Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Candice Y. Johnson Date

Exposure Misclassification and Selection Bias in a Case-Control Study of Prepregnancy Body Mass Index and Neural Tube Defects

By

Candice Y. Johnson
Doctor of Philosophy

Epidemiology

_____	_____
W. Dana Flanders Advisor	Godfrey P. Oakley, Jr. Advisor
_____	_____
Margaret A. Honein Committee Member	Penelope P. Howards Committee Member
_____	_____
Sonja A. Rasmussen Committee Member	Matthew J. Strickland Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Exposure Misclassification and Selection Bias in a Case-Control Study of Prepregnancy Body Mass Index and Neural Tube Defects

By

Candice Y. Johnson
B.Sc., Carleton University, 2005
M.Sc., University of Ottawa, 2008

Advisor: W. Dana Flanders, M.D., Sc.D., Godfrey P. Oakley, Jr., M.D., M.S.P.M.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2012

Abstract

Exposure Misclassification and Selection Bias in a Case-Control Study of Prepregnancy Body Mass Index and Neural Tube Defects

By Candice Y. Johnson

We explored potential contributions of exposure misclassification, selection bias, and confounding to a study of prepregnancy body mass index (BMI) and neural tube defects (NTDs), severe birth defects of the brain and spinal cord. Over a dozen studies have found associations between BMI and NTDs, with obese mothers the most likely to have an affected pregnancy. Investigators have suggested that exposure misclassification or selection bias could account for the observed associations; however, no previous study has quantitatively addressed the potential effects of these three biases together. We investigated hypothesized mechanisms for selection bias, examined effects of making inaccurate assumptions of nondifferential or differential misclassification when adjusting for exposure misclassification, and proposed a method to simultaneously adjust for exposure misclassification, selection bias, and confounding using weighted logistic regression. Using information from these studies, we simultaneously adjusted for these three biases in a case-control study of prepregnancy BMI and two common NTDs, anencephaly and spina bifida, using data from the National Birth Defects Prevention Study. Given our assumptions, adjustment for multiple biases had little effect on associations between BMI and anencephaly. However, associations between obesity and spina bifida were attenuated following multiple bias analysis; it is possible that reported associations between obesity and spina bifida that do not take into account the potential effects of exposure misclassification or selection bias are overestimates, partially driven by bias. Although misclassification, selection bias, and confounding have the potential to affect results, multiple bias analysis remains uncommonly used. Our proposed method is one option to incorporate adjustment for multiple biases into epidemiologic studies.

Exposure Misclassification and Selection Bias in a Case-Control Study of Prepregnancy Body Mass Index and Neural Tube Defects

By

Candice Y. Johnson
B.Sc., Carleton University, 2005
M.Sc., University of Ottawa, 2008

Advisor: W. Dana Flanders, M.D., Sc.D. and Godfrey P. Oakley, Jr., M.D., M.S.P.M.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2012

ACKNOWLEDGMENTS

If I could list everyone who has in some way contributed to this dissertation, the acknowledgements section would be longer than the dissertation itself. I am indebted to everyone who has helped along the way. There are a few people who need an extra special thank you:

First and foremost, my long-suffering dissertation committee (Dana Flanders, Godfrey Oakley, Jr, Peggy Honein, Penny Howards, Sonja Rasmussen, and Matt Strickland) has had infinite patience dealing with overly ambitious projects, overly ambitious deadlines, and overdue last-minute drafts. Matt played an instrumental role in this development of this dissertation by introducing me to Peggy at the SER annual meeting in 2007 -- a small gesture with long-lasting effects. As a fellow PhD student and later as faculty, he has always been generous with his advice and his time. Peggy and Sonja suggested the topic for this dissertation, but I'm not sure if it ended up anything like what they had in mind. I am so fortunate to have had the opportunity to work with such knowledgeable and encouraging mentors (they will always be "Peggy and Sonja" to me -- I never seem to see one without the other!). Godfrey has taught me the importance of enthusiasm and persistence, but also the need to ask the right questions and tackle the right problems. He and Mary Ann also took on the very important responsibility of making sure I had some semblance of a life, inviting me out with them on a regular basis. Dana and I first got to know each other when presenting a journal club on exposure misclassification, perhaps a foreshadowing of the dissertation to come; it is an event he might have come to regret, now that he has to deal with my snarky and antagonistic comments on a regular basis. In

between our bickering, I have learned to (almost) never accept (almost) anything as truth, and to be as non-committal as possible in (almost) every response. Penny and I still can't agree on how and when we met, but I hope she will agree that working together has always been fun and productive. She might lose points on certain mentoring checklists for not promoting enthusiasm, but in doing so she has made me a better scientist.

It has been a pleasure to work with the scientists and staff of the National Birth Defects Prevention Study and the Birth Defects Branch (CDC/NCBDDD). Chris Cosper, Suzanne Gilboa, Mechele Lynch, Jennita Reefhuis, Sarah Tinker, and Tineka Yowe-Conley have proved to be a wealth of information and assistance. The NBDPS Data Sharing Committee has consistently provided excellent feedback and ideas. Bill Paradies can finally breathe a sigh of relief now that he no longer has to process my security clearance. Kim Waller somehow managed to keep up with us, even though 90% of the time I forget to copy her on relevant emails.

The Department of Epidemiology at Emory University has been my home away from home for the past 5 years. Pam Mink and Michael Kramer have been (without their informed consent) fabulous surrogate committee members. Jena Black, Bridgett Figueroa, Deanna Murray, Missy Page, and Janet Qi have been always ready with an answer or a helping hand. The department is made up of a truly wonderful group of faculty, staff, and students who I have enjoyed getting to know during my time here.

The FUCHSIA Women's Study team allowed me to fulfill my data collection requirements with them, eat their snacks, and force upon them the wonders of Canadian marshmallow cookies.

Proving that lunch is truly the most important meal of the day, the "Lunch Bunch" (Matt Magee, Matt Page, and Rachel Patzer) has been an invaluable source of culinary exploration, advice, and procrastination.

The Laney Graduate School provided the fellowship that allowed me the flexibility to develop a dissertation not tied to a faculty member's grant and the peace of mind of not having to worry about finding financial support for 5 years.

Thank you all!

Although they say it takes a village to raise a child, it appears to have taken a moderately-sized metropolitan area to graduate this PhD student.

TABLE OF CONTENTS

Chapter	Title	Page
1	Introduction	1
2	Epidemiologic Evidence for an Association Between Maternal Obesity and Neural Tube Defects: a Systematic Review and Meta-Analysis of the Published Literature	37
3	Pregnancy Termination Following Prenatal Diagnosis of Anencephaly or Spina Bifida: a Systematic Review of the Literature	66
4	Prenatal Diagnosis of Spina Bifida	93
5	Potential Sensitivity of Bias Analysis Results to Incorrect Assumptions of Nondifferential or Differential Exposure Misclassification	127
6	Weighted Logistic Regression for Multiple Bias Analysis	162
7	Potential Impacts of Biases on Associations Between Prepregnancy Body Mass Index and Neural Tube Defects	216
8	Discussion and Conclusions	258

LIST OF TABLES

Table	Title	Page
1.1	Study Sites, Years of Case and Control Recruitment, and Study Catchment Areas for the National Birth Defects Prevention Study, 1997-2007	22
1.2	Source of Control Mothers in the National Birth Defects Prevention Study, 1997-2007	23
1.3	Associations Between Maternal Prepregnancy Obesity and Neural Tube Defects Reported by the National Birth Defects Prevention Study and a Recent Meta-Analysis	24
2.1	Electronic Search Strategy Used to Identify Articles From Embase and Medline Through the end of July 2010	46
2.2	Characteristics of Studies Included in the Meta-Analysis of Maternal Obesity and Neural Tube Defects	47
2.3	Results from Random-Effects Meta-Analyses of Associations Between Maternal Obesity and Neural Tube Defects	51
2.4	Comparison of Results From Random-Effects Meta-Analyses of Associations Between Maternal Obesity and Neural Tube Defects Using two Different Reference Groups for Exposure	52
2.5	Influence of Individual Studies on Results of Random-Effects Meta-Analyses of Obesity (Obesity Versus Normal Weight) and Anencephaly or Spina Bifida	53

Table	Title	Page
2.6	Comparison of Adjusted to Crude Analyses of Maternal Obesity and Anencephaly or Spina Bifida	54
2.7	Summary of Results From Three Meta-Analyses of Obesity and Neural Tube Defects (Obese Versus Normal Weight)	56
3.1	Proportion of Pregnancies Ending in Termination of Pregnancy Following Prenatal Diagnosis of Anencephaly or Spina Bifida, by Geographic Region	81
3.2	Proportion of Pregnancies Affected by Spina Bifida Ending in Termination of Pregnancy Following Prenatal Diagnosis, by Fetal and Study Characteristics	83
3.3	Proportion of Pregnancies Affected by Spina Bifida Ending in Termination of Pregnancy Following Prenatal Diagnosis by Gestational Age at Prenatal Diagnosis	85
4.1	Agreement Between Maternal Report and Medical Record Abstract of Prenatal Diagnosis of Spina Bifida	110
4.2	Agreement Between Maternal Report and Medial Record Abstract of Whether Prenatal Diagnosis of Spina Bifida was Made <24 or ≥ 24 Weeks of Gestation	111
4.3	Frequency of Maternal Report of Prenatal Diagnosis of Spina Bifida, National Birth Defects Prevention Study, 1998-2005	112

Table	Title	Page
4.4	Frequency of Maternal Report of Prenatal Diagnosis and Early Prenatal Diagnosis of Spina Bifida, by National Birth Defects Prevention Study Site, 1998-2005	113
4.5	Maternal Characteristics Investigated in Association With Maternal Report of Prenatal Diagnosis of Spina Bifida	114
4.6	Pregnancy-Related Characteristics Investigated in Association With Maternal Report of Prenatal Diagnosis of Spina Bifida	116
4.7	Maternal Characteristics Investigated in Association With Maternal Report of Timing of Prenatal Diagnosis of Spina Bifida	118
4.8	Pregnancy-Related Characteristics Investigated in Association With Maternal Report of Timing of Prenatal Diagnosis of Spina Bifida	121
5.1	Contingency Table for True (Correctly Classified) and Observed (Misclassified) Exposures	143
5.2	Ratio of Misclassification-Adjusted Odds Ratio to True Odds Ratio Over 1,000 Iterations of Simulation When Making Correct and Incorrect Assumptions About Nondifferential and Differential Misclassification	144
5.3	Results From 1,000 Iterations of Probabilistic Adjustment for Exposure Misclassification Under Various Assumptions in a Study of Obesity and Diabetes, National Health and Nutrition Examination Survey, 1999-2008	146

Table	Title	Page
5.4	Contingency Tables and True Values of Sensitivity and Specificity of Obesity Classification for a Study of Self-Reported and Measured Obesity and Diabetes, Non-Pregnant Females Aged 18-49, National Health and Nutrition Examination Survey, 1999-2008	148
5.5	Results of Adjustment for Exposure Misclassification in a Study of Obesity and Diabetes using Estimates of Sensitivity and Specificity From Literature Review, Non-Pregnant Females Aged 18-19, National Health and Nutrition Examination Survey, 1999-2008	149
5.B1	Identified Published Studies Reporting Sensitivity and Specificity of Self-Reported Obesity Status in American Adult Females of Reproductive Age	154
5.B2	Triangular Distributions for Sensitivity and Specificity of Obesity Classification for Non-Cases, Created From Literature Review	157
6.1	Contingency Tables for the Source and Selected Populations and the Observed Data, Stratified by a Confounder: Simulated Data	183
6.2	Exposure Classification Table From the Selected Population, Stratified by Disease and Confounder: Simulated Data	184
6.3	Adjustment for Exposure Misclassification in the Observed Data: Simulated Data	185
6.4	Adjustment for Selection Bias in the Selected Population: Simulated Data	187

Table	Title	Page
6.5	Example of Predictive Values for Exposure Classification, Non-Pregnant Hispanic Women Aged 16-49, National Health and Nutrition Examination Survey, 1999-2008	188
6.6	Associations Between Prepregnancy Obesity and Cleft Lip With or Without Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007	189
6.7	Associations Between Prepregnancy Obesity and Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007	191
6.A1	The Source and Selected Populations	200
6.A2	Observed Data With Exposure Misclassification and Selection Bias. This Represents Data Usually Available in Epidemiologic Studies	201
6.A3	Exposure Classification Table	202
6.A4	Adjustment for Exposure Misclassification	203
6.A5	Adjustment for Selection Bias in Addition to Exposure Misclassification	204
6.A6	Exposure Weights Using Sensitivity and Specificity	205
6.B1	Associations Between Prepregnancy Underweight and Overweight and Cleft Lip With or Without Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007	206

Table	Title	Page
6.B2	Associations Between Prepregnancy Underweight and Overweight and Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007	209
7.1	Characteristics of Participants, National Birth Defects Prevention Study, 1997-2007	235
7.2	Confounding-Adjusted Associations Between Prepregnancy Obesity and Anencephaly and Spina Bifida in Study Sites Including and Excluding Cases Among Terminations of Pregnancy, National Birth Defects Prevention Study, 1997-2007	238
7.3	Associations Between Prepregnancy Body Mass Index and Anencephaly, Adjusted for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.	239
7.4	Associations Between Prepregnancy Body Mass Index and Spina Bifida, Adjusted for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.	241
7.A1	Ratio of Stillbirths and Terminations of Pregnancy to Livebirths by Study Site, National Birth Defects Prevention Study, 1997-2007	246
7.B1	Hypothetical Data Showing an Example of a Classification Table for Body Mass Index Categories	251

LIST OF FIGURES

Figure	Legend	Page
1.1	Directed acyclic graph illustrating a potential mechanism for selection bias in studies of prepregnancy body mass index and neural tube defects	25
2.1	Flow of studies into and out of the systematic review of maternal obesity and neural tube defects	57
2.2	Forest plot of studies reporting associations between maternal obesity (obese compared to normal weight mothers) and anencephaly, with summary estimate from random-effects meta-analysis	58
2.3	Forest plot of studies reporting associations between maternal obesity (obese compared to non-obese mothers) and anencephaly, with summary estimate from random-effects meta-analysis	59
2.4	Forest plot of studies reporting associations between maternal obesity (obese compared to normal weight mothers) and spina bifida, with summary estimate from random-effects meta-analysis	60
2.5	Forest plot of studies reporting associations between maternal obesity (obese compared to non-obese mothers) and spina bifida, with summary estimate from random-effects meta-analysis	61
5.1	Creation of the 12 datasets for simulation	152
6.1	Directed acyclic graph illustrating relationships between exposure, disease, confounder, and selection	193

Figure	Legend	Page
6.2	Estimation of bias parameters when adjusting for both exposure misclassification and selection bias	194
7.1	Directed acyclic graph illustrating assumed relationships between variables in a study of prepregnancy body mass index and neural tube defects	243

CHAPTER 1

Introduction

OBESITY, NEURAL TUBE DEFECTS, AND EPIDEMIOLOGIC BIAS

Obesity is a well-recognized risk factor for chronic diseases such as type 2 diabetes and cardiovascular disease (1). The potential harmful effects of obesity during pregnancy are now receiving considerable attention given the growing number of women entering pregnancy who are classified as obese (2).

Neural tube defects (NTDs) are serious defects of the brain or spinal cord consistently found to be associated with prepregnancy obesity. Over a dozen epidemiologic studies have reported that obese mothers are almost twice as likely as normal weight mothers to have a child with an NTD (3, 4). Although the consistency of study results suggests that prepregnancy obesity might cause NTDs, investigators have also hypothesized that the association could be non-causal and attributable to bias (5-7). In most studies of prepregnancy obesity and NTDs, investigators have addressed potential bias from confounding using statistical methods such as multivariable modeling. However, the potential effects of other biases such as exposure misclassification and selection bias have infrequently been addressed quantitatively.

In epidemiologic studies of birth defects, body mass index (BMI) is a widely used measure of obesity and is commonly categorized into 4 groups: underweight (BMI < 18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2), and obese ($\geq$30 kg/m^2). BMI is often calculated from maternal report of height and prepregnancy weight;

however, these are sometimes inaccurately self-reported, leading to exposure misclassification (5, 8). Previous studies have shown that on average American women tend to over-report height and under-report weight, resulting in underestimation of BMI; however, the direction of error is also dependent on the woman's true BMI (9, 10). Because obesity is defined using a categorized form of an imprecisely measured continuous variable, BMI, the expected direction of bias from misclassification is difficult to predict (11, 12).

In studies of prepregnancy obesity and NTDs, selection bias is thought to arise from differences in the likelihood of prenatal diagnosis and pregnancy termination between obese and non-obese mothers (7). The directed acyclic graph in Figure 1.1 illustrates the pathway through which this potential source of bias might occur. An abnormal result from maternal serum alpha-fetoprotein (MSAFP) screening is often the first indication that a fetus has an open NTD, but ultrasonography is commonly used for both screening and diagnosis of these defects. Visualization of fetal structures and prenatal detection of birth defects by ultrasound are more difficult in obese than in non-obese mothers (13, 14). Following prenatal diagnosis of an NTD, many pregnancies end in termination (15). Because obese mothers might be less likely to have the defect detected prenatally, they might also be more likely to carry the pregnancy to term than non-obese mothers, who might have opted for termination of pregnancy following prenatal diagnosis. If all cases of NTDs among terminations were included in epidemiologic studies, no bias would result; however, fetuses with NTDs among terminations are difficult to ascertain, and are often incompletely ascertained or excluded from studies for this reason. As a result, non-

obese mothers might be under-represented among ascertained NTD case mothers, leading to a spurious association. Adjustment for selection bias is simple if selection probabilities for each exposure-outcome combination are available; however, selection probabilities, or reasonable estimates, are usually unknown (16).

No study to date has presented a quantitative assessment of the potential effects of both exposure misclassification and selection bias on results from a study of prepregnancy BMI and NTDs. The overarching goal of this dissertation is to quantitatively evaluate the contributions of these biases, in addition to confounding, to associations between prepregnancy BMI and the two most common NTDs, anencephaly and spina bifida, and to investigate if the previously observed associations might be explained by bias.

PREPREGNANCY OBESITY

It is estimated that in 2009-2010 over one-third of American women aged 20-39 were obese (17). In the United States, the prevalence of self-reported prepregnancy obesity among mothers delivering liveborn infants has been estimated at 22% using data from the Behavioral Risk Factor Surveillance System in 9 states in 2003 (2) and 19% using data from the Pregnancy Risk Assessment Monitoring System in 26 states and New York City in 2004 and 2005 (18).

Epidemiologic studies have reported associations between prepregnancy obesity and preterm birth (19, 20), gestational diabetes (21), stillbirth (22), birth defects (4), and other adverse pregnancy outcomes and complications (23). Clinical guidelines now recommend

that women achieve a healthy weight prior to pregnancy to prevent these adverse outcomes (23-25).

NEURAL TUBE DEFECTS

NTDs occur during the third and fourth week of embryonic development if the neural tube, precursor of the central nervous system, fails to fully close (26). The prevalence of NTDs varies geographically, although it is difficult to determine the extent to which this represents differences in occurrence versus differences in use of prenatal diagnosis and pregnancy termination or differences in coding or case classification (27). The two most common types of NTDs are anencephaly and spina bifida. Rarer NTDs include encephalocele, a herniation of brain tissue through a defect in the skull; craniorachischisis, a lethal failure of neural tube closure at multiple points which results in an exposed brain and spinal cord (combination of anencephaly and spina bifida); and iniencephaly, a typically fatal defect of the spine and brain that results in a shortened or absent neck (28-30).

Anencephaly

Anencephaly is thought to result from a defect in closure of the neural tube at the cranial end of the developing embryo during the fourth week of gestation (26, 31). It is characterized by lack of skull bones and scalp above the level of the eyes and absence of most of the brain tissue. The abnormally developing brain is thought to degenerate after continued exposure to the amniotic fluid so that at the time of birth anencephalic infants have only small amounts of brain tissue remaining (26). Anencephaly is a lethal defect,

and if a prenatal diagnosis is made, pregnancy termination is the most common outcome of pregnancy (15). If pregnancy termination is not sought, intrauterine death is common. Liveborn infants might survive for hours or days, and rarely to several weeks (32, 33).

Spina Bifida

Failure of the neural tube to close at the caudal end of the neural tube is believed to result in spina bifida (31). The two main types are spina bifida occulta and spina bifida cystica. Spina bifida cystica is further classified as meningocele and myelomeningocele. Spina bifida occulta involves incomplete formation of the vertebrae only, and because it is typically asymptomatic it is excluded from epidemiologic studies of spina bifida. Spina bifida cystica (henceforth referred to simply as "spina bifida") is a more severe form of the defect. Both meningocele and myelomeningocele involve herniation of a sac containing meninges and cerebrospinal fluid through the incompletely formed vertebrae (26). In myelomeningocele, the spinal cord is also involved in the herniation and the defect is frequently not covered by skin ("open" spina bifida) (28).

In the United States, it is estimated that over 90% of liveborn infants with spina bifida survive the first year of life, and nearly 80% of affected children survive to adulthood (34). Herniation of neural tissues, particularly when the spinal cord is involved, results in various degrees of paralysis and mobility impairment. Myelomeningocele is also associated with anomalies of the brain known as the Arnold-Chiari II malformation, in which a part of the cerebellum is positioned downward out of the skull and toward the spine (35). Although the Arnold-Chiari II malformation can be asymptomatic, mortality

rates are high among infants and young children who exhibit symptoms such as difficulty breathing and swallowing (36). Hydrocephaly is frequently observed in spina bifida patients and cerebrospinal fluid shunts are placed for its treatment. Surgeries to replace malfunctioning shunts and shunt infections are common complications of shunt placement (36). Both urinary and bowel incontinence affect the quality of life of many individuals with spina bifida and neurogenic bladder can lead to kidney disease and end-stage renal failure (36, 37).

Although the herniation of nervous tissue itself is believed to be responsible for much of the paralysis and mobility impairment observed in patients with myelomeningocele, further damage to the unprotected spinal cord is thought to occur *in utero* from continued exposure of the spinal cord to the amniotic fluid during fetal development (36). Recently, a randomized controlled trial of *in utero* spina bifida repair has been completed (38). Compared to infants receiving postnatal intervention only, infants randomized to fetal surgery had a lower incidence of death and shunt placement, and in childhood had better cognitive and motor function. However, prenatal surgery resulted in a high incidence of complications to both mother and child, including preterm birth, oligohydramnios, and separation of the chorioamniotic membrane.

RISK FACTORS FOR NEURAL TUBE DEFECTS

Several risk factors for NTDs have been identified, including genetic syndromes and use of certain anticonvulsants, such as valproic acid, during early pregnancy (39, 40),,the risk factor that has received the most attention and that has been translated into successful

population-level intervention is folic acid supplementation. Results from randomized controlled trials in the 1980s and 1990s demonstrated that folic acid supplementation during the periconceptional period could prevent a substantial proportion of NTDs (41-43). Since then, folic acid fortification programs have been introduced in over 50 countries. Studies comparing the pre- and post-fortification prevalence of NTDs in these areas have found reductions in prevalence following fortification (44-48).

In the United States in 2004-2006, the birth prevalence of NTDs was estimated to be over 6/10,000 live births in the presence of folic acid fortification (49). Because the proportion of NTDs attributable to genetic syndromes, maternal use of anticonvulsants, or other recognized causes remains small, investigators continue to search for risk factors for NTDs to better understand their etiology and develop strategies for their prevention.

PREPREGNANCY OBESITY AND NEURAL TUBE DEFECTS

Epidemiologic Evidence

Numerous epidemiologic studies have found that obese mothers are more likely to have an NTD-affected pregnancy than normal weight mothers. Meta-analyses (3, 4) summarizing the evidence have come to this same conclusion and have found that the association is stronger for spina bifida (odds ratio [OR] 2.24, 95% confidence interval [CI] 1.86, 2.69) than anencephaly (OR 1.39, 95% CI 1.03, 1.87) (4). Both meta-analyses investigated effects of study-level variables on meta-analysis results (study design, study size, year of publication, control for confounders, inclusion of cases among terminations

of pregnancy, exclusion of mothers with pregestational diabetes); however, there was no evidence that these variables affected the observed associations (3, 4).

Hypothesized Mechanisms Explaining the Association Between Maternal Obesity and NTDs

The mechanism whereby maternal obesity might cause NTDs remains unknown. Several lines of research have been investigated, most focusing on differences in nutritional status between obese and normal weight women. Given the known role of folate and folic acid in NTD prevention, study of the differences in folic acid supplementation and folate intake and metabolism between obese and non-obese women has been an area of specific research interest. Several studies have reported that obese women of childbearing age are less likely to consume multivitamin supplements (50, 51), have lower dietary folate consumption (50, 52), and have lower serum folate levels (50) than normal weight women.

Prepregnancy obesity has been reported to be associated with a generally poorer quality of diet during pregnancy (52). Intakes of sweets (53), sugars (54), and foods with high glycemic index or glycemic load (54-56) during the periconceptional period have been reported to be higher among mothers of NTD cases than mothers of infants without birth defects. Investigators have suggested that insulin and glycemic control might be the mechanisms underlying the reported associations of both obesity and diabetes with NTDs, although how hyperinsulinemia or hyperglycemia might cause NTDs is not understood (54, 55, 57).

The aforementioned non-causal mechanisms (misclassification of BMI category, selection bias through prenatal diagnosis and termination of pregnancy) have also been proposed to explain the association between obesity and NTDs (5-7). However, to date this possibility has not been explored in depth.

BIAS ANALYSIS

The three main threats to internal validity of epidemiologic studies are confounding, information bias (measurement error and misclassification), and selection bias (16). Although confounding is the bias whose effect is most commonly explored in epidemiologic studies, measurement error and selection bias could have important effects on results. Quantitative adjustment for the effects of these biases is rarely presented (58, 59).

Bias (sensitivity) analysis is a technique that allows investigators to make assumptions about the magnitude and direction of biases thought to be occurring, and to use these assumptions to quantitatively explore the potential effect of these biases on results of a study (59). Bias analyses can have varying levels of complexity (60). A simple bias analysis involves choosing one or a few plausible values for bias parameters (selection probabilities for analyses of selection bias; sensitivity, specificity or predictive values for misclassification; strength of association with exposure and disease for unmeasured confounding) and using these parameters to make an “adjustment” for the bias. Probabilistic bias analysis is an extension of simple bias analysis in which a probability

distribution is created for the bias parameter. A value is randomly sampled from this distribution and used in the analysis to adjust for bias and this process is repeated over many iterations to generate a distribution of the adjusted measure of association. Multiple bias analysis is a further extension of the bias analyses described above that involves taking into account, in the same analysis, potential effects of more than one type of bias (60, 61).

Many different approaches for conducting bias analysis have been described, including methods applying formulae for bias adjustment to data from contingency tables (59), adjusting observed measures of association by multiplication with error terms (62, 63), and making record-level adjustments (as opposed to contingency table-level adjustments) to simulate the "true" data for individual study participants (64). Although the results of sensitivity analyses are highly dependent on the choice of bias parameters, evidence-based choices and transparent reporting of methods and results allow readers to judge for themselves the appropriateness of the assumptions made (59).

Quantitative Analyses for Exposure Misclassification and Selection Bias: Previous Studies of Prepregnancy BMI and NTDs

To date, no study of prepregnancy BMI and NTDs has presented quantitative assessment of the potential impact of misclassification of BMI on study results. Two studies have attempted to quantify the direction and magnitude of selection bias in case-control studies of prepregnancy obesity and NTDs.

A simple bias analysis was presented for an association between obesity and spina bifida in a study from the Atlanta Birth Defects Case-Control Study (7). The authors assumed that obese mothers were 17% less likely to receive a prenatal diagnosis than non-obese mothers and that 26% of affected pregnancies ended in termination. They concluded that selection bias could result in an overestimation of the strength of association. Given their assumptions, the crude OR for obesity and spina bifida decreased from 2.56 to 2.02 after adjustment for selection bias.

A similar bias analysis was conducted using data from the National Birth Defects Prevention Study (5). The authors assumed that the true proportion of pregnancies ending in termination was twice that observed in the data. They found little effect on results, given their assumptions. The crude OR for obesity and spina bifida (obese versus normal weight) decreased from 2.25 to 2.12 after adjustment for selection bias, and the crude OR for anencephaly changed from 0.97 to 0.94.

In both of these bias analyses, a single scenario was considered. Because the results of bias analyses are dependent on assumptions made in the analysis and on the bias parameters used as inputs, consideration of a wider range of plausible scenarios using probabilistic bias analysis could give a better sense of the variability in magnitude of bias that might reasonably occur (60). Such an analysis has not been conducted to date. In addition, no study has simultaneously combined adjustment for exposure misclassification, selection bias, and confounding to investigate the potential joint effects of these three biases on results.

SOURCE OF DATA FOR THE DISSERTATION

The National Birth Defects Prevention Study (NBDPS) is a large, multi-site case-control study of genetic and environmental risk factors for birth defects, with participant enrollment at multiple sites in the United States beginning October 1997 and continuing to date. Detailed descriptions of NBDPS study design are available elsewhere (65-68).

NBDPS ascertains cases of over 30 types of major structural birth defects from existing population-based birth defect surveillance systems in 10 states which have active case-finding methods and use multiple sources of ascertainment: Arkansas, California, Georgia, Iowa, Massachusetts, New Jersey, New York, North Carolina, Texas, and Utah (Table 1.1). Eight of the 10 study sites systematically ascertain cases of birth defects among live births, stillbirths (fetal deaths ≥20 gestational weeks), and terminations of pregnancy (any gestational age); New York began case ascertainment among terminations of pregnancy in 2000. New Jersey only systematically includes live births and stillbirths, and Massachusetts only systematically includes live births.

Mothers who are residents of the study catchment area at the time of the birth and whose child has an NBDPS-eligible birth defect are invited to serve as case mothers for the study. Cases of birth defects believed to be caused by single-gene conditions or chromosome abnormalities are excluded by clinical geneticists who perform a detailed review of data abstracted from medical records of every potential case to determine eligibility, and if eligible, to further classify the case (68). Cases can be classified as

isolated (no other major unrelated birth defect), multiple (one or more other, unrelated defects), or as part of a sequence or association (several defects believed to be related) (68).

Methods for control selection vary by study site, but in all sites only liveborn infants with no major birth defect and whose mothers are residents of the study catchment area at the time of the birth are eligible. Controls are randomly selected from either birth certificates or birth hospitals (Table 1.2). Between 1997 and 2003, control participation rates ranged from 65% to 77% across sites, with some differences in maternal characteristics such as race/ethnicity, education, and timing of entry into prenatal care observed between control participants and the source population (66).

Between 6 weeks and 24 months following the estimated date of delivery, mothers of eligible cases and controls are contacted by mail and are provided an introductory letter explaining the study. Trained interviewers contact the mothers by telephone to administer the computer-assisted telephone interview, which includes questions on sociodemographic factors, illnesses, medication use, prenatal care, household and occupational exposures, and family history of birth defects, as well as a food frequency questionnaire. Following the interview, mothers are invited to participate in the genetic component of the study and are sent a buccal (cheek) cell collection kit to sample DNA from themselves, their baby (if living) and the baby's father (67).

Prepregnancy BMI and NTDs in NBDPS Data

NBDPS data on associations between maternal prepregnancy BMI and NTDs have been published previously; results from a simple bias analysis to investigate the impact of selection bias using NBDPS data were described in the previous section (5). Consistent with the results of other studies and meta-analyses, the analysis suggested that obese mothers are approximately twice as likely to have a pregnancy affected by spina bifida as normal weight mothers (Table 1.3). In contrast to several other studies, no association with anencephaly was observed. The association between prepregnancy BMI and anencephaly has been less consistently observed in the literature (4).

The biases hypothesized to be contributing to associations between prepregnancy BMI and NTDs are also thought to occur in NBDPS. Selection bias is possible because some NBDPS study sites exclude pregnancy terminations from their case definition; even among sites including terminations, case ascertainment among pregnancy terminations is believed to be incomplete. NBDPS collects maternal height and prepregnancy weight by self-report up to 2 years following the mother's expected date of delivery, making misclassification of BMI category possible.

Our overarching goal is to re-analyze NBDPS data on prepregnancy BMI and NTDs, conducting a probabilistic multiple bias analysis taking into account exposure misclassification, selection bias, and confounding to determine how adjusting for these biases might affect study results. Prior to conducting the bias analysis, we will need sufficient information on the strength of associations between variables of interest to assign plausible probability distributions to bias parameters. In particular, we will need to

determine whether or not prenatal diagnosis of NTDs is less common in obese mothers than normal weight mothers, and what proportion of pregnancies end in termination of pregnancy following prenatal diagnosis of NTDs.

EPIDEMIOLOGY OF PRENATAL DIAGNOSIS AND TERMINATION OF PREGNANCY

Prenatal Screening for and Diagnosis of Neural Tube Defects

Prenatal screening for NTDs has been possible since the early 1970s when it was first found that high amniotic fluid levels of alpha-fetoprotein (AFP) were predictive of open NTDs (NTD types in which the brain or spinal cord come into contact with the amniotic fluid). Several years later, screening programs using AFP measured from maternal serum were introduced (69-71). In the years following introduction of serum screening programs, ultrasonography has become widely used for prenatal detection of birth defects, and the continued need for serum screening for NTDs in centers in which all women are also offered ultrasonography has been questioned (72). Although the performance of ultrasound-only screening for NTDs has not been well-studied, in some centers serum screening has been discontinued in favor of ultrasound (72-74).

Anencephaly

Anencephaly might be suspected when elevated MSAFP levels are detected through serum screening. Prenatal diagnosis of anencephaly can be easily achieved by ultrasonography because of the conspicuous absence of skull bones that characterizes this defect. Because skull bones do not fully develop until the 12th gestational week,

anencephaly cannot be reliably prenatally diagnosed by ultrasound before then (75). A multi-site study from the EUROCAT registry in 18 regions of Europe estimated that 96% (range: 83-100%) of fetuses with anencephaly in the registry had been first diagnosed prenatally (76). In many parts of Europe, mothers have better access to and uptake of prenatal care than in other regions such as the United States, meaning that the proportion of anencephalic fetuses detected prenatally in the Untied States might be substantially lower than in Europe, despite how readily these defects are seen on ultrasound.

Spina bifida

Elevated MSAFP levels could also indicate the presence of open spina bifida, but screening for closed spina bifida (when the defect is covered by skin) must be accomplished through ultrasonography. Ultrasound screening for spina bifida is achieved in two main ways. The first is by observing changes in the shape of the skull and cerebellum commonly seen in affected fetuses between the 13^{th} and 24^{th} gestational weeks; these ultrasound signs are referred to the “lemon sign” and “banana sign” (75). The second is by visualization of the defect itself. Ultrasound visualization of the defect can be useful for providing information about its severity. The location of the defect on the spine is fairly predictive of the level of impairment, with defects lower on the spine associated with less physical impairment than defects higher on the spine (75). Although prenatal diagnosis of spina bifida is typically fairly common, a EUROCAT study found more variability in the proportion of spina bifida cases prenatally diagnosed than anencephaly, estimating that 68% of all cases of spina bifida in the registry had been

prenatally diagnosed (range: 38-100%) (76). As discussed above, the prevalence of prenatal diagnosis might be higher in Europe than other regions.

Maternal Obesity and Prenatal Detection of Birth Defects

Difficulty in ultrasound visualization of the fetus among obese mothers is widely recognized (77). Images from ultrasound are less clear in obese mothers because extra adipose tissue on the abdomen means a greater distance between the ultrasound transducer and the fetus (78). Studies attempting to quantify differences in suboptimal visualization of the fetus between obese and normal weight mothers have consistently found greater likelihood of suboptimal visualizations in mothers with greater BMI.

Visualization of the fetal anatomy

Standard ultrasound examination of the fetal anatomy can be conducted reliably after 18 weeks' gestation and involves visualization of the head, face, neck, heart, abdomen, spine and limbs (79). A study of prenatal ultrasounds performed in the late 1980s suggested that visualization was poorest among mothers in the 90^{th} percentile of BMI (> 36 kg/m^2) (78). Similar results were shown in studies from Texas, in which inadequate visualization occurred in 28% of normal weight mothers compared to 43-70% of obese mothers (13), and from Toronto, in which the fetal anatomical survey was inconclusive in 3% of normal weight but 26% of obese mothers (80).

The same trends are shown for visualization of specific parts of the fetal anatomy. A multicenter study of visualization of ultrasound markers that might indicate genetic

disorders found that most markers were more difficult to visualize among obese mothers (77). For example, inadequate estimation of nuchal translucency occurred in 3-8% of obese mothers but only 1% of normal weight mothers. Suboptimal visualization of the fetal heart was observed more commonly among obese compared to normal weight mothers in studies from Michigan (37% versus 19%) (81) and Ohio (51% versus 27%) (82). It has been suggested that better visualization of the fetal heart might occur when the ultrasound is repeated later in the pregnancy, although it is estimated that 12-17% of obese mothers will have suboptimal visualization on both ultrasounds (83).

Because ultrasound visualization of fetal structures appears to be more difficult among obese mothers, one study in Michigan compared two types of ultrasound equipment (standard equipment versus a more advanced system) to determine if visualization could be improved through use of more advanced technology. Although results suggested that the advanced ultrasound equipment improved visualization of the fetal heart among non-obese women, among obese women visualization was generally similar using both instruments (84).

Detection of birth defects

Because visualization of the fetal anatomy is reduced for obese mothers, studies have investigated whether or not obesity also affects the likelihood of prenatal diagnosis. Most studies follow a retrospective study design in which records of cases with postnatally confirmed birth defects are reviewed to determine what proportion had been prenatally diagnosed by ultrasound, and if this proportion differed between obese and normal

weight mothers. One study from Texas found that 25-49% of cases had been prenatally diagnosed among obese mothers, compared to 66% among normal weight mothers (14). The investigators reported similar results when restricting to targeted ultrasounds (ultrasounds conducted for high-risk pregnancies or when anomalies are suspected on routine ultrasound): 75-88% of birth defects had been detected prenatally among obese mothers compared to 97% in normal weight mothers. One multicenter study from the United States showed that obese mothers were less likely than normal weight mothers to report prenatal diagnosis of cleft lip with or without cleft palate (adjusted OR 0.59, 95% CI 0.41-0.85) (85).

Of the studies conducted to date, most have combined different types of birth defects in the analysis and none has investigated how maternal obesity might affect prenatal diagnosis of NTDs specifically. Unlike other birth defects such as cleft lip with or without cleft palate in which ultrasound is the only method for prenatal screening and diagnosis, anencephaly and open spina bifida might be suspected prior to ultrasound on the basis of MSAFP screening. In addition, some birth defects are more readily seen on ultrasound than others. For defects such as cleft lip with or without cleft palate, which are difficult to see on ultrasound even among normal weight mothers, prenatal diagnosis in obese mothers could be challenging. For defects such as anencephaly, however, presence of the defect might be noticeable enough that obesity would not affect ability to make a prenatal diagnosis. For these reasons, we expect that the association between obesity and prenatal diagnosis might differ among defects.

Pregnancy Termination Following Prenatal Diagnosis of NTDs

When an NTD is prenatally diagnosed, termination of pregnancy is a common pregnancy outcome. One review article estimated that over 80% of pregnancies in which anencephaly was prenatally diagnosed ended in termination, as did over 60% of those in which spina bifida was detected prenatally (15). Most of the recent studies of pregnancy termination following prenatal diagnosis have been conducted in Europe. In the few studies conducted in the United States over the last 20 years, the proportions of pregnancies ending in termination have been on average lower than estimates from Europe (86-88). The reasons for these differences have not been well explored. It is possible that in Europe, families choose termination of pregnancy more often than in North America. However, these results could also be explained by mothers in Europe having better access to prenatal care and earlier prenatal diagnosis. If prenatal diagnosis is made late in pregnancy, termination is often no longer an available option.

Little is known about what maternal characteristics (e.g., race/ethnicity, household income, education) are associated with having a prenatal diagnosis and having a prenatal diagnosis early in pregnancy. Understanding factors associated with prenatal diagnosis will help to identify subgroups of women who might not have access to or are not using prenatal diagnostic services, or who use these services, but not until late in the pregnancy. Prenatal diagnosis not only allows family to decide whether or not to continue the pregnancy, but provides an opportunity for the family to learn about the defect, discuss options for fetal or postnatal surgical intervention, plan for the delivery to occur at a

facility better equipped to treat the newborn, and make other decisions in anticipation of the infant's hospital stay.

CONTRIBUTIONS OF THE DISSERTATION

This dissertation will provide quantitative evaluation of evidence concerning bias in studies of prepregnancy BMI and NTDs, fill gaps in the literature regarding the epidemiology of prenatal diagnosis and pregnancy termination, and demonstrate a method to simultaneously adjust for exposure misclassification, selection bias, and confounding using logistic regression. An important result will be providing bias-corrected estimates of the association, if any, between prepregnancy BMI and NTDs, and characterizing the associated uncertainty in these estimates. Additional evidence concerning whether or not this association is attributable to bias will aid health care professionals to more accurately counsel women about their risk of having a pregnancy affected by an NTD.

Table 1.1. Study Sites, Years of Case and Control Recruitment, and Study Catchment Areas for the National Birth Defects Prevention Study, 1997-2007.

Study Site	Recruitment Years	Catchment Area
Arkansas	1997 – 2007	State-wide
California	1997 – 2007	8 counties
Georgia	1997 – 2007	Five-county metropolitan Atlanta
Iowa	1997 – 2007	State-wide
Massachusetts	1998 – 1999	State-wide
	2000 – 2007	State-wide except 5 western counties
New Jersey	1997 – 2002	Random sample of births state-wide
New York	1997 – 2007	Western New York, lower Hudson valley
North Carolina	2003 – 2007	19 counties
Texas	1997 – 1998	State-wide except Houston, Galveston, Nacogdoches, Beaumont
	1998 – 2001	San Antonio, West Texas, Panhandle
	2002 – 2003	West Texas, Panhandle
	2004 – 2007	South Texas
Utah	2003 – 2007	State-wide

Table 1.2. Source of Control Mothers in the National Birth Defects Prevention Study, 1997-2007.

Study Site	Recruitment Years	Source of Control Mothers
Arkansas	1997 – 2000	Hospitals
	2001 – 2007	Birth certificates
California	1997 – 2007	Hospitals
Georgia	1997 – 1999	Hospitals
	2000 – 2007	Birth certificates
Iowa	1997 – 2007	Birth certificates
Massachusetts	1997 – 2007	Birth certificates
New Jersey	1997 – 2002	Birth certificates
New York	1997 – 2007	Hospitals
North Carolina	2003 – 2007	Hospitals
Texas	1997 – 2007	Hospitals
Utah	2003 – 2007	Birth certificates

Table 1.3. Associations Between Maternal Prepregnancy Obesity and Neural Tube Defects Reported by the National Birth Defects Prevention Study (5) and a Recent Meta-Analysis (4).

	Odds Ratio (95% Confidence Interval)	
Neural Tube Defect	NBDPS[a]	Meta-Analysis[b]
Anencephaly	0.96 (0.62, 1.48)	1.39 (1.03, 1.87)
Spina bifida	2.10 (1.63, 2.71)	2.24 (1.86, 2.69)

Abbreviations: NBDPS, National Birth Defects Prevention Study.

[a] BMI $\geq$ 30 kg/m^2 versus BMI 18.5-24.9 kg/m^2; adjusted for age, ethnicity, education, parity, smoking, folic acid use; excluded mothers with prepregnancy diabetes.

[b] NBDPS results are included in the meta-analysis; comparison is 'obese' versus 'normal weight' as defined by the authors of each study.

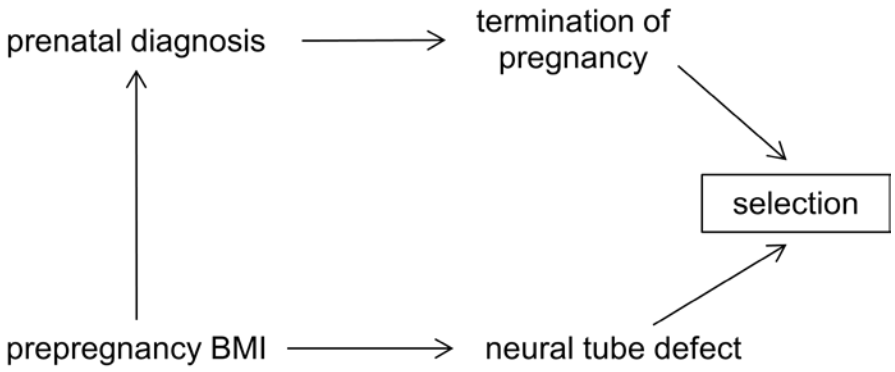

Figure 1.1. Directed acyclic graph illustrating a potential mechanism for selection bias in studies of prepregnancy body mass index and neural tube defects. Abbreviation: BMI, body mass index.

REFERENCES

1. Guh DP, Zhang W, Bansback N, et al. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health*. 2009;9:88.
2. Kim SY, Dietz PM, England L, et al. Trends in pre-pregnancy obesity in nine states, 1993-2003. *Obesity*. 2007;15:986-993.
3. Rasmussen SA, Chu SY, Kim SY, et al. Maternal obesity and risk of neural tube defects: a metaanalysis. *Am J Obstet Gynecol*. 2008;198(6):611-619.
4. Stothard KJ, Tennant PWG, Bell R, et al. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. *JAMA*. 2009;301(6):636-650.
5. Waller DK, Shaw GM, Rasmussen SA, et al. Prepregnancy obesity as a risk factor for structural birth defects. *Arch Pediatr Adolesc Med*. 2007;161(8):745-750.
6. Peek M, Nanan R. Comment on: Association of maternal pre-pregnancy weight with birth defects: evidence from a case-control study in Western Australia. *Aust N Z J Obstet Gynaecol*. 2009;49:564-565.
7. Cragan JD, Khoury MJ. Effect of prenatal diagnosis on epidemiologic studies of birth defects. *Epidemiology*. 2000;11(6):695-699.
8. Shaw GM, Carmichael SL. Prepregnant obesity and risks of selected birth defects in offspring. *Epidemiology*. 2008;19(4):616-620.
9. Stommel M, Schoenborn CA. Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001-2006. *BMC Public Health*. 2009;2009:421.

10. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001-2006. *Prev Chronic Dis*. 2009;6(4):A121.

11. Gustafson P. *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. Boca Raton, FL: Chapman & Hall/CRC; 2003.

12. Rothman KJ. BMI-related errors in the measurement of obesity. *Int J Obes*. 2008;32(Suppl 3):S56-S59.

13. Dashe JS, McIntire DD, Twickler DM. Maternal obesity limits the ultrasound evaluation of fetal anatomy. *J Ultrasound Med*. 2009;28(8):1025-1030.

14. Dashe JS, McIntire DD, Twickler DM. Effect of maternal obesity on the ultrasound detection of anomalous fetuses. *Obstet Gynecol*. 2009;113:1001-1007.

15. Mansfield C, Hopfer S, Marteau TM, et al. Termination rates after prenatal diagnosis of Down syndrome, spina bifida, anencephaly, and Turner and Klinefelter syndromes: a systematic literature review. *Prenat Diagn*. 1999;19:808-812.

16. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology, 3rd edition*. Philadelphia: Lippincott Williams & Wilkins, 2008.

17. Flegal KM, Carroll MD, Kit BK, et al. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *JAMA*. 2012;307(5):491-497.

18. Chu SY, Kim SY, Bish CL. Prepregnancy obesity prevalence in the United States, 2004-2005. *Matern Child Health J*. 2009;13(5):614-620.

19. McDonald SD, Han Z, Mulla S, et al. Overweight and obesity in mothers and risk of preterm birth and low birth weight infants: systematic review and meta-analyses. *BMJ*. 2010;341:c3428.

20. Torloni MR, Betran AP, Daher S, et al. Maternal BMI and preterm birth: a systematic review of the literature with meta-analysis. *J Matern Fetal Neonat Med*. 2009;22(11):957-970.

21. Torloni MR, Betran AP, Horta BL, et al. Prepregnancy BMI and the risk of gestational diabetes: a systematic review of the literature with meta-analysis. *Obes Rev*. 2009;10(2):194-203.

22. Chu SY, Kim SY, Lau J, et al. Maternal obesity and risk of stillbirth: a metaanalysis. *Am J Obstet Gynecol*. 2007;197(3):223-228.

23. Society of Obstetricians and Gynaecologists of Canada. SOGC Clinical Practice Guideline No. 239: Obesity in pregnancy. *J Obstet Gynaecol Can*. 2010;32(2):165-173.

24. American College of Obstetricians and Gynecologists. ACOG Committee Opinion No. 315: Obesity in pregnancy. *Obstet Gynecol*. 2005;106:671-675.

25. Centre for Maternal and Child Enquiries and the Royal College of Obstetricians and Gynaecologists. CMACE/RCOG Joint Guideline: Management of obesity in pregnancy. 2010.

26. Moore KL, Persaud TVN. *The Developing Human: Clinically Oriented Embryology, 8th edition*. Philadelphia: Saunders; 2008.

27. Borman B, Cryer C. Fallacies of international and national comparisons of disease occurrence in the epidemiology of neural tube defects. *Teratology*. 1990;42:405-412.

28. Botto LD, Moore CA, Khoury MJ, et al. Neural-tube defects. *New Engl J Med.* 1999;341(20):1509-1519.

29. Johnson KMK, Suarez L, Felkner MM, et al. Prevalence of craniorachischisis in a Texas-Mexico border population. *Birth Defects Res A Clin Mol Teratol*. 2004;70:92-94.

30. Rowland CA, Correa A, Cragan JD, et al. Are encephaloceles neural tube defects? *Pediatrics*. 2006;118:916-923.

31. Copp AJ, Greene NDE. Genetics and development of neural tube defects. *J Pathol*. 2010;220:217-230.

32. Obeidi N, Russell N, Higgins JR, et al. The natural history of anencephaly. *Prenat Diagn*. 2010;30:357-360.

33. Jaquier M, Klein A, Boltshauser E. Spontaneous pregnancy outcome after prenatal diagnosis of anencephaly. *BJOG*. 2006;113(8):951-953.

34. Wong LY, Paulozzi LJ. Survival of infants with spina bifida: a population study, 1979-94. *Paediatr Perinat Epidemiol*. 2001;15(4):374-378.

35. Juranek J, Salman MS. Anomalous development of brain structure and function in spina bifida myelomeningocele. *Dev Disabil Res Rev*. 2010;16(1):23-30.

36. Thompson DN. Postnatal management and outcome for neural tube defects including spina bifida and encephalocoeles. *Prenat Diagn*. 2009;29(4):412-419.

37. Leibold SR. Achieving continence with a neurogenic bowel. *Pediatr Clin North Am*. 2010;57(4):1013-1025.

38. Adzick NS, Thom EA, Spong CY, et al. A randomzied trial of prenatal versus postnatal repair of myelomeningocele. *N Engl J Med*. 2011;364(11):993-1004.

39. Jentink J, Loane MA, Dolk H, et al. Valproic acid monotherapy in pregnancy and major congenital malformations. *N Engl J Med*. 2010;362(23):2185-2193.

40. Lynch SA. Non-multifactorial neural tube defects. *Am j Med Genet C Semin Med Genet*. 2005;135C(1):69-76.

41. Smithells RW, Sheppard S, Schorah CJ, et al. Possible prevention of neural-tube defects by periconceptional vitamin supplementation. *Lancet*. 1980;1(8164):339-340.

42. Czeizel AE, Dudas I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med*. 1992;327(26):1832-1835.

43. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet*. 1991;338(8760):131-137.

44. Centers for Disease Control and Prevention. Trends in wheat-flour fortification with folic acid and iron -- worldwide, 2004 and 2007. *MMWR Morb Mortal Wkly Rep*. 2008;57(1):8-10.

45. Lopez-Camelo JS, Castilla EE, Orioli IM, et al. Folic acid flour fortification: impact on the frequencies of 52 congenital anomaly types in three South American countries. *Am J Med Genet A*. 2010;152A:2444-2458.

46. Amarin ZO, Obeidat AZ. Effect of folic acid fortification on the incidence of neural tube defects. *Paed Perinat Epidemiol*. 2010;24(4):349-351.

47. De Wals P, Tairou F, Van Allen MI, et al. Reduction in neural-tube defects after folic acid fortification in Canada. *N Engl J Med*. 2007;357(2):135-142.

48. Canfield MA, Collins JS, Botto LD, et al. Changes in the birth prevalence of selected birth defects after grain fortification with folic acid in the United States: findings from a multi-state population-based study. *Birth Defects Res A Clin Mol Teratol*. 2005;73(10):679-689.

49. Parker SE, Mai CT, Canfield MA, et al. Updated national birth prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth Defects Res A Clin Mol Teratol*. 2010;88(12):1008-1016.

50. Mojtabai R. Body mass index and serum folate in childbearing age women. *Eur J Epidemiol*. 2004;19:1029-1036.

51. Case AP, Ramadhani T, Canfield MA, et al. Folic acid supplementation among diabetic, overweight, or obese women of childbearing age. *J Obstet Gynecol Neonatal Nurs*. 2007;36(4):335-341.

52. Laraia BA, Bodnar LM, Siega-Riz AM. Pregravid body mass index is negatively associated with diet quality during pregnancy. *Public Health Nutr*. 2007;10(9):920-926.

53. Friel JK. Nutritional patterns of mothers of children with neural tube defects in Newfoundland. *Am J Med Genet*. 1995;55(2):195-199.

54. Shaw GM, Quach T, Nelson V, et al. Neural tube defects associated with maternal periconceptional dietary intake of simple sugars and glycemic index. *Am J Clin Nutr*. 2003;78:972-978.

55. Yazdy MM, Liu S, Mitchell AA, et al. Maternal glycemic intake and the risk of neural tube defects. *Am J Epidemiol*. 2010;171:407-414.

56. Carmichael SL, Yang W, Shaw GM. Periconceptional nutrient intakes and risks for neural tube defects in California. *Birth Defects Res A Clin Mol Teratol*. 2010;88(8):670-678.

57. Hendricks KA, Nuno OM, Suarez L, et al. Effects of hyperinsulinemia and obesity on risk of neural tube defects among Mexican Americans. *Epidemiology*. 2001;12(6):630-635.

58. Jurek AM, Maldonado G, Greenland S, et al. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol*. 2006;21:871-876.

59. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology, 3rd edition*. Philadelphia: Lippincott Williams & Wilkins, 2008.

60. Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data*. New York: Springer; 2009.

61. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Statist Soc A*. 2005;168(2):267-306.

62. Maldonado G. Adjusting a relative-risk estimate for study imperfections. *J Epidemiol Commun Health*. 2008;62:655-663.

63. Lash TL. Methods to apply probabilistic bias analysis to summary estimates of association. *Pharmacoepidemiol Drug Saf*. 2010;19(6):638-644.

64. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34:1370-1376.

65. Yoon PW, Rasmussen SA, Lynberg MC, et al. The National Birth Defects Prevention Study. *Public Health Rep*. 2001;116(Suppl 1):32-40.

66. Cogswell ME, Bitsko RH, Anderka M, et al. Control selection and participation in an ongoing, population-based, case-control study of birth defects: the National Birth Defects Prevention Study. *Am J Epidemiol*. 2009;170:975-985.

67. Rasmussen SA, Lammer EJ, Shaw GM, et al. Integration of DNA sample collection into a multi-site birth defects case-control study. *Teratology*. 2002;66:177-184.

68. Rasmussen SA, Olney RS, Holmes LB, et al. Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Defects Res A Clin Mol Teratol*. 2003;67:193-201.

69. Krantz DA, Hallahan TW, Sherwin JE. Screening for open neural tube defects. *Clin Lab Med*. 2010;30:721-725.

70. Brock DJ, Sutcliffe RG. Alpha-fetoprotein in the antenatal diagnosis of anencephaly and spina bifida. *Lancet*. 1972;2(7770):197-199.

71. Wald NJ, Cuckle H, Brock JH, et al. Maternal serum-alpha-fetoprotein measurement in antenatal screening for anencephaly and spina bifida in early pregnancy. Report of U.K. Collaborative Study on alpha-fetoprotein in relation to neural-tube defects. *Lancet*. 1977;1(8026):1323-1332.

72. Wald N. Neural tube defects. In: Wald N, Leck I, eds. *Antenatal and neonatal screening, second edition*. New York: Oxford University Press, 2000.

73. Dashe JS, Twickler DM, Santos-Ramos R, et al. Alpha-fetoprotein detection of neural tube defects and the impact of standard ultrasound. *Am J Obstet Gynecol*. 2006;195:1623-1628.

74. Norem CT, Schoen EJ, Walton DL, et al. Routine ultrasonography compared with maternal serum alpha-fetoprotein for neural tube defect screening. *Obstet Gynecol*. 2005;106(4):747-452.

75. Cameron M, Moran P. Prenatal screening and diagnosis of neural tube defects. *Prenat Diagn*. 2009;29(4):402-411.

76. Boyd PA, Wellesley DG, De Walle HEK, et al. Evaluation of the prenatal diagnosis of neural tube defects by fetal ultrasonographic examination in different centres across Europe. *J Med Screen*. 2000;7:169-174.

77. Aagaard-Tillery KM, Porter TF, Malone RD, et al. Influence of maternal BMI on genetic sonography in the FaSTER trial. *Prenat Diagn*. 2010;30:14-22.

78. Wolfe HM, Sokol RJ, Martier SM, et al. Maternal obesity: a potential source of error in sonographic prenatal diagnosis. *Obstet Gynecol*. 1990;76:339-342.

79. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin No. 101: Ultrasonography in pregnancy. *Obstet Gynecol*. 2009;113(2 Pt 1):451-461.

80. Maxwell C, Dunn E, Tomlinson G, et al. How does maternal obesity affect the routine fetal anatomic ultrasound? *J Matern Fetal Neonat Med*. 2010;epub January 21.

81. Hendler I, Blackwell SC, Bujold E, et al. The impact of maternal obesity on midtrimester sonographic visualization of fetal cardiac and craniospinal structures. *Int J Obes*. 2004;28:1607-1611.

82. Khoury FR, Ehrenberg HM, Mercer BM. The impact of maternal obesity on satisfactory detailed anatomic ultrasound image acquisition. *J Matern Fetal Neonat Med*. 2009;22(4):337-341.

83. Hendler I, Blackwell SC, Bujold E, et al. Suboptimal second-trimester ultrasonographic visualization of the fetal heart in obese women: should we repeat the examination? *J Ultrasound Med*. 2005;24(9):1205-1209.

84. Hendler I, Blackwell SC, Treadwell MC, et al. Does advanced ultrasound equipment improve the adequacy of ultrasound visualization of fetal cardiac structures in the obese gravid woman? *Am J Obstet Gynecol*. 2004;190:1616-1620.

85. Johnson CY, Honein MA, Hobbs CA, et al. Prenatal diagnosis of orofacial clefts, National Birth Defects Prevention Study, 1998-2004. *Prenat Diagn*. 2009;29:833-839.

86. Forrester MB, Merz RD. Prenatal diagnosis and elective termination of neural tube defects in Hawaii, 1986-1997. *Fetal Diagn Ther*. 2000;15:146-151.

87. Harmon JP, Hiett AK, Palmer CG, et al. Prenatal ultrasound detection of isolated neural tube defects: is cytogenetic evaluation warranted? *Obstet Gynecol*. 1995;86(4 Pt 1):595-599.

88. Waller DK, Pujazon MA, Canfield MA, et al. Frequency of prenatal diagnosis of birth defects in Houston, Galveston and the Lower Rio Grande Valley, Texas 1995. *Fetal Diagn Ther*. 2000;15(6):348-354.

CHAPTER 2

Epidemiologic Evidence for an Association Between Maternal Obesity and Neural Tube Defects: a Systematic Review and Meta-Analysis of the Published Literature

INTRODUCTION

Based on self-reported data from the Behavioral Risk Factor Surveillance System and the Pregnancy Risk Assessment Monitoring System, it is estimated that approximately 20% of American women enter pregnancy obese (1, 2). Prepregnancy obesity is associated with adverse pregnancy outcomes and complications such as gestational diabetes (3), preterm birth (4, 5), stillbirth (6), and birth defects (7). Among the most consistently reported adverse pregnancy outcomes associated with prepregnancy obesity is the increased risk of having a pregnancy affected by a neural tube defect (NTD) (7, 8). Due to the strength of the epidemiologic evidence to date, clinical guidelines recommend that obese women be counseled prior to pregnancy regarding their increased risk of having an affected pregnancy (9, 10).

Two meta-analyses investigating associations between maternal obesity and NTDs have been published recently: one reporting associations between maternal overweight and obesity and NTDs (8), and the other considering anencephaly and spina bifida separately (7). Both meta-analyses found that obese mothers were more likely than normal weight mothers to have an NTD-affected pregnancy.

In the present analysis, we extend the previously published meta-analyses in three ways. First, we update the results with more recently published studies of prepregnancy obesity and NTDs. Second, because the previous meta-analyses used crude effect estimates in the analysis, we conduct the analysis using covariate-adjusted estimates and investigate the extent to which covariate adjustment affects results. Third, we explore an alternate categorization of obesity. Both previous meta-analyses compared obese to normal weight mothers. Here, we additionally provide a comparison between obese mothers and non-obese mothers, an exposure categorization reported in several studies of obesity and NTDs.

METHODS

Identification of Studies

Medline and Embase were searched through the end of July 2010 according to the search strategy presented in Table 2.1. Titles and abstracts were screened, and full texts of relevant articles were retrieved. We included epidemiologic studies reporting associations between prepregnancy body mass index (BMI) and NTDs (anencephaly, spina bifida, or a combined category of NTDs which could include other, rarer, NTD types). We also included studies not reporting these associations, but from which sufficient information was available to calculate the necessary measures. We excluded review articles, commentaries or editorials without original data, studies reporting weight but not BMI, studies reporting BMI as a continuous variable, and studies not separating NTDs from other defects of the central nervous system. There was no restriction by language or study design. When more than one study reported data from the same population, the article

with the primary hypothesis most relevant to prepregnancy obesity and NTDs was retained. If there was more than one of these, the largest or most recent study was included. Reference lists of included studies and the two previously published meta-analyses were searched for additional studies. Google Scholar (http://scholar.google.com) was searched through the end of August 2010 to identify articles that had cited each included study.

Data Abstraction

Information abstracted from each article included location, years of participant recruitment, study design, source of BMI information, BMI cutpoints, NTD phenotypes, covariates in multivariable models, exclusion criteria, and number, source and pregnancy outcomes of cases and controls.

Meta-Analysis

Three outcomes were investigated separately: anencephaly, spina bifida, and a combined category of NTDs. When studies only provided separate estimates for anencephaly and spina bifida, these categories were combined and included in the meta-analysis of combined NTDs.

We investigated two exposure categorizations: obese compared to normal weight mothers, and obese compared to non-obese mothers. We defined obese as BMI $\geq$30 kg/m^2, non-obese as BMI <30kg/m^2, and normal weight as BMI 18.5-24.9 kg/m^2. If the article did not provide these exact cutpoints, the closest available cutpoint was used (e.g.

BMI $\geq$ 29 instead of $\geq$ 30 kg/m^2). Adjusted odds ratios were abstracted from the articles whenever possible; otherwise, crude estimates were abstracted or calculated from available information.

Random-effects meta-analysis was used for quantitative synthesis of results using Stata version 10 (Statacorp, College Station, TX). We calculated summary odds ratios (ORs) and 95% confidence intervals (CIs) for each meta-analysis, and Cochran Q and I^2 statistics were used to quantify between-study heterogeneity (11). Potential for publication bias was assessed using Egger's test (12). The influence of individual studies was assessed by repeating the meta-analysis with each of the studies removed in turn.

Review of Potential Confounders

Articles were included in this analysis if both crude and adjusted ORs were available. We abstracted the list of stratification variables used in the article, including covariates entered into the multivariable model and covariates controlled for by restriction. We calculated two metrics: the ratio of the adjusted to the crude OR, which showed the effect of covariate adjustment on the magnitude of the association, and the ratio of the adjusted to the crude precision, which showed its effect on precision. Precision was defined as the ratio of the upper to the lower 95% confidence bound. This analysis was conducted separately for anencephaly and spina bifida to investigate possible differences in the effect of covariate adjustment between the two phenotypes.

RESULTS

Search Strategy

Using the electronic search strategy, we identified 377 articles whose titles and abstracts were screened for inclusion (Figure 2.1). Twelve met inclusion criteria (13-24). An additional 3 articles meeting inclusion criteria were found by searching reference lists (25-27), and a fourth was found by searching for articles citing included articles (28). Three of these four studies were likely not found by the electronic search strategy because results for several birth defects were presented in the article and NTDs were not specifically mentioned in the abstract (25, 26, 28). In the fourth, the word "weight" was used in the title and abstract, and words "body mass index" and "obesity" were only included in the full text of the article (27).

Characteristics of the 16 included studies are shown in Table 2.2. Most studies used a case-control design, were conducted in the United States, used birth defect surveillance as a source of cases, and obtained information on BMI from maternal self-report.

Meta-Analyses

There were 13 studies of NTDs (13-18, 20-23, 25, 27, 28), 6 studies of spina bifida (15, 16, 18, 22, 23, 28), and 5 studies of anencephaly (16, 18, 22, 23, 28) in which infants of obese mothers were compared to those of normal weight mothers, and 15 studies of NTDs (13-23, 25-27), 7 studies of spina bifida (15, 16, 18, 20-23), and 6 studies of anencephaly (16, 18, 20-23) in which infants of obese mothers were compared to those of non-obese mothers. All studies published after 2000 used a BMI cutpoint of 30 kg/m^2 for obesity, except one that used a cutpoint of 27 kg/m^2 (26). Studies published in 2000 and

prior used cutpoints of 28 kg/m^2 (25), 29 kg/m^2 (15, 19, 20, 24), 31 kg/m^2 (21), and 32 kg/m^2 (27) for obesity.

When obese mothers were compared to normal weight mothers, we found a weak association between prepregnancy obesity and anencephaly (summary OR 1.12, 95% CI: 0.80, 1.58; Figure 2.2, Table 2.3). The association was stronger when using non-obese mothers as the reference group (summary OR 1.34, 95% CI 1.07-1.70, Figure 2.3). Restricting only to studies that had contributed to both meta-analyses did not change results; the association was again stronger when using the dichotomized version of obesity (Table 2.4).

For spina bifida, the association was stronger when using non-obese compared to normal weight mothers as the reference group (summary OR using normal weight reference group: 1.66, 95% CI: 1.20, 2.29, Figure 2.4; summary OR using non-obese reference group: 1.88, 95% CI 1.47-2.40, Figure 2.5). When restricting the analysis to studies appearing in both meta-analyses, the magnitudes of the associations were similar to each other (Table 2.4).

Between-study heterogeneity was low for studies of anencephaly and moderate for studies of spina bifida, but no one study appeared to explain the heterogeneity (Table 2.5). The study by Waller et al. (22) was the largest study included in the meta-analysis and had the greatest influence on results. Based on the results of Egger's test, publication bias was not a likely explanation for the observed association.

Review of Potential Confounders

Three studies of anencephaly (16, 18, 22) and 4 of spina bifida (15, 16, 18, 22) presented both adjusted and crude estimates (Table 2.6). In 6 of the 7 studies (16, 18, 22), adjustment for covariates moved the OR downward, and in the seventh, there was no difference (15). Adjustment had more effect on precision in studies of anencephaly than studies of spina bifida. Every study adjusted for maternal education, and most for maternal age, ethnicity, and folic acid or vitamin use.

DISCUSSION

Similar to results from previous meta-analyses, we found that obese mothers were more likely to have a child with an NTD than non-obese or normal weight mothers. Low to moderate between-study heterogeneity was detected in each meta-analysis, reflecting the consistency with which most studies have reported this association. In the present analyses, we additionally found that adjusting for some commonly-investigated maternal sociodemographic and behavioral covariates had little effect on results.

Overall, our meta-analysis results are similar to those of Rasmussen et al. (8) but the magnitudes of the association are not as strong as those of Stothard et al. (7) (Table 2.7). In particular, the association reported by Stothard et al. between maternal obesity and anencephaly was not observed in the present study when comparing obese to normal weight mothers, although we did find an association when dichotomizing obesity. Each meta-analysis included different sets of studies, based on variations in inclusion and

exclusion criteria and the studies that had been published at the time the meta-analysis was conducted. In addition, the present study used adjusted estimates in the meta-analysis while crude estimates were used in the previous studies. This could account for the attenuation of results because we found that control for confounders decreased the OR estimates.

If we assume the covariates included in each study were truly confounders of the association, we can conclude that confounding led to a positive bias (i.e., adjusted estimates were smaller than crude estimates). However, the magnitude of the bias was not large. It is possible that other combinations of covariates, not investigated in any of the included studies, could have greater effects on the magnitude of the association.

BMI is a continuous variable that is commonly categorized in epidemiologic studies. We presented results using two common categorizations of BMI: a 4-level version in which obese mothers are compared to normal weight mothers (with underweight and overweight being the other 2 categories), and a dichotomized version in which obese mothers are compared to non-obese mothers. Choice of categorization for BMI will depend on which contrast is more relevant for the research question being investigated. Using normal weight mothers as the reference group would be more appropriate if one were interested in the potential effects of interventions that would move all mothers from the obese category to the normal weight category. Using non-obese mothers as the reference group would be more appropriate if the interest was in an intervention which

would move all mothers from the obese category and distribute them into the other three BMI categories.

The results from this meta-analysis are consistent with previous work demonstrating an association between maternal obesity and NTDs; however the specific mechanisms underlying this association remain poorly understood.

Table 2.1. Electronic Search Strategy Used to Identify Articles From Embase and Medline Through the end of July 2010.

(‘obesity’/exp OR obesity OR obese OR overweight OR (body AND mass AND index)) AND (‘neural tube defect’/exp OR anencephaly OR anencephalus OR (spina AND bifida) OR (neural AND tube AND (defect OR defects)) OR ntd OR meningocele OR myelomeningocele)

Table 2.2. Characteristics of Studies Included in the Meta-Analysis of Maternal Obesity and Neural Tube Defects.

				Cases			Controls or Non-Cases			
Study	Location	Years	Design	N	Source	Outcomes	N	Source	Outcomes	Source of BMI
Waller 1994 (21)	Illinois and California, USA	1985-1987	Case-control	499 NTD	Multiple sources	Live births, stillbirths, TOP	534	Birth certificates, hospitals, mothers undergoing serum screening, ultrasonography, amniocentesis	Live births, stillbirths, TOP	Self-reported height, weight
Shaw 1996 (20)	California, USA	1989-1991	Case-control	538 NTD 217 AN 296 SB	Surveillance (multiple-source)	Live births, stillbirths, TOP	539	Hospitals	Live births	Self-reported height, weight
Watkins 1996 (24)	Atlanta, USA	1968-1980	Case-control	307 NTD	Surveillance (multiple-source)	Live births, stillbirths	2,755	Birth certificates	Live births	Self-reported height, weight: (weight at delivery minus weight gain)

Werler 1996 (27)	Boston, USA, Philadelphia, USA, Toronto, Canada	1992-1994	Case-control	45 NTD	Birth and tertiary care hospitals	Live births, fetal deaths, TOP	91	Hospitals (non-malformed control group)	"Infants"	Self-reported height, weight
Kallen 1998 (15)	Sweden	1983-1993	Cohort	338 NTD	2 registries	"Infants"		665,706 births	Live births, stillbirths	Recorded height, weight (weight at delivery minus self-reported weight gain)
Moore 2000 (25)	USA (multiple sites)	1984-1987	Cohort	49 NTD	Mothers undergoing serum screening	Live births, stillbirths, TOP		22,951 births	Live births, stillbirths, TOP	Self-reported height, weight
Shaw 2000 (19)	California, USA	1987-1988	Case-control	247 NTD	Surveillance (multiple-source)	Live births, stillbirths, TOP	461	Same geographic area	Live births	Self-reported height, weight

Hendricks 2001 (14)	Texas, USA	1995-2000	Case-control	149 NTD	Surveillance; Mexican-American women	"Delivered" infants, TOP	178	Hospitals, birthing centers; Mexican-American women	Live births	Self-reported height, weight
Mikhail 2002 (26)	Chicago, USA	1981-1994	Case-control	17 NTD	Clinical records; African-American women	"Delivered" infants, TOP	144	Clinical records; African-American women	"Delivered" infants	Medical record
Watkins 2003 (23)	Atlanta, USA	1993-1997	Case-control	43 NTD 12 AN 22 SB	Surveillance (multiple-source)	Live births, stillbirths, TOP	330	Hospitals	"Births"	Self-reported height, weight
Waller 2007 (22)	USA (8 sites)	1997-2002	Case-control	193 AN 425 SB	Surveillance	Live births, stillbirths, TOP	3,904	Birth certificates, hospitals	Live births	Self-reported height, weight
Shaw 2008 (18)	California, USA	1999-2004	Case-control	125 AN 164 SB	Surveillance (multiple-source)	Live births, stillbirths, TOP	541	Hospitals	Live births	Self-reported height, weight

Oddy 2009 (17)	Western Australia, Australia	1997-2000	Case-control	27 NTD	Registry (multiple-source)	Live births, stillbirths, TOP	418	Midwives' Notification System in Western Australia	Live births	Height recorded in Midwives' Notification System, self-reported weight
Blomberg 2010 (13)	Sweden	1995-2007	Cohort	389 NTD	3 registries	Live births, stillbirths		1,235,877 total births	Live births, stillbirths	Recorded at first prenatal visit
Li 2010 (16)	Shanxi, China	2003-2007	Case-control	511 NTD 232 AN 238 SB	Surveillance	Live births, stillbirths, TOP	687		"Infants"	Self-reported height, weight
Rankin 2010 (28)	North of England	2003-2005	Cohort	23 NTD 7 AN 15 SB	5 maternity hospitals; registry	Live births, stillbirths, TOP		30,703 births	Live births, stillbirths, TOP	"BMI at the first antenatal visit"

Abbreviations: AN, anencephaly; BMI, body mass index; N, number; NTD, neural tube defect; SB, spina bifida; TOP, termination of pregnancy.

Table 2.3. Results From Random-Effects Meta-Analyses of Associations Between Maternal Obesity and Neural Tube Defects.

		Summary Estimate		Heterogeneity		Cochran Q	Egger's
Meta-Analysis	N	OR	95% CI	I^2	95% UI	p-value	p-value
Obese versus normal weight[a]							
Anencephaly	5	1.12	0.80, 1.58	6	0, 80	0.37	0.54
Spina bifida	6	1.66	1.20, 2.29	43	0, 77	0.12	0.55
All NTDs	13	1.72	1.47, 2.01	21	0, 58	0.22	0.65
Obese versus non-obese[b]							
Anencephaly	6	1.34	1.07, 1.70	0	0, 64	0.63	0.37
Spina bifida	7	1.88	1.47, 2.40	42	0, 76	0.11	0.43
All NTDs	15	1.73	1.55, 1.93	0	0, 49	0.54	0.43

Abbreviations: BMI, body mass index; CI, confidence interval; N, number of included studies; NTDs, neural tube defects; OR, summary odds ratio; UI, uncertainty interval.

[a] As defined in each study; most common contrast: BMI $\geq$ 30 versus 18.5-24.9 kg/m^2.

[b] As defined in each study; most common contrast: BMI $\geq$ 30 versus < 30 kg/m^2.

Table 2.4. Comparison of Results From Random-Effects Meta-Analyses of Associations Between Maternal Obesity and Neural Tube Defects Using two Different Reference Groups for Exposure.

		Normal Weight Reference Group		Non-Obese Reference Group	
Meta-Analysis	N	OR	95% CI	OR	95% CI
Anencephaly	4	1.12	0.75, 1.68	1.31	0.96, 1.79
Spina bifida	5	1.61	1.12, 2.31	1.66	1.17, 2.35
All NTDs	13	1.71	1.45, 2.02	1.73	1.54, 1.95

Abbreviations: CI, confidence interval; N, number of included studies; NTDs, neural tube defects; OR, summary odds ratio.

Table 2.5. Influence of Individual Studies on Results of Random-Effects Meta-Analyses of Obesity (Obese Versus Normal Weight) and Anencephaly or Spina Bifida.

	Summary Estimate		Heterogeneity		Cochran Q
Study Removed	OR	95% CI	I^2	95% UI	p-value
Anencephaly					
None	1.12	0.80, 1.58	6	0, 80	0.37
Watkins 2003 (23)	1.06	0.77, 1,47	0	0, 81	0.48
Waller 2007 (22)	1.31	0.78, 2.19	9	0, 86	0.35
Shaw 2008 (18)	1.02	0.64, 1.61	9	0, 86	0.35
Li 2010 (16)	1.18	0.85, 1.64	0	0, 84	0.40
Rankin 2010 (28)	1.12	0.75, 1.68	25	0, 71	0.26
Spina bifida					
None	1.66	1.20, 2.29	43	0, 77	0.12
Kallen 1998 (15)	1.71	1.13, 2.59	35	0, 74	0.11
Watkins 2003 (23)	1.56	1.12, 2.17	30	0, 72	0.13
Waller 2007 (22)	1.44	1.00, 2.07	0	0, 74	0.29
Shaw 2008 (18)	1.80	1.28, 2.55	18	0, 63	0.19
Li 2010 (16)	1.77	1.31, 2.39	19	0, 64	0.19
Rankin 2010 (28)	1.61	1.12, 2.31	42	0, 77	0.07

Abbreviations: CI, confidence interval; OR, summary odds ratio; UI, uncertainty interval.

Table 2.6. Comparison of Adjusted to Crude Analyses of Maternal Obesity and Anencephaly or Spina Bifida.

Study	Adjusted[a]		Crude[a]		Ratio of OR[b]	Ratio of Precision[c]	Stratification Variables (Covariates in Multivariable Model; Exclusion Criteria)
	OR	95% CI	OR	95% CI			
Anencephaly							
Waller 2007 (22)	0.96	0.62, 1.48	1.06	0.69, 1.59	0.91	1.04	Age, ethnicity, education, parity, smoking, folic acid use; diabetes
Shaw 2008 (18)	1.4	0.8, 2.4	1.6	1.0, 2.6	0.88	1.15	Race/ethnicity, education, vitamin use, energy intake, height, dietary folate intake; diabetes, use of seizure medications, history of birth defect in previous pregnancy
Li 2010 (16)	0.62	0.22, 1.81	0.84	0.35, 2.02	0.74	1.43	Age, education, occupation, parity, history of birth defect in previous pregnancy, folic acid use; not Han ethnicity

Spina bifda							
Kallen 1998 (15)	1.49	0.92, 2.43	1.48	0.91, 2.32	1.01	1.04	Year of birth, age, parity, education, smoking, immigrant status
Waller 2007 (22)	2.10	1.63, 2.71	2.25	1.76, 2.87	0.93	1.02	Age, ethnicity, education, parity, smoking, folic acid use; diabetes
Shaw 2008 (18)	1.2	0.7, 2.0	1.4	0.8, 2.2	0.86	1.04	Race/ethnicity, education, vitamin use, energy intake, height, dietary folate intake; diabetes, use of seizure medications, history of birth defect in previous pregnancy
Li 2010 (16)	0.78	0.28, 2.23	0.84	0.33, 2.13	0.93	1.23	Age, education, occupation, parity, history of birth defect in previous pregnancy, folic acid use; not Han ethnicity

Abbreviations: CI, confidence interval; OR, odds ratio.

[a] All contrasts are body mass index ≥ 30 versus 18.5-24.9 kg/m^2 except Kallen et al. (15) which is ≥ 29 versus < 29 kg/m^2.

[b] Ratio of adjusted OR to crude OR.

[c] Ratio of adjusted OR precision to crude OR precision, where precision is defined as the ratio of the upper to lower confidence bound.

Table 2.7. Summary of Results From Three Meta-Analyses of Obesity and Neural Tube Defects (Obese Versus Normal Weight).

	Neural Tube Defects		Anencephaly		Spina Bifida	
Meta-Analysis	OR	95% CI	OR	95% CI	OR	95% CI
Rasmussen 2008 (8)	1.70	1.34, 2.15				
Stothard 2009 (7)	1.87	1.62, 2.15	1.39	1.03, 1.87	2.24	1.86, 2.69
Present meta-analysis	1.71	1.48, 1.97	1.12	0.80, 1.58	1.66	1.20, 2.29

Abbreviations: CI, confidence interval; OR, summary odds ratio.

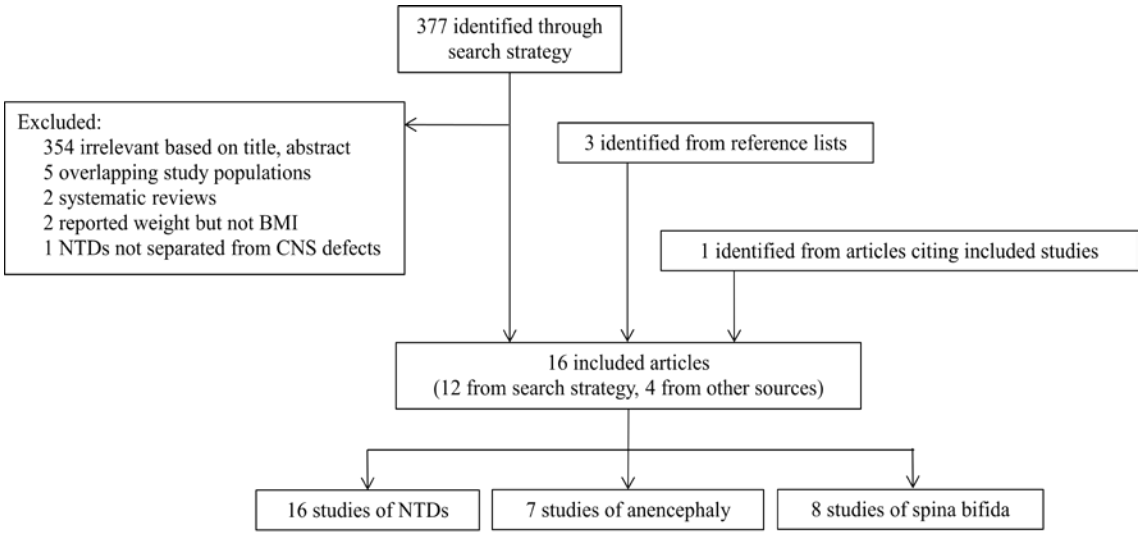

Figure 2.1. Flow of studies into and out of the systematic review of maternal obesity and neural tube defects. Abbreviations: BMI, body mass index; CNS, central nervous system; NTD, neural tube defect.

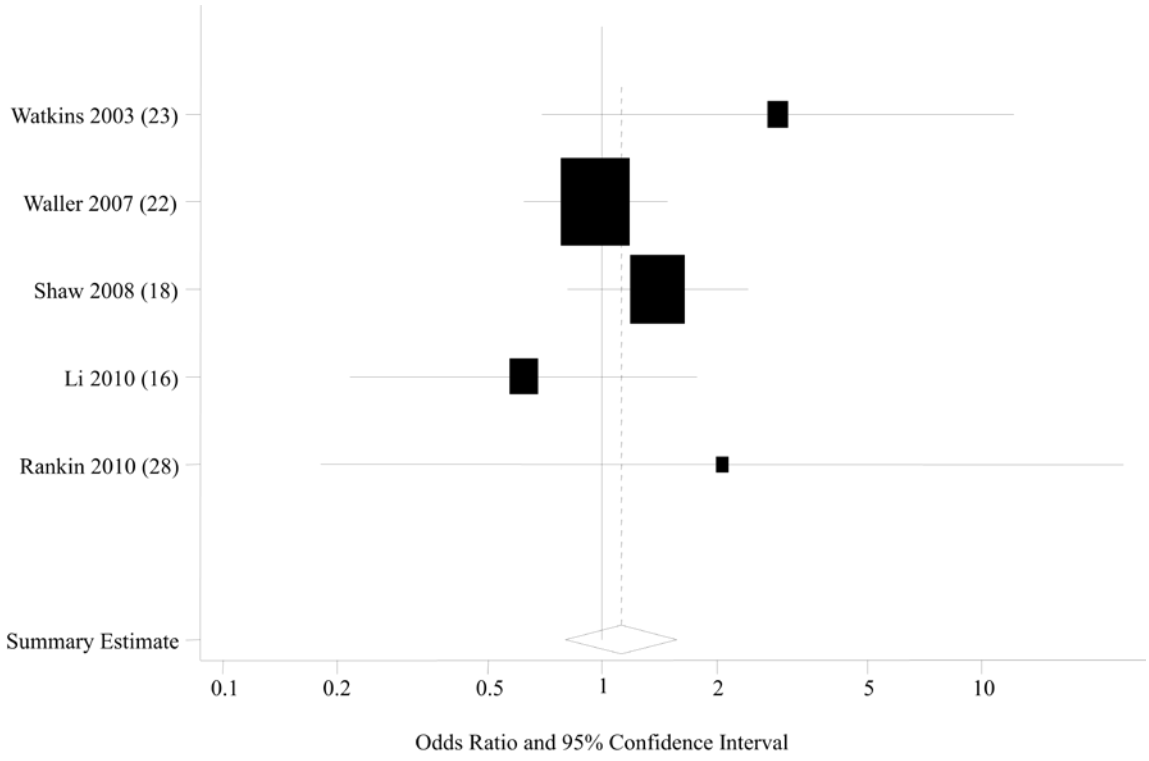

Figure 2.2. Forest plot of studies reporting associations between maternal obesity (obese compared to normal weight mothers) and anencephaly, with summary estimate from random-effects meta-analysis.

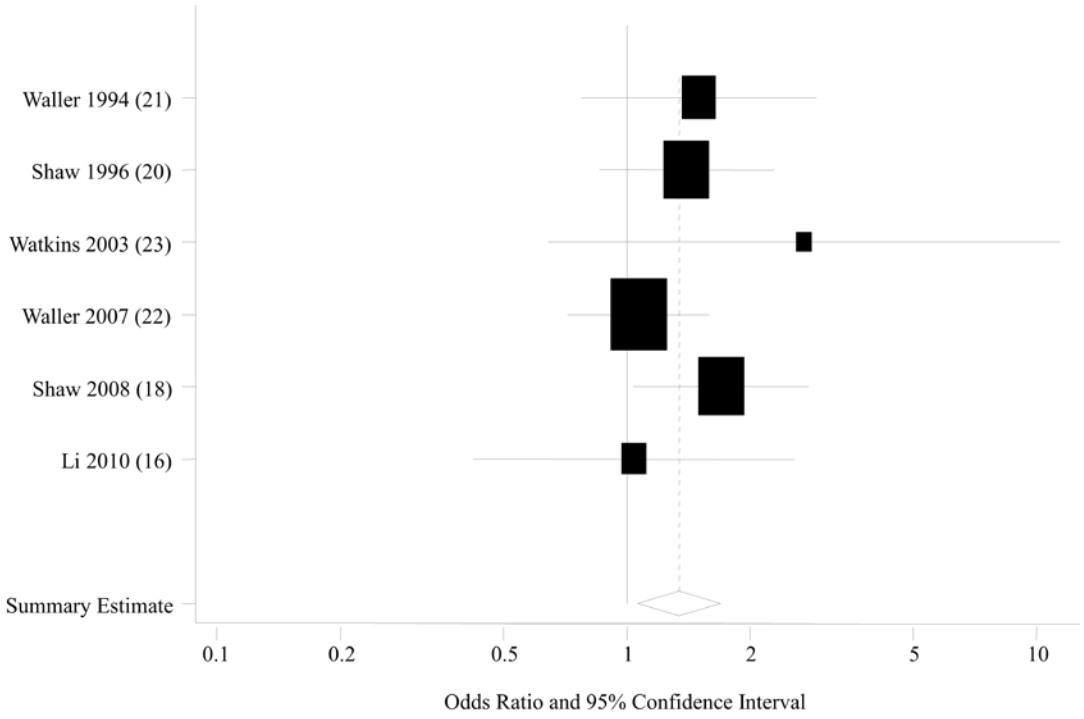

Figure 2.3. Forest plot of studies reporting associations between maternal obesity (obese compared to non-obese mothers) and anencephaly, with summary estimate from random-effects meta-analysis.

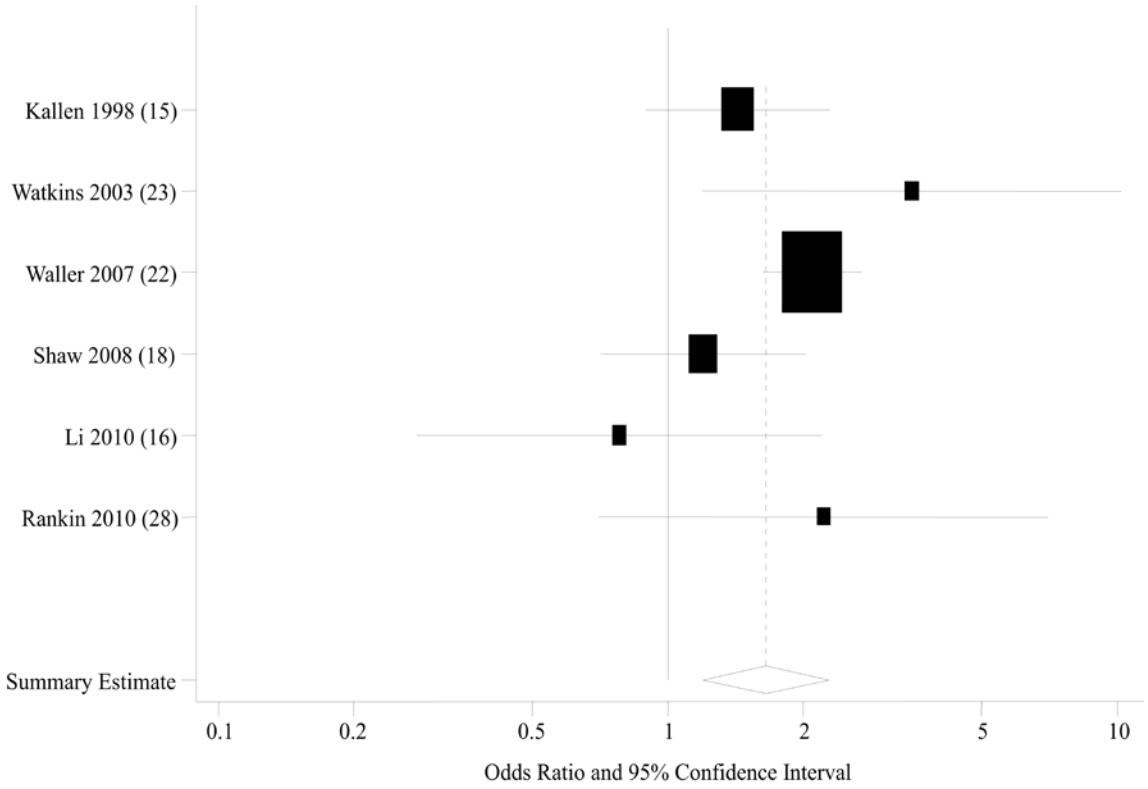

Figure 2.4. Forest plot of studies reporting associations between maternal obesity (obese compared to normal weight mothers) and spina bifida, with summary estimate from random-effects meta-analysis.

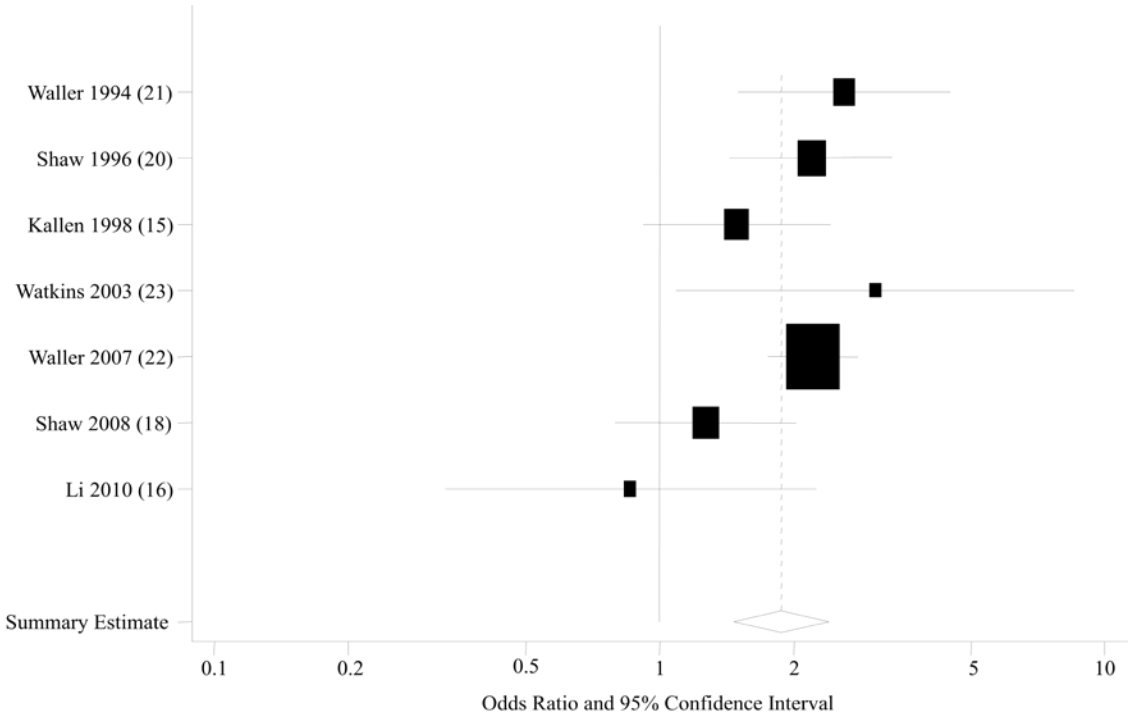

Figure 2.5. Forest plot of studies reporting associations between maternal obesity (obese compared to non-obese mothers) and spina bifida, with summary estimate from random-effects meta-analysis.

REFERENCES

1. Kim SY, Dietz PM, England L, et al. Trends in pre-pregnancy obesity in nine states, 1993-2003. *Obesity*. 2007;15:986-993.
2. Chu SY, Kim SY, Bish CL. Prepregnancy obesity prevalence in the United States, 2004-2005. *Matern Child Health J*. 2009;13(5):614-620.
3. Torloni MR, Betran AP, Horta BL, et al. Prepregnancy BMI and the risk of gestational diabetes: a systematic review of the literature with meta-analysis. *Obes Rev*. 2009;10(2):194-203.
4. Torloni MR, Betran AP, Daher S, et al. Maternal BMI and preterm birth: a systematic review of the literature with meta-analysis. *J Matern Fetal Neonat Med*. 2009;22(11):957-970.
5. McDonald SD, Han Z, Mulla S, et al. Overweight and obesity in mothers and risk of preterm birth and low birth weight infants: systematic review and meta-analyses. *BMJ*. 2010;341:c3428.
6. Chu SY, Kim SY, Lau J, et al. Maternal obesity and risk of stillbirth: a metaanalysis. *Am J Obstet Gynecol*. 2007;197(3):223-228.
7. Stothard KJ, Tennant PWG, Bell R, et al. Maternal overweight and obesity and the risk of congenital anomalies: A systematic review and meta-analysis. *JAMA*. 2009;301(6):636-650.
8. Rasmussen SA, Chu SY, Kim SY, et al. Maternal obesity and risk of neural tube defects: a metaanalysis. *Am J Obst Gynecol*. 2008;198(6):611-619.

9. Centre for Maternal and Child Enquiries and the Royal College of Obstetricians and Gynaecologists. CMACE/RCOG Joint Guideline: Management of obesity in pregnancy. 2010.

10. Society of Obstetricians and Gynaecologists of Canada. SOGC Clinical Practice Guideline No. 239: Obesity in pregnancy. *J Obstet Gynaecol Can*. 2010;32(2):165-173.

11. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.

12. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634.

13. Blomberg MI, Kallen B. Maternal obesity and morbid obesity: The risk for birth defects in the offspring. *Birth Defects Res A Clin Mol Teratol*. 2010;88(1):35-40.

14. Hendricks KA, Nuno OM, Suarez L, et al. Effects of hyperinsulinemia and obesity on risk of neural tube defects among Mexican Americans. *Epidemiology*. 2001;12(6):630-635.

15. Kallen K. Maternal smoking, body mass index, and neural tube defects. *Am J Epidemiol*. 1998;147(12):1103-1111.

16. Li Z, Liu J, Ye R, et al. Maternal prepregnancy body mass index and risk of neural tube defects: A population-based case-control study in Shanxi Province, China. *Birth Defects Res A Clin Mol Teratol*. 2010;88(7):570-574.

17. Oddy WH, De Klerk NH, Miller M, et al. Association of maternal pre-pregnancy weight with birth defects: Evidence from a case-control study in Western Australia. *Aust N Z J Obst Gynaecol*. 2009;49(1):11-15.

18. Shaw GM, Carmichael SL. Prepregnant obesity and risks of selected Birth Defects in Offspring. *Epidemiology*. 2008;19(4):616-620.

19. Shaw GM, Todoroff K, Schaffer DM, et al. Maternal height and prepregnancy body mass index as risk factors for selected congenital anomalies. *Paed Perinat Epidemiol*. 2000;14(3):234-239.

20. Shaw GM, Velie EM, Schaffer D. Risk of neural tube defect-affected pregnancies among obese women. *JAMA*. 1996;275(14):1093-1096.

21. Waller DK, Mills JL, Simpson JL, et al. Are obese women at higher risk for producing malformed offspring? *Am J Obst Gynecol*. 1994;170(2):541-548.

22. Waller DK, Shaw GM, Rasmussen SA, et al. Prepregnancy obesity as a risk factor for structural birth defects. *Arch Pediatr Adolesc Med*. 2007;161(8):745-750.

23. Watkins ML, Rasmussen SA, Honein MA, et al. Maternal obesity and risk for birth defects. *Pediatrics*. 2003;111(5 II):1152-1158.

24. Watkins ML, Scanlon KS, Mulinare J, et al. Is maternal obesity a risk factor for anencephaly and spina bifida? *Epidemiology*. 1996;7(5):507-512.

25. Moore LL, Singer MR, Bradlee ML, et al. A prospective study of the risk of congenital defects associated with maternal obesity and diabetes mellitus. *Epidemiology*. 2000;11(6):689-694.

26. Mikhail LN, Walker CK, Mittendorf R. Association between maternal obesity and fetal cardiac malformations in African Americans. *J Natl Med Assoc*. 2002;94(8):695-700.

27. Werler MM, Louik C, Shapiro S, et al. Prepregnant weight in relation to risk of neural tube defects. *JAMA*. 1996;275(14):1089-1092.

28. Rankin J, Tennant PWG, Stothard KJ, et al. Maternal body mass index and congenital anomaly risk: a cohort study. *Int J Obes (Lond)*. 2010;epub April 6.

CHAPTER 3

Pregnancy Termination Following Prenatal Diagnosis of Anencephaly or Spina Bifida: a Systematic Review of The Literature

Candice Y. Johnson, Margaret A. Honein, W. Dana Flanders, Penelope P. Howards, Godfrey P. Oakley, Jr., and Sonja A. Rasmussen

Author affiliations: Departments of Epidemiology (Candice Y. Johnson, W. Dana Flanders, Penelope P. Howards, Godfrey P. Oakley, Jr.) and Biostatistics and Bioinformatics (W. Dana Flanders), Emory University, Atlanta, Georgia; National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia (Candice Y. Johnson, Margaret A. Honein, Sonja A. Rasmussen).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Corresponding author: Candice Y. Johnson, cyjohnson@alum.emory.edu.

ABSTRACT

In regions where screening for anencephaly and spina bifida is widespread, most cases of these defects are first diagnosed prenatally. The purpose of this study was to collect contemporary estimates of the frequency of termination of pregnancy (TOP) following prenatal diagnosis of anencephaly or spina bifida and to investigate factors associated with TOP which might contribute to selection bias in epidemiologic studies. Medline and Embase were searched from 1990 to 2011 for studies reporting the frequency of TOP following prenatal diagnosis of anencephaly or spina bifida. We included studies with English-language abstracts, with ≥ 20 prenatally diagnosed cases of either defect, and in which the majority of cases were ascertained in 1990 or later. Among the 15 studies identified, 7 included anencephaly and 14 included spina bifida. Six were from North America and 9 from Europe. The overall frequency of TOP following prenatal diagnosis was 86% for anencephaly and 64% for spina bifida. TOP for spina bifida was more common when the prenatal diagnosis occurred before 24 weeks gestation, in cases with greater severity, and in Europe versus North America. Most pregnancies in which anencephaly or spina bifida is prenatally diagnosed will end in TOP. Because epidemiologic studies and surveillance systems might be more likely to underascertain birth defects when the pregnancy ends in TOP and the frequency of TOP following prenatal diagnosis might vary by maternal and case characteristics, investigators should be alert to the possibility of selection bias in epidemiologic studies of these defects.

INTRODUCTION

Neural tube defects (NTDs) are birth defects caused by failure of the neural tube to close completely, resulting in incomplete formation of the brain or spinal cord (1, 2). The two most common types of NTDs are anencephaly, characterized by absence of much of the skull and brain, and spina bifida, a herniation of neural tissue through an incompletely formed spine (1). Anencephaly is a lethal condition and liveborn infants typically survive less than one day (3, 4). The severity of spina bifida is more variable, often corresponding to the location of the defect on the spine (5). In the most severe cases, complications of spina bifida can lead to death; however, this is not the most common outcome, with over 90% of liveborn infants with spina bifida in the United States surviving the first year of life, although with varying levels of sensory loss and paralysis (5, 6). There is marked geographic variability in reported prevalence of NTDs, depending on whether or not the pregnancy occurs in a country requiring folic acid fortification of grain products and whether or not cases among termination of pregnancy (TOP) are included in the estimate (7-10).

Prenatal detection of NTDs is now common in many countries. Screening for elevated maternal serum alpha-fetoprotein levels in the second trimester of pregnancy can identify over two-thirds of cases of open neural tube defects, including almost all cases of anencephaly (11, 12). Ultrasound visualization of the defect has also become a common and effective method for prenatal detection of NTDs and other defects, with use of second and third trimester ultrasound rapidly increasing since the 1970s (13). In one European study, 80% of postnatally confirmed NTD cases had been identified prenatally

in countries in which ultrasound screening was routine (14). Given the often severe nature of NTDs, TOP is common following prenatal diagnosis if the diagnosis is made early enough for this to be an available option. A systematic review of studies published between 1987 and 1995 estimated that 84% and 64% of pregnancies known to be affected with anencephaly and spina bifida, respectively, ended in TOP (15). Since that review was published, the frequency of prenatal diagnosis has continued to increase but no updated comprehensive summary of the frequency of TOP has been performed to determine if these estimates accurately reflect the present-day situation.

The increasing use of prenatal diagnosis and TOP has important implications for the interpretation of results from epidemiologic studies of birth defects such as NTDs, for which both prenatal diagnosis and TOP are relatively common. Not all cases of NTDs are able to be included in epidemiologic studies because cases from pregnancies ending in TOP are more difficult to ascertain than those ending in live birth and typically require inclusion of additional case ascertainment sources. Descriptive studies underestimate the number of pregnancies with recognized NTDs when only live births are included or some proportion of affected pregnancies resulting in TOPs is missed (8, 16, 17). In etiologic studies, exclusion or incomplete ascertainment of NTDs among TOPs can lead to selection bias when the exposure of interest is associated with likelihood of TOP (18). In addition, clinical studies of long-term outcomes observed in cases followed from birth might not be useful for counseling parents with prenatally diagnosed fetuses about prognosis if liveborn cases represent only a small, selected subset of all cases.

Quantifying the frequency of TOP and factors associated with TOP is important for understanding how underascertainment of cases might affect study results. The purpose of the present review is to collect contemporary estimates of the proportion of pregnancies ending in TOP following prenatal diagnosis of anencephaly or spina bifida and to investigate factors associated with TOP that could contribute to selection bias.

METHODS

Search Strategy

We searched Medline and Embase from 1990 through May 2011 for epidemiologic studies reporting both the number of cases of anencephaly or spina bifida prenatally diagnosed in a specific time period and the number of these cases in which the pregnancy outcome was TOP. The search strategy included search terms and synonyms for “neural tube defect”, “anencephaly”, “spina bifida”, “prenatal diagnosis”, and “pregnancy termination” (Appendix). We also identified studies by searching reference lists of included articles and by using Google Scholar to search for more recently published articles citing the included studies. Information abstracted from each article included location and dates of participant recruitment, number of prenatally diagnosed cases of anencephaly or spina bifida, the number of these cases with TOP as the pregnancy outcome, and characteristics investigated in association with TOP.

Inclusion and Exclusion Criteria

Two types of studies were eligible for inclusion: studies following a prospective or retrospective cohort of prenatally diagnosed fetuses to determine outcome of pregnancy

and studies using birth defects surveillance or registries that ascertain prenatally diagnosed cases and pregnancies ending in TOP. Additional inclusion criteria were: an English-language abstract, pregnancy outcome known for at least 20 prenatally diagnosed cases of anencephaly or of spina bifida, and at least half the study years in 1990 or later. We restricted our analysis to studies of at least 20 prenatally diagnosed cases to ensure the estimates were fairly stable. The restriction to studies mostly conducted in 1990 or later was made because after this time fetal ultrasound was in widespread use and the decision to continue or end an affected pregnancy would have likely involved not only serum screen results but also ultrasound confirmation of the specific defect.

We excluded studies restricted to fetuses with NTDs in combination with other specific non-NTD diagnoses (e.g., studies of fetuses with both NTDs and chromosomal abnormalities) or specific indications on ultrasound (e.g. studies of fetuses with both NTDs and increased nuchal translucency). We also excluded studies conducted exclusively in non-singletons. When two studies included information from overlapping populations, we included the most recent study or the study with the largest catchment area (e.g., a national study would be chosen over a regional study). Studies were also excluded if they were conducted in a location where TOP was not legally permitted at any gestational age.

Statistical Analyses

We calculated the frequency of TOP as the number of pregnancies ending in TOP among those in which a prenatal diagnosis was made and pregnancy outcome was known. When

reporting results summarized across studies we divided the sum of the number of pregnancies ending in TOP across studies by the sum of the prenatally diagnosed pregnancies and provided 95% confidence intervals (CI).

In some studies, pregnancies were lost to follow-up and the outcome of pregnancy was unknown. When this occurred, we restricted the analysis to the subset of pregnancies with known outcomes to make these studies comparable to studies which reported no pregnancies lost to follow-up; it was often not possible to determine if a study truly had no pregnancies lost to follow-up or if these pregnancies were excluded prior to analysis.

When fetuses undergoing surgery for *in utero* spina bifida repair had been excluded from the original study we added them back into our analysis and categorized them as pregnancies not ending in TOP.

In our analysis of factors associated with TOP, we categorized studies according to study design (cohort vs. surveillance or registry), case type (all cases vs. isolated defects), defect type (open vs. closed, for spina bifida only), geographic region (Europe vs. North America), and gestational age at prenatal diagnosis (<24 weeks vs. ≥24 weeks). If studies reported information for more than one stratum (e.g., results for open and closed defects presented separately within the same article) they were included once in each category.

RESULTS

We identified 13 articles meeting inclusion criteria using the search strategy (19-31). One additional article was found using Google Scholar (this article had cited one of the articles identified using the search strategy) (32) and one article was known to the authors and included (33). Of these 15 included articles, 7 reported information on anencephaly and 14 on spina bifida. Six studies were from North America and 9 were from Europe.

Frequency of TOP in Included Studies

The overall frequency of TOP following prenatal diagnosis in the 7 studies of anencephaly was 86% (95% CI: 83-89%) and ranged from 64% to 97% in individual studies (Table 3.1). In the 14 studies of spina bifida, the overall frequency was 64% (95% CI: 61-67%) and estimates from individual studies ranged from 31% to 89%.

Factors Associated With Frequency of TOP

There were too few studies of anencephaly to investigate factors associated with frequency of TOP between studies; therefore, only results for spina bifida are shown (Table 3.2). No study reported associations between maternal sociodemographic characteristics and TOP after prenatal diagnosis of spina bifida.

Geographic region

In both North America and Europe, the frequency of TOP following prenatal diagnosis of spina bifida was variable. Estimates ranged from 31% to 82% in North America and from 41% to 89% in Europe. Overall, frequency of TOP was higher in Europe (69%) than North America (49%).

Study design

Of the 14 studies, 3 used data from birth defect surveillance or registries and the remainder followed a cohort of prenatally diagnosed fetuses for pregnancy outcome. Frequency of TOP in studies using surveillance or registries was higher than in those using a cohort design; however, results from the surveillance and registry studies were highly influenced by one study with large sample size (n = 405) and high prevalence of TOP (78%).

Case type

Five studies presented analyses restricted to fetuses with isolated spina bifida and the remainder of the studies included all types of cases. The frequency of TOP was higher for studies including all types of cases than those restricted to fetuses with isolated defects (67% vs. 55%).

Defect type

Two studies presented results for closed spina bifida and five for open spina bifida. In the two studies of closed spina bifida, the frequency of TOP was 22% and 50%, but both estimates were based on fewer than 10 prenatally diagnosed fetuses. For open spina bifida, the frequency of TOP ranged from 36% to 91% with an overall frequency of 57%.

Gestational age at prenatal diagnosis

Four studies (all from Europe) reported frequency of TOP stratified by gestational age at prenatal diagnosis (< 24 versus ≥ 24 gestational weeks) with frequency of TOP higher when prenatal diagnosis was made before 24 weeks rather than later (88% vs. 31%). Gestational age at prenatal diagnosis appeared to be responsible for some of the between-study variability in frequency of TOP (Table 3.3). For example, the overall frequency of TOP was lower in the Netherlands (49%) than other European countries (78%); however, once restricted to prenatal diagnoses made <24 weeks, the Netherlands and other European countries had similar estimates (92% vs. 91%). In the Netherlands over half of prenatal diagnoses were made ≥24 gestational weeks, compared to 26% in the other European countries.

DISCUSSION

Among the studies identified in this review, 86% of pregnancies known to be affected with anencephaly and 64% of those known to be affected with spina bifida ended in TOP. However, none of the included studies presumably had 100% sensitivity for ascertaining NTDs and sensitivity likely varied between included studies. Because epidemiologic studies and surveillance programs are more likely to underascertain pregnancies ending in TOP than those ending in live births (34), these are likely to be underestimates of the prevalence of TOP.

These estimates are similar to those from a previous systematic review of the frequency of TOP published over a decade ago which reported frequencies of 84% (95% CI: 82-86%) for anencephaly and 64% (95% CI: 61-67%) for spina bifida (15). The similarity

between estimates in the present and previous studies suggests that the frequency of TOP following prenatal diagnosis of anencephaly or spina bifida has not appreciably changed over time; however, an analysis of time trends within the same study would be needed to rule out the similarities being explained by systematic differences in study design or characteristics of included studies between the previous and the present reviews.

With a substantial proportion of pregnancies in which a prenatal diagnosis was made ending in TOP, investigators should be aware that epidemiologic studies conducted only among live births include a highly selected sample of the total population of fetuses with NTDs. Previous studies have reported that maternal characteristics such as education, age, and race/ethnicity are associated with pregnancy outcome of NTD-affected pregnancies (35, 36); however, these studies have not separated the effects of differences in whether or not a pregnancy ends in termination following prenatal diagnosis from differences in whether or not a prenatal diagnosis was made. Only one study included in this review investigated maternal characteristics associated with TOP following prenatal diagnosis, but it did not present results separately for the NTDs included (anencephaly, spina bifida, and encephalocele) (20). This study found that TOP following prenatal diagnosis of NTDs was more common in older than younger mothers, in Asian compared to white mothers, and in certain areas of their study catchment area in Hawaii. In studies of other types of birth defects, the frequency of TOP following prenatal diagnosis has also been shown to vary by maternal characteristics such as education (37), age (37, 38), and race/ethnicity (38). Because these characteristics are associated with TOP and therefore with which cases will be included in epidemiologic studies, selection bias is

possible in studies investigating these factors in relation to NTD etiology (18). No study included in the present review reported associations between maternal characteristics and TOP following prenatal diagnosis of anencephaly or spina bifida separately. Additional studies are warranted to further address how maternal characteristics relate to TOP for these subtypes of NTDs and if other maternal and fetal characteristics are associated with TOP following prenatal diagnosis.

We found an association between gestational age at prenatal diagnosis and frequency of TOP. Differences in average gestational age at prenatal diagnosis are a possible explanation for the wide variability in frequency of TOP observed between studies. An association with gestational age is expected because many regions have laws governing the gestational ages at which TOP may be performed, meaning that if a prenatal diagnosis is made at a late gestational age, TOP may no longer be an available option. These results suggest the importance of considering characteristics that could delay prenatal diagnosis as explanations for variability in the frequency of TOP following prenatal diagnosis and as potential sources of selection bias. As an example, ultrasound visualization of the fetus can be difficult when mothers are obese, with prenatal diagnosis of birth defects less likely in obese mothers than normal weight mothers (39, 40). The ultrasound examination might need to be repeated later in pregnancy, as the fetus grows larger, to complete the fetal anatomic examination (41, 42). This might cause obese mothers to have on average a later gestational age at prenatal diagnosis and therefore be less likely to be able to consider a TOP. As a result, obese mothers might be more likely to have a live birth and to be included in epidemiologic studies of these birth defects than non-obese mothers.

Another important consideration in the likelihood of continuning or terminating a pregnancy following prenatal diagnosis is the severity of the defect identified prenatally (13, 38, 43). A more severe NTD or one accompanied by multiple major malformations might be more likely to end in TOP than an isolated NTD or a less severe case. Severity of the defect might be more relevant for spina bifida than anencephaly since the latter is uniformly lethal. In our review there was little information on the effect of severity of spina bifida on likelihood of TOP, but the results suggested a higher frequency of TOP for fetuses with multiple malformations than isolated cases and for those with open defects than closed defects. This point is important when results from studies reporting clinical outcomes such as shunting or mobility impairment are used to counsel families with a prenatally diagnosed fetus on long-term prognosis. Consideration should be given to the possibility that the fetuses most likely to be liveborn and to have follow-up information available are those with the least severe defects; thus, results from studies based on liveborn infants may not be generalizable to all prenatally diagnosed fetuses.

Incomplete ascertainment of cases of anencephaly or spina bifida among pregnancies ending in TOP also pose problems for evaluating population-based interventions for the prevention of these defects. Using data from surveillance to evaluate the success of folic acid fortification programs is difficult because of the high frequency of prenatal diagnosis and TOP and the difficulty in separating the effects of the intervention from changes in prenatal diagnosis and TOP over time (44).

One limitation of this study was the inability of our search strategy to identify all relevant articles. Restricting the search databases to Medline and Embase likely resulted in missed articles in languages other than English and articles from journals not indexed by these databases, particularly those outside North America and Europe. Because reporting the proportion of pregnancies with prenatal diagnosis ending in TOP is not a common study objective, this information is often presented in the text and not the abstract. There might be other articles reporting the frequency of TOP following prenatal diagnosis that are not captured by a search strategy that exclusively searches abstracts; this could affect our results if study results systematically differ between studies identified and not identified by our search strategy. A second limitation of our analysis was the exclusion of prenatally diagnosed fetuses with unknown pregnancy outcomes by us or by authors of the included studies. This exclusion would likely produce underestimates of the frequency of TOP if pregnancies with unknown outcomes are more likely to represent TOP than the more easily ascertained live births.

Our results suggest, in accordance with previous studies, that TOP is the most common outcome of pregnancy following prenatal diagnosis of anencephaly and spina bifida, particularly when the prenatal diagnosis is made prior to 24 weeks of gestation. The relatively small proportion of fetuses with NTDs presenting as live births will present challenges to investigators conducting studies in which not all NTD-affected pregnancies among TOPs are included. For some types of studies, such as evaluation of interventions for prevention, use of alternate sources of data might be warranted; however, a better understanding of factors associated with TOP following prenatal diagnosis of

anencephaly or spina bifida will provide much needed information on the potential for selection bias in etiologic studies and generalizability in studies of the prognosis of prenatally diagnosed fetuses.

Table 3.1. Proportion of Pregnancies Ending in Termination of Pregnancy Following Prenatal Diagnosis of Anencephaly or Spina Bifida, by Geographic Region.

			Anencephaly		Spina Bifida	
Study	Location	Years	n/N[a]	%	n/N[a]	%
North America						
Tairou 2006	Quebec City, Canada	1993-2002	27/40	68	56/85	66
Biggio 2004	Birmingham, USA	1996-2000			20/56	36
Forrester 2000	Hawaii, USA	1986-1997	64/78	82	32/65	49
Waller 2000	Texas, USA	1995	23/36	64	10/27	37
Harmon 1995	Indianapolis, USA	1988-1994			19/61	31
Shulman 1994	Memphis, USA	1988-1993			18/22	82
Europe						
Amari 2010	Lübeck, Germany	1993-2008			68/103	66
Aguilera 2009	Bristol, UK	1999-2007			53/74	72
D'Addario 2008	Bari, Italy	2005-2006			38/49	78

Poretti 2008	Switzerland[b]	2001-2007	20/22	91	35/85	41
Ghi 2006	Bologna, Italy	1997-2004			59/66	89
Nikkila 2006	Malmöhus County	1984-1999	63/69	91		
Garne 2005	Europe[c]	1995-1999	421/469	90	314/405	78
Adama van Scheltema 2003	Leiden, Netherlands	1993-1998	19/24	79	11/26	42
Olde Scholtenhuis 2003	Netherlands[d]	1996-1999			43/88	49

[a] Number of pregnancies ending in termination of pregnancy/number of fetuses prenatally diagnosed.

[b] Basle, Geneva, Lucerne, Zurich.

[c] Belgium (Antwerp, Hainaut), Bulgaria (Sofia), Croatia, Denmark (Funen County), France (Paris, Strasbourg), Germany (Mainz, Saxony-Anhalt), Italy (Campania, Tuscany), Malta, Portugal (South), Spain (Asturias, Basque Country), Switzerland (Vaud), and the United Kingdom (Wales).

[d] Amsterdam, Rotterdam, Utrecht.

Table 3.2. Proportion of Pregnancies Affected by Spina Bifida Ending in Termination of Pregnancy Following Prenatal Diagnosis, by Fetal and Study Characteristics.

		Proportion Ending in Termination			
	Number of Studies[a]	n/N[b]	Percentage	95% CI	Range of Estimates[c]
All studies	14	776/1212	64	61, 67	31-89
Geographic region					
Europe	8	621/896	69	66, 72	41-89
North America	6	155/316	49	44, 55	31-82
Study design					
Cohort	11	420/715	59	55, 62	31-89
Surveillance/registry	3	356/497	72	68, 75	37-78
Case type					
All cases	10	635/941	67	64, 70	41-82
Isolated defects	5	185/334	55	50, 61	31-89
Defect type					
Open	5	197/347	57	52, 62	36-91
Closed	2	4/13	31	11, 59	22-50
Gestational age at prenatal diagnosis					
<24 weeks	5	419/477	88	85, 91	77-92
≥24 weeks	4	60/195	31	25, 38	16-41

Abbreviations: CI, confidence interval.

[a] Studies do not sum to total because studies can be counted in none or more than one category.

[b] Number of pregnancies ending in termination of pregnancy/number of fetuses prenatally diagnosed.

[c] Studies with lowest and highest estimates.

Table 3.3. Proportion of Pregnancies Affected by Spina Bifida Ending in Termination of Pregnancy Following Prenatal Diagnosis, by Gestational Age at Prenatal Diagnosis.

		All Fetuses		Prenatal Diagnosis < 24 Weeks		Prenatal Diagnosis ≥ 24 Weeks	
Study	Country	n/N[a]	%	n/N[a]	%	n/N[a]	%
Garne 2005	Europe[b]	314/405	78	253/278	91	44/107	41
Aguilera 2009	United Kingdom	53/74	72	50/65	77	3/9	33
Amari 2010	Germany	68/103	66	63/74	85	5/29	17
Olde Scholtenhuis 2003	Netherlands	43/88	49	35/38	92	8/50	16

[a] Number of pregnancies ending in termination of pregnancy/number of fetuses prenatally diagnosed.

[b] Belgium, Bulgaria, Croatia, Denmark, France, Germany, Italy, Malta, Portugal, Spain, Switzerland, and the United Kingdom.

APPENDIX

Electronic Search Strategy

The following search strategy was used to identify articles eligible for inclusion using http://www.embase.com (simultaneous search of Embase and Medline) for articles published between 1990 and 12 May 2011:

('neural tube defect'/exp OR 'neural tube defect' OR ntd* OR 'anencephalus'/exp OR anencephaly OR anencephalus OR 'spina bifida'/exp OR 'spina bifida' OR 'meningomyelocele'/exp OR myelomeningocele OR meningomyelocele OR 'meningocele'/exp OR meningocele) AND ('prenatal diagnosis'/exp OR 'prenatal diagnosis' OR prenatal* OR antenatal* OR ultraso* OR sonogra* OR amniocentes* OR chorion*) AND ('abortion'/exp OR 'pregnancy termination'/exp OR 'pregnancy termination' OR terminat* OR abort* OR interrupt*)

REFERENCES

1. Botto LD, Moore CA, Khoury MJ, et al. Neural-tube defects. *N Engl J Med*. 1999;341:1509-1519.

2. Mitchell LE, Adzick NS, Melchionne J, et al. Spina bifida. *Lancet*. 2004;364:1885-1895.

3. Jaquier M, Klein A, Boltshauser E. Spontaneous pregnancy outcome after prenatal diagnosis of anencephaly. *BJOG*. 2006;113:951-953.

4. Obeidi N, Russell N, Higgins JR, et al. The natural history of anencephaly. *Prenat Diagn*. 2010;30:357-360.

5. Doherty D, Shurtleff DB. Pediatric perspective on prenatal counseling for myelomeningocele. *Birth Defects Res A Clin Mol Teratol*. 2006;76:645-653.

6. Bol KA, Collins JS, Kirby RS, et al. Survival of infants with neural tube defects in the presence of folic acid fortification. *Pediatrics*. 2006;117:803-813.

7. Borman B, Cryer C. Fallacies of international and national comparisons of disease occurrence in the epidemiology of neural tube defects. *Teratology*. 1990;42:405-412.

8. Roberts HE, Moore CA, Cragan JD, et al. Impact of prenatal diagnosis on the birth prevalence of neural tube defects, Atlanta, 1990-1991. *Pediatrics*. 1995;96:880-883.

9. Honein MA, Paulozzi LJ, Mathews TJ, et al. Impact of folic acid fortification of the US food supply on the occurrence of neural tube defects. *JAMA*. 2001;285:2981-2986.

10. De Wals P, Tairou F, Van Allen MI, et al. Reduction in neural-tube defects after folic acid fortification in Canada. *N Engl J Med*. 2007;357:135-142.

11. Cameron M, Moran P. Prenatal screening and diagnosis of neural tube defects. *Prenat Diagn*. 2009;29:402-411.

12. Driscoll DA, Gross SJ, Professional Practice Guidelines Committee. Screening for fetal aneuploidy and neural tube defects. *Genet Med*. 2009;11:818-821.

13. Peller AJ, Westgate MN, Holmes LB. Trends in congenital malformations, 1974-1999: effect of prenatal diagnosis and elective termination. *Obstet Gynecol*. 2004;104:957-964.

14. Boyd PA, Devigan C, Khoshnood B, et al. Survey of prenatal screening policies in Europe for structural malformations and chromosome anomalies, and their impact on detection and termination rates for neural tube defects and Down's syndrome. *BJOG*. 2008;115:689-696.

15. Mansfield C, Hopfer S, Marteau TM, et al. Termination rates after prenatal diagnosis of Down syndrome, spina bifida, anencephaly, and Turner and Klinefelter syndromes: a systematic literature review. *Prenat Diagn*. 1999;19:808-812.

16. Bower C, Ryan A, Rudy E. Ascertainment of pregnancies terminated because of birth defects: effect of completeness of adding a new source of data. *Teratology*. 2001;63:23-25.

17. Cragan JD, Gilboa SM. Including prenatal diagnoses in birth defects monitoring: experience of the Metropolitan Atlanta Congenital Defects Program. *Birth Defects Res A Clin Mol Teratol*. 2009;85:20-29.

18. Cragan JD, Khoury MJ. Effect of prenatal diagnosis on epidemiologic studies of birth defects. *Epidemiology*. 2000;11:695-699.

19. Harmon JP, Hiett AK, Palmer CG, et al. Prenatal ultrasound detection of isolated neural tube defects: Is cytogenetic evaluation warranted? *Obstet Gynecol*. 1995;86(4 I):595-599.

20. Forrester MB, Merz RD. Prenatal diagnosis and elective termination of neural tube defects in Hawaii, 1986-1997. *Fetal Diagn Ther*. 2000;15(3):146-151.

21. Waller DK, Pujazon MA, Canfield MA, et al. Frequency of prenatal diagnosis of birth defects in Houston, Galveston and the Lower Rio Grande Valley, Texas 1995. *Fetal Diagn Ther*. 2000;15(6):348-354.

22. Olde Scholtenhuis MAG, Cohen-Overbeek TE, Offringa M, et al. Audit of prenatal and postnatal diagnosis of isolated open spina bifida in three university hospitals in The Netherlands. *Ultrasound Obstet Gynecol*. 2003;21(1):48-52.

23. Biggio JR, Wenstrom KD, Owen J. Fetal open spina bifida: A natural history of disease progression in utero. *Prenat Diagn*. 2004;24(4):287-289.

24. Garne E, Loane M, Dolk H, et al. Prenatal diagnosis of severe structural congenital malformations in Europe. *Ultrasound Obstet Gynecol*. 2005;25(1):6-11.

25. Ghi T, Pilu G, Falco P, et al. Prenatal diagnosis of open and closed spina bifida. *Ultrasound Obstet Gynecol*. 2006;28(7):899-903.

26. Tairou F, De Wals P, Bastide A. Validity of death and stillbirth certificates and hospital discharge summaries for the identification of neural tube defects in Quebec City. *Chronic Dis Can*. 2006;27(3):120-124.

27. D'Addario V, Rossi AC, Pinto V, et al. Comparison of six sonographic signs in the prenatal diagnosis of spina bifida. *J Perinat Med*. 2008;36(4):330-334.

28. Poretti A, Anheier T, Zimmermann R, et al. Neural tube defects in Switzerland from 2001 to 2007: Are periconceptual folic acid recommendations being followed? *Swiss Med Wkly*. 2008;138(41-42):608-613.

29. Aguilera S, Soothill P, Denbow M, et al. Prognosis of spina bifida in the era of prenatal diagnosis and termination of pregnancy. *Fetal Diagn Ther*. 2009;26(2):68-74.

30. Amari F, Junkers W, Hartge D, et al. Prenatal course and outcome in 103 cases of fetal spina bifida: A single center experience. *Acta Obstet Gynecol Scand*. 2010;89(10):1276-1283.

31. Nikkila A, Rydhstrom H, Kallen B. The incidence of spina bifida in Sweden 1973-2003: The effect of prenatal diagnosis. *Eur J Public Health*. 2006;16(6):660-662.

32. Adama van Scheltema PN, Nagel HT, Brouwer OF, et al. Outcome of children with prenatally diagnosed central nervous system malformations. *Ultrasound Obstet Gynecol*. 2003;21:41-47.

33. Shulman LP, Grevengood C, Phillips OP, et al. Family planning decisions after prenatal detection of fetal abnormalities. *Am J Obstet Gynecol*. 1994;171:1373-1376.

34. Cragan JD, Roberts HE, Edmonds LD, et al. Surveillance for anencephaly and spina bifida and the impact of prenatal diagnosis -- United States, 1985-1994. *MMWR CDC Surveill Summ*. 1995;44:1-13.

35. Velie EM, Shaw GM. Impact of prenatal diagnosis and elective termination on prevalence and risk estimates of neural tube defects in California, 1989-1991. *Am J Epidemiol*. 1996;144(5):473-479.

36. Parks SE, Canfield MA, Ramadhani TA. Importance of including all pregnancy outcomes to reduce bias in epidemiologic studies of neural tube defects -- Texas, 1999 to 2005. *Birth Defects Res A Clin Mol Teratol*. 2011;91:185-191.

37. Schechtman KB, Gray DL, Baty JD, et al. Decision-making for termination of pregnancies with fetal anomalies: analysis of 53,000 pregnancies. *Obstet Gynecol*. 2002;99:216-222.

38. Shaffer BL, Caughey AB, Norton ME. Variation in the decision to terminate pregnancy in the setting of fetal aneuploidy. *Prenat Diagn*. 2006;26:667-671.

39. Dashe JS, McIntire DD, Twickler DM. Effect of maternal obesity on the ultrasond detection of anomalous fetuses. *Obstet Gynecol*. 2009;113(5):1001-1007.

40. Johnson CY, Honein MA, Hobbs CA, et al. Prenatal diagnosis of orofacial clefts, National Birth Defects Prevention Study, 1998-2004. *Prenat Diagn*. 2009;29(9):833-839.

41. Hendler I, Blackwell SC, Bujold E, et al. The impact of maternal obesity on midtrimester sonographic visualization of fetal cardiac and craniospinal structures. *Int J Obes Relat Metab Disord*. 2004;28:1607-1611.

42. Phatak M, Ramsay J. Impact of maternal obesity on procedure of mid-trimester anomaly scan. *J Obstet Gynaecol*. 2010;30:447-450.

43. Evans MI, Sobiecki MA, Krivchenia EL, et al. Parental decisions to terminate/continue following abnormal cytogenetic prenatal diagnosis: "what" is still more important than "when". *Am J Med Genet*. 1996;61:353-355.

44. Besser LM, Williams LJ, Cragan JD. Interpreting changes in the epidemiology of anencephaly and spina bifida following folic acid fortification of the U.S. grain supply in the setting of long-term trends, Atlanta, Georgia, 1968-2003. *Birth Defects Res A Clin Mol Teratol*. 2007;79:730-739.

CHAPTER 4

Prenatal Diagnosis of Spina Bifida

Candice Y. Johnson, Margaret A. Honein, W. Dana Flanders, Penelope P. Howards, Godfrey P. Oakley, Jr., Sonja A. Rasmussen, and The National Birth Defects Prevention Study

Author affiliations: Departments of Epidemiology (Candice Y. Johnson, W. Dana Flanders, Penelope P. Howards, Godfrey P. Oakley, Jr.) and Biostatistics and Bioinformatics (W. Dana Flanders), Emory University, Atlanta, Georgia; National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention (Candice Y. Johnson, Margaret A. Honein, Sonja A. Rasmussen).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Corresponding author: Candice Y. Johnson, cyjohnson@alum.emory.edu.

ABSTRACT

With routine use of maternal serum screening and prenatal ultrasound, spina bifida is commonly diagnosed prenatally. Our aims were to investigate characteristics associated with prenatal diagnosis of spina bifida and to evaluate the agreement between maternal report and medical record report of prenatal diagnosis. We included 714 mothers of infants with spina bifida from the National Birth Defects Prevention Study, 1998-2005. We assessed sensitivity of maternal report compared to medical record abstract of prenatal diagnosis (reported vs. not reported) and timing of prenatal diagnosis (<24 vs. ≥24 gestational weeks) and used logistic regression to identify characteristics associated with prenatal diagnosis and timing of prenatal diagnosis. Sensitivity of maternal report was 83% for prenatal diagnosis and 98% for timing of prenatal diagnosis. Prenatal diagnosis and early prenatal diagnosis were associated with maternal serum screening, periconceptional folic acid supplementation, household income ≥$50,000, and having more than a high school education. Maternal report could be a reliable source of data for future studies investigating factors associated with prenatal diagnosis and timing of prenatal diagnosis such as health-seeking behaviors and socioeconomic status.

INTRODUCTION

Spina bifida is a birth defect resulting from failure of the neural tube to close completely during embryonic development (1). Use of prenatal screening for the detection of spina bifida and other birth defects has increased rapidly over the last several decades (2). Maternal serum alpha-fetoprotein (MSAFP) screening in the second trimester of pregnancy can identify open spina bifida and fetal imaging studies such as ultrasonography are commonly used for screening or for confirmation of the defect (3).

The frequency of prenatal diagnosis of spina bifida has been estimated at 42-59% in the United States (4-6). Prenatal diagnosis allows families who want this information the opportunity to make decisions about continuing the pregnancy or to plan for the birth, including arranging to deliver in a facility better equipped to treat the newborn and exploring options such as fetal surgery (5). Understanding factors associated with prenatal diagnosis is important for identifying subgroups of mothers who are less likely to have a prenatal diagnosis and therefore would not be able to consider all available options prior to the infant's birth. Previous studies have reported that the frequency of prenatal diagnosis of birth defects varies by maternal race/ethnicity, prepregnancy body mass index, household income, and other characteristics (4-9). Few studies, however, have examined if any of these characteristics are specifically associated with prenatal diagnosis of spina bifda.

Investigating prenatal diagnosis in population-based studies can be challenging if data from medical records are not available to confirm if a prenatal diagnosis was made.

Using maternal self-report of prenatal diagnosis is an alternative to medical records data. Although studies comparing maternal report to medical record report of pregnancy-related events have shown good agreement (10-13), the accuracy of maternally reported prenatal diagnosis has not been well studied.

The purpose of this study is two-fold: to evaluate agreement between maternal and medical record abstract of prenatal diagnosis of spina bifida and to investigate maternal and pregnancy characteristics associated with prenatal diagnosis.

METHODS

Population

The National Birth Defects Prevention Study is an ongoing, population-based case-control study of risk factors for major structural birth defects (14). Cases are identified through active birth defect surveillance in Arkansas, California, Georgia, Iowa, Massachusetts, New York, New Jersey, North Carolina, Texas, and Utah. Eight sites systematically ascertain cases among live births, stillbirths (fetal deaths $\geq$ 20 gestational weeks), and terminations of pregnancy (any gestational age). New Jersey includes cases among live births and stillbirths only, and Massachusetts only systematically includes live births.

Clinical geneticists review data from medical records for all potentially eligible cases to determine eligibility and to classify the case into the appropriate defect group (15). Cases with confirmed or suspected chromosomal anomalies or single-gene disorders are

excluded. Mothers agreeing to participate in the study are contacted 6 weeks to 2 years following delivery to complete a telephone interview including questions on maternal sociodemographics, pregnancy history, and chronic and acute illnesses, medication use, diet, smoking, alcohol use, and occupation during pregnancy. Information on the mother's reproductive and pregnancy histories are also abstracted from medical records. Institutional review board approval was obtained from all participating sites.

For this study, eligible mothers were case mothers who completed the interview and who had a child with spina bifida with a date of delivery on or after January 1, 1998 and with an expected date of delivery on or before December 31, 2005.

Definition of Prenatal Diagnosis

In the interview, mothers were asked, "Did you have any ultrasounds which showed any abnormalities with the fetus, placenta, or fluid?" and were then asked to describe the abnormality. Fetuses were classified as prenatally diagnosed if the maternal report fell into at least one of the following categories:

- The defect was identified by name (e.g., "spina bifida" or synonym)
- The defect was not identified by name but an anatomical description consistent with the defect was given (e.g., "herniation of spinal cord")
- The defect was not identified by name but an anomaly was reported that most likely represents the defect (e.g., "there was a lump on the spine")

- Fetal surgery for *in utero* spina bifida repair was reported, implying a prenatal diagnosis.

Determination of prenatal diagnosis status was made with knowledge of the true (postnatally confirmed) diagnosis to give the benefit of the doubt that a prenatal diagnosis was made when the description would have otherwise been too vague to determine the exact type of defect. We also classified as prenatal diagnosis those diagnoses reported in the interview in response to questions about amniocentesis, chorionic villus sampling, MSAFP screening, and other prenatal imaging and tests. Although most of these test results would not be diagnostic for spina bifida, we assume that prenatal diagnoses reported in response to these questions reflect that a prenatal diagnosis was made but the mother did not remember which test provided the definitive diagnosis.

Gestational Age at Prenatal Diagnosis

Gestational age at prenatal diagnosis was only calculated for mothers who self-reported that a prenatal diagnosis was made. In some pregnancies, the defect was only suspected at first and was confirmed at a later date, after further testing. Because it was difficult to evaluate the level of certainty for any given ultrasound result, the earliest prenatal test result with any mention of the defect was used to determine the date of prenatal diagnosis.

If the number of completed gestational weeks at prenatal diagnosis was reported, this was used as the gestational age at prenatal diagnosis. Otherwise, we calculated gestational age

by counting the number of completed weeks elapsed between the estimated date of the last menstrual period (due date - 280 days) and the date of the ultrasound. We categorized gestational age as an early prenatal diagnosis (<24 completed gestational weeks) or late prenatal diagnosis (≥24 weeks). We chose 24 weeks as the cutpoint because in many regions of the United States and other countries, this is the gestational age after which termination of pregnancy might not be an available option.

If the mother reported only the month and year of the ultrasound, we assigned a date if she indicated the time of the month (beginning, middle, end) when the test was conducted. Ultrasounds at the beginning of the month were assigned the 5^{th} day of the month, in the middle the 15^{th} day, and at the end, the 25^{th} day. If the mother reported only the month of pregnancy during which the prenatal diagnosis was made, we classified months 1 to 6 as <24 completed gestational weeks and months 7 to 10 as ≥24 weeks.

Agreement Between Maternal and Medical Record Abstract

Prenatal diagnosis and gestational age at prenatal diagnosis from the medical record abstract were defined in the same way as for maternal report. We included prenatal diagnoses reported in the medical record abstractor's notes about the case even when no abstracted record of an ultrasound or prenatal test was available. When gestational age at prenatal diagnosis was estimated from both ultrasound and date of last menstrual period, we used the ultrasound estimate of gestational age.

Studying prenatal diagnosis was not a specific aim of NBDPS and as a result prenatal diagnoses were not systematically abstracted from medical records. For analysis of agreement between maternal report and medical abstracts, we excluded mothers from California and New Jersey because medical record abstracts were available for only a small subset of mothers. For the other study sites, the completeness of medical record abstracts for prenatal diagnoses was unknown. Because of this, we measured agreement using sensitivity of maternal report compared to medical record abstract. Sensitivity of maternal report of prenatal diagnosis was calculated as the proportion of mothers with a medical record abstract of prenatal diagnosis who reported that a prenatal diagnosis was made. Sensitivity of maternal report of timing of prenatal diagnosis was defined as the proportion of mothers with an early prenatal diagnosis in the medical record abstract who reported that an early prenatal diagnosis was made.

These metrics uses as a denominator only medical record abstracts in which prenatal diagnoses are recorded; in this subset of mothers, medical records can be considered the gold standard (because there is confirmation that a prenatal diagnosis was made) and sensitivity of maternal report can be assessed. We assume that mothers with and without medical record abstracts are comparable.

Characteristics Associated With Prenatal Diagnosis

We investigated associations between prenatal diagnosis (reported versus not reported) and timing of prenatal diagnosis (early versus late) with the following maternal characteristics: age at delivery (<20, 20-24, 25-29, 30-34, ≥35), race/ethnicity (non-

Hispanic white, non-Hispanic black, Hispanic, other), maternal education (<12 years, 12 years, >12 years), household income (<$50,000, ≥$50,000), folic acid supplementation in the month before and 1st month of pregnancy (yes, no), smoking between the month prior to conception and the 3rd month of pregnancy (yes, no), alcohol consumption during this same period (yes, no), and prepregnancy body mass index (BMI; underweight [<18.5 kg/m^2], normal weight [18.5-24.9 kg/m^2], overweight [25.0-29.9 kg/m^2], obese [≥30 kg/m^2]). We also investigated associations between the same outcomes and pregnancy-related characteristics: year of due date (1998-1999, 2000-2001, 2002-2003, 2004-2005), initiation of prenatal care (1st trimester, 2nd or 3rd trimester, no prenatal care), parity (0, ≥1), MSAFP screening (yes, no), plurality (singleton, twins or higher order multiples), and presence of other major birth defects (yes, no).

Statistical Analyses

We used logistic regression to estimate crude odds ratios (OR) and 95% confidence intervals (CI) for associations between each characteristic and prenatal diagnosis and timing of prenatal diagnosis. For the analysis of timing of prenatal diagnosis, we performed one analysis among all mothers (mothers not reporting a prenatal diagnosis were classified as having a prenatal diagnosis ≥24 weeks) and another restricting to mothers who reported having a prenatal diagnosis.

RESULTS

Agreement

Agreement between maternal report and medical record abstract was better for timing of prenatal diagnosis (sensitivity of maternal report: 98%) than for whether or not a prenatal diagnosis was made (sensitivity of maternal report: 83%) (Tables 4.1 and 4.2). Agreement was similar between live births, stillbirths, and terminations (data not shown).

For the 220 of mothers who reported prenatal diagnosis and had gestational age at prenatal diagnosis available from both the interview and medical record abstracts (i.e., sufficient information was available to determine the gestational week at prenatal diagnosis), 77 (35%) had exact agreement on gestational age at prenatal diagnosis, 103 (47%) agreed within 1 week, 139 (63%) agreed within 2 weeks, and 176 (80%) agreed within 4 weeks.

Frequency of Prenatal Diagnosis

In our dataset, 389 of 714 mothers (54%) reported that spina bifida was identified prenatally (Table 4.3). Overall, 40% of mothers reported prenatal diagnosis <24 weeks; when restricting to mothers who reporting having a prenatal diagnosis, 82% of these diagnoses were made <24 weeks.

Frequency of maternally-reported prenatal diagnosis varied somewhat by study side, ranging from 43-68% for prenatal diagnosis, 29-56% for prenatal diagnosis <24 weeks overall, and 71-100% for prenatal diagnosis <24 weeks among mothers reporting a prenatal diagnosis (Table 4.4).

Characteristics Associated With Prenatal Diagnosis

Prenatal diagnosis was less likely in Hispanic mothers compared to non-Hispanic white mothers (OR 0.59, 95% CI 0.43, 0.83), mothers with <12 years of education compared to those with 12 years of education (OR 0.62, 95% CI 0.40, 0.95), and overweight mothers compared to normal weight mothers (OR 0.56, 95% CI 0.39, 0.85). Prenatal diagnosis was more likely in mothers reporting periconceptional folic acid supplementation (OR 1.69, 95% CI 1.26, 2.28), mothers who reported having MSAFP screening (OR 1.92, 95% CI 1.41, 2.63), and mothers of twins or higher order multiples (OR 1.95, 95% CI 0.79, 4.81) (Tables 4.5 and 4.6).

Timing of prenatal diagnosis overall was associated with these same variables, with similar magnitudes and directions of the association. However, differences were noted when restricting to mothers who reported prenatal diagnosis. In this restricted subgroup, timing of prenatal diagnosis was no longer inversely associated with Hispanic ethnicity (OR 0.92, 95% CI 0.48, 1.76), less than high school education (OR 0.95, 95% CI 0.40, 2.26), and overweight (OR 0.70, 95% CI 0.46, 1.06). In this subgroup, mothers reporting smoking during pregnancy were less likely to have an early prenatal diagnosis (OR 0.52, 95% CI 0.26, 1.02), as were mothers entering prenatal care after the 1st trimester of pregnancy (OR 0.52, 95% CI 0.25, 1.09), and mothers with prepregnancy obesity (OR 1.46, 95% CI 0.62, 3.43) (Tables 4.7 and 4.8).

DISCUSSION

In this study, we found moderate agreement between maternal report and medical record abstract of prenatal diagnosis of NTDs and high agreement for timing of prenatal diagnosis among mothers reporting prenatal diagnosis. Using maternal report in our further analyses, we found that several characteristics, such as MSAFP screening, folic acid supplementation, household income, and maternal education were associated with prenatal diagnosis and timing of prenatal diagnosis.

Understanding factors associated with prenatal diagnosis and early prenatal diagnosis is important to determine if there are certain subgroups of mothers less likely to have a prenatal diagnosis and therefore be unable to consider all available options for the pregnancy. This could include deciding whether or not to continue the pregnancy, choosing to deliver at a hospital better equipped to treat the newborn, researching postnatal interventions, or investigating options for fetal surgery. A recent trial of *in utero* spina bifida repair found that infants randomized to fetal surgery had lower rates of death and shunt placement than those randomized to receive no surgery, although the surgery was associated with important adverse outcomes such as preterm birth (16). Investigators suggest that surgeries be performed between 19 and 25 gestational weeks (16, 17), highlighting the potential benefits of an early prenatal diagnosis.

In studies in which data from medical records are unavailable, using maternally-reported prenatal diagnosis could be an option for studying factors related to prenatal diagnosis. In our population, 83% of mothers reported prenatal diagnosis when a prenatal diagnosis was indicated in the medical record abstract. Of the remaining 17% of mothers, a small

proportion could represent mothers who truly did not know a prenatal diagnosis of spina bifida had been made. More likely, the remaining mothers are those whose answers to the questions on prenatal diagnosis were too vague for us to confidently code as a prenatal diagnosis of spina bifida. In future analyses, we will determine if this is the case and if there are specific maternal characteristics associated with quality of the maternal report.

Among mothers reporting a prenatal diagnosis, 98% of maternal reports agreed with medical record abstracts indicating an early prenatal diagnosis. When gestational ages at prenatal diagnosis were compared between the two sources, there was a difference of 5 or more weeks for 20% of mothers. This suggests that studies requiring exact week of prenatal diagnosis might not find maternal report to be reliable, but that mothers can provide an approximate gestational age sufficient for categorization of prenatal diagnosis as early or late in pregnancy. It should be noted that in this analysis medical record abstracts were not necessarily the gold standard; it is possible that the ultrasound date abstracted from the medical record was not the earliest ultrasound that identified the defect.

There was variability in the frequency of prenatal diagnosis and timing of prenatal diagnosis between NBDPS study sites, which could indicate differences in prenatal care access or use. Despite availability of second trimester MSAFP screening and ultrasonograpy, in all sites aside from Utah, fewer than half of mothers reported that a prenatal diagnosis was made before 24 weeks. Differences in ascertainment of cases of spina bifida among terminations of pregnancy are other possible reasons for variability

between sites. Terminations of pregnancy are difficult to identify and require multiple sources of ascertainment; some study sites do not systematically ascertain cases among terminations, and in the other study sites, ascertainment is likely incomplete. Because terminations of pregnancy occur following prenatal diagnosis and are only <24 weeks in some regions, underascertainment of terminations would result in an underestimation of the proportion of fetuses prenatally diagnosed and an underestimation of the proportion of prenatal diagnoses occurring <24 weeks. In future analyses, the number of terminations of pregnancy missed will be estimated and used to adjust results for underascertainment.

Because of this underascertainment, in the analysis of chraracteristics associated with prenatal diagnosis, selection bias could arise if termination of pregnancy is associated with the characteristic under investigation (18). In future analyses, we will adjust for this potential selection bias. Until then, selection bias could be hypothesized to explain some of the results we discuss next.

Mothers reporting MSAFP screening during pregnancy were more likely to report prenatal diagnosis and early prenatal diagnosis (both in the overall sample and among mothers who reported prenatal diagnosis). MSAFP screening is recommended for all women during pregnancy and can identify most fetuses with open spina bifida (19). High MSAFP levels would likely prompt a thorough sonographic examination of the fetal anatomy and increase the likelihood a prenatal diagnosis is made; prenatal diagnosis might therefore occur earlier with MSAFP screening than without. Because both MSAFP

screening and prenatal diagnosis by ultrasound depend on mothers presenting for prenatal care at gestational ages during which these tests can be conducted, the association could be confounded by patterns of prenatal care use; future analyses conducted separately for open and closed spina bifida will provide evidence if this might be the case.

Previous studies have shown that prenatal diagnosis of birth defects is less likely in obese mothers than mothers who are normal weight because of suboptimal visualization of the fetus (8, 9, 20). In this study, we did not find the clear dose-response effect that has been observed in studies of other types of birth defects; overweight but not obesity was associated with decreased likelihood of prenatal diagnosis. Unlike most other birth defects, open spina bifida can be detected not only by ultrasound but through MSAFP screening, which is less susceptible to differences in detection rates by BMI, and so we might not expect to see the same relationship between BMI and prenatal diagnosis for these defects.

Household income, maternal education, and folic acid supplementation were fairly strongly associated with prenatal diagnosis, but none of these characteristics or behaviors would directly affect whether or not a prenatal diagnosis is made by the sonographer. Instead, these variables could be markers of health literacy, acceptance of prenatal testing, or attendance at scheduled prenatal care visits. Because maternal education was not associated with timing of prenatal diagnosis when restricted to mothers who reported prenatal diagnosis, it is possible that maternal education affects only the likelihood a mother has prenatal testing, but not the time at which the test is performed. In contrast,

household income and folic acid supplementation were associated both with prenatal diagnosis and the gestational age at which the prenatal diagnosis was made.

One limitation of our dataset was that investigation of prenatal diagnosis was not a specific aim of NBDPS, and as a result information on prenatal diagnosis from the interview was limited. We had no information about ultrasounds returning normal results and so we could not distinguish between mothers who had multiple ultrasounds during pregnancy but the defect was missed and mothers who had no ultrasound at all or no ultrasound during gestational ages at which prenatal diagnosis would be most likely. In addition, medical record abstraction of prenatal diagnoses was likely incomplete. Having no prenatal diagnosis reported in the medical record abstract could indicate either that the defect was not seen on ultrasound or that the prenatal diagnosis was not abstracted. Although agreement was good between maternal report and medical record abstract, it was not perfect, raising the possibility that associations we observed are characteristics associated with *reporting* a prenatal diagnosis and not prenatal diagnosis itself.

Although in our analyses we identified several characteristics associated with prenatal diagnosis and timing of prenatal diagnosis, the reasons for these associations remain unclear. Further studies will be needed to determine the reasons for the associations with prenatal diagnosis; for example, if they are markers for compliance with recommended prenatal care visits, access to health care, health literacy, or other factors related to prenatal diagnosis. In addition, the potential roles for confounding, misclassification, and selection bias should be further explored. In future studies, maternally-reported prenatal

diagnosis might be a useful and convenient measure of prenatal diagnosis and could allow exploration of characteristics associated with prenatal diagnosis in population-based studies.

Table 4.1. Agreement Between Maternal Report and Medical Record Abstract of Prenatal Diagnosis of Spina Bifida.

	Medical Record Abstract	
	Reported	Not Reported
Maternal Report		
Reported	243	62
Not Reported	49	170
Sensitivity[a], % (95% CI)	83 (79, 87)	

Abbreviations: CI, confidence interval.

[a] Sensitivity of maternal report compared to medical record abstract.

Table 4.2. Agreement Between Maternal Report and Medical Record Abstract of Whether Prenatal Diagnosis of Spina Bifida was Made <24 or ≥24 Weeks of Gestation.

	All Mothers (n = 449)		Mothers Reporting Prenatal Diagnosis (n = 194)	
	Medical Record Abstract			
	<24	≥24	<24	≥24
Maternal Report				
<24 weeks	141	52	141	19
≥24 weeks	28	228	3	31
Sensitivity[a], % (95% CI)	83 (77, 88)		98 (94, 99)	

Abbreviations: CI, confidence interval.

[a] Sensitivity of maternal report compared to medical record abstract.

Table 4.3. Frequency of Maternal Self-Report of Prenatal Diagnosis of Spina Bifida, National Birth Defects Prevention Study, 1998-2005.

	n/N	%
Prenatal diagnosis[a]	389/714	54
Diagnosis < 24 weeks		
Overall[b]	252/633	40
Prenatal diagnoses[c]	252/308	82

[a] Number of mothers reporting prenatal diagnosis over total number of mothers.

[b] Number of mothers reporting prenatal diagnosis <24 gestational weeks over total number of mothers.

[c] Number of mothers reporting prenatal diagnosis <24 gestational weeks over total number of mothers reporting prenatal diagnosis.

Table 4.4. Frequency of Maternal Report of Prenatal Diagnosis and Early Prenatal Diagnosis of Spina Bifida, by National Birth Defects Prevention Study Site, 1997-2005.

			Prenatal Diagnosis < 24 Gestational Weeks			
	Prenatal diagnosis		All Fetuses		Prenatally Diagnosed[a]	
Study Site	n/N	%	n/N	%	n/N	%
Arkansas	47/90	52	26/78	33	26/35	74
California	60/138	43	35/120	29	35/42	83
Georgia	49/80	61	30/65	46	30/34	88
Iowa	59/96	61	42/87	48	42/50	84
Massachusetts	26/47	55	15/42	36	15/21	71
New Jersey	24/52	46	16/47	34	16/19	84
New York	30/44	68	18/38	47	18/24	75
North Carolina	12/28	43	11/27	41	11/11	100
Texas	49/89	55	32/81	40	32/41	78
Utah	33/50	66	27/48	56	27/31	87

[a] Mothers reporting prenatal diagnosis.

Table 4.5. Maternal Characteristics Investigated in Association With Maternal Report of Prenatal Diagnosis of Spina Bifida.

	n/N	%	OR	95% CI
Age at delivery				
<20	41/77	53	1.09	0.65, 1.83
20-24	92/167	55	1.18	0.79, 1.75
25-29	119/233	51	1.00	Referent
30-34	86/146	59	1.37	0.90, 2.09
≥35	51/91	56	1.22	0.75, 1.99
Race/ethnicity				
NH white	227/382	59	1.00	Referent
NH black	33/62	53	0.78	0.45, 1.33
Hispanic	106/228	46	0.59	0.43, 0.83
Other	21/40	53	0.76	0.39, 1.45
Missing	2/2	100		
Education				
<12 years	59/141	42	0.62	0.40, 0.95
12 years	110/204	54	1.00	Referent
>12 years	220/369	60	1.26	0.89, 1.78
Household income				
<$50,000	260/499	52	0.60	0.41, 0.86
≥$50,000	104/161	65	1.00	Referent
Missing	25/54	46		

Folic acid				
No	179/371	48	1.00	Referent
Yes	210/343	61	1.69	1.26, 2.28
Smoking				
No	314/483	54	1.00	Referent
Yes	73/129	57	1.12	0.76, 1.64
Missing	2/2	100		
Alcohol use				
No	250/473	53	1.00	Referent
Yes	137/239	57	1.20	0.88, 1.64
Missing	2/2	100		
Prepregnancy BMI				
Underweight	16/25	64	1.28	0.55, 2.99
Normal weight	183/315	58	1.00	Referent
Overweight	72/162	44	0.56	0.39, 0.85
Obese	99/171	58	0.99	0.68, 1.45
Missing	19/41	46		

Abbreviations: BMI, body mass index; CI, confidence interval; n/N, mothers reporting prenatal diagnosis over total number of mothers; OR, odds ratio.

Table 4.6. Pregnancy-Related Characteristics Investigated in Association With Maternal Report of Prenatal Diagnosis of Spina Bifida.

	n/N	%	OR	95% CI
Due date[a]				
1998-1999	85/160	53	1.00	Referent
2000-2001	87/160	54	1.05	0.68, 1.63
2002-2003	75/151	53	1.00	0.64, 1.58
2004-2005	73/123	59	1.29	0.80, 2.07
Prenatal care entry				
1st trimester	316/581	54	1.00	Referent
2nd or 3rd trimester	62/109	57	1.11	0.73, 1.67
No prenatal care	0/10	0		
Missing	11/14	79		
Parity				
0	137/247	55	1.00	Referent
≥1	252/467	54	0.94	0.69, 1.28
MSAFP screening				
No	137/294	47	1.00	Referent
Yes	228/364	63	1.92	1.41, 2.63
Missing	24/56	43		
Plurality				
Singleton	371/688	54	1.00	Referent
Twins or higher	16/23	70	1.95	0.79, 4.81

Missing	2/3	67		
≥1 major defect				
No	348/633	55	1.00	Referent
Yes	41/81	51	0.84	0.53, 1.33

Abbreviations: BMI, body mass index; CI, confidence interval; MSAFP, maternal serum alpha-fetoprotein; n/N, mothers reporting prenatal diagnosis over total number of mothers; OR, odds ratio.

Table 4.7. Maternal Characteristics Investigated in Association With Maternal Report of Timing of Prenatal Diagnosis of Spina Bifida.

	All Fetuses				Prenatally Diagnosed			
	n/N[a]	%	OR	95% CI	n/N[b]	%	OR	95% CI
Age at delivery								
<20	22/65	34	0.95	0.53, 1.72	22/29	76	0.95	0.36, 2.51
20-24	61/146	42	1.34	0.87, 2.07	61/71	86	1.84	0.81, 4.18
25-29	73/209	35	1.00	Referent	73/95	77	1.00	Referent
30-34	63/135	47	1.63	1.05, 2.54	63/75	84	1.58	0.73, 3.45
≥35	33/78	42	1.37	0.80, 2.33	33/38	87	1.99	0.69, 5.71
Race/ethnicity								
Non-Hispanic white	150/339	44	1.00	Referent	150/184	82	1.00	Referent
Non-Hispanic black	21/54	39	0.80	0.45, 1.44	21/25	84	1.19	0.38, 3.69
Hispanic	69/208	33	0.63	0.44, 0.90	69/86	80	0.92	0.48, 1.76
Other	10/30	33	0.63	0.29 ,1.39	10/11	91	2.27	0.28, 18.31

Missing	2/2	100			2/2	100		
Education								
<12 years	35/127	28	0.61	0.38, 1.00	35/45	78	0.95	0.40, 2.26
12 years	70/183	38	1.00	Referent	70/89	79	1.00	Referent
>12 years	147/323	46	1.35	0.93, 1.95	147/174	84	1.48	0.77, 2.84
Household income								
<$50,000	164/443	37	0.51	0.35, 0.75	164/204	80	0.53	0.25, 1.12
≥$50,000	77/144	53	1.00	Referent	77/87	89	1.00	Referent
Missing	11/46	24			11/17	65		
Folic acid								
No	97/327	30	1.00	Referent	97/135	72	1.00	Referent
Yes	155/306	51	2.43	1.76, 3.37	155/173	90	3.37	1.82, 6.24
Smoking								
No	211/521	41	1.00	Referent	211/252	84	1.00	Referent
Yes	40/111	36	0.83	0.54, 1.27	40/55	73	0.52	0.26, 1.02

Missing	1/1	100			1/1	100		
Alcohol use								
No	157/413	38	1.00	Referent	157/190	83	1.00	Referent
Yes	94/218	43	1.24	0.89, 1.73	94/116	81	0.90	0.49, 1.63
Missing	1/2	50			1/2	50		
Prepregnancy BMI								
Underweight	13/22	59	1.95	0.81, 4.71	13/13	100		
Normal weight	118/277	43	1.00	Referent	118/145	81	1.00	Referent
Overweight	51/149	34	0.70	0.46, 1.06	51/59	86	1.46	0.62, 3.43
Obese	55/144	38	0.83	0.55, 1.26	55/72	76	0.74	0.37, 1.47
Missing	15/41	37			15/19	79		

Abbreviations: BMI, body mass index; CI, confidence interval; OR, odds ratio.

[a] Mothers reporting early prenatal diagnosis over total number of mothers.

[b] Mothers reporting early prenatal diagnosis over mothers reporting prenatal diagnosis.

Table 4.8. Pregnancy-Related Characteristics Investigated in Association With Maternal Report of Timing of Prenatal Diagnosis of Spina Bifida.

	All Fetuses				Prenatally Diagnosed			
	n/N[a]	%	OR	95% CI	n/N[b]	%	OR	95% CI
Due date[c]								
1998-1999	44/134	33	1.00	Referent	44/59	75	1.00	Referent
2000-2001	58/146	40	1.35	0.83, 2.20	58/73	79	1.32	0.58, 2.98
2002-2003	48/125	38	1.28	0.77, 2.12	48/59	81	1.49	0.62, 3.58
2004-2005	48/106	45	1.69	1.00, 2.86	48/56	86	2.05	0.79, 5.29
Prenatal care entry								
1st trimester	215/523	41	1.00	Referent	215/258	83	1.00	Referent
2nd or 3rd trimester	31/90	34	0.75	0.47, 1.20	31/43	72	0.52	0.25, 1.09
No prenatal care	0/10	0						
Missing	6/10	60			6/7	86		
MSAFP screening								

No	84/265	32	1.00	Referent	84/108	78	1.00	Referent
Yes	156/319	49	2.06	1.47, 2.90	156/183	85	1.65	0.90, 3.04
Missing	12/49	24			12/17	71		
Parity								
0	87/216	40	1.00	Referent	87/106	82	1.00	Referent
≥1	165/417	40	0.97	0.69, 1.36	165/202	82	0.97	0.53, 1.79
Plurality								
Singleton	238/609	39	1.00	Referent	238/292	82	1.00	Referent
Twins or higher	12/21	57	2.08	0.86, 5.01	12/14	86	1.36	0.30, 6.26
Missing	2/3	67			2/2	100		
≥1 major defect								
No	224/560	40	1.00	Referent	224/275	81	1.00	Referent
Yes	28/73	38	0.93	0.57, 1.54	28/33	85	1.28	0.47, 3.46

Abbreviations: CI, confidence interval; MSAFP, maternal serum alpha-fetoprotein; OR, odds ratio.

[a] Mothers reporting early prenatal diagnosis over total number of mothers.

[b] Mothers reporting early prenatal diagnosis over mothers reporting prenatal diagnosis.

[c] Restricted to study sites contributing cases and controls in all study years: Arkansas, California, Georgia, Iowa, Massachusetts, New York, Texas.

REFERENCES

1. Botto LD, Moore CA, Khoury MJ, et al. Neural-tube defects. *N Engl J Med*. 1999;341:1509-1519.

2. Peller AJ, Westgate MN, Holmes LB. Trends in congenital malformations, 1974-1999: effect of prenatal diagnosis and elective termination. *Obstet Gynecol*. 2004;104:957-964.

3. Cameron M, Moran P. Prenatal screening and diagnosis of neural tube defects. *Prenat Diagn*. 2009;29(4):402-411.

4. Forrester MB, Merz RD. Prenatal diagnosis and elective termination of neural tube defects in Hawaii, 1986-1997. *Fetal Diagn Ther*. 2000;15:146-151.

5. Waller DK, Pujazon MA, Canfield MA, et al. Frequency of prenatal diagnosis of birth defects in Houston, Galveston and the Lower Rio Grande Valley, Texas 1995. *Fetal Diagn Ther*. 2000;15:348-354.

6. Weedn AE, Mosley BA, Cleves MA, et al. Prenatal ultrasound diagnosis of birth defects in the National Birth Defects Prevention Study. Unpublished manuscript.

7. Montana E, Khoury MJ, Cragan JD, et al. Trends and outcomes after prenatal diagnosis of congenital cardiac malformations by fetal echocardiography in a well defined birth population, Atlanta, Georgia, 1990-1994. *J Am Coll Caridol*. 1996;28(7):1805-1809.

8. Dashe JS, McIntire DD, Twickler DM. Effect of maternal obesity on the ultrasound detection of anomalous fetuses. *Obstet Gynecol*. 2009;113:1001-1007.

9. Johnson CY, Honein MA, Hobbs CA, et al. Prenatal diagnosis of orofacial clefts, National Birth Defects Prevention Study, 1998-2004. *Prenat Diagn*. 2009;29(9):833-839.

10. Olson JE, Shu XO, Ross JA, et al. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Oncology Group. *Am J Epidemiol*. 1997;145(1):58-67.

11. Yawn BP, Suman VJ, Jacbosen SJ. Maternal recall of distant pregnancy events. *J Clin Epidemiol*. 1998;51(5):399-405.

12. Tomeo CA, Rich-Edwards JW, Michels KB, et al. Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology*. 1999;10:774-777.

13. Rice F, Lewis A, Harold G, et al. Agreement between maternal report and antenatal records for a range of pre and peri-natal factors: the influence of maternal and child characteristics. *Early Hum Dev*. 2007;83:497-504.

14. Yoon PW, Rasmussen SA, Lynberg MC, et al. The National Birth Defects Prevention Study. *Public Health Rep*. 2001;116(Suppl 1):32-40.

15. Rasmussen SA, Olney RS, Holmes LB, et al. Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Defects Res A Clin Mol Teratol*. 2003;67(3):193-201.

16. Adzick NS, Thom EA, Spong CY, et al. A randomized trial of prenatal versus postnatal repair of myelomeningocele. *N Engl J Med*. 2011;364(11):993-1004.

17. Adzick NS. Fetal myelomeningocele: natural history, pathophysiology, and in-utero intervention. *Semin Fetal Neonatal Med*. 2010;15(1):9-14.

18. Cragan JD, Khoury MJ. Effect of prenatal diagnosis on epidemiologic studies of birth defects. *Epidemiology*. 2000;2000(11):695-699.

19. Cheschier N, ACoPB-O. ACOG practice bulletin. Neural tube defects. Number 44, July 2003. (Replaces committee opinion number 252, March 2001). *Int J Gynaecol Obstet*. 2003;83(1):123-133.

20. Hendler I, Blackwell SC, Bujold E, et al. The impact of maternal obesity on midtrimester sonographic visualization of fetal cardiac and craniospinal structures. *Int J Obes*. 2004;28:1607-1611.

CHAPTER 5

Potential Sensitivity of Bias Analysis Results to Incorrect Assumptions of Nondifferential or Differential Exposure Misclassification

Candice Y. Johnson, Penelope P. Howards, Margaret A. Honein, Matthew J. Strickland, and W. Dana Flanders

Author affiliations: Departments of Epidemiology (Candice Y. Johnson, Penelope P. Howards, Matthew J. Strickland, W. Dana Flanders), Environmental Health (Matthew J. Strickland), and Biostatistics and Bioinformatics (W. Dana Flanders), Emory University, Atlanta, Georgia; National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia (Candice Y. Johnson, Margaret A. Honein).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Corresponding author: Candice Y. Johnson, cyjohnson@alum.emory.edu.

ABSTRACT

Results of bias analyses (sensitivity analyses) for exposure misclassification are dependent on assumptions made in the analysis. Few studies have described how adjustment for misclassification is affected by incorrect assumptions about whether sensitivity and specificity were the same (nondifferential) or different (differential) for cases and non-cases. The authors evaluated the effects of making incorrect assumptions about differential and nondifferential misclassification when adjusting for exposure misclassification. First, they used simulated datasets in which nondifferential and differential misclassification were introduced. When incorrect assumptions were made (for example, assuming nondifferential misclassification when it was truly differential), the median misclassification-adjusted odds ratios (OR) from simulation were biased, ranging from 51% to 315% of the true OR, given the authors' assumptions. Then, they used data on obesity and diabetes from the National Health and Nutrition Examination Survey in which both self-reported (misclassified) and measured (true) obesity were available. The true OR between obesity and diabetes was 6.00, but incorrect assumptions produced misclassification-adjusted ORs ranging from 7.49 to 9.48, farther from the truth than the OR unadjusted for misclassification, 5.55. Choice of nondifferential or differential misclassification is an important consideration when adjusting for exposure misclassification because an incorrect assumption can lead to biased results.

INTRODUCTION

Bias analysis (also referred to as sensitivity analysis) has been proposed as an improvement over the qualitative descriptions of study limitations and potential sources of bias typically provided by investigators. The quantitative nature of these analyses allows a more transparent assessment of the potential direction and magnitude of bias and guards against the tendency of investigators to favor causation over bias as the most likely explanation for observed results (1, 2). Some investigators have advocated for greater incorporation of quantitative analyses for exposure misclassification and other forms of bias into epidemiologic studies (3-7) and many examples are now available in the published literature (8-12).

Bias analysis for exposure misclassification involves identifying potential sources of misclassification, estimating bias parameters (for example, sensitivity [Se] and specificity [Sp] of exposure classification) from validation studies or literature review, and using this information to adjust study results, often using simple algebraic manipulations of the contingency table. Probabilistic bias analysis extends this basic approach by allowing the investigator to assign a probability distribution to each bias parameter, sample randomly from the distribution, and perform the bias analysis repeatedly to produce a distribution of the adjusted measure of association. These probabilistic methods allow investigators to acknowledge uncertainty in choice of bias parameters and are more frequently used now that they are available in widely used software such as SAS, Stata, and Excel (1, 8, 13).

Much of the literature on bias analysis for exposure misclassification focuses on choosing values or creating distributions for Se and Sp (1, 2). Less emphasis has been given to the importance of correctly specifying whether misclassification is nondifferential or differential. In many studies, it might not be obvious whether nondifferential misclassification (Se and Sp are the same for cases and non-cases) or differential misclassification (Se and Sp differ between cases and non-cases) is the more appropriate assumption unless internal validation data are available, in which case these parameters can be estimated directly, albeit often with error. Assuming nondifferential misclassification in a bias analysis when misclassification is truly differential might not produce an estimate closer to the truth than the unadjusted estimate (14, 15). However, investigators might be hesitant to assume differential misclassification unless outcome-specific estimates of Se and Sp are available or the investigator has some indication of how they differ between cases and non-cases.

The purpose of this study is to illustrate the potential sensitivity of bias analysis results to incorrect assumptions of nondifferential or differential misclassification. Using simulation, we create datasets with nondifferential and differential exposure misclassification and adjust for misclassification using correct and incorrect assumptions about nondifferential and differential misclassification. We then use data on obesity and diabetes from the National Health and Nutrition Examination Survey (NHANES) to provide an example of how correct and incorrect assumptions might affect the results of a bias analysis in an epidemiologic study. In the simulated datasets and NHANES, both

true and misclassified versions of exposure are known so we can evaluate the success of correct and incorrect assumptions on the adjustment for misclassification.

MATERIALS AND METHODS

Example 1: Simulated Data

We created 12 datasets for the analysis (Figure 5.1). For each, we created a population of 10,000 simulated study participants in which 10% of participants were randomly assigned to be exposed using Bernoulli trials and disease status was randomly assigned using Bernouilli trials with probability of disease (D) following a logistic model conditional on exposure (E) of the form logit p(D) = E.

Three types of exposure misclassification were then introduced, with each type of misclassification applied to 4 datasets. For the first type of misclassification, we used Bernoulli trials to randomly misclassify 10% of the population, independent of disease status (nondifferential misclassification); accuracy of exposure classification between cases and non-cases could differ by chance. For the second type of misclassification, cases were misclassified with 5% probability and non-cases with 10% probability. For convenience, any misclassification scenario in which cases have more accurate classification than non-cases will be referred to as “differential A” in this paper. For the third type of misclassification, cases were misclassified with 10% probability and non-cases with 5% probability. Any misclassification scenario in which cases have less accurate classification than non-cases will be referred to as “differential B” in this paper.

After creating datasets with misclassified exposure, we attempted to adjust for exposure misclassification in these datasets to determine if we could obtain a valid estimate of the true odds ratio (OR) even when using incorrect assumptions about nondifferential or differential misclassification in the analysis. We used a common algebraic method that involves calculating expected cell counts for the correctly classified contingency table given cell counts for the misclassified contingency table and estimates of Se and Sp (2). The formulae in Table 5.1 are applied to back-calculate the true (correctly classified) data from the observed (misclassified) data given estimates of Se and Sp, assuming these estimates are accurate.

For each of the 3 misclassification types discussed previously, we made 4 different assumptions about exposure misclassification in the analysis: exactly nondifferential misclassification, approximately nondifferential misclassification, differential A, and differential B (described in further detail below). Correct assumptions were those in which the assumption made in the analysis matched the true type of misclassification in the population. For example, a correct assumption would be assuming differential A in the analysis when the misclassification truly was differential A; an incorrect assumption would be assuming differential A in the analysis when misclassification was truly nondifferential. Both “exactly nondifferential” and “approximately nondifferential” misclassification were considered to be correct assumptions for the nondifferential misclassification type.

Se and Sp values for each assumption were calculated as follows:

Nondifferential misclassification

Two types of nondifferential misclassification were investigated: exactly nondifferential misclassification and approximately nondifferential misclassification. The distinction between these types are explained further below. For both, Se and Sp were calculated from the total simulated population (cases and non-cases combined). Both cases and non-cases were assigned these values in the adjustment.

Differential A

Se and Sp for non-cases were calculated from the non-cases in the simulated population. The Se and Sp values used for cases depended on the analysis. When making a correct assumption (i.e., for the population with differential A misclassification), Se and Sp for cases were calculated from the cases in the simulated population. When making incorrect assumptions (i.e., for populations with nondifferential or differential B misclassification), the Se for cases was assigned to be the Se for non-cases + 0.05, and the Sp for cases to be the Se for non-cases + 0.02.

Differential B

Se and Sp for non-cases were calculated from the non-cases in the simulated population. When making a correct assumption (i.e., for populations with differential B misclassification), Se and Sp for cases was calculated from the cases in the simulated population. When making incorrect assumptions (i.e., for populations with

nondifferential or differential A misclassification), the Se for cases was assigned to be the Se for non-cases - 0.05, and the Sp for cases to be the Se for non-cases - 0.02.

We implemented a probabilistic analysis by specifying triangular distributions for Se and Sp (1, 2). Se and Sp were calculated from the simulated data as described previously at each of 1,000 iterations of the simulation and the calculated values were used as the mode of the respective distributions. The maximum and minimum values for the triangular distributions were assigned to be +/- 0.05 of the mode for Se and +/- 0.02 for Sp. The distributions were truncated when necessary so all values of Se and Sp fell between 0.5 and 1, inclusive. For each of the 1,000 iterations, one value of Se and one value of Sp were randomly chosen for cases and for non-cases and these values were used to calculate the misclassification-adjusted OR.

The ratio of the misclassification-adjusted OR to the true OR (calculated from the source population without misclassification at each iteration) was calculated and will be referred to as the ratio of odds ratios (ROR). Results are presented as the median ROR and 95% simulation interval (SI). The 95% SI represents the 2.5^{th} and 97.5^{th} percentiles of the ROR distribution generated by simulation.

We make a distinction between two types of nondifferential misclassification in the analyses, which we refer to as “exactly nondifferential misclassification” and “approximately nondifferential misclassification”. When exactly nondifferential misclassification was assumed, one value of Se and one value of Sp were selected from

the distribution and those Se and Sp values were assigned to both cases and to non-cases; as a result, at every iteration of the simulation, cases and non-cases had exactly than same values of Se and Sp. For approximately nondifferential misclassification, values of Se and Sp for cases and for non-cases were independently chosen from the same distributions, so that the values were similar but not necessarily the same between cases and non-cases at each iteration of the simulation. We use the term "approximately nondifferential" to mean that the distributions of Se and Sp for cases and non-cases were the same (nondifferential), even though the values used in the analysis could differ (differential).

Example 2: Data From an Epidemiologic Study

We included non-pregnant women aged 18 to 49 participating in the National Health and Nutrition Examination Survey (NHANES) between 1999 and 2008. NHANES uses a complex, multistage, probability sampling design to select participants from the civilian, non-institutionalized population of the United States (16). NHANES participants complete an in-person interview during which they self-report height and weight. One or two weeks later, they visit a mobile examination center during which their height and weight are measured. Women with missing values for self-reported or measured height or weight were excluded.

Obesity (exposure) was defined as body mass index ≥ 30 kg/m^2, calculated as weight in kilograms divided by squared height in meters. We will refer to obesity status calculated from self-reported height and weight as "self-reported obesity" (misclassified exposure),

and obesity status calculated from measured height and weight as “measured obesity” (true exposure).

Self-reported diagnosis of diabetes (outcome) was obtained by questionnaire. No distinction was made between type 1 and type 2 diabetes. Women who reported “borderline” diabetes were categorized as having no diabetes diagnosis and women with missing data on diabetes status were excluded.

We conducted a literature review (Appendix A) to identify estimates of Se and Sp for obesity misclassification from studies comparing self-reported to measured obesity among adult females in the United States. We excluded studies conducted exclusively in children, teenagers, or the elderly as well as estimates from published NHANES data because our purpose was to approximate an adjustment for misclassification when internal validation data were unavailable.

Based on the results of the literature review, we created triangular distributions for Se and Sp (further described in Results). We conducted the bias analysis under the same 4 assumptions as in the example with simulated data: exactly nondifferential misclassification, approximately nondifferential misclassification, differential misclassification A, and differential misclassification B. Probabilistic adjustment for misclassification was conducted over 1,000 iterations to generate a distribution of the misclassification-adjusted OR. Results are presented as the median OR and 95% SI. For simplicity, in this example we did not take into account in the analysis the complex

sampling design of NHANES and as such, the results should not be interpreted as being representative of the United States population.

RESULTS

Example 1: Simulated Data

In the simulated population, when correct assumptions about nondifferential or differential misclassification were made, results were on average unbiased; however, when incorrect assumptions were made, the results were on average biased (Table 5.2).

Truth: nondifferential misclassification

When misclassification was truly nondifferential, assuming either exactly nondifferential misclassification or approximately nondifferential misclassification produced results that were on average unbiased. The 95% SI was wider when assuming approximately nondifferential misclassification than exactly nondifferential misclassification due to the added variability in Se and Sp values between cases and non-cases. Assuming either differential A or differential B produced biased results, but the bias was in different directions (median ROR, differential A: 1.12; median ROR, differential B: 0.89). All 95% SIs, however, included the null (unbiased) value.

Truth: differential A

When differential A was the true pattern of misclassification, all incorrect assumptions produced RORs that were too small (median ROR range: 0.51 to 0.55) and none of the

95% SIs included the null value. Correctly assuming differential A produced unbiased results on average.

Truth: differential B

When differential B was the truth, all incorrect assumptions produced RORs that were too large (median ROR range: 2.25 to 3.15) and none of the 95% SIs included the null value. Correctly assuming differential B produced unbiased results on average.

Example 2: Data From an Epidemiologic Study

We identified Se and Sp for obesity classification from five published studies meeting inclusion criteria (Appendix B, Table 5.B1). Because most estimates of Se were near 0.90, we chose this as the mode of the triangular distribution and assigned a minimum and maximum of 0.85 and 0.95 to allow for uncertainty. For Sp, we chose a mode of 0.97 based on the average of all estimates and the minimum (0.94) and maximum (1.00) values of the distribution based on the highest and lowest estimates obtained through literature review. No study provided diabetes-specific estimates of Se or Sp.

In our NHANES population of 6,243 women, the true OR between measured obesity and diabetes was 6.00 and the misclassified OR between self-reported obesity and diabetes was 5.55. Misclassification-adjusted median ORs ranged from 6.25 to 9.48 (Table 5.3); all adjustments over-estimated the magnitude of the association.

The misclassification assumption producing estimates nearest to the truth was differential misclassification A, in which cases had higher Se and Sp than non-cases (OR = 6.25, 95% SI: 5.21, 7.55). In the dataset, the true misclassification pattern most closely resembled differential misclassification A; however, Se was 0.10 higher in cases compared to non-cases, greater than the 0.05 we assumed in the analysis, and Sp was 0.01 lower, not 0.02 higher as we had assumed (Table 5.4).

We repeated the analyses using Se and Sp estimates abstracted from each individual study (Table 5.5). Under the correct assumption (differential misclassification A), misclassification-adjusted ORs ranged from 5.96 to 8.02. All but one included the true OR of 6.00 in the 95% SI. However, when making incorrect assumptions, misclassification adjustment overestimated the magnitude of the association, with median misclassification-adjusted ORs ranging from 7.20 to 19.17. Only 1 95% SI included the true value of 6.00.

DISCUSSION

We presented examples in which adjustment for exposure misclassification was undertaken using several different assumptions about differential and nondifferential misclassification. Using simulations and data from an epidemiologic study, we highlight that making incorrect assumptions about exposure misclassification can produce “adjusted” results that are biased, and in some cases more biased than the estimates that were not adjusted for misclassification.

In the simulations, we observed that when correct assumptions about nondifferential or differential misclassification were made, the results were on average unbiased. However, in these simulations we knew the true values of Se and Sp in the population. If incorrect estimates of Se and Sp were used to create the triangular distribution, results could have been biased even when making the correct assumption. In our NHANES example, our estimates of Se and Sp from literature review were not exactly the same as the true values. As a result, none of the adjustments produced an unbiased estimate on average. However, results closest to the truth were obtained when the correct assumption about differential misclassification was made.

Although we have not defined it explicitly, in this study we have used the term “bias” to mean any difference between a given estimate and the “truth”. This definition does not distinguish random error from systematic error when comparing point estimates in a single study, and it might have been more appropriate to use the more general term “error” instead of “bias” to describe differences between the observed estimate and the truth. Using simulation, however, we have shown that inaccurate assumptions about nondifferential and differential misclassification can systematically produce “adjusted” estimates that do not converge to the true value.

Investigators are encouraged to be cautious when presenting and interpreting results from bias analyses because results are only valid if the assumptions used in the analysis at least approximate the truth (1). In discussions of bias analysis in the literature, more emphasis has been given to choosing distributions of Se and Sp for bias analysis than choosing the

correct assumption regarding nondifferential or differential misclassification. In our examples using simulated data and NHANES data, making an incorrect assumption about nondifferential or differential misclassification had important effects on results. Taking uncertainty into account in the analysis by assigning probability distributions to Se and Sp was not sufficient to make up for an incorrect misclassification assumption; in the simulations and the NHANES example, when incorrect assumptions were made, the 95% SI often did not include the true value.

In the absence of validation data (and even when validation data are available, because results of validation studies are themselves subject to error) the rationale for choosing nondifferential versus differential misclassification in the analysis is often left to the investigator's perception of how misclassification occurred in the study. This commonly consists of a qualitative description of the possible sources of bias without presentation of evidence supporting the decision (17). This is similar to the qualitative discussion of the direction and magnitude of bias that quantitative bias analysis is meant to guard against. Unfortunately, this situation is difficult to avoid because there is rarely sufficient information available to determine whether nondifferential or differential misclassification is most likely for a given study design and method of exposure measurement. Even if a certain misclassification process is strongly suspected (for example, assuming nondifferential misclassification in a prospective cohort study in which exposure is measured before disease occurs), there is no guarantee that this type of misclassification actually occurred in the study (18). By chance, Se and Sp could have differed between cases and non-cases, producing differential misclassification instead of

nondifferential misclassification, or vice versa (19). Factors aside from chance are also important. For example, when exposure categories are combined, differential misclassification can be produced even if the measurement error or misclassification process on the original variable was nondifferential (14, 18, 20).

An important role for bias analysis in epidemiologic studies is producing ranges of plausible estimates rather than providing a single bias-adjusted effect estimate as the final result. Without knowing whether misclassification was truly differential or nondifferential in our NHANES example, we would have no evidence for choosing the results of one assumption over the others as the most likely. However, we might conclude with some confidence that exposure misclassification does not account for the observed association, with no 95% SI covering the null value (OR = 1), given our assumptions.

In this study, we have presented examples demonstrating that making inaccurate assumptions about nondifferential or differential misclassification has the potential to produce biased results when adjusting for exposure misclassification. Investigators should recognize the likelihood of making one or more incorrect assumptions during bias adjustment and consider reporting results based on more than one assumption about misclassification in their analysis. Although this strategy might not provide a single point estimate as the result, it remains a useful method for providing plausible ranges of the effect estimate in the absence of exposure misclassification.

Table 5.1. Contingency Tables for True (Correctly Classified) and Observed (Misclassified) Exposures.

	True Exposure		Misclassified Exposure	
	Exposure	No Exposure	Exposure	No Exposure
Disease	A	B	$a = Se_1*A + (1-Sp_1)*B$	$b = (1-Se_1)*A + Sp_1*B$
No disease	C	D	$c = Se_0*C + (1-Sp_0)*D$	$d = (1-Se_0)*C + Sp_0*D$

Abbreviations: Se_i, sensitivity in cases (i = 1) and non-cases (i = 0); Sp_i, specificity in cases (i = 1) and non-cases (i = 0).

Table 5.2. Ratio of Misclassification-Adjusted Odds Ratio to True Odds Ratio Over 1,000 Iterations of Simulation When Making Correct and Incorrect Assumptions About Nondifferential and Differential Misclassification.

		Ratio of Odds Ratios	
Truth[a]	Adjustment Assumption[b]	Median	95% Simulation Interval[c]
Nondifferential	Exactly nondifferential[d]	1.00	0.79, 1.43
	Approximately nondifferential[d]	1.00	0.67, 1.69
	Differential A	1.12	0.78, 1.80
	Differential B	0.89	0.59, 1.42
Differential A	Exactly nondifferential	0.55	0.47, 0.66
	Approximately nondifferential	0.55	0.40, 0.76
	Differential A[d]	1.01	0.70, 1.54
	Differential B	0.51	0.30, 0.87
Differential B	Exactly nondifferential	2.28	1.51, 5.62
	Approximately nondifferential	2.25	1.32, 6.33
	Differential A	3.15	1.14, 5.35
	Differential B[d]	1.01	0.71, 1.46

[a] Nondifferential misclassification: cases and non-cases had 10% probability of misclassification. Differential A: cases had 5% probability of misclassification, non-cases had 10% probability of misclassification. Differential B: cases had 10% probability of misclassification, non-cases had 5% probability of misclassification.

[b] Exactly nondifferential misclassification: cases and non-cases share identical values for sensitivity (Se) and specificity (Sp); the modes of the distributions are the actual values of

Se and Sp calculated from the data. Approximately nondifferential misclassification: cases and non-cases have the same distributions of Se and Sp, but not necessarily the same values of Se and Sp; the modes of the distributions are the actual values of Se and Sp calculated from the data. Differential A: for incorrect assumptions, the modes of the Se and Sp distributions for cases are 0.05 and 0.02 higher than for non-cases and the modes of the non-case distributions are the actual values of Se and Sp calculated from the data; for correct assumptions, the modes for Se and Sp for cases take on the true values from the simulated population. Differential B: for incorrect assumptions, the modes of the Se and Sp distributions for cases are 0.05 and 0.02 lower than for non-cases and the modes of the non-case distributions are the actual values of Se and Sp calculated from the data; for correct assumptions, the modes for Se and Sp for cases take on the true values from the simulated population.

[c] Lower and upper bounds are the 2.5^{th} and 97.5^{th} percentiles of the ratio of odds ratio distribution generated by simulation.

[d] In the text, these are referred to as the “correct” assumptions. All others are referred to as “incorrect” assumptions.

Table 5.3. Results From 1,000 Iterations of Probabilistic Adjustment for Exposure Misclassification Under Various Assumptions in a Study of Obesity and Diabetes, National Health and Nutrition Examination Survey, 1999-2008.

	Odds Ratio Estimate[a]	95% Simulation Interval[b]
Truth - measured obesity[c]	6.00	
Misclassified - self-reported obesity[d]	5.55	
Adjusted for misclassification assuming:		
Exactly nondifferential[e,f]	7.49	6.63, 8.66
Approximately nondifferential[e,g]	7.50	6.07, 9.38
Differential A[e,h]	6.25	5.21, 7.55
Differential B[e,i]	9.48	7.35, 12.42

[a] Estimates for misclassification-adjusted odds ratios are the median of the odds ratio distribution generated by simulation.

[b] Lower and upper bounds are the 2.5th and 97.5th percentiles of the odds ratio distribution generated by simulation.

[c] Obesity (body mass index $\geq$ 30 kg/m^2) calculated from measured height and weight.

[d] Obesity calculated from self-reported height and weight.

[e] Adjustment for exposure misclassification using Se and Sp estimates obtained a summary of the literature review estimates. Triangular distributions for non-cases (minimum, mode, maximum) were Se (0.85, 0.90, 0.95) and Sp (0.94, 0.97, 1.00).

[f] One value of Se and one value of Sp chosen from the non-case distributions at each iteration of the simulation. Cases and non-cases were assigned the same values for the analysis.

[g] Cases and non-cases assigned the non-case distributions. Values of Se and Sp were chosen independently from these distributions for cases and non-cases.

[h] Case distributions were Se (0.90, 0.95, 1.00) and Sp (0.96, 0.99, 1.00).

[i] Case distributions were Se (0.80, 0.85, 0.90) and Sp (0.92, 0.95, 0.98).

Table 5.4. Contingency Tables and True Values of Sensitivity and Specificity of Obesity Classification for a Study of Self-Reported and Measured Obesity and Diabetes, Non-Pregnant Females Aged 18-49, National Health and Nutrition Examination Survey, 1999-2008.

	Self-reported Obesity		Measured Obesity		Sensitivity[a]		Specificity[b]	
	Obese	Not Obese	Obese	Not Obese				
All participants	1,845	4,398	2,092	4,151	1,756/2,092	0.84	4,062/4,151	0.98
Diabetes								
Diabetes	152	70	164	58	151/164	0.93	57/58	0.98
No diabetes	1,693	4,328	1,928	4,093	1,605/1,928	0.83	4,005/4,093	0.99

[a] Proportion of individuals truly exposed who reported exposure.

[b] Proportion of individuals truly unexposed who reported not being exposed.

Table 5.5. Results of Adjustment for Exposure Misclassification in a Study of Obesity and Diabetes Using Estimates of Sensitivity and Specificity From Literature Review, Non-Pregnant Females Aged 18-49, National Health and Nutrition Examination Survey, 1999-2008.

Adjustment Assumption and Source of Se and Sp	Median Odds Ratio[a]	95% Simulation Interval[b]
Truth - measured obesity[c]	6.00	
Misclassified - self-reported obesity[d]	5.55	
Exactly nondifferential[e]		
Krul et al., 2010	7.43	6.63, 8.57
Johnson et al., 2009	7.45	6.69, 8.48
Brunner Huber, 2007	7.20	6.46, 8.23
Hussain et al., 2007	8.10	7.23, 9.31
Nieto-Garcia et al., 1990	11.25	9.05, 15.70
Approximately nondifferential[f]		
Krul et al., 2010	7.43	6.14, 9.15
Johnson et al., 2009	7.46	6.18, 9.09
Brunner Huber, 2007	7.20	5.99, 8.78
Hussain et al., 2007	8.12	6.67, 9.99
Nieto-Garcia et al., 1990	11.26	8.33, 16.86
Differential A[g]		
Krul et al., 2010	6.09	5.16, 7.18

Johnson et al., 2009	6.29	5.34, 7.46
Brunner Huber, 2007	5.96	5.07, 6.99
Hussain et al., 2007	6.79	5.71, 8.12
Nieto-Garcia et al., 1990	8.02	6.33, 10.43
Differential B[h]		
Krul et al., 2010	9.54	7.51, 12.68
Johnson et al., 2009	9.29	7.43, 11.89
Brunner Huber, 2007	9.10	7.25, 11.88
Hussain et al., 2007	10.23	8.08, 13.34
Nieto-Garcia et al., 1990	19.17	11.99, 46.62

[a] Estimates for misclassification-adjusted odds ratios are the median of the odds ratio distribution generated by simulation.

[b] Lower and upper bounds are the 2.5th and 97.5th percentiles of the odds ratio distribution generated by simulation.

[c] Obesity (body mass index $\geq$ 30 kg/m^2) calculated from measured height and weight.

[d] Obesity calculated from self-reported height and weight.

[e] One value of Se and one value of Sp chosen from the non-case distributions (Table 5.B2) at each iteration of the simulation. Cases and non-cases were assigned these values for the analysis.

[f] Cases and non-cases assigned the non-case distributions (Table 5.B2). Values of Se and Sp were chosen independently from these distributions for cases and non-cases.

[g] Case distributions assigned as the non-case distribution (Table 5.B2) shifted upward by 0.05 for Se, upward by 0.02 for Sp.

[h] Case distributions assigned as the non-case distribution (Table 5.B2) shifted downward by 0.05 for Se, downward by 0.02 for Sp.

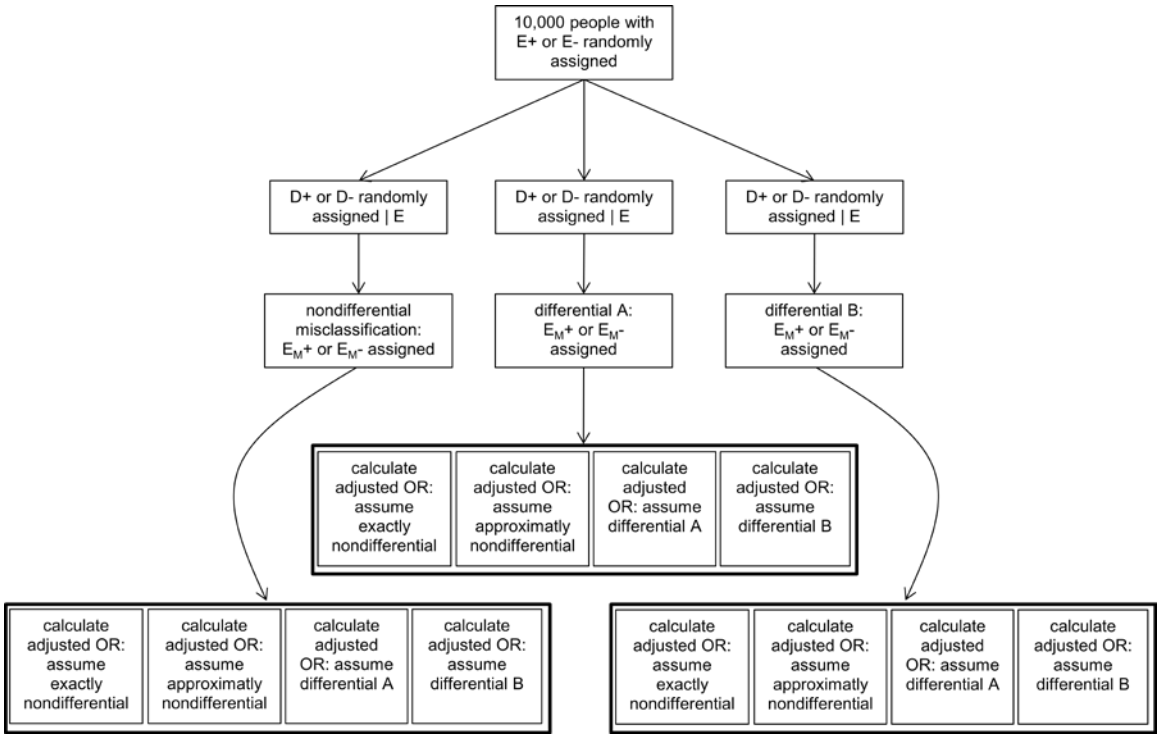

Figure 5.1. Creation of the 12 datasets for simulation. At each iteration of the simulation, exposure, disease, and misclassified exposure status are assigned and misclassification-adjusted odds ratios are calculated under 4 different assumptions. The process is repeated 1,000 times. Abbreviations: D, disease; E, true exposure; E_M, misclassified exposure; OR, odds ratio.

APPENDIX A

Literature Review Methods

We searched Embase and Medline for studies published between 1985 and April 2011 (approximately last 25 years) reporting sensitivity and specificity (or data sufficient so these could be calculated) of obesity classification in American adult females. Search terms included self-report, height, weight, BMI (body mass index), sensitivity, and specificity (the exact search strategy is available upon request). We identified 62 articles in the search and excluded articles for the following reasons: irrelevant topic (n = 18), commentary without original data (n = 1), included only children or the elderly (n = 12), included only males (n = 1), or not conducted in the United States (n = 23). Of the 7 remaining studies, we excluded 4 because they provided estimates from NHANES; this exclusion was made because we wanted to approximate scenarios in which internal validation data was unavailable.

To the 3 studies identified through the search strategy (21-23) we added 2 studies that we were aware of, but that were not identified by the search strategy (24, 25).

APPENDIX B

Studies Identified in Review of the Literature

Table 5.B1. Identified Published Studies Reporting Sensitivity and Specificity of Self-Reported Obesity Status in American Adult Females of Reproductive Age.

	Study	Population	Years	N	Se	Sp	Diabetes
1	Krul et al., 2010 (24)	North American females aged 18-65 participating in CAESAR project	1999-2000	1,248	0.885	0.996	No information
2	Johnson et al., 2009 (23)	Females aged 18-65, participants in the Pennington Center Longitudinal Study	1992-2008	9,797	EA: 0.912 AA: 0.909	EA: 0.964 AA: 0.947	No information

3	Brunner Huber 2007 (25)	Females 18-45 using birth control, attending suburban Family Medicine clinic in Atlanta	2004	250	0.900	0.994	No information
4	Hussain et al., 2007 (22)	English-speaking, nonpregnant females aged 18-44, predominantly African-American, attending urban community health center for medical appointments	2003	231	0.896	0.944	7.4% diagnosed, population reported to be at high risk
5	Nieto-Garcia et al., 1990 (21)	Females aged 20-79, participants in Collaborative Lipid Research Clinics Family Study	1975-1978	572	0.80	0.99[a]	No information

Abbreviations: AA, African-American; EA, European-American; N, number; Se, sensitivity; Sp, specificity.

[a] Includes males, but reported to be similar for all participant subgroups

Estimates of Sensitivity and Specificity from Literature Review

Table 5.B2. Triangular Distributions for Sensitivity and Specificity of Obesity Classification for Non-Cases, Created From Literature Review.

	Sensitivity			Specificity		
Source of Estimates	Min	Mode	Max	Min	Mode	Max
Krul et al., 2010	0.84	0.89	0.94	0.97	0.99	1.00
Johnson et al., 2009	0.86	0.91	0.96	0.94	0.96	0.98
Brunner Huber, 2007	0.85	0.90	0.95	0.97	0.99	1.00
Hussain et al., 2007	0.85	0.90	0.95	0.92	0.94	0.96
Nieto-Garcia et al., 1990	0.75	0.80	0.85	0.97	0.99	1.00

Abbreviations: max; maximum; min, minimum.

REFERENCES

1. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer; 2009.

2. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology, 3rd edition*. Philadelphia: Lippincott Williams & Wilkins, 2008.

3. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology*. 2003;14:459-466.

4. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol*. 2004;160:384-392.

5. Greenland S. Multiple-bias modelling for analysis of observational data (with discussion). *J R Stat Soc A*. 2005;168:267-306.

6. Maldonado G. Adjusting a relative-risk estimate for study imperfections. *J Epidemiol Community Health*. 2008;62:655-663.

7. Fox MP. Creating a demand for bias analysis in epidemiologic research. *J Epidemiol Community Health*. 2009;63:91.

8. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34:1370-1376.

9. Jurek AM, Maldonado G, Spector LG, et al. Periconceptional maternal vitamin supplementation and childhood leukaemia: an uncertainty analysis. *J Epidemiol Community Health*. 2009;63:168-172.

10. MacLehose RF, Olshan AF, Herring AH, et al. Bayesian methods for correcting misclassification: an example from birth defects epidemiology. *Epidemiology*. 2009;20:27-35.

11. 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;18:2184-2190.

12. Stott-Miller M, Heike CL, Kratz M, et al. Increased risk of orofacial clefts associated with maternal obesity: case-control study and Monte Carlo-based bias analysis. *Paediatr Perinat Epidemiol*. 2010;24:502-512.

13. Orsini N, Bellocco R, Bottai M, et al. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. *The Stata Journal*. 2008;8:29-48.

14. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*. 1991;134:1233-1244.

15. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton, FL: Chapman & Hall/CRC; 2004.

16. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Serves, Centers for Disease Control and Prevention, 1999-2008.

17. Jurek AM, Maldonado G, Greenland S, et al. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol*. 2006;21:871-876.

18. Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent differential misclassification. *Am J Epidemiol*. 1991;134:433-437.

19. Jurek AM, Greenland S, Maldonado G, et al. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol*. 2005;34:680-687.

20. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology, Third Edition*. Philadephia: Lippincott Williams & Wilkins, 2008.

21. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: Sensitivity and specificity using self-reported weight and height. *Epidemiology*. 1990;1(2):146-152.

22. Hussain S, Taylor M, Waltermaurer E, et al. Computer-administered screening of reproductive-aged women for diabetes risk in primary care settings, feasibility and acceptability of such screening, and validity of risk assessments based on self-reported weight. *Prev Chronic Dis*. 2007;4(3):A54.

23. Johnson WD, Bouchard C, Newton Jr RL, et al. Ethnic differences in self-reported and measured obesity. *Obesity*. 2009;17(3):571-577.

24. Krul AJ, Daanen HA, Choi H. Self-reported and measured weight, height and body mass index (BMI) in Italy, the Netherlands and North America. *Eur J Public Health*. 2011;21(4):414-419.

25. Brunner Huber LR. Validity of self-reported height and weight in women of reproductive age. *Matern Child Health*. 2007;11:137-144.

CHAPTER 6

Weighted Logistic Regression for Multiple Bias Analysis

Candice Y. Johnson, Penelope P. Howards, Matthew J. Strickland, D. Kim Waller, W. Dana Flanders, and The National Birth Defects Prevention Study

Author affiliations: Departments of Epidemiology (Candice Y. Johnson, Penelope P. Howards, Matthew J. Strickland, W. Dana Flanders), Environmental Health (Matthew J. Strickland), and Biostatistics and Bioinformatics (W. Dana Flanders), Emory University, Atlanta, Georgia; National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia (Candice Y. Johnson); University of Texas at Houston, School of Public Health, Houston, Texas (D. Kim Waller).

Corresponding author: Candice Y. Johnson, cyjohnson@alum.emory.edu.

ABSTRACT

Exposure misclassification, selection bias, and confounding are important sources of bias in epidemiologic studies, yet only confounding is routinely addressed quantitatively. The authors describe a method to simultaneously adjust for these biases using weighted logistic regression. Selection probabilities and predictive values for exposure classification are used as weights to re-balance the joint distribution of exposure and disease to what the distribution would have been without bias. The method was applied to a case-control study of prepregnancy obesity (obese: body mass index ≥ 30 kg/m^2 versus normal weight: 18.5-24.9 kg/m^2) and isolated cleft lip with or without cleft palate (CL/P) and cleft palate (CP) using data from the National Birth Defects Prevention Study. Adjusting for confounding only, associations were observed between prepregnancy obesity and both CL/P (odds ratio [OR] 1.20, 95% confidence interval [CI] 1.05, 1.38) and CP (OR 1.27, 95% CI 1.05, 1.52). After adjusting for exposure misclassification, selection bias, and confounding, given the authors' assumptions, associations were attenuated (CL/P median OR range from simulation: 0.99 to 1.04; CP median OR range: 1.04 to 1.09). Considering the potential effects of biases other than confounding is important in epidemiologic studies. This approach allows simultaneous adjustment for multiple biases using logistic regression.

INTRODUCTION

Results of an epidemiologic study might not accurately reflect the direction or magnitude of the true association if during the study's design, execution, or analysis they have been distorted by bias (1). Confounding is often addressed in the analysis using statistical techniques such as multiple variable regression, but even though biases such as misclassification and selection bias might have equally important effects on results, they are not usually addressed quantitatively (2, 3). There are published examples of studies adjusting for more than one type of bias (3-8) and methods for multiple bias analysis have been implemented using widely-available software such as Excel, Stata, and SAS (4, 6, 9). Some methods adjust for biases using sequential contingency tables, although adjusting for confounding is difficult with this approach (6, 9). Other methods adjust for multiple biases using regression models and simulation (4, 10) or by multiplying effect estimates by adjustment factors (8, 11).

In this paper, we describe a method for multiple bias analysis allowing simultaneous adjustment for exposure misclassification, selection bias, and confounding in a manner parallel to multiple variable logistic regression. Weights are created to rebalance the joint distribution of exposure and disease to the distribution that would have existed in the absence of bias, and the analysis is undertaken using a weighted regression model. For simple (non-probabilistic) multiple bias analysis, no simulation is involved and investigators can apply the method in standard statistical software packages without need for advanced programming. The approach is based on existing methods that use weighted logistic regression to adjust for exposure misclassification (12) and selection bias (13).

We will describe the method and demonstrate its application using both hypothetical and applied examples.

EXAMPLE 1: HYPOTHETICAL CASE-CONTROL STUDY

Study Population

We created a source population of 100,000 individuals with dichotomous exposure (E), disease (D), and confounder (C), assuming that E causes D, and C causes both E and D (Figure 6.1, Table 6.1). We randomly assigned C with a prevalence of 0.10 and assigned values of E conditional on C with $p(E = 1|C = 1) = 0.10$ and $p(E = 1|C = 0) = 0.05$. We assigned D using Bernouilli trials in which p(D) followed a logistic model with E and C as predictors of D, in the form $\text{logit } p(D) = -3 + 1(E) + 1(C)$. We will refer to this population of 100,000 individuals as the "source population".

We simulated a case-control study with the goal of selecting from the source population 100% of affected individuals as cases and a 10% random sample of unaffected individuals as controls. However, we introduced selection bias making selection (S) dependent on both E and D (Figure 6.1) so that selection of participants was not truly random for cases and non-cases, and was incomplete for cases incomplete. Selection probabilities (π_{ij}, i = disease, j = exposure) were $\pi_{11} = 0.900$, $\pi_{10} = 0.950$, $\pi_{01} = 0.080$ and $\pi_{00} = 0.099$. We will refer to this population as the "selected population".

We misclassified exposure differentially with respect to D and assigned each individual in the selected population a potentially misclassified exposure value, E_M. Probability of

misclassification for each E-D combination (p_{ij}) was assigned to be $p_{11} = 0.02$, $p_{10} = 0.05$, $p_{01} = 0.04$, and $p_{00} = 0.10$. These data will be referred to as the "observed data" and represent the data investigators typically have available for analysis in epidemiologic studies.

For simplicity, we assume in this example that C and D were measured without error and no participant had missing data for any variable.

Estimating Exposure Misclassification Weights

Predictive values for exposure classification can be used as the weights for a weighted logistic regression to adjust for exposure misclassification (12). Predictive values are obtained from cross-tabulations of misclassified and true exposure in the selected population: $p(E|D,E_M,S)$. If control for confounding is desired, predictive values should be estimated conditional on C as well. Because in this example we have simulated data, we can calculate predictive value proportions directly; in practice, predictive value probabilities can be estimated from validation studies. As an example of calculation of these proportions in the observed data, the proportion of participants with C = 1, D = 1, and $E_M = 1$ who are truly exposed, $p(E = 1|C = 1, D = 1, E_M = 1, S = 1)$, is 260/308 and the proportion of participants who are truly unexposed, $p(E = 0|C = 1, D = 1, E_M = 1, S = 1)$, is 48/308 (Table 6.2). These exposure misclassification weights sum to 1 within strata of E_M. In Appendix A, we show how to create weights when predictive values are unavailable but estimates of sensitivity and specificity of exposure classification are known.

Estimating Selection Weights

Selection weights are estimated as the inverse of the probability of selection into the study (13). Because selection must be affected by both E and D for selection bias to occur, the selection probabilities are conditional on these variables: p(S|E,D). In some cases, to be discussed later (see "Notes on Weight Estimation"), selection probabilities might instead be conditional on misclassified exposure status: $p(S|E_M, D)$. If one is interested in adjusting for confounding in addition to selection bias, selection probabilities should also be conditional on C. In our simulated data, we can calculate the selection proportions directly; in practice, the true selection probabilities are unknown but can be estimated from participation rates or similar information. As an example of calculation of these proportions, there are 263 individuals with C = 1, D = 1, and E = 1 in the selected population, of 296 originally eligible for inclusion from the source population (Table 6.1). The selection proportion is 263/296; the inverse selection proportion is 296/263 and this is the value used as the selection weight in the analysis for individuals with C = 1, D = 1, and E = 1.

Calculation of the Point Estimate

We start by creating a dataset in which each participant appears twice. One copy of the participant is assigned to be exposed ($E_A = 1$) and the other to be unexposed ($E_A = 0$), because we do not know which is the true exposure status. We use notation E_A to distinguish the assigned exposure status from the true exposure status E (unknown in practice). Each copy of the participant is then assigned an exposure misclassification

weight for E_A. For example, the weight for a participant with C = 1, D = 1, and E_M = 1 would be $p(E_A = 1|C = 1, D = 1, E_M = 1) = 260/308$ for the copy with E_A = 1 and $p(E_A = 0|C = 1, D = 1, E_M = 1) = 48/308$ for the copy with E_A = 0, based on the predictive value proportions calculated previously (Table 6.3). The selection weights are then applied; for example, each participant copy with C = 1, D = 1, E_A = 1 would be given a weight of 296/263, as previously described (Table 6.4). The exposure misclassification weight and selection weight are multiplied to create the final weight used in the analysis. An algebraic demonstration of the method is provided in Appendix A and shows that the contingency table counts from the source population can be reproduced exactly when the values of the weights are known with certainty.

The method can be implemented using standard statistical software. In the regression model, assigned exposure (E_A) is used as the exposure variable instead of the exposure reported by the study participant (E_M). The final weight (product of exposure and selection weights) is used as the weight. Confounding is adjusted for by entering suspected confounders into the multivariable model. In SAS (SAS Institute, Cary, NC), the multiple bias analysis can be performed using any procedure for logistic regression. For example, in PROC LOGISTIC the code for multiple bias analysis is as follows:

```
proc logistic data = datasetname;
        model disease (event = “1”) = assigned_exposure confounder;
        weight finalweight;
run;
```

The confidence interval produced from logistic regression should *not* be used because it does not properly take into account additional variability introduced during the adjustment process (6). Incorporating uncertainty using probabilistic analyses is discussed in the next Example.

Results

In the source population, the confounder-adjusted odds ratio (OR) for the association between D and E is 2.75 and in the selected population it is 3.09 (Table 6.1). The confounder-adjusted OR is 1.18 using the observed (misclassified) data. The change in the estimated OR demonstrates the cumulative effects of these biases on study results.

After adjustment for exposure misclassification, selection bias, and confounding using the weighted analysis, the OR between exposure and disease is 2.75, which matches the source population.

EXAMPLE 2: PREPREGNANCY OBESITY AND ISOLATED OROFACIAL CLEFTS

In this example, we provide an application of this method to an epidemiologic study to simultaneously adjust for exposure misclassification, selection bias, and confounding using weighted logistic regression. These data provide a more realistic example and involve a multi-level exposure, missing data on exposure, and literature-based estimation of weights.

Prepregnancy obesity is associated with numerous adverse pregnancy outcomes including birth defects such as cleft lip with or without cleft palate (CL/P) and cleft palate (CP) (14, 15). Studies have reported weak associations between prepregnancy obesity and isolated CL/P and CP with confounding-adjusted ORs ranging from 1.01 to 1.29 for CL/P and 1.08 to 1.21 for CP for obese compared to normal weight mothers (16-18). One of these studies adjusted for nondifferential exposure misclassification and reported CL/P ORs ranging from 1.38 to 2.94, given various assumptions about sensitivity and specificity of exposure classification (18). No study to date has adjusted for exposure misclassification, selection bias, and confounding in the same analysis.

Study Population

We used data from the National Birth Defects Prevention Study (NBDPS), a multi-site population-based case-control study investigating genetic and environmental risk factors for major structural birth defects (19-21). Cases were identified from birth defect surveillance programs and controls were identified from birth certificates or hospital birth records in the same catchment areas as cases. Participating mothers were interviewed by telephone after delivery. For this analysis we included 1,990 mothers of infants with isolated CL/P, 943 mothers of infants with isolated CP, and 8,177 mothers of control infants who completed the interview. We excluded 36 mothers with missing data on race/ethnicity, a potential confounder. All mothers had infants born on or after October 1, 1997 with an estimated date of delivery on or before December 31, 2007. All participating sites received institutional review board approval.

The exposure of interest was prepregnancy obesity, defined following standard cutpoints of body mass index (BMI): underweight (<18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2) and obese (≥30 kg/m^2). BMI was calculated as self-reported prepregnancy weight in kilograms divided by squared self-reported height in meters. BMI was missing for 101 CL/P case mothers (5.1%), 28 CP case mothers (3.0%), and 337 control mothers (4.1%).

Potential confounders included maternal race/ethnicity, age, and education, and smoking status during pregnancy. In preliminary analyses, adjusting for these variables did not appreciably change the effect estimate from the crude result. Therefore, for simplicity, we adjusted only for potential confounding by maternal race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other) in this example.

Estimation of Exposure Misclassification Weights

Lacking internal validation data for BMI classification, we chose the National Health and Nutrition Examination Survey (NHANES) as the source of external validation data (22). NHANES participants completed an in-person interview during which they self-reported weight and height. One or two weeks later, they underwent a physical examination during which their height and weight were measured. We restricted our validation sample to non-pregnant females aged 16-49 who participated in NHANES between 1999 and 2008 and who had both measured height and weight recorded. Predictive values for 4-level BMI classification were estimated conditional on race/ethnicity and accounting for the

complex sampling design using a cross-tabulation of self-reported BMI category and measured BMI category. An example of a classification table for Hispanic women is shown in Table 6.5.

In NHANES, measured BMI was available for most women who had missing values for self-reported BMI. By treating missing data as a misclassification problem, we used the same approach to adjust for missing data as to adjust for misclassification. We calculated predictive values for exposure classification among women with missing BMI and used these as weights in the analysis to adjust for missing data.

We had no information about whether exposure misclassification was differential or nondifferential by case/control status and so conducted three analyses. In the first, we assumed misclassification was nondifferential and assigned both cases and controls to have the same NHANES predictive values, conditional on maternal race/ethnicity. In the second, we assumed that classification was better for cases than controls. We assigned controls to have the NHANES predictive values, and assigned the matching predictive values (i.e., the predictive values corresponding to women in which the self-reported and measured BMI categories were the same) to be 0.02 higher for cases than controls. In the final analysis, we assumed classification was better for controls than cases, and assigned the matching predictive values to be 0.02 higher for controls than cases.

Because we are using an exposure variable with 4 levels (underweight, normal weight, overweight, obese), each participant is entered into the analysis 4 times, once for each potential true exposure status.

Estimation of Selection Weights

A study from the Danish National Birth Cohort (DNBC) found that mothers self-reporting as underweight, overweight, or obese were less likely to participate in the study than mothers self-reporting as normal weight (23). The ratio of relative frequencies (frequency of the characteristic in the study divided by frequency in the source population) was 0.84 for underweight, 1.04 for normal weight, 0.96 for overweight, and 0.89 for obese mothers.

Participation rates from NBDPS show that case mothers were 6% more likely to agree to participate in the study than control mothers. No information was available, however, for the joint selection probabilities for BMI categories and case/control status.

Because we did not have information on true selection probabilities, we used selection ratios as substitutes. Selection ratios were defined as the ratio of the frequency of a characteristic (e.g., obesity) among study participants to the expected frequency of this characteristic in the source population. We used the ratio of relative frequencies from DNBC as estimates of selection ratios for BMI categories and constructed selection ratios for case/control status based on NBDPS participation rates. In the analysis, the weights were equal to the inverse of the selection ratios.

Without further information on selection ratios, we assumed that participation of control mothers would differ between BMI categories, but that participation of case mothers would not. Selection ratios for controls were assigned to vary by BMI (underweight = 0.84, normal weight = 1.04, overweight = 0.96, obese = 0.89) based on estimates from the DNBC (23). Cases were assigned selection ratios that did not vary by BMI. All were given a selection ratio of 1.10, indicating that they were 6% more likely to participate than normal weight controls (1.04*1.06 = 1.10). We assigned cases with missing BMI to have a selection ratio of 1.10 and controls with missing BMI to have a selection ratio of 0.94, the simple average of the lowest (0.84) and highest (1.04) selection ratios in the DNBC. For simplicity, we assumed that selection ratios did not differ by race/ethnicity.

We conducted a second analysis in which selection bias was assumed to be of a lesser magnitude. Cases were assigned a selection ratio of 1.10 as described above and controls were assigned selection ratios 0.05 higher than in the first analysis (underweight = 0.89, normal weight 1.09, overweight = 1.01, obese = 0.94, missing = 0.99).

Incorporation of Uncertainty

Uncertainty about the true values of the bias parameters (predictive values and selection probabilities) can be incorporated into bias analyses by assigning a probability distribution to each parameter, sampling randomly from the distribution over many iterations, and using the chosen values in the analysis (3, 6). Unlike a simple bias analysis in which a single value of the parameter is chosen and a single adjusted OR is obtained,

the result of a probabilistic bias analysis is a distribution of the adjusted OR, summarized as the median adjusted OR and 95% simulation interval (SI), the 2.5^{th} and 97.5^{th} percentiles of the adjusted OR distribution.

For the analysis of exposure misclassification, we first defined triangular distributions around the matching predictive values (i.e., women who correctly reported their BMI category). The mode of the distribution was set to be the NHANES value and the minimum and maximum values of the distributions were set to be 0.05 lower or higher than the mode, with distributions truncated to fall between 0 and 1, inclusive. All subsequent distributions within the same category of self-reported BMI were created constrained so the sum within self-reported BMI categories was 1 and the ratios of the predictive values were maintained. For example, the predictive values for Hispanic women self-reporting as obese were 0, 0.01, 0.07, and 0.92 for the underweight, normal weight, overweight, and obese categories (Table 6.5). The first value chosen was from the distribution for obesity with mode 0.92. Subsequent distributions were created conditional on the value drawn for the specific simulation run (e.g., if 0.95 was drawn, the sum of the values drawn from the other distributions was limited to 0.05). The mode of the distribution for overweight (original mode 0.07) was made to be larger than the mode for the distribution for normal weight (original mode 0.01) to preserve the same ratios of predictive values. Additional detail on creation of the distributions for predictive values is provided in Chapter 7 (Appendix B).

For selection ratios, we incorporated uncertainty using log-triangular distributions (log-triangular instead of triangular because they are on the ratio scale) with the mode equal to the natural log of the selection ratio chosen based on DNBC and NBDPS values. The minimum of the distribution was defined as 0.05 less than the mode and the maximum was set to be 0.05 greater than the mode. After randomly selecting a value from the distribution, it was exponentiated to obtain the selection ratio.

Each simulation was run 1,000 times, each time drawing new values for exposure misclassification weights and selection weights from their respective distributions. Median adjusted ORs and 95% SIs were produced to describe the variability in the point estimates attributable to bias.

To additionally incorporate random error into the simulation intervals, variability was added to the point estimate distribution. At each iteration of the simulation, we scaled the standard error from the conventional analysis (unweighted logistic regression) by multiplying it by a randomly selected value from a standard normal distribution. This randomly-scaled standard error was then subtracted from the log multiple bias-adjusted OR and this result was exponentiated to produce the random error-added point estimate (6). This distribution was summarized as the median OR and random error-added 95% SI.

Results

Crude and bias-adjusted OR estimates for associations between obesity and CL/P are shown in Table 6.6 and between obesity and CP are shown in Table 6.7. Associations between underweight and overweight and CL/P and CP are not discussed here, but are presented in Appendix Tables 6.B1 and 6.B2. For all analyses, the normal weight BMI category is the reference group.

Adjusting for confounding moved the OR up and away from the null for associations between obesity and both CL/P and CP, but results were overall similar to the crude analyses, showing a weak association.

When results were only adjusted for selection bias according to our assumptions, ORs for CL/P and CP were attenuated compared to the crude and confounding-adjusted estimates and associations were approximately null.

Adjustment for misclassification and missing data also moved the magnitude of the association down and towards the null, but weak associations between obesity and CL/P and CP remained.

After adjusting for exposure misclassification, selection bias, and confounding, given our assumptions about the values of the weights, associations between obesity and both CL/P and CP moved toward the null. Associations were approximately null for CL/P and very weak for CP.

NOTES ON WEIGHT ESTIMATION

For multiple bias analysis methods in which biases are adjusted serially, the order in which biases are adjusted is important (3, 6). Because adjustment for all types of bias occurs simultaneously using our approach, order itself is unimportant. However, bias parameters must be estimated in a specific way.

When both exposure misclassification and selection bias are present, there are four hypothetical datasets from which bias parameters could be estimated. In Figure 6.2, these hypothetical datasets are shown and their corresponding bias parameters are labeled A_1, A_2, B_1, and B_2. Parameters A_1 and A_2 correspond to Example 1 (hypothetical study). Predictive values for exposure classification conditional on selection (A_1) were estimated using correctly classified exposure in the selected population. Selection probabilities (A_2) were estimated independent of exposure misclassification (from the source population). Parameters B_1 and B_2 correspond to Example 2 (data from NBDPS). Selection weights (B_1) were estimated from a dataset most closely resembling a source population with misclassified exposure (DNBC) and predictive values (B_2) were estimated from a dataset most closely resembling a source population with correctly classified exposure (NHANES). When estimating B_2, we implicitly assumed that misclassification was nondifferential with respect to selection so that predictive values would be the same in people selected into the study and in people eligible but not selected, conditional on D and C.

Final weights for the analysis were calculated by multiplying the predictive values by the selection probabilities: A_1*A_2 in Example 1 and B_1*B_2 in Example 2. Using either of these final weights (A_1*A_2 or B_1*B_2) will, on average, produce results equal to the associations observed in the source population, given the weights are estimated correctly. These weights were estimated in a way that leads from the subset of the population with misclassified exposure (observed data) back to the complete source population with correctly classified exposure (source population). Any other combination of weights (i.e., A_1*B_1, A_2*B_2) might not produce results equal to associations observed in the source population, unless they are numerically equivalent to A_1*A_2 or B_1*B_2.

DISCUSSION

Using data from simulations, we have demonstrated the use of weighted logistic regression to simultaneously adjust for exposure misclassification, selection bias, and confounding. We have also provided an example of how adjustment for exposure misclassification and selection bias has the potential to affect study results.

One of the main advantages of this approach over several previously described methods for multiple bias analysis is that it is implementation uses logistic regression, a commonly used analytic method in epidemiology. Unlike multiple bias analysis methods that rely on contingency tables, control of multiple confounders is easily accomplished through modeling. In addition, adjustment for multiple biases can be done simultaneously instead of serially. Methods for multiple bias analysis using regression models have been previously described, but these have relied on simulations to adjust for biases (4, 10). For

epidemiologists who are uncomfortable writing their own simulations, creating weights might be a simpler alternative. This method can be extended to conduct probabilistic multiple bias analysis. Although the extension is straightforward, implementing probabilistic analyses requires more complex programming. In our example , results from the simple bias analysis were generally similar to results from the probabilistic analysis.

In our example using NBDPS data, we found that associations observed in the confounding-adjusted analyses were attenuated when adjustment for exposure misclassification and selection bias were incorporated, given our assumptions. Other case-control studies of obesity and isolated and non-isolated CL/P and CP have found crude or confounding-adjusted associations with similar magnitudes as the confounding-adjusted estimate from NBDPS (15, 18, 24-27). The literature suggests a weak association between prepregnancy obesity and CL/P and CP when the effects of exposure misclassification or selection bias are not considered in the analysis. If adjustment for these biases were accounted for, our conclusions about the strength of the evidence as a whole could change.

Two cohort studies of prepregnancy obesity and CL/P and CP (isolated and non-isolated cases combined) have been conducted in which measured height and weight were used, reducing misclassification. Both of these studies found that obesity was associated with increased risk of CL/P and CP (27, 28) suggesting that an association might exist when misclassification is absent or negligible. However, BMI was missing for 15% and 25% of

participants, making it difficult to draw conclusions about the potential roles for bias because BMI might not be missing at random (29, 30).

The success of any adjustment for bias is limited by lack of knowledge of the true values of bias parameters. In our example using NBDPS data, there was little information to confirm that our chosen bias parameters accurately reflected the truth. Predictive value estimates from NHANES might not have generalized to NBDPS, and we had no information to decide if misclassification was nondifferential or differential. Joint selection probabilities were unavailable, and we had to assume that selection bias occurred with a particular direction and magnitude. In the probabilistic analyses, we incorporated uncertainty in parameter choice in the analyses but we cannot be certain our triangular distributions included the true values or were centered around them. We can conclude, however, that given the information available in the literature, the associations previously observed between obesity and CL/P and CP could be entirely attributable to exposure misclassification, selection bias, and confounding. The bias parameters needed to produce a null effect are not so extreme as to be outside the realm of possibility.

Although we adjusted for three types of bias in our analyses, we did not include a comprehensive analysis for all types of error that could have occurred in the study, such as outcome misclassification, covariate misclassification, and unmeasured confounding. We also did not explore other issues important for making causal inference, such as the appropriateness of categorizing the continuous exposure measure, BMI, into four levels, of assuming no effect measure modification. In addition, if an association between BMI

and CL/P or CP exists, BMI might only be a proxy for an unknown causal exposure responsible for the effect such as adiposity, diet, or physical activity (31).

With use of bias analysis limited by available information on bias parameters, validation studies and other investigations into sources of bias will be important for obtaining more accurate bias parameters to use in future bias analyses. Even when there is uncertainty about the accuracy of bias parameters, performing bias analyses using plausible assumptions provides quantitative estimates of the potential direction and magnitude of biases in the study. Weighted logistic regression is a straightforward method that can be used to incorporate multiple bias analysis into epidemiologic studies to investigate the potential impact of biases on study results.

Table 6.1. Contingency Tables for the Source and Selected Populations and the Observed Data, Stratified by a Confounder: Simulated Data.

	C = 1		C = 0		Adjusted OR[a]
Source population	E = 1	E = 0	E = 1	E = 0	
D = 1	296	1,100	528	4,092	2.75
D = 0	727	7,982	3,937	81,338	
Selected population	E = 1	E = 0	E = 1	E = 0	
D = 1	263	1,056	477	3,900	3.09
D = 0	67	826	317	8,026	
Observed data	$E_M = 1$	$E_M = 0$	$E_M = 1$	$E_M = 0$	
D = 1	308	1,011	644	3,733	1.18
D = 0	142	751	1,137	7,206	

Abbreviations: C, confounder; D, disease; E, true exposure; E_M, misclassified exposure; OR, odds ratio.

[a] Mantel-Haenszel odds ratio adjusted for C.

Table 6.2. Exposure Classification Table From the Selected Population, Stratified by Disease and Confounder: Simulated Data.

	C = 1			C = 0		
	E = 1	E = 0	Total	E = 1	E = 0	Total
D = 1						
$E_M = 1$	260	48	308	470	174	644
$E_M = 0$	3	1,008	1,011	7	3,726	3,733
Total	263	1,056	1,319	477	3,900	4,377
D = 0						
$E_M = 1$	65	77	142	303	834	1,137
$E_M = 0$	2	749	751	14	7,192	7,206
Total	67	826	893	317	8,026	8,343

Abbreviations: C, confounder; D, disease; E, true exposure; E_M, misclassified exposure.

Table 6.3. Adjustment for Exposure Misclassification in the Observed Data: Simulated Data.

C	D	E_M	E_A	Observed N[a]	$p(E\|C,D,E_M,S)$[b]	Misclassification Adjusted N[c]
1	1	1	1	308	260/308	260
1	1	1	0	308	48/308	48
1	1	0	1	1,011	3/1,011	3
1	1	0	0	1,011	1,008/1,011	1,008
1	0	1	1	142	65/142	65
1	0	1	0	142	77/142	77
1	0	0	1	751	2/751	2
1	0	0	0	751	749/751	749
0	1	1	1	644	470/644	470
0	1	1	0	644	174/644	174
0	1	0	1	3,733	7/3,733	7
0	1	0	0	3,733	3,726/3,733	3,726
0	0	1	1	1,137	303/1,137	303
0	0	1	0	1,137	834/1,137	834
0	0	0	1	7,206	14/7,206	14
0	0	0	0	7,206	7,192/7,206	7,192

Abbreviations: C, confounder; D, disease; E_A, assigned exposure; E_M, misclassified exposure; N, number of observations; $p(E|C,D,E_M,S)$, probability of E given C, D, and E_M among those selected (S) into the study.

[a] Number of participants in the observed data given values of C, D, and E_M (from Table 6.1). Each participant is entered into the analysis twice: once with $E_A = 1$ and once with $E_A = 0$.

[b] Predictive values proportions calculated from the selected population (from Table 6.2).

[c] Product of observed N and $p(E|C,D,E_M,S)$.

Table 6.4. Adjustment for Selection Bias in the Selected Population: Simulated Data.

C	D	E_A	Misclassification-Adjusted N[a]	IPSW[b]	Final N[c]
1	1	1	260 + 3 = 263	296/263	296
1	1	0	48 + 1,008 = 1,056	1,100/1,056	1,100
1	0	1	65 + 2 = 67	727/67	727
1	0	0	77 + 749 = 826	7,982/826	7,982
0	1	1	470 + 7 = 477	528/477	528
0	1	0	174 + 3,726 = 3,900	4,092/3,900	4,092
0	0	1	303 + 14 = 317	3,937/317	3,937
0	0	0	834 + 7,192 = 8,026	81,338/8,026	81,338

Abbreviations: C, confounder; D, disease; E_A, assigned exposure; IPSW, inverse probability of selection weight; N, number of observations.

[a] Number of observations in strata of C, D, and E_A (from Table 6.3, summed over E_M).

[b] Inverse of selection probabilities (from Table 6.1).

[c] When predictive value and selection proportions are known with certainty, the final N is equal to the number of individuals in the source population (Table 6.1). Outside of simulation, obtaining the same counts as the source population would be unlikely.

Table 6.5. Example of Predictive Values for Exposure Classification, Non-Pregnant Hispanic Women Aged 16-49, National Health and Nutrition Examination Survey, 1999-2008.

	Measured BMI			
Self-reported BMI	Underweight	Normal weight	Overweight	Obese
Underweight	0.66	0.34	0	0
Normal weight	0.03	0.78	0.18	0.01
Overweight	0	0.05	0.74	0.21
Obese	0	0.01	0.07	0.92
Missing	0.01	0.32	0.26	0.41

Abbreviations: BMI, body mass index.

Table 6.6. Associations Between Prepregnancy Obesity and Cleft Lip With or Without Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.

Bias Adjustment	Simple Analysis[a]	Probabilistic Analysis (Bias Only)[b]		Probabilistic Analysis (Bias + Random Error)[b,c]	
	OR[d]	Median OR[d]	95% SI	Median OR[d]	95% SI + Error
Unadjusted[e]				1.15	1.00, 1.32
Confounding only[e,f]				1.20	1.05, 1.38
Selection bias only					
Selection 1	0.98	0.99	0.91, 1.07	0.99	0.84, 1.15
Selection 2	0.99	0.99	0.91, 1.08	1.00	0.85, 1.16
Misclassification only[g]					
Nondifferential	1.05	1.11	1.01, 1.17	1.10	0.95, 1.27
Differential 1	1.03	1.09	1.03, 1.15	1.08	0.93, 1.25
Differential 2	1.07	1.13	1.06, 1.19	1.12	0.96, 1.30
Multiple biases[h]					
Selection 1					
Nondifferential	0.96	1.02	0.93, 1.10	1.01	0.86, 1.19
Differential 1	0.94	1.00	0.93, 1.08	0.99	0.85, 1.17
Differential 2	0.98	1.03	0.95, 1.13	1.03	0.88, 1.21
Selection 2					
Nondifferential	0.96	1.02	0.94, 1.11	1.02	0.87, 1.20

Differential 1	0.95	1.00	0.92, 1.09	1.00	0.85, 1.18
Differential 2	0.98	1.04	0.95, 1.13	1.04	0.88, 1.22

Abbreviations: OR, odds ratio; SI, simulation interval.

[a] Single values of bias parameters chosen for the analysis.

[b] Triangular distributions of bias parameters sampled over 1,000 iterations.

[c] Random error incorporated as described in the text.

[d] Obese (body mass index ≥ 30 kg/m^2) versus normal weight (body mass index 18.5-24.9 kg/m^2).

[e] Point estimate and 95% confidence interval from conventional logistic regression analysis.

[f] Adjusted for maternal race/ethnicity.

[g] Adjusted for exposure misclassification and missing exposure data.

[h] Adjusted for exposure misclassification, missing exposure data, selection bias, and confounding by maternal race/ethnicity.

Table 6.7. Associations Between Prepregnancy Obesity and Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.

Bias Adjustment	Simple Analysis[a]	Probabilistic Analysis (Bias Only)[b]		Probabilistic Analysis (Bias + Random Error)[b,c]	
	OR[d]	Median OR[d]	95% SI	Median OR[d]	95% SI + Error
Unadjusted[e]				1.20	1.00, 1.45
Confounding only[e,f]				1.27	1.05, 1.52
Selection bias only					
Selection 1	1.03	1.03	0.95, 1.12	1.03	0.85, 1.26
Selection 2	1.04	1.04	0.95, 1.13	1.04	0.85, 1.27
Misclassification only[g]					
Nondifferential	1.08	1.13	1.07, 1.20	1.13	0.93, 1.37
Differential 1	1.06	1.11	1.06, 1.17	1.11	0.91, 1.34
Differential 2	1.09	1.15	1.09, 1.22	1.15	0.94, 1.39
Multiple biases[h]					
Selection 1					
Nondifferential	1.01	1.06	0.98, 1.16	1.06	0.87, 1.30
Differential 1	0.99	1.04	0.96, 1.14	1.04	0.85, 1.28
Differential 2	1.02	1.08	1.00, 1.18	1.08	0.89, 1.33
Selection 2					
Nondifferential	1.01	1.07	0.98, 1.17	1.07	0.88, 1.31
Differential 1	0.99	1.05	0.96, 1.15	1.05	0.86, 1.29

Differential 2	1.03	1.09	1.00, 1.19	1.09	0.89, 1.34

Abbreviations: OR, odds ratio; SI, simulation interval.

[a] Single values of bias parameters chosen for the analysis.

[b] Triangular distributions of bias parameters sampled over 1,000 iterations.

[c] Random error incorporated as described in the text.

[d] Obese (body mass index $\geq$ 30 kg/m^2) versus normal weight (body mass index 18.5-24.9 kg/m^2).

[e] Point estimate and 95% confidence interval from conventional logistic regression analysis.

[f] Adjusted for maternal race/ethnicity.

[g] Adjusted for exposure misclassification and missing exposure data.

[h] Adjusted for exposure misclassification, missing exposure data, selection bias, and confounding by maternal race/ethnicity.

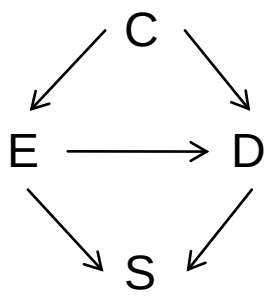

Figure 6.1. Directed acyclic graph illustrating relationships between exposure (E), disease (D), confounder (C), and selection (S).

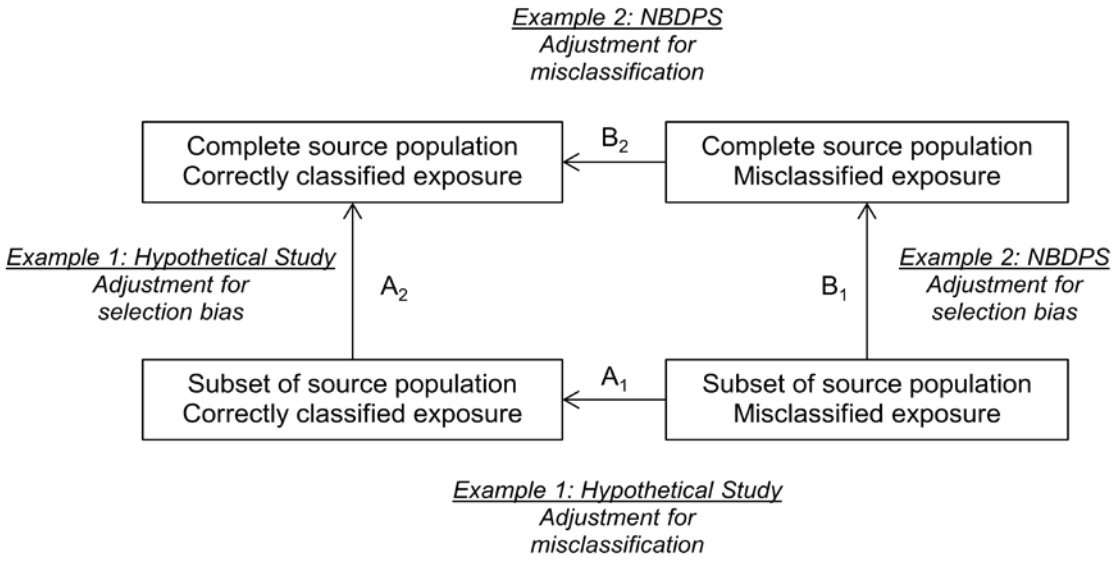

Figure 6.2. Estimation of bias parameters (A_1, A_2, B_1, B_2) when adjusting for both exposure misclassification and selection bias. Combinations of bias parameters that will produce an unbiased estimate of the correctly classified source population are A_1 and A_2, or B_1 and B_2. Other combinations (e.g., A_1 and B_2 or B_1 and A_2) might produce biased estimates. Abbreviations: NBDPS, National Birth Defects Prevention Study.

APPENDIX A

In this Appendix, we show algebraically that multiple bias analysis using a weighted analysis, when applied to observed data with exposure misclassification and selection bias, leads, on average, to the same joint distribution of exposure and disease as the source population when predictive values for exposure and selection probabilities are known. For demonstration purposes only, misclassification and selection bias are adjusted for sequentially; the same result would be obtained if the weights were applied simultaneously. Adjustment for confounding can also be incorporated, as shown in the Examples in the main text.

Table 6.A1 shows the joint distribution of true exposure (E) and disease (D) in the source and selected populations. We assume that the selected population was obtained by sampling from the source population with selection probabilities π_{ij} (i = disease, j = exposure). Table 6.A1 shows the expected number of participants in the selected population, based on this selection.

We assume that exposure misclassification (E_M) was introduced by measuring exposure in the selected population with sensitivities and specificities Se_i and Sp_i. This process leads to the expected number of participants in the observed data seen in Table 6.A2.

We now demonstrate that the expected cell counts after adjustment for exposure misclassification and selection bias using a weighted analysis in the observed data leads,

on average, back to the cell counts of the source population contingency table when the predictive values for exposure classification and the selection probabilities are known. The exposure classification table is shown in Table 6.A3. The right marginal total is set to equal the expected number of participants in the selected population stratified by D and summed over E_M. The middle columns show the expected number of participants in the selected population with E = 1 and E = 0 for each category of D and E_M, from which the predictive values (exposure misclassification weights) can be calculated as the cell-specific expected values divided by the marginal totals.

To adjust for exposure misclassification, each participant is entered into the analysis twice: one copy of the participant is assigned to be exposed ($E_A = 1$) and the other to be unexposed ($E_A = 0$) (Table 6.A4). Each observation is then multiplied (weighted) by its corresponding predictive value, $p(E|D,E_M,S)$, calculated from the exposure classification table (Table 6.A3), to produce the expected values for the misclassification-adjusted counts (labeled “Misclassification Adjusted N” in Table 6.A4).

Table 6.A5 shows the misclassification-adjusted counts from Table 6.A4 summed over E_M; these expected counts are equal to the expected counts of the contingency table for the selected population. To adjust for selection bias, each observation is multiplied (weighted) by the inverse of its selection probability (inverse probability of selection weight).

After adjustment for both exposure misclassification and selection bias using the weighted analysis, the final expected counts are equal to the contingency table counts from the source population.

A nearly identical argument shows that the source population counts can be replicated exactly if the exact predictive proportions and exact selection proportions for these populations are known (we define “predictive proportion positive” as the proportion of those measured as exposed who are truly exposed, with a similar definition for the unexposed).

Calculation of Exposure Weights Using Sensitivity and Specificity

If predictive values for exposure are unavailable but values of Se and Sp are available, these can be used to calculate the weights. Here, we derive these weights and show that they are equivalent to using predictive values in the weighted analysis. We use the a cell (D = 1, E = 1) as demonstration, but similar arguments can be used to derive weights for the other cells.

Positive predictive value (PPV), the exposure weight for the a’ cell, can be expressed in terms of Se and Sp:

$$PPV = p(E)Se_1 / [p(E)Se_1 + (1 - p(E))(1-Sp_1)]$$

Because p(E), prevalence of E, is unknown in the observed data, we must express it in terms of known quantities. To do this, we first express $\pi_{11}a$ in terms of known quantities. Rearranging from Tables 6.A1 and 6.A2:

$a' = Se_1(\pi_{11}a) + (1-Sp_1)(M_1 - \pi_{11}a)$

$a' = Se_1(\pi_{11}a) + M_1 - M_1Sp_1 + Sp_1(\pi_{11}a) - \pi_{11}a$

$\pi_{11}a = [a' - (1-Sp_1)M_1] / (Se_1 + Sp_1 - 1)$

Prevalence of E becomes:

$p(E) = \pi_{11}a / M_1$

$= [a' - (1-Sp_1)M_1] / (Se_1 + Sp_1 - 1)M_1$

$= [a'/M_1 - (1-Sp_1)] / (Se_1 + Sp_1 - 1)$

We now substitute this quantity into the equation for PVP and rearrange:

$PPV = [(a'/M_1 - (1-Sp_1))/(Se_1 + Sp_1 + 1)]Se_1 /$

$[((a'/M_1 - (1-Sp_1))/(Se_1 + Sp_1 + 1))Se_1 + ((Se_1 – a'/M_1)/(Se_1 + Sp_1 - 1))(1-Sp_1)]$

$= (a'Se_1/M_1 - Se_1 + Se_1Sp_1) / (a'Se_1/M_1 - Se_1 + Se_1Sp_1 + Se_1 – a'/M_1 + a'Sp_1/M_1 - Se_1Sp_1)$

$= a'/M_1(Se_1 - Se_1M_1/a' + Se_1Sp_1M_1/a') / (a'/M_1)(Se_1 + Sp_1 - 1)$

$= (Se_1 - Se_1M_1/a' + Se_1Sp_1M_1/a') / (Se_1 + Sp_1 - 1)$

$= Se_1/a'(a' - M_1 + Sp_1M_1) / (Se_1 + Sp_1 - 1)$

$= Se_1 (a' - M_1 + Sp_1M_1) / a'(Se_1 + Sp_1 - 1)$

This quantity can be used as the exposure weight. From Table 6.A4, we see that the misclassification adjusted $N = Se_1(\pi_{11}a)$. This is equivalent to a'PVP:

$a'PVP = a' [Se_1(a' - M_1 + Sp_1M_1) / a'(Se_1 + Sp_1 - 1)]$

$= Se_1[(a' - M_1 + Sp_1M_1) / (Se_1 + Sp_1 - 1)]$

$= Se_1(\pi_{11}a)$

Similar derivations can be used to show that Se and Sp can be substituted for other predictive values for the weights (Table 6.A6).

Table 6.A1. The Source and Selected Populations.

	Exposure Measure	
Source Population	E = 1	E = 0
D = 1	a	b
D = 0	c	d
Selected Population	E = 1	E = 0
D = 1	$\pi_{11}a$	$\pi_{10}b$
D = 0	$\pi_{01}c$	$\pi_{00}d$

Abbreviations: D, disease; E, true exposure; E_M, misclassified exposure.

Table 6.A2. Observed Data With Exposure Misclassification and Selection Bias. This Represents Data Usually Available in Epidemiologic Studies.

Observed Data	$E_M = 1$	$E_M = 0$	Total
D = 1	$a' = (Se_1)\,\pi_{11}a + (1\text{-}Sp_1)\,\pi_{10}b$	$b' = (1\text{-}Se_1)\,\pi_{11}a + (Sp_1)\,\pi_{10}b$	M_1
D = 0	$c' = (Se_0)\,\pi_{01}c + (1\text{-}Sp_0)\,\pi_{00}d$	$d' = (1\text{-}Se_0)\,\pi_{01}c + (Sp_0)\,\pi_{00}d$	M_0

Abbreviations: D, disease; E, true exposure; E_M, misclassified exposure; Se_i, sensitivity; Sp_i, specificity.

Table 6.A3. Exposure Classification Table.

	E = 1	E = 0	Total
D = 1			
$E_M = 1$	$(Se_1)\ \pi_{11}a$	$(1\text{-}Sp_1)\ \pi_{10}b$	$a' = (Se_1)\ \pi_{11}a + (1\text{-}Sp_1)\ \pi_{10}b$
$E_M = 0$	$(1\text{-}Se_1)\ \pi_{11}a$	$(Sp_1)\ \pi_{10}b$	$b' = (1\text{-}Se_1)\ \pi_{11}a + (Sp_1)\ \pi_{10}b$
Total	$\pi_{11}a$	$\pi_{10}b$	M_1
D = 0			
$E_M = 1$	$(Se_0)\ \pi_{01}c$	$(1\text{-}Sp_0)\ \pi_{00}d$	$c' = (Se_0)\ \pi_{01}c + (1\text{-}Sp_0)\ \pi_{00}d$
$E_M = 0$	$(1\text{-}Se_0)\ \pi_{01}c$	$(Sp_0)\ \pi_{00}d$	$d' = (1\text{-}Se_0)\ \pi_{01}c + (Sp_0)\ \pi_{00}d$
Total	$\pi_{01}c$	$\pi_{00}d$	M_0

Abbreviations: D, disease; E, true exposure; E_M, misclassified exposure; Se_i, sensitivity; Sp_i, specificity.

Table 6.A4. Adjustment for Exposure Misclassification

D	E_M	E_A	Observed N	p(E|D,E_M,S)	Misclassification Adjusted N[a]
1	1	1	a’	$(Se_1)\ \pi_{11}a/a'$	$(Se_1)\ \pi_{11}a$
1	1	0	a’	$(1\text{-}Sp_1)\ \pi_{10}b/a'$	$(1\text{-}Sp_1)\ \pi_{10}b$
1	0	1	b’	$(1\text{-}Se_1)\ \pi_{11}a/b'$	$(1\text{-}Se_1)\ \pi_{11}a$
1	0	0	b’	$(Sp_1)\ \pi_{10}b/b'$	$(Sp_1)\ \pi_{10}b$
0	1	1	c’	$(Se_0)\ \pi_{01}c/c'$	$(Se_0)\ \pi_{01}c$
0	1	0	c’	$(1\text{-}Sp_0)\ \pi_{00}d/c'$	$(1\text{-}Sp_0)\ \pi_{00}d$
0	0	1	d’	$(1\text{-}Se_0)\ \pi_{01}c/d'$	$(1\text{-}Se_0)\ \pi_{01}c$
0	0	0	d’	$(Sp_0)\ \pi_{00}d/d'$	$(Sp_0)\ \pi_{00}d$

Abbreviations: D, disease; E, exposure; E_M, misclassified exposure; N, number of observations; p(E|D,E_M,S), probability of true exposure, given disease, misclassified exposure, and selection into the study; Se_i, sensitivity; Sp_i, specificity.

[a] Product of observed N and p(E|D,E_M,S).

Table 6.A5. Adjustment for Selection Bias in Addition to Exposure Misclassification

D	E_A	Misclassification Adjusted N[a]	IPSW	Final N[b]
1	1	$(Se_1)\ \pi_{11}a + (1\text{-}Se_1)\ \pi_{11}a = \pi_{11}a$	$1/\ \pi_{11}$	$\pi_{11}a/\pi_{11} = a$
1	0	$(1\text{-}Sp_1)\ \pi_{10}b + (Sp_1)\ \pi_{10}b = \pi_{10}b$	$1/\ \pi_{10}$	$\pi_{10}b/\pi_{10} = b$
0	1	$(Se_0)\ \pi_{01}c + (1\text{-}Se_0)\ \pi_{01}c = \pi_{01}c$	$1/\ \pi_{01}$	$\pi_{01}c/\pi_{01} = c$
0	0	$(1\text{-}Sp_0)\ \pi_{00}d + (Sp_0)\ \pi_{00}d = \pi_{00}d$	$1/\ \pi_{00}$	$\pi_{00}d/\pi_{00} = d$

Abbreviations: D, disease; E_A, assigned exposure; IPSW, inverse probability of selection weight; N, number of observations.

[a] From Table 6.A3, summed over E_M.

[b] Product of misclassification adjusted N and IPSW.

Table 6.A6. Exposure Weights Using Sensitivity and Specificity.

D	E_M	E_A	Observed N	Exposure Weight	Misclassification Adjusted N
1	1	1	a’	Se_1 (a’ - M_1 + Sp_1M_1) / a’(Se_1 + Sp_1 - 1)	$(Se_1)\pi_{11}a$
1	1	0	a’	1 - [Se_1 (a’ - M_1 + Sp_1M_1) / a’(Se_1 + Sp_1 - 1)]	$(1\text{-}Sp_1)\ \pi_{10}b$
1	0	1	b’	Sp_1(b’ - M_1 + Se_1M_1) / b’(Se_1 + Sp_1 - 1)	$(1\text{-}Se_1)\ \pi_{11}a$
1	0	0	b’	1 - [Sp_1(b’ - M_1 + Se_1M_1) / b’(Se_1 + Sp_1 - 1)]	$(Sp_1)\ \pi_{10}b$
0	1	1	c’	Se_0 (c’ - M_0 + Sp_0M_0) / c’(Se_0 + Sp_0 - 1)	$(Se_0)\ \pi_{01}c$
0	1	0	c’	1 - [Se_0 (c’ - M_0 + Sp_0M_0) / c’(Se_0 + Sp_0 - 1)]	$(1\text{-}Sp_0)\ \pi_{00}d$
0	0	1	d’	Sp_0(d’ - M_0 + Se_0M_0) / d’(Se_0 + Sp_0 - 1)	$(1\text{-}Se_0)\ \pi_{01}c$
0	0	0	d’	1 - [Sp_0(d’ - M_0 + Se_0M_0) / d’(Se_0 + Sp_0 - 1)]	$(Sp_0)\ \pi_{00}d$

Abbreviations: D, disease; E_A, assigned exposure; E_M, misclassified exposure; N, number of observations; Se_i, sensitivity; Sp_i, specificity.

APPENDIX B

Additional Results

Table 6.B1. Associations Between Prepregnancy Underweight and Overweight and Cleft Lip With or Without Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.

Bias Adjustment	Simple Analysis[a]	Probabilistic Analysis (Bias Only)[b]		Probabilistic Analysis (Bias + Random Error)[b,c]	
	OR[d]	Median OR[d]	95% SI	Median OR[d]	95% SI + Error
Underweight					
Unadjusted[e]				1.32	1.07, 1.62
Confounding only[e,f]				1.32	1.07, 1.62
Selection bias only					
Selection 1	1.06	1.06	0.98, 1.15	1.06	0.85, 1.34
Selection 2	1.07	1.07	1.00, 1.17	1.07	0.86, 1.35
Misclassification only[g]					
Nondifferential	1.15	1.21	1.06, 1.36	1.21	0.95, 1.52
Differential 1	1.11	1.17	1.04, 1.33	1.17	0.93, 1.48
Differential 2	1.19	1.24	1.09, 1.40	1.24	0.98, 1.97
Multiple biases[h]					
Selection 1					

Nondifferential	1.01	1.06	0.92, 1.20	1.05	0.82, 1.34
Differential 1	0.98	1.03	0.90, 1.17	1.02	0.80, 1.30
Differential 2	1.05	1.09	0.95, 1.25	1.08	0.84, 1.38
Selection 2					
Nondifferential	1.02	1.07	0.93, 1.21	1.06	0.83, 1.35
Differential 1	0.99	1.04	0.90, 1.18	1.03	0.81, 1.31
Differential 2	1.06	1.10	0.96, 1.25	1.09	0.85, 1.39
Overweight					
Unadjusted[e]				0.99	0.87, 1.13
Confounding only[e,f]				1.01	0.89, 1.15
Selection bias					
Selection 1	0.89	0.92	0.85, 0.99	0.92	0.79, 1.06
Selection 2	0.90	0.92	0.85, 0.99	0.92	0.79, 1.07
Misclassification only[g]					
Nondifferential	0.91	0.99	0.88, 1.10	0.99	0.83, 1.18
Differential 1	0.88	0.95	0.85, 1.06	0.95	0.80, 1.13
Differential 2	0.95	1.03	0.92, 1.15	1.03	0.87, 1.23
Multiple biases[h]					
Selection 1					
Nondifferential	0.87	0.96	0.84, 1.08	0.96	0.80, 1.14
Differential 1	0.84	0.92	0.81, 1.03	0.92	0.77, 1.10
Differential 2	0.91	1.00	0.88, 1.13	1.00	0.84, 1.19

Selection 2					
Nondifferential	0.88	0.96	0.85, 1.08	0.96	0.80, 1.15
Differential 1	0.84	0.92	0.81, 1.03	0.92	0.77, 1.10
Differential 2	0.91	1.00	0.88, 1.13	1.00	0.84, 1.28

Abbreviations: OR, odds ratio; SI, simulation interval.

[a] Single values of bias parameters chosen for the analysis.

[b] Triangular distributions of bias parameters sampled over 1,000 iterations.

[c] Random error incorporated into the simulation interval.

[d] Underweight (body mass index <18.5 kg/m^2) or overweight (body mass index 25.0-29.9 kg/m^2) versus normal weight (body mass index 18.5-24.9 kg/m^2).

[e] Point estimate and 95% confidence interval from conventional logistic regression analysis.

[f] Adjusted for maternal race/ethnicity.

[g] Adjusted for exposure misclassification and missing exposure data.

[h] Adjusted for exposure misclassification, missing exposure data, selection bias, and confounding by maternal race/ethnicity.

Table 6.B2. Associations Between Prepregnancy Underweight and Overweight and Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.

Bias Adjustment	Simple Analysis[a]	Probabilistic Analysis (Bias Only)[b]		Probabilistic Analysis (Bias + Random Error)[b,c]	
	OR[d]	Median OR[d]	95% SI	Median OR[d]	95% SI + Error
Underweight					
Unadjusted[e]				1.18	0.88, 1.59
Confounding only[e,f]				1.18	0.88, 1.58
Selection bias only					
Selection 1	0.95	0.96	0.89, 1.04	0.96	0.71, 1.31
Selection 2	0.96	0.97	0.89, 1.05	0.97	0.72, 1.32
Misclassification only[g]					
Nondifferential	1.06	1.11	0.98, 1.26	1.12	0.81, 1.52
Differential 1	1.02	1.08	0.95, 1.22	1.08	0.79, 1.48
Differential 2	1.10	1.15	1.01, 1.30	1.15	0.84, 1.58
Multiple biases[h]					
Selection 1					
Nondifferential	0.93	0.97	0.85, 1.11	0.96	0.70, 1.31
Differential 1	0.90	0.94	0.82, 1.08	0.93	0.68, 1.27
Differential 2	0.97	1.00	0.87, 1.15	1.00	0.72, 1.36

Selection 2					
Nondifferential	0.94	0.98	0.85, 1.12	0.97	0.70, 1.32
Differential 1	0.90	0.95	0.82, 1.09	0.94	0.68, 1.28
Differential 2	0.97	1.01	0.88, 1.16	1.00	0.73, 1.37
Overweight					
Unadjusted[e]				1.07	0.90, 1.27
Confounding only[e,f]				1.11	0.93, 1.31
Selection bias					
Selection 1	0.96	0.99	0.92, 1.07	0.99	0.82, 1.20
Selection 2	0.97	0.99	0.92, 1.07	1.00	0.82, 1.20
Misclassification only[g]					
Nondifferential	0.95	1.03	0.92, 1.15	1.03	0.83, 1.26
Differential 1	0.91	0.99	0.88, 1.10	0.99	0.80, 1.21
Differential 2	0.99	1.07	0.96, 1.20	1.07	0.86, 1.31
Multiple biases[h]					
Selection 1					
Nondifferential	0.93	1.02	0.90, 1.15	1.02	0.83, 1.26
Differential 1	0.89	0.98	0.86, 1.11	0.98	0.79, 1.21
Differential 2	0.97	1.07	0.94, 1.20	1.06	0.86, 1.31
Selection 2					
Nondifferential	0.93	1.03	0.90, 1.15	1.02	0.83, 1.26
Differential 1	0.89	0.98	0.86, 1.11	0.98	0.79, 1.21

Differential 2	0.97	1.07	0.84, 1.20	1.06	0.86, 1.31

Abbreviations: OR, odds ratio; SI, simulation interval.

[a] Single values of bias parameters chosen for the analysis.

[b] Triangular distributions of bias parameters sampled over 1,000 iterations.

[c] Random error incorporated into the simulation interval.

[d] Underweight (body mass index <18.5 kg/m^2) or overweight (body mass index 25.0-29.9 kg/m^2) versus normal weight (body mass index 18.5-24.9 kg/m^2).

[e] Point estimate and 95% confidence interval from conventional logistic regression analysis.

[f] Adjusted for maternal race/ethnicity.

[g] Adjusted for exposure misclassification and missing exposure data.

[h] Adjusted for exposure misclassification, missing exposure data, selection bias, and confounding by maternal race/ethnicity.

REFERENCES

1. Maclure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology*. 2001;12:114-122.
2. Jurek AM, Maldonado G, Greenland S, et al. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol*. 2006;21:871-876.
3. Greenland S, Lash TL. Bias Analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology, 3rd edition*. Philadelphia: Lippincott Williams & Wilkins, 2008:345-380.
4. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14:451-458.
5. Greenland S. Multiple-bias modelling for analysis of observational data. *J Royal Stat Soc A*. 2005;168(2):267-306.
6. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer; 2009.
7. 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;18:2184-2190.
8. Lash TL, Schmidt M, Jensen AO, et al. Methods to apply probabilistic bias analysis to summary estimates of association. *Pharmacoepidemiol Drug Saf*. 2010;19:638-644.
9. Orsini N, Bellocco R, Bottai M, et al. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. *Stata J*. 2008;8:29-48.

10. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34:1370-1376.

11. Maldonado G. Adjusting a relative-risk estimate for study imperfections. *J Epidemiol Community Health*. 2008;62:655-663.

12. Lyles RH, Lin J. Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. *Stat Med*. 2010;29:2297-2309.

13. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-625.

14. Scialli AR, Public Affairs Committee of the Teratology Society. Teratology Public Affairs Committee position paper: maternal obesity and pregnancy. *Birth Defects Res A Clin Mol Teratol*. 2006;76(2):73-77.

15. Stothard KJ, Tennant PW, Bell R, et al. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. *JAMA*. 2009;301(6):636-650.

16. Shaw GM, Todoroff K, Schaffer DM, et al. Maternal height and prepregnancy body mass index as risk factors for selected congenital anomalies. *Paediatr Perinat Epidemiol*. 2000;14(3):234-239.

17. Waller DK, Shaw GM, Rasmussen SA, et al. Prepregnancy obesity as a risk factor for structural birth defects. *Arch Pediatr Adolesc Med*. 2007;161(8):745-750.

18. Stott-Miller M, Heike CL, Kratz M, et al. Increased risk of orofacial clefts associated with maternal obesity: case-control study and Monte Carlo-based bias analysis. *Paediatr Perinat Epidemiol*. 2010;24(5):502-512.

19. Yoon PW, Rasmussen SA, Lynberg MC, et al. The National Birth Defects Prevention Study. *Public Health Rep*. 2001;116(Supple 1):32-40.

20. Cogswell ME, Bitsko RH, Anderka M, et al. Control selection and participation in an ongoing, population-based, case-control study of birth defects: the National Birth Defects Prevention Study. *Am J Epidemiol*. 2009;170(8):975-985.

21. Rasmussen SA, Olney RS, Holmes LB, et al. Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Defects Res A Clin Mol Teratol*. 2003;67(3):193-201.

22. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey. 1999-2008. (http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm). (Accessed).

23. Nohr EA, Frydenberg M, Henriksen TB, et al. Does low participation in cohort studies induce bias? *Epidemiology*. 2006;17:413-418.

24. Moore LL, Singer MR, Bradlee ML, et al. A prospective study of the risk of congenital defects associated with maternal obesity and diabetes mellitus. *Epidemiology*. 2000;11(6):689-694.

25. Oddy WH, De Klerk NH, Miller M, et al. Association of pre-pregnancy weight and birth defects: evidence from a case-control study in Western Australia. *Aust N Z Obstet Gynaecol*. 2009;49(1):11-15.

26. Blomberg MI, Kallen B. Maternal obesity and morbid obesity: the risk for birth defects in the offspring. *Birth Defects Res A Clin Mol Teratol*. 2010;88(1):35-40.

27. Rankin J, Tennant PW, Stothard KJ, et al. Maternal body mass index and congenital anomaly risk: a cohort study. *Int J Obes (Lond)*. 2010;34(9):1371-1380.

28. Cedergren M, Kallen B. Maternal obesity and risk for orofacial clefts in the offspring. *Cleft Palate Craniofac J*. 2005;42:367-371.

29. Mishra GD, Dobson AJ. Multiple imputation for body mass index: lessons from the Australian Longitudinal Study on Women's Health. *Stat Med*. 2004;23(19):3077-3087.

30. Ramos E, Lopes C, Oliveira A, et al. Unawareness of weight and height - the effect on self-reported prevalence of overweight in a population-based study. *J Nutr Health Aging*. 2009;13(4):310-314.

31. Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes*. 2008;32:S8-S14.

CHAPTER 7

Potential Impacts of Biases on Associations Between Prepregnancy Body Mass Index and Neural Tube Defects

Candice Y. Johnson, Margaret A. Honein, Sonja A. Rasmussen, Penelope P. Howards, Godfrey P. Oakley, Jr., Matthew J. Strickland, D. Kim Waller, W. Dana Flanders, and The National Birth Defects Prevention Study

Author affiliations: Departments of Epidemiology (Candice Y. Johnson, Penelope P. Howards, Godfrey P. Oakley, Jr., Matthew J. Strickland, W. Dana Flanders), Environmental Health (Matthew J. Strickland), and Biostatistics and Bioinformatics (W. Dana Flanders), Emory University, Atlanta, Georgia; National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia (Candice Y. Johnson, Margaret A. Honein, Sonja A. Rasmussen); University of Texas at Houston, School of Public Health, Houston, Texas (D. Kim Waller).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Corresponding author: Candice Y. Johnson, cyjohnson@alum.emory.edu.

ABSTRACT

Epidemiologic studies have consistently reported associations between prepregnancy body mass index (BMI) and neural tube defects such as anencephaly and spina bifida, but few have quantitatively addressed the potential roles of exposure misclassification or selection bias. The authors used data from the National Birth Defects Prevention Study, including 370 mothers of infants with anencephaly (cases), 738 mothers of infants with spina bifida (cases) and 6,030 mothers of infants with no major birth defect (controls). The authors investigated associations between obesity (BMI $\geq$30 kg/m^2 versus 18.5-24.9 kg/m^2) and anencephaly or spina bifida, simultaneously adjusting for exposure misclassification, selection bias, and confounding using probabilistic weighted logistic regression. For anencephaly, the authors observed no association with obesity in the confounding-adjusted analyses (odds ratio [OR] 1.11, 95% confidence interval [CI] 0.83, 1.49). Probabilistic adjustment for exposure misclassification, selection bias, and confounding did not appreciably change results, given the authors' assumptions (median OR range: 1.05 to 1.11). For spina bifida, the confounding-adjusted estimate (OR 1.62, 95% CI 1.33, 1.97) was stronger than when adjusting for all 3 biases (median OR range: 1.26 to 1.46), given their assumptions. It is likely that associations between obesity and spina bifida are not as strong as previously reported.

INTRODUCTION

Neural tube defects (NTDs), such as anencephaly and spina bifida, are severe birth defects of the brain and spinal cord that occur early in embryonic development (1). Anencephaly is characterized by absence of the brain, skull, and scalp and is a defect incompatible with life (2). Spina bifida results from herniation of meninges, with or without the spinal cord, through an incompletely formed spine (3). In the United States, over 90% of infants with spina bifida survive the first year of life (4), but affected individuals often have lifelong mobility impairment and other neurologic problems (5). Periconceptional folic acid supplementation has been shown to prevent NTDs in randomized controlled trials and observational studies (6-10). However, folic acid does not appear to prevent all cases of NTDs, and investigators have continued to search for other risk factors for NTDs.

Over a dozen studies have identified prepregnancy body mass index (BMI) as a potential risk factor for NTDs, with obese mothers almost twice as likely to have a pregnancy affected by an NTD than mothers with normal weight (11-15). Several biologic mechanisms have been proposed to explain this association, including pathways involving nutrient intake, diet quality, and glycemic control (16). A non-biologic mechanism has also been proposed: that exposure misclassification and selection bias, typically not quantitatively addressed in analyses, could be driving the association (17-19).

Height and weight (used to calculate BMI) are not always accurately reported by women (20). American women who are overweight or obese sometimes under-report BMI, and women who are underweight tend to over-report BMI (21). When BMI is categorized for analysis, as opposed to using the continuous variable, the expected direction of bias from misclassification is unpredictable and could be towards or away from the null (22, 23). Self-reported BMI has been used in the analysis of most studies of prepregnancy BMI and NTDs (13, 17, 24-32), making it likely that some degree of bias from misclassification is present in these studies.

Selection bias could arise if BMI and NTDs are predictors of case ascertainment (selection into the study) and case ascertainment is incomplete (19). The mechanism by which selection bias is hypothesized to occur in studies of BMI and NTDs is illustrated by the directed acyclic graph in Figure 7.1. Under this mechanism of selection bias, normal weight case mothers are more likely than obese case mothers to learn prenatally that they are carrying a fetus with an NTD. Studies of other types of birth defects have shown that visualization of the fetal anatomy, and thus ultrasound prenatal diagnosis of a birth defect, is more difficult in obese mothers than normal weight mothers (33-37). Once a prenatal diagnosis is made of a severe birth defect such as anencephaly or spina bifida, termination of pregnancy is common (38). Ascertainment of cases among terminations of pregnancy can be difficult because they are not typically captured in vital records and often occur in outpatient settings. The net result of this process is that normal weight case mothers (more likely to have had a prenatal diagnosis of an NTD and a subsequent termination of pregnancy) are less likely to be included in studies than obese case

mothers (more likely to have carried the pregnancy to term and only learned of the NTD after the infant's birth). This could produce a spurious association between obesity and NTDs.

Although exposure misclassification and selection bias are hypothesized to be important sources of bias in studies of BMI and NTDs, studies to date have mainly focused on bias from confounding. No study to date has incorporated adjustment for both exposure misclassification and selection bias into the statistical analysis. The aim of our study was to perform a probabilistic, quantitative assessment of the potential impacts of exposure misclassification, selection bias, and confounding on associations between prepregnancy BMI and anencephaly and spina bifida.

MATERIALS AND METHODS

Population

We used data from the National Birth Defects Prevention Study (NBDPS), an ongoing, population-based, case-control study of genetic and environmental risk factors for major structural birth defects (39, 40). Potentially eligible cases were identified from birth defect surveillance systems in Arkansas, California, Georgia, Iowa, Massachusetts, New Jersey, New York, North Carolina, Texas, and Utah. Data abstracted from medical records were reviewed by clinical geneticists to determine eligibility and to further classify the case. Cases with known or suspected chromosomal abnormalities or single-gene disorders were excluded. Eight of the 10 study sites included live births, stillbirths (fetal deaths ≥20 gestational weeks), and terminations of pregnancy (any gestational age).

New Jersey included live births and stillbirths only, and Massachusetts included live births only. Live births with no major birth defects were selected as controls from birth certificates or hospital records in the same catchment area and during the same time period as cases. Between 6 weeks and 2 years following delivery, mothers were contacted by telephone to complete an interview including questions on sociodemographics, reproductive history, and nutrition, medication use, chronic and acute illnesses, and occupation during pregnancy. All participating sites obtained institutional review board approval.

For the present study, we included control mothers, anencephaly case mothers, and spina bifida case mothers who completed the interview and who had a delivery date on or after October 1, 1997 and an estimated date of delivery on or before December 31, 2007. Unless otherwise indicated, we excluded cases and controls from study sites and years during which NTD cases among terminations of pregnancy were not eligible for inclusion: Massachusetts and New Jersey (all years) and New York (1997-1999).

Exposure Measurement

Data on height and prepregnancy weight were obtained in answer to the questions “What is your height without shoes?” and “How much did you weigh before your pregnancy with (baby’s name)?”. Prepregnancy BMI was calculated as self-reported weight in kilograms divided by squared self-reported height in meters and categorized into four standard levels: underweight (<18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2),

overweight (25.0-29.9, kg/m^2), and obese (≥30 kg/m^2). In all analyses, normal weight mothers served as the reference group.

Conventional Analysis

We used logistic regression to estimate crude and confounding-adjusted odds ratios (OR) and 95% confidence intervals (CI) for associations between prepregnancy BMI and anencephaly and spina bifida. We initially identified 5 potential confounders: maternal race/ethnicity, household income, maternal age at delivery, maternal education, and periconceptional folic acid supplementation. In preliminary analyses, we found that adjustment for these variables had little effect on results. For simplicity, we included only maternal race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other) and household income (<$50,000, ≥$50,000) in the multivariable regression analyses. Because observations with missing data are dropped from regression models, we excluded mothers with missing data on these 2 potential confounders: 529 control mothers (8%), 40 anencephaly case mothers (10%), and 60 spina bifida case mothers (8%).

If underascertainment of cases among terminations of pregnancy was producing selection bias in the way we predicted, we would expect to see stronger associations in study sites excluding terminations than those including them. We estimated associations between BMI categories and NTDs in study sites including and excluding terminations to investigate whether or not the magnitude of the associations differed.

Analysis of Exposure Misclassification

To adjust for exposure misclassification, we used a method based on weighted logistic regression (41). Four copies of each participant were used in the analysis, with each copy assigned one of the 4 possible true BMI categories (underweight, normal weight, overweight, obese). Predictive values for exposure classification were estimated from available data (described further below). These values represent the probability that each assigned BMI category represented the truth, given the category self-reported by the participant. When used as weights in the weighted logistic regression, the predictive values re-distribute participants into BMI categories in proportion to the probability that those BMI categories were the "truth". The predictive values sum to 1 across the 4 assigned BMI categories so that each study participant is effectively entered into the analysis only once, even though 4 copies of the participant exist. Assigned BMI category is used as the exposure variable in the logistic regression model instead of the BMI category reported by the participant.

We used data from the National Health and Nutrition Examination Survey (NHANES) to estimate predictive values for BMI classification. In NHANES, participants self-reported their weight and height during an initial interview. Within the next 2 weeks, they underwent a physical examination during which their height and weight were measured. The present validation data included the 7,177 non-pregnant females aged 16-49 who had both their height and weight measured during the NHANES cycles between 1999 and 2008 and had no missing data on race/ethnicity or household income.

In NHANES, some participants had missing self-reported BMI but had measured BMI available. In the same way as we used predictive values to adjust for misclassification, we calculated predictive values for exposure classification from participants with missing self-reported BMI to adjust for missing data on BMI.

Predictive values were estimated from NHANES data conditional on potential confounders race/ethnicity and household income. Race/ethnicity was defined in NHANES using the same 4 categories as NBDPS. Because the household income cutpoints used by NHANES differed from those used by NBDPS, we dichotomized household income as <$55,000 and ≥$55,000 in the NHANES dataset.

Three exposure misclassification types were considered in the analysis: nondifferential misclassification (cases and controls have equivalent classification), differential misclassification in which cases had better classification than controls (referred in this manuscript as “differential 1”), and differential misclassification in which controls had better classification than cases (referred in this manuscript as “differential 2”).

To create a probabilistic analysis (i.e., assigning probability distributions to the predictive values instead of using fixed values), we constructed triangular distributions around each predictive value. For all misclassification types, all participants were initially assigned the same triangular distribution: mode of the distribution equal to the predictive values estimated from NHANES, and minimum and maximum of the distribution 0.05 below or above the mode. For nondifferential misclassification, these were the triangular

distributions used in the analysis. These distributions were shifted upward by 0.02 for cases (differential 1) or downward by 0.02 for cases (differential 2). Further details are provided in Appendix B.

At each of 1,000 iterations of the simulation, we selected a value from each predictive value distribution and used it as the weight in the analysis. To incorporate random error into the study results, we used a previously described method involving addition of random error to simulation results (42). We drew a random variable from a standard normal distribution, multiplied it with the crude standard error, added this value to the beta-estimate from the bias-adjusted regression model, and exponentiated it to produce the random error-added OR. The results of the analysis were summarized as the median random error-added OR from simulation and 95% random error-added simulation intervals (SI+RE), the 2.5^{th} and 97.5^{th} percentiles of the random error-added OR distribution from the simulation. For all simulation results, the abbreviation "OR" means the random error-added OR.

Analysis of Selection Bias

We used inverse probability of selection weights to adjust for selection bias (43). In this method, the probability of selection (case ascertainment) is estimated, the inverse is taken, and this value is used as the weight in a weighted logistic regression. The weights are used to make copies of under-represented study participants (number of copies equal to the weight) so that in the analysis, the "pseudo-population" created by weighting approximates the population that would have existed in the absence of selection bias.

The extent to which cases among stillbirths and terminations of pregnancy are underascertained in birth defect surveillance programs is unknown, and there were few estimates to guide our estimation of selection probabilities. We hypothesized 2 types of incomplete case ascertainment. In the first (referred to in this manuscript as “selection 1”), we assumed the proportion of pregnancies ending in stillbirth or termination of pregnancy differed between study sites, possibly due to differences in sociodemographic characteristics of the population or state laws governing termination of pregnancy. In the second (referred to in this manuscript as “selection 2”), we assumed that these proportions were constant across study sites, and any differences between sites were due to differences in case ascertainment.

In the analysis of potential selection bias, we made the following assumptions:

- Case ascertainment among live births is complete.
- Case ascertainment among stillbirths and terminations is incomplete at every study site.
- There is better case ascertainment among stillbirths than terminations of pregnancy.
- Case ascertainment does not vary by BMI.
- The likelihood a mother opts for termination of pregnancy does not vary by BMI.

Because we assumed that ascertainment of livebirths was complete, the number of livebirths observed at each study site represents the true number of liveborn cases and

controls; however, for cases, the number of stillbirths and terminations of pregnancy are underestimated. Instead of estimating selection probabilities, we calculated the ratios of stillbirths to livebirths (SB:LB) and of terminations of pregnancy to livebirths (TOP:LB) for cases at each study site (Appendix Table 7.A1).

To create a probabilistic analysis, we constructed triangular distributions for the SB:LB and TOP:LB ratios; because these are ratio measures, the triangular distributions were based on the ln(SB:LB) and ln(TOP:LB) values, which were then transformed back to their original values prior to analysis. Detailed descriptions of creation of the triangular distributions are provided in Appendix B. Briefly, for selection 1, ln(SB:LB) and ln(TOP:LB) for each site were assigned to be the minimum of the triangular distributions (according to our assumption that all sites underascertained these values) and the mode and maximum of the distributions were assigned based on the prevalence of stillbirths and terminations of pregnancy for anencephaly or spina bifida, as described in Appendix B. Each site therefore had a different distribution. For selection 2, 1 triangular distribution for SB:LB and 1 for TOP:LB were created for the whole dataset. The minimum of the distributions was assigned to be the largest values of the site-specific ln(SB:LB) and ln(TOP:LB) values. The mode and maximum of the distributions were created as described in Appendix B. For simplicity, we assumed that these ratios did not vary by race/ethnicity or income (confounders).

The weight to be used in the weighted logistic regression was calculated as follows:

w_{SB} = exp (ln(SB:LB) drawn from distribution) / observed SB:LB value

w_{TOP} = exp (ln(TOP:LB) drawn from distribution) / observed TOP:LB value

A weight of 2 for TOP:LB, for example, would indicate that the ratio of terminations to livebirths was assumed to be twice as high as observed in the study. In the analysis, all pregnancies ending in termination would be assigned a weight of 2 and the livebirths a weight of 1 (LB:LB ratio = 1) The weights calculated here are substitutes for the inverse probability of selection weights and function in the same manner.

In the same way as for the analysis of misclassification weighted logistic regression, 1,000 simulations were conducted and the results were summarized as median OR and 95% SI+RE, with random error incorporated as described above.

Multiple Bias Analysis

At each of the 1,000 simulations, the weights for exposure misclassification and selection bias were multiplied to obtain an overall weight for the multiple bias analysis (Dissertation Chapter 6). We used this weight in a weighted multivariable logistic regression to simultaneously adjust for exposure misclassification, selection bias, and confounding. We adjusted for confounding by controlling for maternal race/ethnicity and household income in the regression model. Results were summarized as median OR and 95% SI+RE; however, the standard error used to incorporate random error was estimated from a confounding-adjusted (instead of unadjusted) conventional logistic regression model.

RESULTS

Characteristics of study participants are shown in Table 7.1. Case mothers were more often of Hispanic race/ethnicity than control mothers. Case mothers were slightly more likely to have a household income <$50,000 than control mothers.

Table 7.2 shows associations between BMI and NTDs in study sites including and excluding terminations of pregnancy. For both anencephaly and spina bifida, for every BMI category, ORs were higher in the sites excluding terminations than sites including these cases, adjusting for potential confounding by maternal race/ethnicity and household income.

Bias Analyses: Anencephaly

Crude and confounding-adjusted associations were similar (Table 7.3). After adjusting for only confounding, there was a weak association between anencephaly and underweight (OR 1.26, 95% CI 0.80, 1.98), but not with overweight (OR 0.95, 95% CI 0.72, 1.25) or obesity (OR 1.11, 95% CI 0.83, 1.49).

Adjustment for selection bias, given our assumptions, did not appreciably change the magnitude of the associations compared to crude or confounding-adjusted estimates. The greatest change was a strengthening of the association with underweight (median OR range: 1.35-1.39, compared to confounding-adjusted OR 1.26) There was little effect of adjusting for exposure misclassification under any of the 3 assumptions, except for a

small attenuation of the association with underweight (median OR range: 1.07-1.12, compared to confounding-adjusted OR 1.26).

Simultaneous adjustment for three biases had little effect on results, with associations between overweight and obesity changing only slightly. The greatest effect was a mild attenuation of the association with underweight, with multiple bias-adjusted median ORs ranging from 1.14 to 1.21, compared to a crude OR of 1.28.

Bias Analyses: Spina Bifida

In crude and confounding-adjusted analyses (Table 7.4), dose-response associations were observed, with higher BMI categories associated with higher likelihood of having a child with spina bifida (underweight OR 0.73, 95% CI 0.47, 1.12; overweight OR 1.16, 95% CI 0.95, 1.42; obese OR 1.62, 95% CI 1.33, 1.97).

Associations were mostly unchanged after adjustment for only selection bias when assuming the prevalence of stillbirths and terminations differed between sites (selection 1). When assuming the prevalence was the same (selection 2), the dose-response pattern observed in the crude and confounding-adjusted analyses became less obvious.

Adjustment for misclassification had little effect, except for attenuating the association with underweight (median OR range: 0.86 to 0.94, compared to confounding-adjusted OR 0.73).

After adjustment for multiple biases, associations were weaker for the obese category (multiple bias-adjusted median OR range: 1.30-1.48, compared to confounding-adjusted OR 1.62) and the underweight category (multiple bias-adjusted median OR range: 0.79-1.03, compared to confounding-adjusted OR 0.73).

DISCUSSION

Given the assumptions in our analyses, we found that exposure misclassification, selection bias, and confounding had little effect on associations between obesity and anencephaly, but might account for some of the observed association with spina bifida. However, the analysis was complicated by lack of evidence available to guide our assumptions about the potential sources, magnitudes, and directions of bias in the study.

Crude and confounding-adjusted results from NBDPS are fairly similar to the results of other studies of BMI and NTDs, including previously published estimates using NBDPS data from 1997-2002 (17, 44, 45). Many of these studies were conducted in the United States, used case-control designs, likely had underascertainment of cases among terminations of pregnancy, and calculated BMI from self-reported height and prepregnancy weight (13, 24, 25, 27, 28, 30, 31). It is possible that similar magnitudes and directions of bias from exposure misclassification and selection bias could be occurring in these studies as occurred in NBDPS, and results of our multiple bias analysis could be generalizable to other studies.

Our analysis of sites including and excluding selection bias produced results consistent with what we would expect if selection bias resulted in overestimation of the OR; however, there could be other differences between these 2 types of sites that explain why the magnitude of the association tended to be greater in the sites excluding terminations. There is evidence to support an association between prepregnancy BMI and NTDs even in the absence of selection bias: a study conducted between 1968 and 1980, a period during which prenatal diagnosis of NTDs was uncommon, found that obese mothers (BMI $\geq$ 29 kg/m^2) were twice as likely to have an affected child as mothers with normal BMI (26, 46). During this time period, selection bias acting through our hypothesized mechanism would not have occurred.

Under our 2 assumptions about selection bias, we found that each assumption produced different results for spina bifida. In one, the dose-response was preserved and in the other, it was less apparent. Although no information was available to determine which assumption was more accurate, under both assumptions an association between obesity and spina bifida remained. The small proportion of cases of spina bifida in our dataset ending in termination (11%, assumed to be an underestimate) suggests that even if obese mothers are less likely to have a termination of pregnancy, the prevalence of terminations might be too low to create substantial selection bias.

We made 3 assumptions about the direction of misclassification in the study because we did not know which was the most appropriate. An incorrect assumption about nondifferential or differential misclassification could produce biased results when

adjusting for misclassification (Dissertation Chapter 5). Validation data for the accuracy of self-reported BMI were not available from NBDPS, and we had to assume that NHANES was an appropriate validation sample. Because the validation data from NHANES were for current and not prepregnancy BMI, the validity of this assumption is unclear. To our knowledge, no study has reported the accuracy of self-reported prepregnancy BMI categories, in particular none has reported separate estimates for anencephaly case mothers, spina bifida case mothers, and control mothers. However, results of the analyses under the 3 assumptions produced similar results, given our assumptions about the triangular distributions.

Other investigators might disagree with some of the assumptions we have made, or might believe that a bias we did not include in the analysis is too important to ignore. Bias analysis allows investigators the opportunity to present quantitative assessments of the potential effects of bias on study results given their assumptions about sources and magnitudes of bias in the study. However, to resolve differences and determine which assumptions are the most accurate, studies reporting data facilitating bias analysis, such as validation studies or investigations of potential for selection bias, will be particularly important.

Other important considerations that could affect our results were not addressed in this analysis. These include misclassification of outcome or covariates, other types of selection bias (exclusion of mothers with missing data on race/ethnicity and household income, differences in characteristics of mothers participating and not participating in the

study (47), other maternal characteristics associated with pregnancy outcome (48)), residual confounding from not conditioning selection weights on potential confounders, existence of unmeasured confounders, effect measure modification or interaction, or potential etiologic heterogeneity between spina bifida subtypes (49), among others. In addition, BMI is a complex exposure and likely only serves as a proxy for other potentially causal entities such as adiposity, diet, physical activity, genetics, or combinations of these and other variables (50, 51).

The main challenge in our multiple bias analysis was the lack of evidence available to guide our assumptions about how exposure misclassification or selection bias occurred in the study and our inability to determine if our assumptions approximated the truth. The results of the multiple bias analysis therefore might not improve on the confounding-adjusted estimates. However, the results suggest that an overestimation of the association between obesity and spina bifida is plausible, given our assumptions, which are not unrealistic.

Table 7.1. Characteristics of Participants, National Birth Defects Prevention Study, 1997-2007.

	Controls N = 6,030	Anencephaly N = 370	Spina Bifida N = 738
	n (%)[a]	n (%)[a]	n (%)[a]
Pregnancy outcome			
Live birth	6,030 (100)	102 (28)	636 (86)
Stillbirth		94 (25)	21 (3)
Termination of pregnancy		174 (47)[b]	81 (11)[b]
Case type			
Isolated defect		335 (91)	651 (88)
≥2 major birth defects		35 (10)	87 (12)
Study site			
Arkansas	972 (16)	51 (14)	101 (14)
California	894 (15)	92 (25)	145 (20)
Georgia	787 (13)	45 (12)	105 (14)
Iowa	843 (14)	42 (11)	118 (16)
New York	519 (9)	11 (3)	40 (5)
North Carolina	533 (9)	37 (10)	49 (7)
Texas	884 (15)	62 (17)	105 (14)
Utah	598 (10)	30 (8)	75 (10)
Prepregnancy BMI			
Underweight	309 (5)	23 (6)	24 (3)

Normal weight	3,037 (50)	177 (48)	315 (43)
Overweight	1,400 (23)	78 (21)	171 (23)
Obese	1,077 (18)	69 (19)	183 (25)
Missing	207 (3)	23 (6)	45 (6)
Household income			
<$50,000	4,264 (71)	272 (74)	562 (76)
≥$50,000	1,766 (29)	98 (26)	176 (24)
Maternal race/ethnicity			
Non-Hispanic white	3,469 (57)	184 (50)	399 (54)
Non-Hispanic black	701 (12)	25 (7)	61 (8)
Hispanic	1,466 (24)	133 (36)	235 (32)
Other	394 (7)	28 (8)	43 (6)
Age at delivery			
<20	610 (10)	41 (11)	72 (10)
20-24	1,532 (25)	86 (23)	170 (23)
25-29	1,760 (29)	115 (31)	250 (34)
30-34	1,380 (23)	84 (23)	156 (21)
≥35	748 (12)	44 (12)	90 (12)
Folic acid supplementation			
Yes	3,016 (50)	191 (52)	375 (51)
No	3,014 (50)	179 (48)	363 (49)

Abbreviations: BMI, body mass index; n, number of mothers with characteristic; N, total number of mothers.

[a] Percentages might not add exactly to 100 because of rounding.

[b] n = 1 with unknown outcome categorized as termination of pregnancy for the analyses.

Table 7.2. Confounding-Adjusted[a] Associations Between Prepregnancy Obesity and Anencephaly and Spina Bifida in Study Sites Including and Excluding Cases Among Terminations of Pregnancy, National Birth Defects Prevention Study, 1997-2007.

	Underweight		Normal Weight	Overweight		Obese	
	OR	95% CI	OR	OR	95% CI	OR	95% CI
Anencephaly							
All study sites	1.27	0.83, 1.96	1.00 (Referent)	1.02	0.78, 1.32	1.18	0.89, 1.57
Sites including TOP[b]	1.26	0.80, 1.98	1.00 (Referent)	0.95	0.72, 1.25	1.11	0.83, 1.49
Sites excluding TOP[c]	1.60	0.35, 7.24	1.00 (Referent)	1.55	0.62, 3.89	1.34	0.43, 4.19
Spina bifida							
All study sites	0.76	0.51, 1.13	1.00 (Referent)	1.23	1.02, 1.48	1.78	1.48, 2.14
Sites including TOP[b]	0.73	0.47, 1.12	1.00 (Referent)	1.16	0.95, 1.42	1.62	1.33, 1.97
Sites excluding TOP[c]	0.93	0.32, 2.65	1.00 (Referent)	1.53	0.92, 2.54	2.84	1.73, 4.68

Abbreviations: CI, confidence interval; OR, odds ratio; TOP, terminations of pregnancy.

[a] Adjusted for maternal race/ethnicity and household income.

[b] Arkansas, California, Georgia, Iowa, New York, North Carolina, Texas, Utah.

[c] Massachusetts, New Jersey.

Table 7.3. Associations Between Prepregnancy Body Mass Index and Anencephaly, Adjusted for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.

	Underweight		Normal Weight	Overweight		Obese	
	OR	95% SI[a]	OR	OR	95% SI[a]	OR	95% SI[a]
Unadjusted[b]	1.28	0.82, 2.00	1.00 (Referent)	0.96	0.73, 1.26	1.10	0.83, 1.46
Confounding only[b,c]	1.26	0.80, 1.98	1.00 (Referent)	0.95	0.72, 1.25	1.11	0.83, 1.49
Selection bias only							
Selection 1	1.35	0.86, 2.20	1.00 (Referent)	0.99	0.76, 1.27	1.10	0.81, 1.49
Selection 2	1.39	0.90, 2.26	1.00 (Referent)	1.07	0.81, 1.37	1.06	0.78, 1.44
Misclassification only[d]							
Nondifferential	1.09	0.66, 1.69	1.00 (Referent)	1.01	0.76, 1.34	1.08	0.81, 1.43
Differential 1	1.07	0.64, 1.65	1.00 (Referent)	0.98	0.73, 1.29	1.06	0.80, 1.41
Differential 2	1.12	0.68, 1.73	1.00 (Referent)	1.06	0.79, 1.40	1.10	0.83, 1.46
Multiple biases[e]							
Selection 1							

Nondifferential	1.16	0.73, 1.82	1.00 (Referent)	1.00	0.77, 1.34	1.09	0.81, 1.48
Differential 1	1.14	0.72, 1.77	1.00 (Referent)	0.97	0.74, 1.29	1.07	0.79, 1.45
Differential 2	1.19	0.75, 1.86	1.00 (Referent)	1.04	0.80, 1.40	1.11	0.83, 1.51
Selection 2							
Nondifferential	1.18	0.75, 1.85	1.00 (Referent)	1.04	0.79, 1.39	1.07	0.80, 1.46
Differential 1	1.16	0.73, 1.81	1.00 (Referent)	1.00	(0.76, 1.34)	1.05	0.78, 1.43
Differential 2	1.21	0.76, 1.90	1.00 (Referent)	1.08	(0.82, 1.44)	1.09	0.81, 1.49

Abbreviations: OR, odds ratio; SI, simulation interval.

[a] 95% simulation interval with random error incorporated.

[b] 95% confidence interval.

[c] Adjusted for maternal race/ethnicity and household income.

[d] Adjusted for exposure misclassification and missing data.

[e] Adjusted for exposure misclassification, missing data, selection bias, and confounding by maternal race/ethnicity and household income.

Table 7.4. Associations Between Prepregnancy Body Mass Index and Spina Bifida, Adjusted for Different Combinations of Biases, National Birth Defects Prevention Study, 1997-2007.

	Underweight		Normal Weight	Overweight		Obese	
	OR	95% SI[a]	OR	OR	95% SI[a]	OR	95% SI[a]
Unadjusted[b]	0.75	0.49, 1.15	1.00 (Referent)	1.18	0.97, 1.43	1.64	1.35, 1.99
Confounding only[b,c]	0.73	0.47, 1.12	1.00 (Referent)	1.16	0.95, 1.42	1.62	1.33, 1.97
Selection bias only							
Selection 1	0.75	0.50, 1.19	1.00 (Referent)	1.11	0.91, 1.34	1.54	1.26, 1.88
Selection 2	0.99	0.65, 1.57	1.00 (Referent)	1.04	0.85, 1.25	1.36	1.10, 1.66
Misclassification only[d]							
Nondifferential	0.85	0.52, 1.29	1.00 (Referent)	1.20	0.96, 1.48	1.54	1.27, 1.88
Differential 1	0.81	0.50, 1.23	1.00 (Referent)	1.15	0.93, 1.43	1.52	1.25, 1.85
Differential 2	0.88	0.54, 1.34	1.00 (Referent)	1.24	1.00, 1.54	1.57	1.29, 1.91
Multiple biases[e]							
Selection 1							

Nondifferential	0.84	0.54, 1.31	1.00 (Referent)	1.12	0.91, 1.39	1.44	1.18, 1.79
Differential 1	0.81	0.52, 1.26	1.00 (Referent)	1.08	0.88, 1.34	1.41	1.15, 1.75
Differential 2	0.88	0.56, 1.37	1.00 (Referent)	1.17	0.95, 1.45	1.46	1.20, 1.82
Selection 2							
Nondifferential	1.03	0.64, 1.59	1.00 (Referent)	1.07	0.86, 1.32	1.29	1.05, 1.60
Differential 1	0.99	0.62, 1.54	1.00 (Referent)	1.03	0.83, 1.27	1.26	1.03, 1.57
Differential 2	1.06	0.67, 1.65	1.00 (Referent)	1.11	0.90, 1.38	1.31	1.07, 1.63

Abbreviations: OR, odds ratio; SI, simulation interval.

[a] 95% simulation interval with random error incorporated.

[b] 95% confidence interval.

[c] Adjusted for maternal race/ethnicity and household income.

[d] Adjusted for exposure misclassification and missing data.

[e] Adjusted for exposure misclassification, missing data, selection bias, and confounding by maternal race/ethnicity and household income.

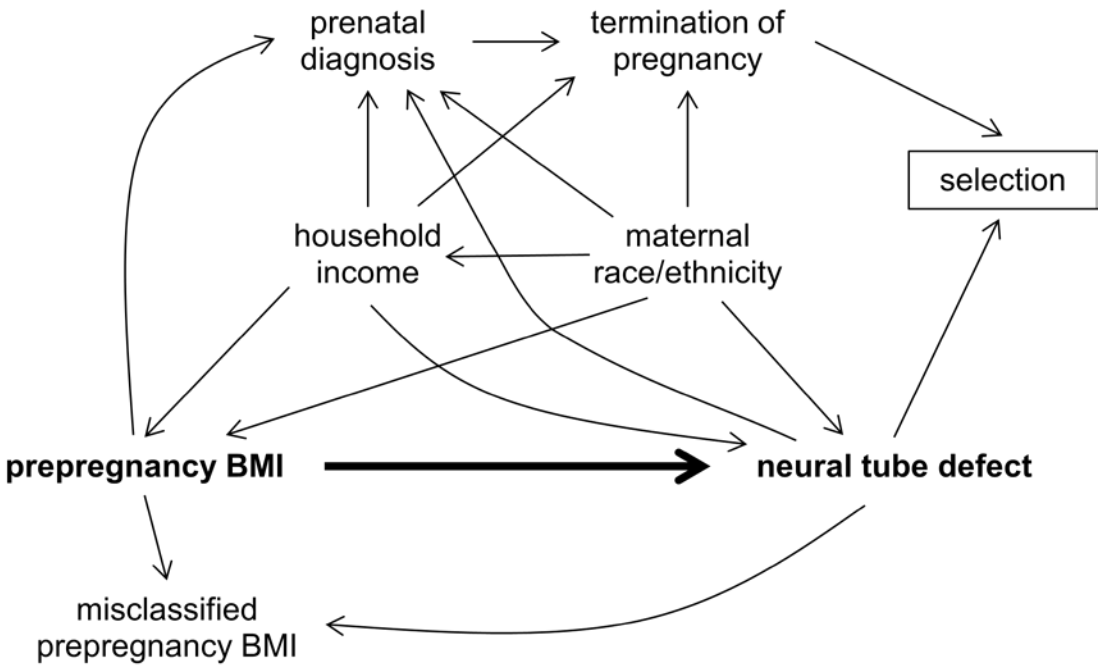

Figure 7.1. Directed acyclic graph illustrating assumed relationships between variables in a study of prepregnancy body mass index and neural tube defects. Potential sources of bias include exposure misclassification (differential, if affected by neural tube defect), selection bias (if the analysis is conditioned on selection, a collider), and confounding (household income and maternal race/ethnicity). Abbreviations: BMI, body mass index.

APPENDIX A

Selection Bias Parameters

In this Appendix, we describe how we created parameters to adjust for selection bias.

We suspected there were differences in ascertainment of stillbirths and terminations of pregnancy between study sites because of differences in case finding methods used by each site. We therefore stratified the cases by outcome and by study site. We assumed that at each site ascertainment of cases among livebirths was 100% but ascertainment of stillbirths and terminations of pregnancy were <100%.

For each site, we calculated the observed ratio of stillbirths to livebirths (SB:LB) among cases and the observed ratio of terminations of pregnancy to livebirths (TOP:LB) among cases (Table 7.A1). For example, for anencephaly, the TOP:LB ratio was 1.00 for New York and 3.50 for Iowa, suggesting that there were more terminations of pregnancy for anencephaly in Iowa than New York or that Iowa had more complete case ascertainment than New York among terminations of pregnancy.

We created log-triangular distributions for SB:LB and TOP:LB to use in a probabilistic analysis (log-triangular instead of triangular because they are ratio measures). We investigated two assumptions about underascertainment in our analysis. First, we assumed that the SB:LB and TOP:LB ratios were truly different by study site. This might be because of differences in sociodemographics, access to or uptake of prenatal screening and prenatal care, or laws governing termination of pregnancy by site. Second, we

assumed that SB:LB and TOP:LB ratios were truly the same between study sites and any differences were due to differences in ascertainment between sites.

For assumption 1, the minimum of the distributions were set to ln(observed SB:LB) and ln(observed TOP:LB), using the study site-specific ratios, because we assumed that the observed values were underestimates (i.e., the true value had to be larger than the observed values). For assumption 2, we chose the highest values of SB:LB and TOP:LB among all the sites. For example, for anencephaly, the minimum of the SB:LB distribution was ln(2.00) (largest value was from North Carolina) and the minimum of the TOP:LB distribution was ln(3.50) (largest value was from Iowa). For both assumptions, the mode of the distributions were set to ln(SB:LB) + 0.25 and ln(TOP:LB) + 0.50 and the maximum of the distributions were ln(SB:LB) + 0.50 and ln(TOP:LB) + 1.00 to create symmetric distributions.

In the probabilistic analysis, after a value was randomly drawn from the SB:LB or TOP:LB log-triangular distribution, it was exponentiated (to undo the log transformation) and then divided by the observed SB:LB or TOP:LB ratio. For example, if a value of ln(4) was drawn from the anencephaly TOP:LB distribution for Iowa, the weight was calculated as 4/3.5 = 1.14. This weight represents that we believe there are 14% more cases among TOP than were included in the study. Weights for stillbirths and terminations of pregnancy were calculated in this way, and live births were assigned weights of 1 for the analysis.

Table 7.A1. Ratio of Stillbirths and Terminations of Pregnancy to Livebirths by Study Site, National Birth Defects Prevention Study, 1997-2007.

	Total N	LB	SB	TOP	SB:LB[a]	TOP:LB[b]
Anencephaly						
Arkansas	51	15	18	18[c]	1.20	1.20
California	92	28	20	44	0.71	1.57
Georgia	45	9	13	23	1.44	2.56
Iowa	42	8	6	28	0.75	3.50
New York	11	5	1	5	0.20	1.00
North Carolina	37	6	12	19	2.00	3.17
Texas	62	21	17	24	0.81	1.14
Utah	30	10	7	13	0.70	1.30
Total	370	102	94	174[c]	0.92	1.71
Spina bifida						
Arkansas	101	85	5	11	0.06	0.13
California	145	124	6	15	0.05	0.12
Georgia	105	83	5	17[c]	0.06	0.20
Iowa	118	96	1	21	0.01	0.22
New York	40	37	0	3	0	0.08
North Carolina	49	43	1	5	0.02	0.12
Texas	105	95	3	7	0.03	0.07
Utah	75	73	0	2	0	0.03
Total	738	636	21	81[c]	0.03	0.13

Abbreviations: LB, live birth; N, number of cases; SB, stillbirth; TOP, termination of pregnancy.

[a] Number of stillbirths for every live birth.

[b] Number of terminations of pregnancy for every live birth.

[c] n = 1 with missing pregnancy outcome categorized as TOP

APPENDIX B

Triangular Distributions for Misclassification

In this Appendix we provide hypothetical data to explain, by way of example, the method by which triangular distributions were created for predictive values to adjust for exposure misclassification. In the analysis, the predictive value distributions were created conditional on household income and maternal race/ethnicity.

Table 7.B1 shows a hypothetical classification table with predictive values corresponding to different combinations of self-reported and assigned BMI categories. The probabilities in each row sum to 1. We assumed that any probability of 0 in the classification table is fixed (i.e., is exactly 0 and cannot vary). This allows us to restrict the values in certain cells of the table. For example, participants reporting as obese could only be assigned to be obese in the analysis, and participants reporting as underweight could only be assigned to be underweight or normal weight in the analysis; the sum of the probabilities in these restricted categories must sum to 1.

For categories in which only one value was possible (e.g., self-reporting as obese, in our example), we fixed the predictive value at 1. For all other categories, we began by creating a triangular distribution around the predictive values in the concordant cells (i.e., self-reported underweight/assigned underweight, self-reported obese/assigned obese, etc.). For the “missing” self-reported BMI category, we began with the self-reported missing/assigned obese cell. The triangular distributions were given the predictive value

in the classification table as the mode and had maxima and minima +/- 0.05 from the mode.

We randomly selected a value from the triangular distribution to be the new predictive value for that cell. We then moved to a neighboring cell in the same row to create the next triangular distribution. If this cell was the only remaining non-zero cell in the row, its value was assigned to be the balance of the probability. For example, in Table 7.B1, once a predictive value for the self-reported underweight/assigned underweight cell was chosen, the value for the self-reported underweight/assigned normal weight cell had to be 1 - (predictive value for self-reported underweight/assigned underweight). If there were at least 2 more non-zero cells in the row, a triangular distribution was created around one of these cells.

To maintain the relationships between cells (e.g., for the self-reported normal weight row, there is higher probability of being assigned overweight than underweight), the predictive value of the cell was restricted to a certain range. In Table 7.B1, suppose a value of 0.75 was chosen from the triangular distribution for the self-reported normal weight/assigned normal weight cell. The remaining probability for the row is 0.25. The assigned overweight cell should have most of the probability (0.12/0.25 = 48%), the underweight next (0.08/0.25 = 32%), and obese the least (0.05/0.25 = 20%). The triangular distribution for the assigned overweight cell was created with a mode of 0.48, and minima and maxima +/- 0.05 of the mode. Once a value was drawn, it was transformed back to predictive values by multiplying by the remaining probability for the row (0.25, in this

example). Predictive values were created for the remainder of the row in the same manner, either fixing values when there was only one non-zero cell remaining, or creating triangular distributions restricted to the balance of the probability remaining when there was more than one non-zero cell.

Table 7.B1. Hypothetical Data Showing an Example of a Classification Table for Body Mass Index Categories.

	Assigned BMI Category			
Self-Reported BMI Category	Underweight	Normal Weight	Overweight	Obese
Underweight	0.80	0.20	0	0
Normal weight	0.08	0.75	0.12	0.05
Overweight	0	0.01	0.89	0.10
Obese	0	0	0	1
Missing	0.10	0.20	0.30	0.40

Abbreviation: BMI, body mass index.

REFERENCES

1. Moore KL, Persaud TVN. *The Developing Human: Clinically Oriented Embryology, 8th edition*. Philadelphia: Saunders; 2008.
2. Obeidi N, Russell N, Higgins JR, et al. The natural history of anencephaly. *Prenat Diagn*. 2010;30:357-360.
3. Botto LD, Moore CA, Khoury MJ, et al. Neural-tube defects. *New Engl J Med*. 1999;341(20):1509-1519.
4. Wong LY, Paulozzi LJ. Survival of infants with spina bifida: a population study, 1979-94. *Paediatr Perinat Epidemiol*. 2001;15(4):374-378.
5. Thompson DN. Postnatal management and outcome for neural tube defects including spina bifida and encephalocoeles. *Prenat Diagn*. 2009;29(4):412-419.
6. Werler MM, Shapiro S, Mitchell AA. Periconceptional folic acid exposure and risk of occurrent neural tube defects. *JAMA*. 1993;269(10):1257-1261.
7. Czeizel AE, Dudas E. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med*. 1992;327(26):1832-1835.
8. Milunsky A, Jick H, Jick SS, et al. Multivitamin/folic acid supplementation in early pregnancy reduces the prevalence of neural tube defects. *JAMA*. 1989;262(20):2847-2852.
9. Mulinare J, Cordero JF, Erickson JD, et al. Periconceptional use of multivitamins and the occurrence of neural tube defects. *JAMA*. 1988;260(21):3141-3145.
10. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet*. 1991;338(8760):131-137.

11. Rasmussen SA, Chu SY, Kim SY, et al. Maternal obesity and risk of neural tube defects: a metaanalysis. *Am J Obstet Gynecol*. 2008;198(6):611-619.

12. Stothard KJ, Tennant PW, Bell R, et al. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. *JAMA*. 2009;301(6):636-650.

13. Shaw GM, Carmichael SL. Prepregnant obesity and risks of selected birth defects in offspring. *Epidemiology*. 2008;19(4):616-620.

14. Oddy WH, De Klerk NH, Miller M, et al. Association of maternal pre-pregnancy weight with birth defects: evidence from a case-control study in Western Australia. *Aust N Z Obstet Gynaecol*. 2009;49(1):11-15.

15. Rankin J, Tennant PW, Stothard KJ, et al. Maternal body mass index and congenital anomaly risk: a cohort study. *Int J Obes (Lond)*. 2010;34(9):1371-1380.

16. Carmichael SL, Rasmussen SA, Shaw GM. Prepregnancy obesity: a complex risk factor for selected birth defects. *Birth Defects Res A Clin Mol Teratol*. 2010;88(10):804-810.

17. Waller DK, Shaw GM, Rasmussen SA, et al. Prepregnancy obesity as a risk factor for structural birth defects. *Archives of Pediatrics and Adolescent Medicine*. 2007;161(8):745-750.

18. Peek M, Nanan R. Comment on: Association of maternal pre-pregnancy weight with birth defects: evidence from a case-control study in Western Australia. *Aust N Z J Obstet Gynaecol*. 2009;49:564-565.

19. Cragan JD, Khoury MJ. Effect of prenatal diagnosis on epidemiologic studies of birth defects. *Epidemiology*. 2000;11(6):695-699.

20. 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(5):338-345.

21. Stommel M, Schoenborn CA. Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001-2006. *BMC Public Health*. 2009;2009:421.

22. 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.

23. Rothman KJ. BMI-related errors in the measurement of obesity. *Int J Obes (Lond)*. 2008;32(Suppl 3):S56-S59.

24. Waller DK, Mills JL, Simpson JL, et al. Are obese women at higher risk for producing malformed offspring? *American Journal of Obstetrics and Gynecology*. 1994;170(2):541-548.

25. Shaw GM, Velie EM, Schaffer D. Risk of neural tube defect-affected pregnancies among obese women. *Journal of the American Medical Association*. 1996;275(14):1093-1096.

26. Watkins ML, Scanlon KS, Mulinare J, et al. Is maternal obesity a risk factor for anencephaly and spina bifida? *Epidemiology*. 1996;7(5):507-512.

27. Watkins ML, Rasmussen SA, Honein MA, et al. Maternal obesity and risk for birth defects. *Pediatrics*. 2003;111(5 II):1152-1158.

28. Werler MM, Louik C, Shapiro S, et al. Prepregnant weight in relation to risk of neural tube defects. *JAMA*. 1996;275(14):1089-1092.

29. Moore LL, Singer MR, Bradlee ML, et al. A prospective study of the risk of congenital defects associated with maternal obesity and diabetes mellitus. *Epidemiology*. 2000;11(6):689-694.

30. Shaw GM, Todoroff K, Schaffer DM, et al. Maternal height and prepregnancy body mass index as risk factors for selected congenital anomalies. *Paediatric and Perinatal Epidemiology*. 2000;14(3):234-239.

31. Hendricks KA, Nuno OM, Suarez L, et al. Effects of hyperinsulinemia and obesity on risk of neural tube defects among Mexican Americans. *Epidemiology*. 2001;12(6):630-635.

32. Li Z, Liu J, Ye R, et al. Maternal prepregnancy body mass index and risk of neural tube defects: A population-based case-control study in Shanxi Province, China. *Birth Defects Res A Clin Mol Teratol*. 2010;88(7):570-574.

33. Hendler I, Blackwell SC, Bujold E, et al. Suboptimal second-trimester ultrasonographic visualization of the fetal heart in obese women: should we repeat the examination? *J Ultrasound Med*. 2005;24(9):1205-1209.

34. Hendler I, Blackwell SC, Bujold E, et al. The impact of maternal obesity on midtrimester sonographic visualization of fetal cardiac and craniospinal structures. *Int J Obes*. 2004;28:1607-1611.

35. Dashe JS, McIntire DD, Twickler DM. Maternal obesity limits the ultrasound evaluation of fetal anatomy. *J Ultrasound Med*. 2009;28(8):1025-1030.

36. Dashe JS, McIntire DD, Twickler DM. Effect of maternal obesity on the ultrasound detection of anomalous fetuses. *Obstet Gynecol*. 2009;113:1001-1007.

37. Johnson CY, Honein MA, Hobbs CA, et al. Prenatal diagnosis of orofacial clefts, National Birth Defects Prevention Study, 1998-2004. *Prenat Diagn*. 2009;29:833-839.

38. Mansfield C, Hopfer S, Marteau TM, et al. Termination rates after prenatal diagnosis of Down syndrome, spina bifida, anencephaly, and Turner and Klinefelter syndromes: a systematic literature review. *Prenat Diagn*. 1999;19:808-812.

39. Yoon PW, Rasmussen SA, Lynberg MC, et al. The National Birth Defects Prevention Study. *Public Health Rep*. 2001;116(Suppl 1):32-40.

40. Rasmussen SA, Olney RS, Holmes LB, et al. Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Defects Res A Clin Mol Teratol*. 2003;67(3):193-201.

41. Lyles RH, Lin J. Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. *Stat Med*. 2010;29(22):2297-2309.

42. Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data*. New York: Springer; 2009.

43. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615-625.

44. Rasmussen SA, Chu SY, Kim SY, et al. Maternal obesity and risk of neural tube defects: a metaanalysis. *Am J Obst Gynecol*. 2008;198(6):611-619.

45. Stothard KJ, Tennant PWG, Bell R, et al. Maternal overweight and obesity and the risk of congenital anomalies: A systematic review and meta-analysis. *JAMA*. 2009;301(6):636-650.

46. Besser LM, Williams LJ, Cragan JD. Interpreting changes in the epidemiology of anencephaly and spina bifida following folic acid fortification of the U.S. grain supply in the setting of long-term trends, Atlanta, Georgia, 1968-2003. *Birth Defects Res A Clin Mol Teratol*. 2007;79(11):730-736.

47. Nohr EA, Frydenberg M, Henriksen TB, et al. Does low participation in cohort studies induce bias? *Epidemiology*. 2006;17(4):413-148.

48. Parks SE, Canfield MA, Ramadhani TA. Importance of including all pregnancy outcomes to reduce bias in epidemiologic studies of neural tube defects -- Texas, 1999 to 2005. *Birth Defects Res A Clin Mol Teratol*. 2011;91(3):185-191.

49. Agopian AJ, Canfield MA, Olney RS, et al. Spina bifida subtypes and sub-phenotypes by maternal race/ethnicity in the National Birth Defects Prevention Study. *Am J Med Genet A*. 2011;epub Dec 2.

50. Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes (Lond)*. 2008;32(Suppl 3):S8-S14.

51. Hernan MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology*. 2011;22(3):368-377.

CHAPTER 8

Discussion and Conclusions

SUMMARY OF STUDIES

In this dissertation, we investigated the potential contributions of exposure misclassification, selection bias, and confounding to results of a case-control study of prepregnancy body mass index (BMI) and two types of neural tube defects (NTDs), anencephaly and spina bifida. We additionally conducted several studies to better understand the potential for these types of bias to occur in the study and to demonstrate how quantitative analyses could be used to investigate the effects of these biases on study results.

Reviewing the literature (Chapter 2), we found that epidemiologic studies conducted during the past 20 years fairly consistently reported associations between prepregnancy BMI and NTDs, with stronger associations observed between obesity and spina bifida than between obesity and anencephaly. Adjustment for potential confounders in these studies had no appreciable effect on results, and one study incorporating adjustment for selection bias using a simple bias analysis found little effect of this bias on the magnitude of the association (1). No study investigated the potential effects of exposure misclassification and none provided an analysis simultaneously incorporating multiple biases to determine the combined effects of these biases.

In studies of prepregnancy BMI and NTDs, selection bias is thought to arise because of differential prenatal diagnosis and subsequent pregnancy termination according to BMI. To better understand if this bias was occurring and if so, to what extent, we first investigated the frequency of termination of pregnancy following prenatal diagnosis of NTDs (Chapter 3). Reviewing the published literature, we found that after prenatal diagnosis of anencephaly, 86% of pregnancies ended in termination; for spina bifida, the frequency was 64%. There was limited information available on factors associated with prenatal diagnosis, but we found that termination was more common in Europe than in North America, for more severe defects, and when the prenatal diagnosis was made before 24 gestational weeks. Although termination of pregnancy following prenatal diagnosis was common for NTDs, there was substantial geographic variability in its occurrence, the reasons for which were not clear from information available in the literature.

In a second study to better understand the potential mechanism for selection bias, we used data from the National Birth Defects Prevention Study (NBDPS) to identify factors associated with prenatal diagnosis (prenatal diagnosis vs. postnatal diagnosis) and timing of prenatal diagnosis (<24 vs. ≥24 gestational weeks) of spina bifida (Chapter 4). Prenatal diagnosis of spina bifida was more likely among mothers who reported having maternal alpha fetoprotein screening during pregnancy than among mothers who reported no screening, and among mothers reporting folic acid supplementation periconceptionally than among mothers not taking supplements periconceptionally or not at all. Socioeconomic status was also associated with prenatal diagnosis, with mothers without a

high school education less likely to report a prenatal diagnosis than mothers with at least a high school education; prenatal diagnosis was also less common among mothers who reported a household income <$50,000 than among mothers with a higher household income. Early prenatal diagnosis was associated with the same variables, in the same direction, and with similar magnitudes of association. The association between prepregnancy BMI and prenatal diagnosis was a specific interest because of its hypothesized role in selection bias in studies of prepregnancy BMI and NTDs. In our analyses, underweight women were more likely than normal weight, overweight, and obese mothers to have a prenatal diagnosis and an early prenatal diagnosis. For normal weight, overweight, and obese mothers, increasing BMI did not correspond with decreasing likelihood of prenatal diagnosis or early prenatal diagnosis, as had been observed in studies of other types of birth defects (2, 3). From this analysis, we concluded that selection bias operating through differential likelihood of prenatal diagnosis by BMI might be of a smaller magnitude than hypothesized based on results from studies of other types of birth defects.

In Chapter 5, we investigated how making an incorrect assumption about nondifferential or differential misclassification could affect the results of a bias analysis. Using simulations, we showed that when making an inaccurate assumption, the odds ratio tended to be less biased when misclassification was truly nondifferential than when misclassification was truly differential.

We then proposed a method to simultaneously adjust for exposure misclassification, selection bias, and confounding using weighted logistic regression (Chapter 6), an approach based on existing methods that use weights to rebalance the joint distribution of exposure and disease to what the distribution would have been in the absence of bias. We showed algebraically that when weights are estimated correctly (i.e., the true values of the weights are known) this method exactly re-creates the odds ratio from the source population (the unbiased odds ratio) from a dataset with bias present.

We used information from Chapters 2-6 to adjust for multiple biases in a study of prepregnancy obesity and anencephaly and spina bifida (Chapter 7). Given our assumptions, we found little effect of bias adjustment on the association between obesity and anencephaly, with the crude and multiple bias-adjusted odds ratios similar in magnitude (odds ratio approximately 1.1). The multiple bias-adjusted association between obesity and spina bifida was attenuated compared to the crude or confounding-adjusted associations, but not null, given our assumptions (odds ratio reduced from approximately 1.6 to 1.3). Although an association between obesity and spina bifida persisted after adjustment for multiple biases, this association could indicate the existence of a causal association, additional sources of bias we did not consider in the analysis, or incorrect choices of bias parameters. Despite the information developed in this dissertation, little information was available about the potential sources of bias and their magnitude in NBDPS data, making evidence-based bias adjustment a challenge.

FUTURE DIRECTIONS

With over a dozen studies consistently reporting crude and confounding-adjusted associations between BMI and NTDs, there is little need for additional studies to be conducted to confirm the existence of this association. There remain some aspects of the association that warrant further attention.

First, a recent study from China found an inverse association between obesity and spina bifida (4), contrary to the results of most previous studies; however, most of studies to date have been conducted in the United States or in predominantly white populations. This raises the possibility that the magnitude or direction of the association could differ by geographic region or race/ethnicity, a hypothesis that has not been formally investigated to date. Further examination of potential differences by race/ethnicity within the United States could be undertaken in studies such as NBDPS, which have a large sample size and include a diverse population of mothers. Valuable information could be obtained from studies of obesity and NTDs in regions where termination of pregnancy is not an available option at any gestational age. In these regions, selection bias operating through prenatal diagnosis and termination of pregnancy would not occur. However, if a substantial proportion of mothers pursued termination of pregnancy by other means, such as by traveling to regions where it was an available option, this selection bias might not be completely avoided.

Second, biological mechanisms explaining why prepregnancy BMI might be associated with NTDs remain unknown (5). If BMI is hypothesized to be a proxy for a complex

exposure such as diet quality or composition, collecting information that would allow better characterization of periconceptional diet could be an important future direction for research. If BMI is thought to be a proxy for adiposity, studies providing validation data could be useful for use in bias analyses. Because obesity is a risk factor for diabetes and diabetes is a risk factor for birth defects, further attention to methods for accurately measuring prepregnancy diabetes are needed, given that a large proportion of women with diabetes remain undiagnosed (5, 6). Additional studies providing evidence in support of any of these mechanisms would be important next steps.

Finally, to determine if a non-biological mechanism (e.g., bias) is a likely explanation for the association between prepregnancy BMI and NTDs, there is a need for further studies characterizing the types of bias that might be occurring in these studies. For adjustment of exposure misclassification, validation data comparing maternally-reported prepregnancy weight to measured prepregnancy weight are needed. Currently, validation studies available in the literature report agreement between self-reported and measured *current* weight, which might not be generalizable to *prepregnancy* weight. To determine whether misclassification is differential or nondifferential, validation studies will need to be conducted separately for cases and non-cases. However, it could prove challenging to identify a population from which to conduct these validation studies. Using a cohort study of women trying to achieve pregnancy might not be generalizable to all women because of the substantial proportion of pregnancies that are unplanned. Weight recorded at the first prenatal care visit might approximate prepregnancy weight if the first visit is early in pregnancy, but this would also potentially be non-generalizable to all women

because of differences between women who seek care early in pregnancy and those who initiate prenatal care in the second and third trimesters.

For adjustment of selection bias, it would be most valuable to know the extent of underascertainment of cases among terminations of pregnancies; with an estimate of the extent of underascertainment, we could adjust for selection bias. In the future, we will pursue development of an approach to estimate the proportion of cases among termination of pregnancy that are missed.

IMPLICATIONS FOR PUBLIC HEALTH

Existing guidelines from the American College of Obstetricians and Gynecologists recommend that women achieve a healthy weight prior to pregnancy (7). Whether or not an association exists between obesity and NTDs, this recommendation is appropriate for not only potentially reducing the risk of having a child with an NTD (if there is a causal relationship) but for reducing the risk of other adverse reproductive and long-term health outcomes. However, specific recommendations that obese mothers should be counseled on their increased risk for having a child with an NTD and that they should be targeted for screening might not be warranted if there is truly no association or only a weak association, as this could unnecessarily elevate maternal anxiety during pregnancy.

Multiple bias analysis becomes particularly important when biases are of a large enough magnitude to affect interpretation of existing studies and prompt reassessment of decisions or recommendations that have been made based on the available literature.

However, it is difficult to determine when this is the case because there is seldom enough information available to identify biases which might have important effects and to confirm that these hypothesized effects are indeed occurring as suspected. Until such information is available, bias analysis remains useful for other reasons, such as providing quantitative assessments of how large potential biases might be under different assumptions.

CONCLUSIONS

From the results of this dissertation, we conclude that, given our assumptions, exposure misclassification, selection bias, and confounding might not fully explain the association between obesity and NTDs. Although selection bias has previously been proposed as an explanation for the observed association, we did not find a substantial difference in the proportion of obese and non-obese mothers reporting prenatal diagnosis (Chapter 4) and did not observe a large difference between the crude and selection bias-adjusted odds ratio (Chapter 7), suggesting that selection bias arising by way of this mechanism might not be great enough to substantially affect the magnitude of the association. From our analyses, it also appeared that exposure misclassification and confounding are unlikely to generate a large enough bias to account for the observed association. In future studies, we will continue to characterize potential sources of exposure misclassification and selection bias hypothesized to occur in studies of prepregnancy BMI and NTDs to better understand the effects of these biases on this association.

REFERENCES

1. Waller DK, Shaw GM, Rasmussen SA, et al. Prepregnancy obesity as a risk factor for structural birth defects. *Arch Pediatr Adolesc Med*. 2007;161(8):745-750.
2. Dashe JS, McIntire DD, Twickler DM. Effect of maternal obesity on the ultrasound detection of anomalous fetuses. *Obstet Gynecol*. 2009;113:1001-1007.
3. Johnson CY, Honein MA, Hobbs CA, et al. Prenatal diagnosis of orofacial clefts, National Birth Defects Prevention Study, 1998-2004. *Prenat Diagn*. 2009;29:833-839.
4. Li Z, Liu J, Ye R, et al. Maternal prepregnancy body mass index and risk of neural tube defects: A population-based case-control study in Shanxi Province, China. *Birth Defects Res A Clin Mol Teratol*. 2010;88(7):570-574.
5. Carmichael SL, Rasmussen SA, Shaw GM. Prepregnancy obesity: a complex risk factor for selected birth defects. *Birth Defects Res A Clin Mol Teratol*. 2010;88:804-810.
6. Rasmussen SA, Galuska DA. Prepregnancy obesity and birth defects: what's next? *Am J Clin Nutr*. 2010;91:1539-1540.
7. American College of Obstetricians and Gynecologists. ACOG Committee Opinion No. 315: Obesity in pregnancy. *Obstet Gynecol*. 2005;106:671-675.