	0.97	0.89, 1.06		0.97	0.79, 1.19	0.93	0.71, 1.22	

†Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.

§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

APPENDIX 2. ADDITIONAL TABLES FOR CHAPTER 3

Table A2.1. Percent Missing in Analytic Variables

	Percent Missing
Race/Ethnicity	0%
Maternal Age	0%
Equivalized Income	10%
Marital Status	4%
Parity	0%
Foreign Born	0%
Employment	6%
Education	5%
Region	6%
Home Ownership	5%
Moving During Birth Year	13%
Length of Residence	10%
Cumulative Times Moved	0%
Kidbirthyear	0%
Total Family Income	0%
Family Size	5%
Rural	6%
Neighborhood Deprivation	17%
Perceived Neighborhood Problems	12%
Gestational Weight Gain Adequacy	27%

Table A2.2. Sensitivity Analyses for Perceived Neighborhood Problems (PNP) Using Different Gestational Weight Gain Categorization Criteria

	Standardized GWG using 2009 IOM Recommendations				**Observed GWG (Not Gestational Age Standardized); Full Birth Sample**				**Observed GWG (Not Gestational Age Standardized) Sample Restricted to Full Term Births Only (n=985)**			
	Crude		Adjusted†		Crude		Adjusted†		Crude		Adjusted†	
	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Inadequate												
Low Problems	1.01	0.75, 1.36	1.03	0.76, 1.38	1.10	0.84, 1.44	1.11	0.86, 1.44	0.98	0.72, 1.34	1.03	0.76, 1.38
Medium Problems	0.86	0.57, 1.28	0.74	0.50, 1.11	1.05	0.72, 1.54	0.89	0.61, 1.29	0.93	0.58, 1.49	0.71	0.45, 1.12
High Problems	1.28	0.94, 1.75	1.09	0.78, 1.51	**1.50**	**1.12, 2.02**	1.17	0.86, 1.59	**1.56**	**1.10, 2.20**	1.12	0.76, 1.65
			Interaction Wald test§: p =0.22				Interaction Wald test§: p =0.59				Interaction Wald test§: p =0.18	
Excessive												
Low Problems	1.00	0.86, 1.16	1.02	0.87, 1.21	1.01	0.86, 1.20	1.05	0.88, 1.26	0.97	0.81, 1.16	0.97	0.79, 1.18
Medium Problems	0.95	0.78, 1.15	0.89	0.72, 1.10	1.06	0.88, 1.28	0.99	0.80, 1.22	0.94	0.76, 1.17	0.84	0.65, 1.10
High Problems	**1.21**	**1.02, 1.44**	1.16	0.95, 1.40	**1.35**	**1.13, 1.62**	**1.24**	**1.01, 1.53**	**1.34**	**1.11, 1.62**	**1.24**	**1.00, 1.55**
			Interaction Wald test§: p =0.55				Interaction Wald test§: p =0.82				Interaction Wald test§: p =0.39	

†Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.
§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

Table A2.3. Sensitivity Analyses for Perceived Neighborhood Problems (PNP) Comparing 2009 Institute of Medicine (IOM) Recommendations to 1990 IOM Recommendations

		Standardized GWG using 2009 IOM Recommendations				Standardized GWG; Adequacy Categories Based on 1990 Standards			
		Crude		Adjusted†		Crude		Adjusted†	
		RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Inadequate									
	Low Problems	1.01	0.75, 1.36	1.03	0.76, 1.38	1.04	0.78, 1.38	1.13	0.86, 1.49
	Medium Problems	0.86	0.57, 1.28	0.74	0.50, 1.11	0.93	0.66, 1.31	0.93	0.67, 1.30
	High Problems	1.28	0.94, 1.75	1.09	0.78, 1.51	**1.36**	**1.01, 1.83**	1.27	0.92, 1.76
				Interaction Wald test§: p =0.22				Interaction Wald test§: p =0.46	
Excessive									
	Low Problems	1.00	0.86, 1.16	1.02	0.87, 1.21	0.98	0.84, 1.14	1.02	0.87, 1.20
	Medium Problems	0.95	0.78, 1.15	0.89	0.72, 1.10	0.95	0.78, 1.14	0.91	0.74, 1.11
	High Problems	**1.21**	**1.02, 1.44**	1.16	0.95, 1.40	**1.27**	**1.07, 1.50**	**1.21**	**1.00, 1.45**
				Interaction Wald test§: p =0.55				Interaction Wald test§: p =0.61	

†Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.
§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

Table A2.4. Sensitivity Analyses for Neighborhood Deprivation Index Using Different Gestational Weight Gain Categorization Criteria

	Standardized GWG using 2009 IOM Recommendations				**Observed GWG (Not Gestational Age Standardized); Full Birth Sample**				**Observed GWG (Not Gestational Age Standardized) Sample Restricted to Full Term Births Only (n=985)**			
	Crude		Adjusted†		Crude		Adjusted†		Crude		Adjusted†	
	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Inadequate												
Quartile 2	1.20	0.86, 1.67	1.12	0.81, 1.54	1.24	0.88, 1.76	1.23	0.87, 1.74	1.45	0.97, 2.17	1.31	0.87, 1.96
Quartile 3	**1.52**	**1.11, 2.10**	1.32	0.93, 1.87	1.40	0.99, 1.98	1.23	0.85, 1.77	**1.69**	**1.13, 2.53**	1.24	0.80, 1.94
Quartile 4	**1.80**	**1.33, 2.44**	1.36	0.93, 2.01	**1.83**	**1.31, 2.56**	1.27	0.80, 2.02	**2.04**	**1.39, 3.00**	1.25	0.71, 2.22
			Interaction Wald test§: p =0.36				Interaction Wald test§: p =0.44				Interaction Wald test§: p =0.32	
Excessive												
Quartile 2	1.00	0.85, 1.18	0.97	0.82, 1.15	1.05	0.87, 1.28	0.99	0.80, 1.21	1.02	0.84, 1.25	0.95	0.77, 1.18
Quartile 3	1.06	0.90, 1.24	1.01	0.84, 1.21	1.05	0.86, 1.29	0.95	0.76, 1.18	1.02	0.81, 1.27	0.93	0.73, 1.19
Quartile 4	**1.20**	**1.02, 1.39**	1.15	0.93, 1.42	**1.37**	**1.16, 1.62**	1.14	0.90, 1.45	**1.30**	**1.07, 1.58**	1.13	0.86, 1.47
			Interaction Wald test§: p =0.69				Interaction Wald test§: p =0.90				Interaction Wald test§: p =0.86	

†Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.

§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if p<0.10.

Table A2.5. Sensitivity Analyses for Neighborhood Deprivation Index Comparing 2009 Institute of Medicine (IOM) Recommendations to 1990 IOM Recommendations

	Standardized GWG using 2009 IOM Recommendations				Standardized GWG; Adequacy Categories Based on 1990 Standards			
	Crude		Adjusted†		Crude		Adjusted†	
	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Inadequate								
Quartile 2	1.20	0.86, 1.67	1.12	0.81, 1.54	1.24	0.91, 1.69	1.19	0.86, 1.62
Quartile 3	**1.52**	**1.11, 2.10**	1.32	0.93, 1.87	**1.50**	**1.09, 2.05**	**1.41**	**1.02, 1.96**
Quartile 4	**1.80**	**1.33, 2.44**	1.36	0.93, 2.01	**1.95**	**1.45, 2.62**	**1.62**	**1.09, 2.39**
			Interaction Wald test§: p =0.36				Interaction Wald test§: p =0.48	
Excessive								
Quartile 2	1.00	0.85, 1.18	0.97	0.82, 1.15	1.03	0.86, 1.22	0.98	0.82, 1.17
Quartile 3	1.06	0.90, 1.24	1.01	0.84, 1.21	1.09	0.92, 1.29	1.03	0.86, 1.23
Quartile 4	**1.20**	**1.02, 1.39**	1.15	0.93, 1.42	**1.35**	**1.17, 1.56**	**1.28**	**1.06, 1.54**
			Interaction Wald test§: p =0.69				Interaction Wald test§: p =0.26	

†Models adjusted for rural/urban, kid's birth year, marital status, employment, education, race/ethnicity, equivalized income, mother's age at birth, parity, region, immigrant status, home ownership, moving in the birth year, cumulative times moved over follow up, length of residence in current census tract.

§Overall tests for significant interaction between race/ethnicity and gestational weight gain were conducted using Wald tests; interaction was considered significant if $p<0.10$.

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured

Author	Weight gain type (pre-pregnancy,GWG, delivery weight, or postpartum)	Length of recall	Source of self-report	Source of "gold standard"	Gold standard rating	Mean Difference (self-reported weight - "gold standard" weight)	Correlation	Misclassification	Variation by demographic factors	Variation by pre-pregnancy BMI[†]
PRE-PREGNANCY WEIGHT										
Stevens-Simon et al (1992)	pre-pregnancy weight	Approximately 14.4 weeks post conception	Maternal report at time of study enrollment	Maternal medical and school records for weight measurements obtained by health care professionals anywhere from 6 months before pregnancy to 2 weeks after conception	***	-1.3 kg	0.96	-	No variation by age, gravidity, parity, SES, and proximity of report to last menstral period	UW: 1.4 kg (2.5) NW: -0.3 kg (4.0) OW: -5.0kg (6.0)
Mandujano (2012)	pre-pregnancy weight, delivery weight, GWG	Time between the first prenatal visit and last provider visit prior to conception	Maternal report at time of first prenatal visit	Measured weight from maternal medical records (+/- 12 weeks of last measured period).	***	-2.94kg	-	NW[‡]: 98% correctly classified; OW: 73% correctly classified; 25% underestimated weight, 2% overestimated OB: 86% correctly classified; 14% underestimated	-	NW: 98% correctly classified OW: 73% correctly classified OB: 86% correctly classified
Ferrara et al. (2006)	pre-pregnancy weight	Time to first prenatal visit (timing not explicitly specified)	Maternal report at time of first prenatal visit	Measured weight recorded in chart prior to last menstral period	***	-1.6kg	0.97	-	-	-
Oken et al. (2007)	Pre-pregnancy weight	≤ 22 weeks post conception	Maternal report via questionnaire and interview at time of study enrollment	Maternal medical record of weight measured within the 3 months before the last menstral period	***	-1kg	0.99	-	No vairation by race/ethnicity, gestational age, or actual weight itself (i.e. gold standard)	-
Herring (2008)	pre-pregnancy weight reported at time of enrollment (at first prenatal visit)	≤ 22 weeks post conception	Maternal report via questionnaire at time of study enrollment	Maternal medical record of weight measured within the 3 months before the last menstral period	***	-1kg	0.99	-	No variation by race/ethnicity, gestational age at study enrollment, and actual weight itslef (i.e. gold standard)	-
Hickey (1993)	pre-pregnancy weight	Time to first prenatal visit (timing not explicitly specified)	Interview at the first prenatal visit	Maternal medical record at first prenatal visit	**	-	0.96	-	-	-
Yu and Nagey (1992)	pre-pregnancy weight	Time to first prenatal visit (timing not explicitly specified)	Maternal report at time of first prenatal visit	Estimated pregravid weight based on measured weight gain during first trimester	**	-1.95kg	0.94	-	-	-
Ledemen and Paxton (1998)	Pre-pregnancy weight	14 weeks post conception	Maternal report at time of body composition visit	Measured weight recorded in clinlical records at the 14 week body composition visit	**	-2.00kg (4.08kg)	-	85% of women were correctly classified; 12% would have been in higher BMI group using measured weight, and 3% in a lower group.	No variation by age, race/ethnicity	UW: -3.18kg (2.45) NW: -2.27kg (3.76) OW: -0.82kg (5.35) OB: -1.32kg (5.00)
Olson (2003)	Pre-pregnancy weight	Time between "midpregnancy" and pre-conception (women were asked to self-report their pre-pregnancy weight on a questionnaire at mid pregnancy)	Mailed report at midpregnancy	Initial weight measured at first prenatal visit during the first trimester	**	-	-	86% of women were correctly classified by BMI category	-	Agreement among low, normal, and obese weight women was 90%; among overweight women 35% of were misclassified as normal weight

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Implications for birth or maternal health outcomes	Sample Size	Sampling approach	Participation Rate	Number of institutions included (e.g. clinics, hospitals ect.)	Years of birth cohort	Country	Micellaneous
PRE-PREGNANCY WEIG								
Stevens-Simon et al (1992)	-	93 in validation study (195 in total study sample)	Consecutive enrollment	47.7% (93/195 of elligible women)	One hospital in Rochester, NY (Strong Memorial Hospital)	1986-1989	USA	
Mandujano (2012)	-	234	Consecutive enrollment of clinic-based sample of women admitted for delivery	Not able to be calcuated; authors did not report total eligible population	One medical center in Cleveland, Ohio (MetroHealth Medical Center)	1990-2009	USA	The validity study was conducted as a secondary analysis to a separate analysis; the main objective of the overall study was to compare glucose tolerant and intolerant women. Only glucose tolerent women are included in the validity study.
Ferrara et al. (2006)	-	695 included in validation study	Convenience sample of women who had both self-report and measured in medical charts (Overall sample was selected through random selection of eligible pregnancies further classified by case/control status)	28.6% (695/2431 pregnancies in total sample)	16 clinics in the Kaiser Permanente Northern California network	Jan 1996-June 1998	USA	Validity study was a secondary analysis to a separate analysis; the main objective of the study was to assess adverse pregnancy outcomes among women with clinically diagnosed gestational diabetes and those who were borderline. There were 2431 pregnancies in the total study sample.
Oken et al. (2007)	-	170 in validation study	Convenience sample of women with medical records within the Project Viva Cohort	8.0% (170/ 2128) of total women in Project Viva cohort (overall study enrollment rate was 64% with 9% withdrawl or loss to followup)	Eight urban and suburban multispecialty practices in Eastern Massachusetts (Harvard Vangaurd Medical Services)	April 1999-July 2002	USA	Validity study was a secondary analysis; the main study investigated associations between maternal gestational weight gain and child's adiposity at age 3. There were 1044 women in total study sample.
Herring (2008)	-	170 for validation study	Convenience sample of women with medical records within the Project Viva Cohort	8.0% (170/ 2128) of total women in Project Viva cohort (overall study enrollment rate was 64% with 9% withdrawl or loss to followup)	Eight urban and suburban multispecialty practices in Eastern Massachusetts (Harvard Vangaurd Medical Services)	April 1999-July 2002	USA	This validation study is the same as the one reported in the Oken et al. (2003) study.Similarly it is a secondary analysis to the main study; the main study investigated association between women's perception of their pre-pregnancy body weight and how much they gained during gestation. There were 1537 women in total study sample.
Hickey (1993)	-	1168	15% random sample of the prenatal population at the clinic of interest	Not able to be calculated because authors did not report exact N for validation study (for original study: 43.9% (1168/2661) elligible women)	One prenatal clinic within the University of Alabama Birmingham and Jefferson County Health Department system of care	1985-1988	USA	The validation study was conducted as a secondary analysis of a separate analysis to assess the association between prenatal weight gain, term/preterm birth and fetal growth restriction between black and white women
Yu and Nagey (1992)	-	1591	Sample from medical record database (specific sampling approach not reported)	Not able to be calcuated as authors did not report total eligible population	Multiple clinics in the University of Maryland Medical System	1986-1989	USA	The gold standard in this study is a mathematically estimated pre-pregnancy weight based on weight at first prenatal visit and extrapolation of weight change in first trimester
Ledemen and Paxton (1998)	-	198 in validation study using measured weight early in pregnancy; 142 in validation study using clinical records with prepregnancy weight;	Convenience sampling	Not able to be calculated, but 71% of total sample had non-missing clinical records and 99% of total sample had non-missing measured weight	3 hospitals and clinics in New York City (Presbyterian Hospital, Harlem Hospital, St. Luke's-Roosevelt Hospital, Maternity Clinic)	Jan 1991-Aug 1993	USA	This study also assesses validity between medical records and self-reported weight (n=142) and finds the following mean differences by pre-pregnancy weight class: UW: 1.09kg (2.4lbs); NW: 0.39kg (0.86lbs); OW: -0.73kg (1.6lbs); OB: 0.23kg (0.51lbs)
Olson (2003)	-	543	convenience sub-sample of women had non-missing values for their weight measured in the first trimester and responded self-reprot questionnaire (overall sample systematically recurited from hosptial system database)	99.2% (543/547 women who were elligible responded to the self-reported questionnaire)	At least one hospital and multiple obstetric clinics serving a 10 county area of Upstate New York (exact number not specified)	1995-1996	USA	This validation study was a secondary analysis to a separate analysis in which the authors were mainly interested in indentifying predictive factors for inadequate and excessive weigth gain.

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Weight gain type (pre-pregnancy,GWG, delivery weight, or postpartum)	Length of recall	Source of self-report	Source of "gold standard"	Gold standard rating	Mean Difference (self-reported weight - "gold standard" weight)	Correlation	Misclassification	Variation by demographic factors	Variation by pre-pregnancy BMI†
Craig (2008)	Pre-pregnancy weight	Time between NHANES interview date and conception; varied for all participants	Maternal report via questionnaire at time of study enrollment	Measured weight at the time of recuritment into study	**	-	kappa-0.44 (SE=0.02)	60% of women were correctly classified	-	-
Shin (2013)	pre-pregnancy weight	Time between interview date and conception (varied based on month of gestation that woman was in)	Maternal report via questionnaire at time of study enrollment	Measured weight at time of interview (sensitivity analysis using first trimester weights only)	**	-2.3kg (SD=0.7)	0.98	72% correct classification (cacluated from n's in Table 3 of paper)	Did find variation by maternal characteristic: age 16-25 had greater reporting error as did Black women	Sensitivity: UW- 50.0; NW- 82.4; OW- 48.0; OB- 83.3; Specificity: UW- 96.6; NW- 75.4; OW- 91.4; OB- 95.4;
Park (2009)	pre-pregnancy weight	Duration of gestation	Maternal report on birth certificate	Measured weight during the first trimester based on woman's first WIC visit	**	-1.93kg (SE= 0.04)	0.95	76% of women were correctly classified Sensitivity: UW- 77.3%; NW-86.0% ; OW-61.1% ; OB-76.4% Specificity: UW- 96.8%; NW- 82.4%; OW- 88.4%; OB-97.5%	Did find variation by maternal characteristics: women age 30-39, blacks, unmarried women, women with tobacco use during pregnancy, women with weights measured more than 6 weeks after gestation had greater than average reporting errors in weight. BMI also varied by age 30-39, less than high school degree, tobacco use during pregnancy, and weight measurements > 6 weeks after pregnancy (all increased reporting error)	UW: -2.58kg (0.13) NW: -2.15kg (0.05) OW: -1.90kg (0.08) OB: -1.41kg (0.09) Maginal prevalences: Self-report: UW- 6%; NW- 45%; OW-25% ; OB- 24%; Measrued: UW- 4%; NW- 40%; OW-27% ; OB- 29%;
Holland et al. (2012)	pre-pregnancy weight	Time to first prenatal visit (timing not explicitly specified)	Maternal report at time of first prenatal visit	Measured weight abstracted from American College of Obstetrics and Gynecology prenatal records from first prenatal visit	**	-1.81 (SD=3.27, range -8.62-15.88)	-	87% of women were correctly classified and 13% were misclassified. Most of the misclassified women were misclassified into a higher weight class when measured weight was used. They predominantly ended up in an adjacent weight category.	-	-
Russell (2013)	pre-pregnancy weight and GWG	Duration of gestation	Maternal report at time of delivery	First measured weight in the maternal medical record before 12 weeks gestation	**	-1.52kg	-	74.7% of women report within 1kg of their actual weight Percent agreement by pre-pregnancy BMI class: UW- 85.7; NW-64 ; OW- 87.5; OB (all)- 84.6	-	UW: -1.19kg NW: -1.12kg OW: -2.15kg; OB I: -2.34 ; OB II: -1.14; OB III: -1.59
Tomeo et al.(1999)	pre-pregnancy weight	32 years after pregnancy	Mailed questionnaire	Maternal medical records collected during pregnancy by the National Collaborative Perinatal Project (NCPP)	*	- 0.66 kg (SD=3.7)	0.86	-	-	-
Stevens-Simon et al. (1986)	pre-pregnancy weight	Time to first prenatal visit (timing not explicitly specified)	Maternal report at time of prenatal visit	Maternal medical and school records for weight obtained by health care professionals anywhere from 6 months before pregnancy to 2 weeks after conception	*	-1kg	0.98	-	No variation by age,	-

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Implications for birth or maternal health outcomes	Sample Size	Sampling approach	Participation Rate	Number of institutions included (e.g. clinics, hospitals ect.)	Years of birth cohort	Country	Micellaneous
Craig (2008)	-	724	Convenience sample (women, overall, were proportionally sampled from the population based on NHANES procedures, but this procedure did not include a specific screen for pregnancy. Only women who happened to be pregnant at the time of interview were included)	Not able to be calcuated as authors did not report total eligible pregnant population	n/a; NHANES is a popualtion based smaple	1999-2004	USA	The women in this study were all participants in the NHANES study. As NHANES, does not restirct timing for when pregnant women could enroll, the "gold standard" measruement could occur at any time during pregnancy. Authors do not report the mean time or range of times during pregnancy at which women's weight were measured.
Shin (2013)	-	504	Proportional sampling from a population-based sample (NHANES) in which pregnant women were oversampled.	75.8% (504/665 ellegible pregnant women)	n/a; NHANES is a popualtion based smaple	2003-2006	USA	
Park (2009)	-	23314	Review and linkage of birth records to WIC reocrds (all linked records included in study)	37.8% (23314/61682 sucessfully linked records with non-missing weight data)	Approximately 200 WIC institutions in Florida	Jan1-Dec31 2005	USA	
Holland et al. (2012)	Misclassification of pre-pregnancy bmi did not lead to a large change in the women who were classified within each GWG adequacy category; Self-reported: Inadequate-16.8% Adequate-28.6% Excessive-54.6% Medically recorded: Inadequate-16.9% Adequate- 27.1% Excessive- 56.0%	307 in the validation study	Randomly selected, eligible women in clinic populations	64.3% (307/477 eligible records)	Multiple clinics in the Umass Memorial Health Care obstetric clinics system	2007-2008	USA	
Russell (2013)	-	99	Hospital-based sample of women admitted for delivery at participating institutions (specific selection procedure not reported)	Not able to be calculated because total number eligible was not reported but 52.4% (99/189) had information on the weight mesaurements in the first trimester	Two hospitals in New Brunswick (St. John's Regional Hospital and The Moncton Hospital)	2011	Canada	
Tomeo et al.(1999)	-	154	Convenience sample of a location-based sub-group of a cohort study	61.6% (154/250 eligible women)	One health center in Providence, Rhode Island (The Child Study Center). Women were participants in the National Collaborative Perinatal Project of he National Institute of Neurological Diseases and Stroke	1959-1965	USA	
Stevens-Simon et al. (1986)	-	76 in the validation study	Convenience sample of stored medical records	49% (76/155 accessed medical records; This is not a true participation rate, authors did not provide N for how many women within the medial records system were eligible)	One hospital centers in Rochester, NY (Strong Memorial Hospital)	1981-1983	USA	Validation stdudy conducted among adolescents.

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Weight gain type (pre-pregnancy,GWG, delivery weight, or postpartum)	Length of recall	Source of self-report	Source of "gold standard"	Gold standard rating	Mean Difference (self-reported weight - "gold standard" weight)	Correlation	Misclassification	Variation by demographic factors	Variation by pre-pregnancy BMI†
Nohr (2008)	pre-pregnancy weight and GWG	Time to first prenatal visit (timing not explicitly specified)	Maternal report at time of first prenatal visit	medically recorded weight at first prenatal visit obtained from medical records	*	-0.66kg (2.42)	-	91% of women correctly classified by BMI category	-	Correct classification varried by BMI class: Correct classification: UW- 86.1%; NW- 96.1%; OW- 83.5%; OB- 84.0%;
Rothberg (2011)	pre-pregnancy weight, pregnanacy weights during gestation (at various times), and postpartum weight	Time to first study interview (average of 18 weeks)	Maternal report at time of prenatal visit	Maternal medical records at time of prenatal visit	*	-	0.96-0.99	-	-	-
Bodnar (2014)	pre-pregnancy weight and GWG	Duration of gestation	Maternal report on birth certificate	Maternal medical records at time of first prenatal visit	*	median difference was close to 0 in all strata, but had wide variability (overall mean difference was not reported in paper)	-	51.7% to 100% of women were correctly classified	Did find variation by maternal characteristics: underweight or obese class 1, black women were misclassified to a higher degree than white women of the same weight classifications; more variability in agreement among preterm (as compared to term births)	Agreement within 5lbs: UW-66.7%; NW-62.7%; OW-43.6%; OB-40.7%
Mendez (2013)	pre-pregnancy weight and weight at delivery	Duration of gestation	Maternal report on birth certificate	Maternal medical records at time of first prenatal visit	*	-	0.95	-	-	-
Alves (2011)	pre-pregnancy weight and delivery weight	Duration of gestation (pre-pregnancy weight self-reported after delivery)	Maternal interview post delivery	Maternal medical records at time of prenatal visit	*	-0.6kg	-	44.2% of women were correctly classified within 1kg of recorded weight; 10.3% of women were misclassified by BMI class	-	-
DELIVERY WEIGHT										
Schive et al. (1999)	Delivery Weight	6-31 months after pregnancy	Maternal report via questionnaire at time of study enrollment	Measured delivery weight from the hospital record	***	-1.28kg	-	61% of women correctly classified within 2.27kg; 28% of women underreported, 11% of women overreported	Did find variation by maternal characteristics: reporting error increased with amount of weight gained (both low and high) and with weight change from delivery to maternal recall. Reporting error was greater among non-white women, less educated, and unmarried women, unintended pregnancy, late response to questionnaire, and for women who became pregnant again before questionnaire	UW: -0.49 NW: -1.13 OW: -2.07 OB: -2.87
Mandujano (2012)	pre-pregnancy weight, delivery weight, and GWG	Time between delivery and interview (length not explicitly reported)	Not explicitly reported (most likely maternal medical recored at time of delivery)	Measured delivery weight from maternal medical records at closest (w/in 2 weeks) visit prior to delivery	***	0.07kg	-	-	-	-

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Implications for birth or maternal health outcomes	Sample Size	Sampling approach	Participation Rate	Number of institutions included (e.g. clinics, hospitals ect.)	Years of birth cohort	Country	Micellaneous
Nohr (2008)	-	5033	Clinic-based comvenience sample of women for whom midwives were able to be contacted to report medically recorded pre-pregnancy weight	Not able to be calculated, but only 5033 (5.0%) of the total records for the DNBC (100419) had available data on "gold standard" prepregnancy weight	Many, based on Danish National Birth Cohort and National Obstetrics Database for North Jutland County	1996-2002	Denmark	Overall, the data comes from a convenience sample of all the births in Demark (the Danish National Birth Cohort; 100419 pregnancies to 92274 women). Data included in review were taken from both the 2008 paper and the first authors dissertaion (Nohr, EA. "Obesity in Pregnancy: Epidemiological studies based on the Danish National Birth Cohort. University of Aarhus, Aarhus, Denmark. 2005) which includes more detail on sample and mean difference
Rothberg (2011)	-	427 in study sample (unclear whether full sample was included in validity analysis)	convenience sample (subsample of a radomized clinical trial)	Not able to be calcluated because did not report exact N for validation study, but 84.9% (427/503) of women at the New Haven, CT site were included in the overall study sample	university affiliated hospital clinics in New Haven, CT (exact N not specified)	2001-2004	USA	This validation study was conducted as a secondary analysis of a separate analysis. The main analysis was looking at racial differences in gestational weight gain and postpartum weight retention.
Bodnar (2014)	-	1207	Hospital-based stratified sampling	86.1% (1204/1440 optimal sample size of eligible records; The optimal sample size was selected from total of 47233 eligible records from Magee women's hospital for balanced design)	One Hospital in Pennsylvania (McGee Women's Hospital)	2003-2010	USA	Some subsample sizes in this analysis are vary small; some of the percent agreement estimates are based on samples of size 2 to 20. (sensitivity analysis using measured weight at first prenatal visit for those who had it)
Mendez (2013)	-	not reported for validation study	convenience sample of women who had medical records and birth records with weight information	Not able to be calcluated because did not report exact N for validation study	One hospital in Pennsylvania (McGee Women's Hospital)	2003-2010	USA	The validation study was a subanalysis of a separate analysis on neighborhood socioeconomic status and weight gain during pregnancy. 55608 women were included in the complete case sample of the overall analysis
Alves (2011)	-	1703 included in validity study (3657 records in the total study sample)	Consecutive enrollment at time of delivery	46.6% (1703/3657 elligible records)	5 maternity units covering the city of Porto, Portugal	April 2005-Sept 2006	Portugal	
DELIVERY WEIGHT								
Schive et al. (1999)	60-70% of women were classified in the same GWG category when self-reported vs measured delivery weight was used (Pre-pregnancy weight was self-reported). Associations between GWG and birth weight persisted using self-reported vs measured weight but were more attneuated using self-reported weight	3518	Stratified sample of births in 48 states, DC and NYC within the 1988 National Maternal and Infant Health survey	52.5% (3518/6695 elligible women)	Multiple institutions, taken from birth records for women in 48 states and the District of Columbia as bart of the National Maternal and Infant Health Survey	1988	USA	GWG for comparison of effect of reporting error on pregnancy outcomes was not categorized using IOM guidelines
Mandujano (2012)	-	234	Consecutive enrollment of clinic-based sample of women admitted for delivery	Not able to be calcuated as authors did not report total eligible population	One hospital in Cleveland, Ohio (MetroHealth Medical Center)	1990-2009	USA	The validity study was conducted as a secondary analysis to a separate analysis; only the glucose tolerent women are included in the validity study. The main objective of the overall study was to compare glucose tolerant and intolerant women.

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Weight gain type (pre-pregnancy,GWG, delivery weight, or postpartum)	Length of recall	Source of self-report	Source of "gold standard"	Gold standard rating	Mean Difference (self-reported weight - "gold standard" weight)	Correlation	Misclassification	Variation by demographic factors	Variation by pre-pregnancy BMI†
Mendez (2013)	prepregnancy weight and weight at delivery	Time between last prenatal visit and delivery	Maternal report on birth certificate	Measured delivery weight from maternal medical records	***	-	0.96	-	-	-
Alves (2011)	pre-pregnancy weight and delivery weight	72 hours post delivery	Maternal interview post delivery	Measured delivery weight from maternal medical records	***	-0.24kg	-	50.6% of women were correctly classified within 1kg	-	-
GESTATIONAL WEIGHT GAIN										
Stevens-Simon et al (1992)	Self-reported GWG	Approximately 14.4 weeks after pregnancy	Maternal report at time of study enrollment. Delivery weight was measured at closest vist prior to delivery.	Maternal medical and school records for weight measurements obtained by health care professionals anywhere from 6 months before pregnancy to 2 weeks after conception	***	1.2kg (SD=5; Range= -2.3 - 20.5)	0.6	-		-
Mandujano (2012)	pre-pregnancy weight, delivery weight, GWG	Time between delivery and interview (length not explicitly reported)	Maternal report at time of first prenatal visit for pre-pregnancy weight; Not explicitly stated for delivery weight (most likely maternal medical recored at time of delivery)	Measured weight from maternal medical records otherwise taken at first prenatal visit if available.	***	3.01kg	-	Using documented weights, 33.3, 19.7, 47.4% of women gained adequately, inadequately, and excessively, respectively. Using self-reported weights, 23.5%, 13.2%, 63.2% were classified as gaining adequately, inadequately, and excessively.	-	Obese women had a significant difference in their documented vs reported GWG, normal weight and overweight women had no significant difference
Russell (2013)	pre-pregnancy weight and GWG	Time between last prenatal visit and delivery	Maternal report at time of delivery	Maternal medical record at first prenatal visit prior to 12 weeks and weight measured at admission for delivery	**	1.61kg	-	70.7% of women correctly classified within 1kg	-	Mean difference: UW: 1.08 kg NW: 0.97 kg OW: 2.59 kg OB I: 2.31 kg OB II: 2.06 kg OB III: 1.59 kg Percent agreement: UW- 71.4%; NW- 68.0%; OW- 81.25%; OB- 50.0;
Buka et al. (2004)	GWG	30 years after pregnancy	Maternal interview	Medical records based on physical examination during pregnancy in the NCPP sample	*	-	kappa=0.15	Sensitivity 56% Specificity 62%	No variation by education, socioeconomic status, or parity	-
Tomeo et al.(1999)	GWG	32 years after pregnancy	Mailed questionnaire	Maternal medical records collected during pregnancy and delivery by the National Collaborative Perinatal Project (NCPP)	*	-	0.42	-	no variation by education, but recall did vary by birth order (r=0.74 for first birth, r=0.29 for later delivery)	-
Reichman (2001)	GWG	Time between last prenatal visit and delivery	Maternal report on birth certificate	Maternal medical records from the HealthStart Maternity Services Summary Data	*	-0.91kg	Chronbach alpha=0.70; Spearman rank correlation = 0.57;	-	-	-
Buescher et al. (1993)	GWG	Time between last prenatal visit and delivery	Maternal report on birth certificate	Maternal medical record	*	-	-	82.8% (95% CI: 78.1-87.5%) agreed exactly	-	-
Hunt et al. (1995)	GWG	10-20 years after pregnancy	Maternal report via questionnaire at time of study enrollment	Maternal medical records recovered from women's delivery hospital	*	Controls: 1.3kg Obese: 0.6kg	-	-	-	-

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Implications for birth or maternal health outcomes	Sample Size	Sampling approach	Participation Rate	Number of institutions included (e.g. clinics, hospitals ect.)	Years of birth cohort	Country	Micellaneous
Mendez (2013)	-	not reported	convenience sample of women who had medical records and birth records with weight information	Not able to be calcluated because did not report exact N for validation study	One hospital in Pennsylvaia (McGee Women's Hospital)	2003-2010	USA	The validation study was a secondary analysis of a separate analysis on neighborhood socioeconomic status and weight gain during pregnancy
Alves (2011)	-	3023 included in validation study (3657 records included in total study sample)	Consecutive enrollment at time of delivery	82.7%(3023/3657 elligible records)	5 maternity units covering the city of Porto, Portugal	April 2005-Sept 2006	Portugal	
GESTATIONAL WEIGHT								
Stevens-Simon et al (1992)	-	93 included in validation study (195 in total study sample)	Consecutive enrollment	47.7% (93/195 elligible women)	One hospital in Rochester, NY (Strong Memorial Hospital)	1986-1989	USA	
Mandujano (2012)	-	234	Consecutive enrollment of clinic-based sample of women admitted for delivery	Not able to be calcuated as authors did not report total eligible population	One health center in Cleveland, Ohio (MetroHealth Medical Center)	1990-2009	USA	The validity study was conducted as a secondary analysis to a separate analysis; only the glucose tolerent women are included in the validity study. The main objective of the overall study was to compare glucose tolerant and intolerant women.
Russell (2013)	-	99	Hospital-based sample of women admitted for delivery at participating institutions (specific selection procedure not reported)	Not able to be calculated because total number eligible was not reported but 52.4% (99/189) had information on the weight mesaurements in the first trimester	Two hospitals in New Brunswick (St. John's Regional Hospital and The Moncton Hospital)	2011	Canada	
Buka et al. (2004)	-	96	Location-based convenience subsample from larger National Collaborative Perinatal Project cohort	Not able to be calcluated because number of elligible women from sample 1 was not reported, but 81.3% (96/118) of the desired sample size participated	Multiple institutions in the Providence, RI and Boston, MA cohorts of the National Collaborative Perinatal Project	1959-1966	USA	
Tomeo et al.(1999)	-	154	Convenience sample of a location-based sub-group of a cohort study	61.6% (154/250 eligible women)	One health center in Providence, Rhode Island (The Child Study Center). Women were participants in the National Collaborative Perinatal Project of he National Institute of Neurological Diseases and Stroke	1959-1965	USA	
Reichman (2001)	-	35687	Record linkage to birth records (all sucessfully linked records included)	76.9% (35687/46437 sucessfully linked records)	Multiple institutions in New Jersey (based on medicaid records; exact number not specified)	1989-1992	USA	The HealthStart program was conducted in New Jersey starting in 1988 through the Medicaid program; HealthStart providers were not medical professionals. Information was taken from prenatal records from women's providers when they were able to be contacted leading to heterogeneity in the way that medically recorded weight was collected.
Buescher et al. (1993)	-	330	Stratified, clustered sample of birth records	83.5% (330/395 elligible birth records)	79 non-military hospitals in North Carolina that had 250 or more births in 1989	December 1989	USA	
Hunt et al. (1995)	-	173 in validation study (110 morbidly obese women, 63 controls)	Convenience based subsample	controls: 32.8% (63/192); morbidly obese: 68.3% (110/161)	Multiple institutions across Utah (154 in total)	1991-1992	USA	Population of women were particpants in the Optifast weight loss program in Utah; the original study was a case-control design

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Weight gain type (pre-pregnancy,GWG, delivery weight, or postpartum)	Length of recall	Source of self-report	Source of "gold standard"	Gold standard rating	Mean Difference (self-reported weight - "gold standard" weight)	Correlation	Misclassification	Variation by demographic factors	Variation by pre-pregnancy BMI[†]
Nohr (2008)	GWG and pre-pregnancy weight	6 months after delivery	Maternal interview at time of study enrollment and 6mo after delivery	Maternal medical records completed by midwives upon admission for delivery	*	0.33kg (3.7kg)	-	64% of women correctly classified by GWG category	-	-
McClure et al. (2011)	GWG	4-12 years after pregnancy	Maternal interview at time of study enrollment	Maternal meidcal receords at first prenatal visit and measured or recorded weight at admissions for delivery	*	1kg (range -19-32kg)	0.63	45% of women were within 2.27kg, 33% overreported by more than 2.27kg, 22% underreported by more than 2.27kg; women who gained inadequately were more likely to overreport, and women who gained excessively were more likely to underreport. There was only fair agreement (kappa=0.43 CI: 0.37-0.49). Recalled GWG misclssified 36% of women; 55% of women who gained inadequately were correctly categorized; 47% of women who gained adequately were correctly categroized.	Did find variation by maternal characteristics: Black women were more likley to underreport by more than 2.27kg, women with less than high school education, unmarried women, and current smokers were more likely to overreport their GWG. Women whose pregnancy was w/in 8y of recall, with a reported GWG ending in 0 or 5, and who were obese at time of enrollment were more likely to overreport GWG.	No variation by pre-pregnancy BMI
Biro et al. (1999)	GWG	0-15 months after pregnancy	Maternal interview at time of study enrollment	Maternal medical records at time of delivery	*	0.54kg	0.99	-	No variation by time since birth	-
Bodnar (2014)	pre-pregnancy weight and GWG	Time between last maternal prenatal visit and delivery (at most 4 weeks)	Maternal report on birth certificate	Maternal medical records at time of first prenatal visit (sensitivity analysis using measured weight at first prenatal visit for those who had it) and last prenatal visit (measured weight taken at this visit)	*	-	-	40.9%-83.3% of women correctly classified	No variation by race/ethnicity	UW: 18.3-59.1%; NW/OW: 16.7-30%; OB I: 16.7-38.1%; OB II/III: 18.3-50% Extremes of GWG had greater misclassification: <20th percentile:16.6-50%; 20-80th percentile: 16.7-23.7%; >80th percentile: 18.3-59.1%.
Wright et al. (2010)	GWG (self-reported on birth certificate)	Time between last prenatal visit and delivery	Maternal report on birth certificate	Electronic medical records	*	-	-	48.2% of women were accurate reporters within 4.5kg; 36.6% were over-reporters, 15.2% were underreporters	Did find variation by maternal characteristic: white women were more likely to accurately report than Black women, women with private insurance were more likely to accuratly report than Medicaid or self-pay, and shorter pregnancies were less likely to accurately recall	Higher baseline BMI were more likely to over-report GWG. Normal weight women had the higheset percent of accurate reporters (78.8%)
POSTPARTUM WEIGHT RETENTION										
Rothberg (2011)	pre-pregnancy weight, pregnanacy weights during gestation (at various times), and postpartum weight	6-12 months postpartum	Maternal interiview after delivery	Maternal medical records	*	-2.66kg (12mo postpartum)	-	-	-	-

Table A3.1. Summary of Studies Included in Final Analysis, Categorized by Time During Pregnancy When Weight was Measured--CONTINUED

Author	Implications for birth or maternal health outcomes	Sample Size	Sampling approach	Participation Rate	Number of institutions included (e.g. clinics, hospitals ect.)	Years of birth cohort	Country	Micellaneous
Nohr (2008)	-	2389	Clinic-based comvenience sample of women for whom midwives were able to be contacted to report medically recorded prepregnancy weight	Not able to be calculated, but only 2389 (2.4%) of the total records for the DNBC (100419) had available data on gestational weight gain	Many, based on Danish National Birth Cohort and National Obstetrics Database for North Jutland County	1996-2002	Denmark	Overall, the data comes from a convenience sample all the births in Demark (the Danish National Birth Cohort; 100419 pregnancies to 92274 women). Data for this review was taken from both the 2008 paper as well as the first authors dissertaion (Nohr, EA. "Obesity in Pregnancy: Epidemiological studies based on the Danish National Birth Cohort. University of Aarhus, Aarhus, Denmark. 2005) which includes more detail on sample and mean difference
McClure et al. (2011)	Association between GWG (inadequate and excessive) and SGA and excessive GWG and PPWR remained the same. Using recalled GWG attenuated the relationship between inadequate GWG and PPWR and the association between excessive GWG and preterm birth.	503	Hospital-based record linkage from available birth records (all eligible women included)	32.1% (503/1569 elligible women)	One hospital in Pennsylvaia (McGee Women's Hospital)	1997-2002	USA	Study conducted as a secondary analysis within the Women and Infant Study of Healthy Hearts (WISH)
Biro et al. (1999)	-	40	Purposive sampling based on length of recall (i.e. time since delivery)	Not able to be calcuated because total elligible women were not reported	One large urban hospital (location not mentioned)	not explicitly reported (1999 most likley)	USA	Study focused on adolescents
Bodnar (2014)	-	1204	Hospital-based stratified sampling	86.1% (1204/1440 optimal sample size of eligible records; The optimal sample size was selected from total of 47233 eligible records from Magee women's hospital for balanced design)	One hospital in Pennsylvania (McGee Women's Hospital)	2003-2010	USA	Some subsample sizes in this analysis are vary small; some of the percent agreement estimates are based on samples of size 2 to 20.
Wright et al. (2010)	-	1223	Hospital-based record linkage of birth records (all sucessfully linked records included)	66.1% (1223/1851 sucessfully linked records)	One large urban academic hospital (location not mentioned)	2007	USA	
POSTPARTUM WEIGHT								
Rothberg (2011)	-	427 in study sample (unclear whether all women included in validity analysis)	convenience sample (subsample of a radomized clinical trial)	Not able to be calcluated because did not report exact N for validation study, but 84.9% (427/503) of women at the New Haven, CT site were included in the overall study sample	Multiple university affiliated hospital clinics in New Haven, CT	2001-2004	USA	This validation study was conducted as a secondary analysis of a separate analysis.

†Abbreviations for pre-pregnancy BMI categories are as follows: UW-underweight; NW-normal weight; OW-overweight; OB-obese