HHS Public Access

Author manuscript

Ann Epidemiol. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

Ann Epidemiol. 2018 August ; 28(8): 510–514. doi:10.1016/j.annepidem.2018.05.009.

Multiple bias analysis using logistic regression: an example from the National Birth Defects Prevention Study

Candice Y. Johnson[a,b,*], **Penelope P. Howards**[a], **Matthew J. Strickland**[c], **D. Kim Waller**[d], **W. Dana Flanders**[a], and **The National Birth Defects Prevention Study**

[a] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

[b] National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, USA

[c] School of Community Health Sciences, University of Nevada Reno, Reno, NV, USA

[d] The University of Texas School of Public Health, Houston, TX, USA

Abstract

Purpose.—Exposure misclassification, selection bias, and confounding are important biases in epidemiologic studies, yet only confounding is routinely addressed quantitatively. We describe how to combine two previously described methods and adjust for multiple biases using logistic regression.

Methods.—Weights were created from selection probabilities and predictive values for exposure classification and applied to multivariable logistic regression models in a case-control study of prepregnancy obesity (body mass index ≥30 versus <30 kg/m^2) and cleft lip with or without cleft palate (CL/P) using data from the National Birth Defects Prevention Study (2,523 cases, 10,605 controls).

Results.—Adjusting for confounding by race/ethnicity, prepregnancy obesity and CL/P were weakly associated (odds ratio 1.10, 95% confidence interval: 0.98, 1.23). After weighting the data to account for exposure misclassification, missing exposure data, selection bias, and confounding, multiple bias-adjusted odds ratios ranged from 0.94 to 1.03 in non-probabilistic bias analyses and median multiple bias-adjusted odds ratios ranged from 0.93 to 1.02 in probabilistic analyses.

Conclusions.—This approach, adjusting for multiple biases using a logistic regression model, suggested that the observed association between obesity and CL/P could be due to the presence of bias.

Keywords

bias; body mass index; cleft lip; regression analysis

[*]**Corresponding author** Candice Johnson, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, 1090 Tusculum Ave MS R-15, Cincinnati OH 45226. Phone: (513) 841-4454. cyjohnson@cdc.gov.

INTRODUCTION

Bias can affect results of epidemiologic studies so that both the direction and magnitude of the observed association can be incorrect.[1] The effects of measurement error (information bias) and selection bias are important, but despite available quantitative bias analysis methods, analyses adjusting for biases other than confounding are rare.[2–7] Bias analyses can be used to determine the likelihood that observed associations are causal, and are particularly useful when policy or interventions are being proposed based on the assumption of causality.[8]

Dozens of studies have found associations between prepregnancy obesity and an increased risk of having a child with a birth defect.[9] Many studies have reported similar, weak associations between prepregnancy obesity and cleft lip with or without cleft palate (CL/P). Three meta-analyses have estimated odds ratios (ORs) of 1.13 (95% confidence interval [CI]: 1.04, 1.23), 1.16 (95% CI: 1.00, 1.34), and 1.20 (95% CI: 1.03, 1.40) for associations between prepregnancy obesity (body mass index (BMI) ≥30 kg/m^2 versus normal weight, 18.5–24.9 kg/m^2) and either CL/P or cleft lip with cleft palate.[9–11]

These associations are small enough that exposure misclassification or selection bias could explain the results. One previous study investigated potential effects of nondifferential exposure misclassification on this association; the OR was 1.25 before accounting for misclassification and ranged from 1.38 to 2.94 after.[12] No other study has attempted to adjust this association for biases other than confounding.

The purpose of this analysis is two-fold. The first is to explore how the association between prepregnancy obesity and CL/P might be affected by exposure misclassification and selection bias. The second is to demonstrate how to combine two previously described methods to adjust for misclassification and selection bias using both non-probabilistic and probabilistic multiple bias analysis.

MATERIALS AND METHODS

Study Population

We used data from the National Birth Defects Prevention Study (NBDPS), a population-based case-control study of birth defects.[13] Cases (live births, still births, terminations of pregnancy) were identified from birth defects surveillance systems in 10 U.S. states (Arkansas, California, Georgia, Iowa, Massachusetts, New Jersey, New York, North Carolina, Texas, Utah) and controls (live births with no major birth defects) from birth certificates or hospital birth records in the same areas. Participating mothers were interviewed 6 weeks to 2 years after their child's birth. Eligible mothers delivered on or after October 1, 1997 with an estimated due date on or before December 31, 2011. Study sites received institutional review board approval and participants provided informed consent.

The outcome of interest was nonsyndromic isolated CL/P; clinical geneticists reviewed medical records to exclude cases possibly caused by genetic or other syndromes.[14] Isolated cases were those occurring in an infant without other major birth defects. The exposure of

interest was prepregnancy BMI, dichotomized as obese (≥30 kg/m^2) or non-obese (<30 kg/m^2). BMI was calculated from prepregnancy weight and height, self-reported during the NBDPS interview.

We included study site, maternal race/ethnicity, and maternal education in the models as potential confounders. Only mothers reporting their race/ethnicity as non-Hispanic white, non-Hispanic black, or Hispanic were included, to correspond with available exposure misclassification validation data (details below).

Among 3,161 CL/P case mothers and 11,692 control mothers, we excluded 382 (12%) case mothers of infants with non-isolated CL/P, 199 (6%) case and 763 (7%) control mothers reporting race/ethnicities not meeting inclusion criteria, 1 (<1%) case and 7 (<1%) control mothers with missing race/ethnicity, and 56 (2%) case and 317 (3%) control mothers with missing data on maternal education. Following exclusions, we included 2,523 case mothers and 10,605 control mothers. Mothers with missing BMI were retained in the analysis so we could account for missing exposure data.

Conventional Analysis

We used logistic regression to estimate crude and confounding-adjusted ORs and 95% CIs for associations between prepregnancy obesity and CL/P. All statistical analyses were conducted in SAS version 9.4 (Cary, NC).

Bias Analysis

In this analysis, we perform both non-probabilistic and probabilistic bias analysis. Probabilistic analyses take into account uncertainty in bias parameter estimates to be taken into account by conducting analyses using a range of values for the bias parameters.[15] For readers wishing more detail on the bias analysis methods, the Supplemental Materials include: (1) a step-by-step worked example in which a non-probabilistic adjustment for multiple biases is conducted by hand, (2) sample SAS code, and (3) an algebraic proof demonstrating the general case. Probabilistic analysis is an extension of the non-probabilistic analysis in the worked example.

Adjustment for Exposure Misclassification and Confounding

We used the method of Lyles and Lin to adjust for exposure misclassification.[16] Predictive values (e.g., positive or negative predictive values) are ideally calculated from cross-tabulations of misclassified and correctly classified exposure categories in a validation dataset. Predictive values represent the probability that the prepregnancy obesity status reported by the participant (obese, not obese) is the true prepregnancy obesity status. If internal validation data are not available, use of external validation data, expert opinion, or educated guesses are other options.[8] If only sensitivity (Se) and specificity (Sp) are available, the formulae included in the Supplemental Materials can be used to convert these to predictive values.

We had no internal validation data on exposure misclassification for NBDPS. We used external validation data from the 1999–2010 National Health and Nutrition Examination

Surveys (NHANES), representative of the civilian noninstitutionalized population of the United States.[17] NHANES participants self-report weight and height during an in-person interview. Later, height and weight are measured during a physical exam. All participants provided informed consent.

We restricted the NHANES analysis to nonpregnant females aged 16–49 years with height and weight measurements. We cross-tabulated self-reported and measured BMI categories conditional on race/ethnicity, accounting for the complex sampling design, to estimate Se and Sp. Although predictive values can be calculated from these data, we estimated Se and Sp to examine nondifferential and differential exposure misclassification (whether or not misclassification is differential depends directly on differences in Se and Sp between cases and controls, not predictive values). Reliable estimates from NHANES were available for non-Hispanic white, non-Hispanic black, and Mexican-American women, and therefore we restricted our NBDPS analysis to these racial/ethnic groups (because approximately two-thirds of Hispanics in the U.S. are of Mexican descent, we used the estimate for Mexican-Americans for all NBDPS Hispanic women).[18]

We assumed that the NHANES Se and Sp were accurate estimates of the Se and Sp in NBDPS. Not knowing if exposure misclassification was differential or nondifferential, we performed three analyses, assuming: (1) nondifferential misclassification, (2) “differential A” misclassification (classification is better for cases than controls), and (3) “differential B” misclassification (classification is better for controls than cases). In the first, we assigned cases and controls to have the same Se and Sp values (NHANES Se and Sp). In the second, we assigned the NHANES Se and Sp to controls and Se + 0.05 and Sp + 0.03 to cases. In the third, we assigned the NHANES Se and Sp to controls and Se – 0.05 and Sp – 0.03 to cases. Se and Sp were restricted to lie between 0.5 and 1.0, inclusive. We converted Se and Sp to predictive values (restricted to lie between 0 and 1, inclusive). Bias parameters were calculated separately for non-Hispanic white, non-Hispanic black, and Hispanic/Mexican-American women. For simplicity, we assumed they did not differ by other variables.

We also used predictive values to account for missing BMI.[16] Among women with missing values for self-reported BMI but measured BMI in NHANES, we determined the probability that a woman with missing BMI was truly obese or not obese; these were the predictive values used in the analysis. Other missing data methods, such as multiple imputation, could also have been used.

For the analysis, we created a dataset with two observations for each participant (participant “copies”): one copy was assigned to have prepregnancy obesity, and the other to not have prepregnancy obesity — these represent the two possible obesity statuses the participant could have had in the absence of exposure misclassification.[16] The assigned prepregnancy obesity status was used as the exposure in the logistic regression models (i.e., not the status reported by the participant). In the model, each participant copy was weighted by the predictive values corresponding to their assigned obesity status (the probability that the exposure assignment was the truth). Potential confounders were included as covariates; any confounder that would be included in a conventional analysis should also be included in the bias analysis. For non-probabilistic analyses, we calculated the OR only. The standard error

from the logistic regression model should not be used to calculate a 95% CI because this does not take into account error introduced by estimating bias parameters.[4]

To conduct probabilistic bias analysis, we assigned triangular distributions to each predictive value using the values calculated above as the mode and +/− 0.10 of the mode as the upper and lower bounds (restricted to fall between 0 and 1). We sampled each parameter 5,000 times and calculated 5,000 ORs. The results were summarized as the median OR and 95% simulation interval (SI), the 2.5th and 97.5th percentile of the OR distribution.

We used the method of Lash et al. to account for random error, but other methods, such as bootstrapping, could also be used.[4] For each bias-adjusted log OR, we multiplied the standard error from the conventional multivariable logistic regression model by a randomly selected value from a standard normal distribution. This value was then subtracted from the log OR and exponentiated. The result is presented as the median OR and 95% random error-added simulation interval (RESI), the 2.5th and 97.5th percentiles of the OR distribution following addition of random error.

Adjustment for Selection Bias and Confounding

We used inverse probability of selection weights (IPSW) to adjust for selection bias.[19] The probability of selection into the study (or participation in the study) is ideally estimated from study records, but if unavailable, external validation data, expert opinion, or educated guesses can be used.[8] The IPSW is the inverse of this probability.

NBDPS participation rates for cases and controls were 67% and 65%.[13] We did not know to what extent these participation rates differed by BMI or to what degree eligible individuals were ascertained. For simplicity, we assumed complete ascertainment of eligible cases and controls. We used external validation data and educated guesses to estimate how participation rates differed by BMI. We found a study showing that mothers self-reporting normal weight were more likely to participate in a pregnancy study than other mothers.[20] (BMI is likely serving as a proxy for sociodemographic differences between women.) We assumed that NBDPS case mothers were motivated to participate regardless of BMI, but control mothers would be more susceptible to sociodemographic determinants of participation. We assigned all case mothers a selection probability of 0.67 (case participation rate), reflecting equal motivation to participate. We assigned selection probabilities of 0.60 for obese control mothers, 0.67 for non-obese control mothers, and 0.65 for control mothers with missing obesity status; the weighted average was 0.65 (control participation rate). For simplicity, we assumed that selection probabilities did not differ by other variables. The IPSW (analysis weights) were the inverse of these probabilities. For non-probabilistic analyses, we estimated the OR using multivariable logistic regression.

For the probabilistic analysis, we assigned triangular distributions to each selection probability, using +/− 0.10 of the mode as the upper and lower bounds, with values restricted to lie between 0 and 1. We selected 5,000 sets of selection probabilities, inverted them to calculate the IPSW, and used these to calculate 5,000 ORs. Results were summarized as median OR and 95% SI. We added random error, producing a median OR and 95% RESI.

Multiple Bias Analysis

To adjust for exposure misclassification, missing exposure data, selection bias, and confounding, we multiplied 5,000 simulated IPSW by 5,000 simulated predictive values to create 5,000 combined weights. Then, as before, to adjust for exposure misclassification we created a dataset with two observations per participant and assigned each observation the combined weight corresponding to the assigned exposure status. The multivariable model regressed assigned exposure (not reported exposure) on the outcome, adjusted for confounders (study site, maternal race/ethnicity, maternal education), and was weighted by the combined weight to estimate the OR. Probabilistic results were summarized as median OR and 95% SI. Random error was added to generate a median OR and 95% RESI.

When biases are adjusted serially in multiple bias analysis, the order of bias adjustment is important; if adjustment is done out of order, incorrect results could be obtained.[4,15] Although we did not adjust biases serially (a single model is used), there is an element of "order" needed to estimate valid parameters.

If we consider exposure misclassification and selection bias, there are four possible datasets: (1) both biases present, (2) selection bias only, (3) exposure misclassification only, and (4) no exposure misclassification or selection bias. Our goal is to move from dataset 1 (two types of bias) to dataset 4 (no bias). This can be done by removing exposure misclassification first (datasets 1 to 2 to 4) or selection bias first (datasets 1 to 3 to 4).

In NBDPS, we removed selection bias first. When estimating IPSWs, we obtained these values from a cohort study, which likely had exposure misclassification and selection bias (dataset 1). Once the IPSW were estimated and applied, this produced dataset 3 (exposure misclassification only). We estimated predictive values from a "dataset 3" (exposure misclassification, no selection bias); this was NHANES. Because NHANES-provided weights accounting for nonresponse and other selection effects, we assumed this represented what NBDPS would have been in absence of selection bias. Once predictive values were estimated and applied, this moved from dataset 3 to 4 (no exposure misclassification or selection bias). Because there was confounding in the underlying source population, we estimated bias parameters conditional on confounders.

RESULTS

The prevalence of prepregnancy obesity was similar between cases and controls (Table 1). There was a lower proportion of non-Hispanic black women and a higher proportion of women of lower educational attainment among cases than controls.

Conventional analysis

The crude OR for the association between prepregnancy obesity and CL/P was 1.09 (95% CI: 0.97, 1.21). After adjusting for study site, maternal race/ethnicity, and maternal education, it was 1.10 (95% CI: 0.98, 1.23). Despite the confidence interval crossing the null, we continued the bias analysis because of evidence in the literature that this weak association might not be due to chance.

Exposure misclassification, missing exposure, and confounding

Using bias parameters from NHANES (Table 2) in the non-probabilistic bias analysis, the adjusted ORs ranged from 1.02 to 1.13 for the three misclassification types (Table 3). In probabilistic analyses, the adjusted median ORs ranged from 1.01 to 1.11.

Selection bias and confounding

In non-probabilistic and probabilistic analyses, the OR adjusted for selection bias and confounding was 0.98 (Table 3).

Multiple bias analysis

In the non-probabilistic multiple bias analyses for exposure misclassification, missing exposure, selection bias, and confounding, the adjusted OR ranged from 0.94 to 1.03 for the three misclassification scenarios. In the probabilistic analyses, it ranged from 0.93 to 1.02 (Table 3).

DISCUSSION

Multiple bias analyses suggest that exposure misclassification and selection bias could account for the weak association between prepregnancy obesity and CL/P, with analyses based on realistic bias parameters compatible with no association. The median multiple bias-adjusted ORs were closer to the null than the confounding-adjusted OR, although the SIs and RESIs spanned values compatible with inverse, positive, or no associations. Selection bias and “differential B” misclassification had the greatest effects in moving the association towards the null.

The ORs for the non-probabilistic and probabilistic analyses were similar, likely because our triangular distributions centered on the bias parameters used in the non-probabilistic analysis. Conducting non-probabilistic bias analyses is a simple way to explore the effects of bias; however, probabilistic analyses might indicate if results are compatible with a wider range of values.

One previous bias analysis investigated the impact of nondifferential exposure misclassification on this association, finding bias towards the null.[12] We explored three misclassification scenarios. Hypotheses about if misclassification is differential or nondifferential are often made in the absence of quantitative evidence, and making the incorrect assumption can cause substantial error in the bias analysis, sometimes even more than choosing inaccurate values of Se and Sp.[21] For prepregnancy obesity and CL/P, misclassification could be differential, because nondifferential error in the measurement of BMI can become differential once the variable is dichotomized.[22]

Combining the methods of Lyles and Lin and Hernán et al. allowed for adjusting combinations of different biases.[16,19] Non-probabilistic analyses using this method do not require advanced statistical programming, but our probabilistic analyses required simulations. These bias analyses are considered semi-Bayesian because distributions are not defined for all model parameters.[4] The analyses could be extended to become fully

Bayesian, although probabilistic analyses are considered approximately Bayesian in some circumstances.[23]

We assumed that our parameters were applicable to the NBDPS study population, but had no evidence to support this. We tested few of many possible parameters, and if our assumptions were incorrect, our results might not reflect the true OR.[21] Although we incorporated uncertainty into probabilistic analyses, our distributions might not have included the true values or been centered around them. Even so, the analyses provided quantitative estimates of the potential direction and magnitude of biases in our results. Justifying our parameter choices during bias analyses made these uncertainties transparent, something not often done in conventional analyses.

We made simplifying assumptions that predictive values varied only by race/ethnicity and case-control status and that selection probabilities varied by exposure and case-control status. Estimating bias parameters conditional on other variables could provide better adjustment for bias, if parameters do vary. However, if bias parameters are estimated from subgroups with small sample size, the random error introduced might bias the analysis, particularly if extreme weights are estimated and cause some participants to carry undue influence.

Additional complexities can be added to our analyses. Outcome misclassification, unmeasured confounding, or covariate misclassification might be integrated using other methods; for example, using the multiple bias-adjusted OR as input for the method of Ding and VanderWeele to examine unmeasured confounders.[24] We did not correlate Se and Sp for cases and controls, which avoids unlikely combinations of values, and used triangular distributions for bias parameters, which are easily implemented but possibly less realistic than other distributions.[8]

Studies of obesity have been criticized when used for causal inference because obesity does not correspond to a well-defined causal question.[25] However, finding that the observed association could be caused by bias might discourage the search for potentially non-existent biologic mechanisms that underlie the association.

The association observed between prepregnancy obesity and CL/P could be attributable to exposure misclassification or selection bias. We encourage epidemiologists to incorporate multiple bias analysis into their research or teaching.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS AND FUNDING

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. The authors wish to thank Regina Simeone and Annette Christianson for assistance with analysis.

This work was supported by the Centers for Disease Control and Prevention (cooperative agreements PA 96043, PA 02081, and FOA DD09–001 to the Centers for Birth Defects Research and Prevention participating in the National Birth Defects Prevention Study).

ABBREVIATIONS AND ACRONYMS

BMI	body mass index
CI	confidence interval
CL/P	cleft lip with or without cleft palate
IPSW	inverse probability of selection weight
NBDPS	National Birth Defects Prevention Study
NHANES	National Health and Nutrition Examination Survey
OR	odds ratio
RESI	random error-added simulation interval
Se	sensitivity
SI	simulation interval
Sp	specificity

REFERENCES

References

1. Maclure M, Schneeweiss S. Causation of bias: the episcope. Epidemiology 2001;12(1):114–122.11138805
2. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. Epidemiology 2003;14(4):451–458.12843771
3. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34(6):1370–1376.16172102
4. Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer, 2009.
5. Orsini N, Bellocco R, Bottai M, Wolk A, Greenland S. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. Stata J 2008;8(1):29–48.
6. Lash TL, Schmidt M, Jensen AO, Engebjerg MC. Methods to apply probabilistic bias analysis to summary estimates of association. Pharmacoepidemiol Drug Saf 2010;19(6):638–644.20535760
7. Maldonado G Adjusting a relative-risk estimate for study imperfections. J Epidemiol Community Health 2008;62(7):655–663.18559450
8. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. Int J Epidemiol 2014;43(6):1969–1985.25080530
9. Stothard KJ, Tennant PW, Bell R, Rankin J. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. JAMA 2009;301(6):636–650.19211471
10. Blanco R, Colombo A, Suazo J. Maternal obesity is a risk factor for orofacial clefts: a meta-analysis. Br J Oral Maxillofac Surg 2015;53(8):699–704.26073906

11. Izedonmwen OM, Cunningham C, Macfarlane TV. What is the risk of having offspring with cleft lip/palate in pre-maternal obese/overweight women when compared to pre-maternal normal weight women? A systematic review and meta-analysis. J Oral Maxillofac Res 2015;6(1):e1.
12. Stott-Miller M, Heike CL, Kratz M, Starr JR. Increased risk of orofacial clefts associated with maternal obesity: case-control study and Monte Carlo-based analysis. Paediatr Perinatl Epidemiol 2010;24(5):502–512.
13. Reefhuis J, Gilboa SM, Anderka M, Browne ML, Feldkamp ML, Hobbs CA, Jenkins MM, Langlois PH, Newsome KB, Olshan AF, Romitti PA, Shapira SK, Shaw GM, Tinker SC, Honein MA, National Birth Defects Prevention Study. The National Birth Defects Prevention Study: a review of the methods. Birth Defects Res A Clin Mol Teratol 2015;103(8):656–669.26033852
14. Rasmussen SA, Olney RS, Holmes LB, Lin AE, Keppler-Noreuil KM, Moore CA, National Birth Defects Prevention Study. Guidelines for case classification for the National Birth Defects Prevention Study. Birth Defects Res A Clin Mol Teratol 2003;67(3):193–201.12797461
15. Greenland S, Lash TL. Bias analysis In: Rothman, Greenland, Lash TL, eds. Modern Epidemiology, 3rd Edition. Philadelphia: Lippincott Williams & Wilkins, 2008;345–380.
16. Lyles RH, Lin J. Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. Stat Med 2010;29(22):2297–2309.20552681
17. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National Health and Nutrition Examination Survey: Plan and Operations, 1999–2010. In: National Center for Health Statistics, ed. Vital Health Stat. Vol. 1, 2013.
18. Stepler R, Brown A. Statistical portrait of Hispanics in the United States. 2016.
19. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15(5):615–625.15308962
20. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? Epidemiology 2006;17(4):413–418.16755269
21. Johnson CY, Flanders WD, Strickland MJ, Honein MA, Howards PP. Potential sensitivity of bias analysis results to incorrect assumptions of nondifferential or differential binary exposure misclassification. Epidemiology 2014;25(6):902–909.25120106
22. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. Am J Epidemiol 1991;134(10):1233–1244.1746532
23. MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? Epidemiology 2012;23(1):151–158.22157311
24. Ding P, VanderWeele T. Sensitivity analysis without assumptions. Epidemiology 2016;27(3):368–377.26841057
25. 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.

Table 1.

Characteristics of case and control mothers in the analysis, National Birth Defects Prevention Study, 1997–2011.

	Case Mothers (n = 2,523)		Control Mothers (n = 10,605)		Odds Ratio (95% Confidence Interval)
	Number	%	Number	%	
Prepregnancy obesity					
No	1,919	76	8,259	78	1.00 (Ref)
Yes	478	19	1,896	18	1.09 (0.97, 1.21)
Missing	126	5	450	4	
Race/ethnicity					
Non-Hispanic white	1,671	66	6,601	62	1.00 (Ref)
Non-Hispanic black	154	6	1,251	12	0.49 (0.41, 0.58)
Hispanic	698	28	2,753	26	1.00 (0.91, 1.11)
Maternal education					
0–11 years	503	20	1,800	17	1.41 (1.24, 1.60)
12 years	684	27	2,545	24	1.36 (1.21, 1.53)
13–15 years	659	26	2,843	27	1.17 (1.04, 1.32)
≥16 years	677	27	3,417	32	1.00 (Ref)
Study site					
Arkansas	297	12	1,384	13	1.00 (Ref)
California	388	15	1,118	11	1.62 (1.36, 1.92)
Georgia	246	10	1,117	11	1.03 (0.85, 1.24)
Iowa	270	11	1,225	12	1.03 (0.86, 1.23)
Massachusetts	296	12	1,284	12	1.07 (0.90, 1.28)
New Jersey	92	4	526	5	0.82 (0.63, 1.05)
New York	208	8	891	8	1.09 (0.89, 1.32)
North Carolina	173	7	862	8	0.94 (0.76, 1.15)
Texas	283	11	1,279	12	1.03 (0.86, 1.23)
Utah	270	11	919	9	1.37 (1.14, 1.65)

Table 2.

Bias parameters used for adjustment of exposure misclassification by racial/ethnic group, National Health and Nutrition Examination Survey, 1999–2010.

	Sensitivity	Specificity	P(obese|missing)[a]	P(not obese|missing)[b]
Mexican-American	0.817	0.968	0.458	0.542
Non-Hispanic black	0.859	0.959	0.521	0.479
Non-Hispanic white	0.841	0.991	0.335	0.665

[a] Proportion of women with missing data on body mass index who were truly obese based on body measurements.

[b] Proportion of women with missing data on body mass index who were truly not obese based on body measurements.

Table 3.

Associations Between Prepregnancy Obesity and Cleft Lip With or Without Cleft Palate, Adjusting for Different Combinations of Biases, National Birth Defects Prevention Study, 1997–2011.

	Conventional Analysis		Non-Probabilistic Bias Analysis[a]	Probabilistic Bias Analysis[b]		Probabilistic Bias Analysis Plus Random Error[b]	
	OR	95% CI	OR	Median OR	95% SI	Median OR	95% RESI
Unadjusted	1.09	0.97, 1.21					
Confounding only[c]	1.10	0.98, 1.23					
Selection bias and confounding[c]			0.98	0.98	0.82, 1.17	0.98	0.80, 1.21
Exposure misclassification and confounding[c,d]							
Nondifferential ($Se_{ca} = Se_{co}$, $Sp_{ca} = Sp_{co}$)			1.12	1.11	0.84, 1.45	1.11	0.83, 1.48
Differential A ($Se_{ca} = Se_{co}+0.05$, $Sp_{ca} = Sp_{co}+0.03$)			1.09	1.09	0.84, 1.41	1.09	0.82, 1.44
Differential B ($Se_{ca} = Se_{co}-0.05$, $Sp_{ca} = Sp_{co}-0.03$)			1.02	1.01	0.76, 1.34	1.01	0.74, 1.36
All biases combined[c,e]							
Nondifferential ($Se_{ca} = Se_{co}$, $Sp_{ca} = Sp_{co}$)			1.03	1.02	0.76, 1.37	1.02	0.75, 1.39
Differential A ($Se_{ca} = Se_{co}+0.05$, $Sp_{ca} = Sp_{co}+0.03$)			1.01	1.01	0.76, 1.33	1.00	0.74, 1.36
Differential B ($Se_{ca} = Se_{co}-0.05$, $Sp_{ca} = Sp_{co}-0.03$)			0.94	0.93	0.69, 1.26	0.93	0.67, 1.28

Abbreviations: CI, confidence interval; OR, odds ratio; RESI, random error-added simulation interval; Se_{ca}, sensitivity for cases; Se_{co}, sensitivity for controls; SI, simulation interval; Sp_{ca}, specificity for cases; Sp_{co}, specificity for controls.

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

[b] Triangular distributions of bias parameters sampled over 5,000 iterations.

[c] Adjusted for confounding by maternal race/ethnicity.

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

[e] Adjusted for selection bias, exposure misclassification, missing exposure, and confounding.