



Published in final edited form as:

Paediatr Perinat Epidemiol. 2018 May ; 32(3): 281–286. doi:10.1111/ppe.12459.

Assessment of Recording Bias in Pregnancy Studies using Health Care Databases: An application to neurologic conditions

Sarah C MacDonald^A, Miguel A. Hernán^{A,B}, Thomas F. McElrath^C, and Sonia Hernández-Díaz^A

^ADepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

^BDepartment of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts

^CDivision of Maternal-Fetal Medicine, Department of Obstetrics & Gynecology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

Abstract

BACKGROUND—Pre-existing conditions are imperfectly recorded in healthcare databases. We assessed whether pre-existing neurologic conditions (epilepsy, multiple sclerosis [MS]) were differentially recorded in the presence of major obstetric outcomes (Caesarean delivery, preterm delivery, preeclampsia) in delivery records. We also evaluated the impact of differential recording on measures of frequency and association between the conditions and outcomes.

METHODS—The 2011-2014 Truven Health MarketScan® Commercial Claims Dataset was used to identify deliveries. We calculated the relative recording of epilepsy and MS at delivery compared with a 270-day pre-delivery interval, both overall and by the presence of major obstetric outcomes. We estimated risk ratios of the association between epilepsy and MS with the outcomes for each ascertainment window.

RESULTS—We identified 909,065 deliveries in women continuously enrolled from 270-days before the delivery date. Of women with epilepsy identified in the pre-delivery interval, 73% had the condition coded at delivery. For MS, the proportion was 60%. MS recording at delivery did not vary by obstetric outcomes, however delivery-coded epilepsy was less likely confirmed in the pre-delivery interval in the presence of preeclampsia. Generally, the period of ascertainment did not meaningfully impact risk ratios, however the risk ratio for preeclampsia associated with epilepsy was 1.67 [95% CI: 1.47, 1.90] when epilepsy was ascertained at delivery and 1.26 [95% CI: 1.07, 1.48] when epilepsy was ascertained in the pre-delivery interval (heterogeneity, $p=0.007$).

CONCLUSIONS—Ascertainment of epilepsy and MS in delivery records underestimated prevalence. However the window of recording generally did not impact associations with obstetric outcomes.

Keywords

Administrative Claims; Healthcare; Epilepsy; Multiple Sclerosis; Pregnancy; Bias (Epidemiology)

INTRODUCTION

Pregnancy studies are increasingly based on health care databases due to their relative ease of access, large sample size, and relative low cost compared to ad hoc data collection. Accurate ascertainment of pre-existing health conditions during pregnancy is vital for the validity of these studies.

While it is well known that retrospective data collection in interview-based studies can suffer from differential misclassification of pre-existing conditions¹, less attention has been paid to misclassification in database studies. However, administrative databases pose unique challenges for disease ascertainment as well, as chronic conditions are often under-recorded² and this recording may not be differential with respect to outcomes of interest.

The approaches to deal with misclassification of pre-existing conditions depend on the type of database. In longitudinal databases (e.g., administrative claims), a natural strategy to capture most conditions is to ascertain them during a long interval. However, this approach reduces sample size because only individuals enrolled for the entire duration of the interval are eligible. In most pregnancy studies, the ascertainment interval is typically the pregnancy period.^{3,4}

In cross-sectional databases (e.g. Nationwide Inpatient Sample), researchers are restricted to ascertaining both the pre-existing conditions *and* the outcome of interest in the same medical encounter. This is problematic because the presence of obstetric complications may affect the recording of pre-existing conditions.⁵⁻⁷

We used administrative claims to (1) assess the relative recording accuracy of pre-existing epilepsy and multiple sclerosis (MS) at the delivery hospitalization and in a longer pregnancy interval, both overall and by commonly associated major obstetric outcomes (Caesarean delivery, preterm delivery, and preeclampsia)^{5,6,8,9} and (2) compare risk ratios for the association between the pre-existing conditions and obstetric outcomes (Caesarean delivery, preterm delivery, and preeclampsia) when the ascertainment window of the pre-existing conditions varies.

METHODS

Study Population

Truven Health MarketScan® (Truven Health Analytics Inc., Ann Arbor, MI, USA) Commercial Claims and Encounters Database is a convenience sample of health care claims from approximately 350 payers across the United States. Unique insurance enrollee ID numbers are available to link inpatient, outpatient, and pharmaceutical claims from the same individual. Using the *International Statistical Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) and *Current Procedural Terminology* procedure codes

previously used in Medicaid⁴ and MarketScan® data,¹⁰ we identified a cohort of women aged 12-55 who delivered in-hospital from January 1 2011 – December 31 2014. Delivery date was the date associated with the first delivery-related procedure code in the inpatient services file or, when a date for the procedure code was unavailable, the admission date. Using the unique family ID variable, women were linked to infants whose first healthcare encounter was within 30 days of the mother’s estimated delivery date.

We restricted the cohort to women continuously enrolled with prescription claim data for at least 270 days before the delivery until, and including, the delivery date (Information on how the 270-day interval length was chosen is provided in Appendix A). Continuous enrollment was defined as evidence of enrollment in a health insurance plan for at least 28 days in each month in the study period. From this cohort we compared women with the pre-existing conditions identified in the full 270 days before and *including* the date of delivery (‘pre-delivery interval’) to women with the pre-existing conditions identified only at the delivery hospitalization (‘delivery interval’; Median length of delivery hospitalization was 2 days).

To simulate studies from longitudinal and cross-sectional data, we based definitions of the conditions on definitions commonly used in the literature for each ascertainment interval.^{5,6,11–15} In the pre-delivery interval, a strict definition was used to maximize the proportion of true cases selected. MS was defined as the ICD-9-CM code 340.xx on two separate days and epilepsy was defined as any of the ICD-9-CM codes 345.0x-345.5x, 345.7x-345.9x, 649.4x, on two separate days plus at least one prescription claim for the antiepileptic medications (Supplemental Table 1). In the delivery hospitalization interval, MS and epilepsy were both defined as at least one ICD-9-CM code for their respective conditions. We described maternal characteristics in women with epilepsy, women with MS, and women without either of these conditions as defined in the pre-delivery interval using the codes listed in Appendix B.

Measurements of agreement

We calculated the proportion of women with the condition recorded at the delivery hospitalization, among those classified with the pre-existing condition in the pre-delivery interval (“Relative Sensitivity”). We also calculated the proportion of women classified with the pre-existing condition in the pre-delivery interval, among those with the condition recorded at the delivery hospitalization (“Relative Predictive Value”).

We presented these measurements overall and stratified by three obstetric outcomes ascertained in the maternal record during the delivery hospitalization: 1) Caesarean delivery (ICD-9-CM: 74.0-74.2, 74.99, 669.7x; Current Procedural Terminology: 59510, 59514, 59515), 2) Preterm delivery (ICD-9-CM: 644.2x), and 3) Preeclampsia (ICD-9-CM: 642.4x-642.7x).

Impact on measures of association

We calculated the risk ratios for the three obstetric outcomes when the pre-existing conditions were defined i) during the pre-delivery interval and ii) at delivery. Risk ratios were estimated using generalized linear models with a log link, Poisson distribution, and robust variances. All models were adjusted for maternal age, region, and year of delivery.

Heterogeneity of risk ratios from each ascertainment interval was tested using the Cochrane Q test and p-values were reported. Statistical analyses were conducted using SAS Enterprise Guide version 6.1 (SAS Institute Inc.).

RESULTS

Of 1,894,935 deliveries between January 1, 2011 and December 31, 2014, 909,065 (48%) were to women continuously enrolled in a healthcare plan with prescription claim data available for 270 days before the delivery date. Of these, 2,248 (0.2%) were classified with epilepsy and 974 (0.1%) with MS in the pre-delivery interval. Maternal characteristics of women with and without the pre-existing conditions are presented in Appendix C.

Measurements of agreement

Among those meeting the criteria for epilepsy in the pre-delivery interval, 73% were also recorded as having epilepsy at delivery (Relative Sensitivity). For MS, the Relative Sensitivity was 60%. Among those with epilepsy recorded at delivery, 63% also met the definition for epilepsy in the pre-delivery interval (Relative Predictive Value). For MS, the Relative Predictive Value was 78%.

Figure 1 shows the Relative Sensitivity (Figure 1a–Figure 1b) and Relative Predictive Value (Figure 1c–Figure 1d) among those with and without the obstetric outcomes of interest. The proportion of women with the pre-existing conditions recorded at delivery tended to be similar or lower for those *with* the obstetric outcomes than those without the obstetric outcomes. For MS, the differences were small whereas for epilepsy there was a large difference in Relative Sensitivity for those delivering preterm (62%) compared with those delivering at term (75%).

The Relative Predictive Value of epilepsy was lower among those experiencing preterm delivery (56% vs. 64%), or preeclampsia (47% vs. 65%) than among those without the obstetric outcomes. As women with preeclampsia are also more likely to have a preterm delivery,¹⁶ we hypothesized that part of this finding could be due to neurologic symptoms associated with preeclampsia, including eclamptic seizures, being misclassified as epilepsy at delivery. To address this, we conducted an analysis where we re-calculated the Relative Predictive Value for the remaining outcomes in a cohort excluding those with preeclampsia. After this exclusion, the Relative Predictive Value increased for those with preterm delivery (Appendix D).

In contrast to epilepsy, the Relative Predictive Values for MS recorded at delivery were similar or greater in those with the obstetric outcomes as compared with those without them.

Impact of misclassification on measures of association

Figure 2 presents the risk ratios for the obstetric outcomes when the pre-existing condition was ascertained in the pre-delivery interval and at the delivery hospitalization. Epilepsy defined in the pre-delivery interval was associated with an increased risk of Caesarean delivery (risk ratio: 1.17 [95% CI 1.11, 1.22]), preterm delivery (1.23 [95% CI 1.09, 1.38]), and preeclampsia (1.26 [95% CI 1.07, 1.34]). Epilepsy defined at delivery had a similar

association with Caesarean delivery (p-value for heterogeneity: 0.31) and preterm delivery (p-value for heterogeneity: 0.77) and a stronger association with preeclampsia (1.67 [95% CI: 1.47, 1.90], p-value for heterogeneity: 0.007).

MS defined in the pre-delivery interval was associated with an increased risk of preterm delivery (1.33 [95% CI 1.13, 1.58]) but not for Caesarean delivery (1.03 [95% CI 0.95, 1.11]) or preeclampsia (1.03 [95% CI 0.78, 1.35]). MS at delivery was similarly associated with Caesarean delivery (p-value for heterogeneity: 0.53), preterm delivery (p-value for heterogeneity: 0.37), and preeclampsia (p-value for heterogeneity: 0.24). Analyses with varying pre-delivery definitions of both epilepsy and MS found overall similar patterns as the primary analyses (Appendix E).

COMMENT

Principal Findings

The prevalence of epilepsy and MS was under-estimated when ascertainment occurred during the delivery hospitalization as compared with the pre-delivery interval. This misclassification generally appeared non-differential by major obstetric outcomes and most measures of association did not vary by the ascertainment window.

Interpretation

For epilepsy, 73% of women with the condition in the pre-delivery interval were also recorded as having epilepsy at delivery. Recording was *lower* in those delivering preterm than in those delivering at term, suggesting that more complicated deliveries may distract from the recording of chronic conditions in the insurance claims. Only 63% of those recorded at delivery met the stricter condition definition in the pre-delivery interval. This proportion was 47% in women with preeclampsia which may be explained by (1) neurologic symptoms associated with preeclampsia, including eclamptic seizures, being incorrectly recorded using epilepsy codes, or (2) more complete documentation of distant seizure disorders in women with preeclampsia (less likely because the Relative Sensitivities were virtually equivalent in those with and without preeclampsia). Regardless of the reason, which cannot be definitely determined from our data, differential misclassification led to an overestimate of the risk ratio for preeclampsia when epilepsy was ascertained at delivery. This finding suggests that previous studies in women with epilepsy defined at delivery^{5,6} may have overestimated the risk ratio as a result of recording bias.

For MS, 78% of women with the condition in the pre-delivery interval were also recorded as having MS at delivery when using stricter criteria. Recording did not appear to vary by Caesarean delivery, preterm delivery, and preeclampsia; and risk ratios for these outcomes were similar whether MS was identified at delivery or throughout pregnancy.

Strengths of the Study

The greatest strength of this study is the large sample size. Indeed, epilepsy and MS are relatively rare in pregnant women^{6,8,17–23} and so many studies are underpowered to assess these conditions. The studies that *have* looked are therefore often large administrative

databases that may be cross-sectional, and thus particularly vulnerable to misclassification biases.^{5,6}

Limitations of the Data

Database restrictions prevented us from comparing ascertainment at delivery with a true gold standard (e.g. medical records) in pregnancy. Instead, we compared ascertainment at delivery to ascertainment before (and including) delivery using stringent definitions. While these definitions were adapted from prior literature^{11–15} and the length of the interval was determined from comprehensive claims review (Appendix A), some error likely still exists. To emphasize the lack of a gold standard, we calculated an *estimate* of sensitivity of the delivery hospitalization using the “Relative Sensitivity” and an *estimate* of the positive predictive value using the “Relative Predictive Value”. At least one study has used a typical gold standard measure (medical records) and found that conditions such as asthma, thyroid disorders, and cardiovascular disease generally had low sensitivities (10-58%) but high positive predictive values (86-100%),² a pattern similar to MS here. Taken together, it may be reasonable to conclude that while many pre-existing conditions are under-recorded at delivery, when the conditions *are* recorded, often they can be confirmed by medical records² or more stringent definitions in pregnancy (current study).

Finally, an underlying assumption was that the misclassification of the conditions during the *pre-delivery* period is non-differential with regards to pregnancy outcomes. While the temporal relation of events certainly makes a direct impact of obstetric outcomes on past recordings unlikely, certain factors (e.g., health care utilization) may still affect the recording during pregnancy and the risk of adverse delivery outcomes (Figure 3). While we cannot completely exclude the possibility of unmeasured confounders on recording, we do not believe that such covariates would exert *differential* effects in the two intervals.

Conclusions

Our findings support the hypothesis that pre-existing conditions are incompletely captured on healthcare databases restricted to delivery, but did not suggest that major obstetric outcomes improve the coding of these conditions. Use of cross-sectional records at delivery did not result in recording bias for most of the associations considered. However, the differential positive predictive value for epilepsy by preeclampsia suggests that epilepsy should be captured in intervals preceding the delivery hospitalization (or simply before preeclampsia onset). In general, when only delivery records are available, investigators should consider whether the outcome of interest can trigger recording of the specific exposure of interest. Indeed, the potential exists and should be examined when using cross-sectional medical records.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Jacqueline Cohen for her help developing the chronic disease algorithms.

Sources of Funding: SCM was supported by the Canadian Institutes for Health Research (CIHR) Doctoral Foreign Study Award. SHD and MAH were supported by the R01 MH100216 grant from the National Institute of Mental Health.

References

1. Drews CD, Kraus JF, Greenland S. Recall bias in a case-control study of sudden infant death syndrome. *Int J Epidemiol.* 1990; 19(2):405–411. [PubMed: 2376455]
2. Yasmeen S, Romano PS, Schembri ME, et al. Accuracy of obstetric diagnoses and procedures in hospital discharge data. *Am J Obstet Gynecol.* 2006; 194(4):992–1001. [PubMed: 16580288]
3. Bateman BT, Huybrechts KF, Fischer MA, et al. Chronic hypertension in pregnancy and the risk of congenital malformations: A cohort study. *Am J Obstet Gynecol.* 2015; 212(3):337.e1–337.e14. [PubMed: 25265405]
4. Palmsten, K., Huybrechts, KF., Mogun, H., et al. Harnessing the Medicaid Analytic eXtract (MAX) to evaluate medications in pregnancy: Design considerations. In: Rogers, LK., editor. *PLoS One.* Vol. 8. 2013. p. e67405
5. MacDonald SC, Bateman BT, Mcelrath TF, et al. Mortality and Morbidity During Delivery Hospitalization Among Pregnant Women With Epilepsy in the United States. *JAMA Neurol.* 2015; 72(9):981–988. [PubMed: 26147878]
6. Kelly VM, Nelson LM, Chakravarty EF. Obstetric outcomes in women with multiple sclerosis and epilepsy. *Neurology.* 2009; 73(22):1831–1836. [PubMed: 19923552]
7. Bansil P, Kuklina EV, Meikle SF, et al. Maternal and fetal outcomes among women with depression. *J Women's Heal.* 2010; 19(2):329–334.
8. Chen Y, Lin HL, Lin HC. Does multiple sclerosis increase risk of adverse pregnancy outcomes? A population-based study. *Mult Scler.* 2009; 15(5):606–612. [PubMed: 19318510]
9. Lu E, Zhao Y, Zhu F, et al. Birth hospitalization in mothers with multiple sclerosis and their newborns. *Neurology.* 2013; 80(5):447–452. [PubMed: 23303853]
10. Thomson Healthcare. *The Healthcare Costs of Having a Baby.* The March of Dimes. 2007
11. Marrie RA, Yu N, Blanchard J, et al. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba.[Erratum appears in *Neurology.* 2011 Sep 13;77(11):1105]. *Neurology.* 2010; 74:465–471. [PubMed: 20071664]
12. Culpepper WJ, Ehrmantraut M, Wallin MT, et al. Veterans Health Administration multiple sclerosis surveillance registry: The problem of case-finding from administrative databases. *J Rehabil Res Dev.* 2006; 43(1):17. [PubMed: 16847768]
13. Pugh MJV, Van Cott AC, Cramer JA, et al. Trends in antiepileptic drug prescribing for older patients with new-onset epilepsy: 2000-2004. *Neurology.* 2008; 70(22, Part 2):2171–2178. [PubMed: 18505996]
14. Jetté N, Reid AY, Quan H, et al. How accurate is ICD coding for epilepsy? *Epilepsia.* 2010; 51(1): 62–69. [PubMed: 19682027]
15. Holden EW, Grossman E, Nguyen HT, et al. Developing a computer algorithm to identify epilepsy cases in managed care organizations. *Dis Manag.* 2005; 8(1):1–14. [PubMed: 15722699]
16. Duley L. The Global Impact of Pre-eclampsia and Eclampsia. *Semin Perinatol.* 2009; 33(3):130–137. [PubMed: 19464502]
17. Katz O, Levy A, Wiznitzer A, et al. Pregnancy and perinatal outcome in epileptic women: a population-based study. *J Matern Neonatal Med.* 2006; 19(1):21–25.
18. Richmond JR, Krishnamoorthy P, Andermann E, et al. Epilepsy and pregnancy: an obstetric perspective. *Am J Obstet Gynecol.* 2004; 190(2):371–379. [PubMed: 14981376]
19. Fairgrieve SD, Jackson M, Jonas P, et al. Population based, prospective study of the care of women with epilepsy in pregnancy. *BMJ.* 2000; 321(7262):674–675. [PubMed: 10987772]
20. Viinikainen K, Heinonen S, Eriksson K, et al. Community-based, prospective, controlled study of obstetric and neonatal outcome of 179 pregnancies in women with epilepsy. *Epilepsia.* 2006; 47(1):186–192. [PubMed: 16417548]
21. Hvas CL, Henriksen TB, Ostergaard JR, et al. Epilepsy and pregnancy: effect of antiepileptic drugs and lifestyle on birthweight. *BJOG.* 2000; 107(7):896–902. [PubMed: 10901562]

22. Veiby G, Daltveit AK, Engelsen BA, et al. Pregnancy, delivery, and outcome for the child in maternal epilepsy. *Epilepsia*. 2009; 50(9):2130–2139. [PubMed: 19490036]
23. Fong A, Chau CT, Quant C, et al. Multiple sclerosis in pregnancy: Prevalence, sociodemographic features, and obstetrical outcomes. *J Matern Neonatal Med*. 2017; 0(0):1–6.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

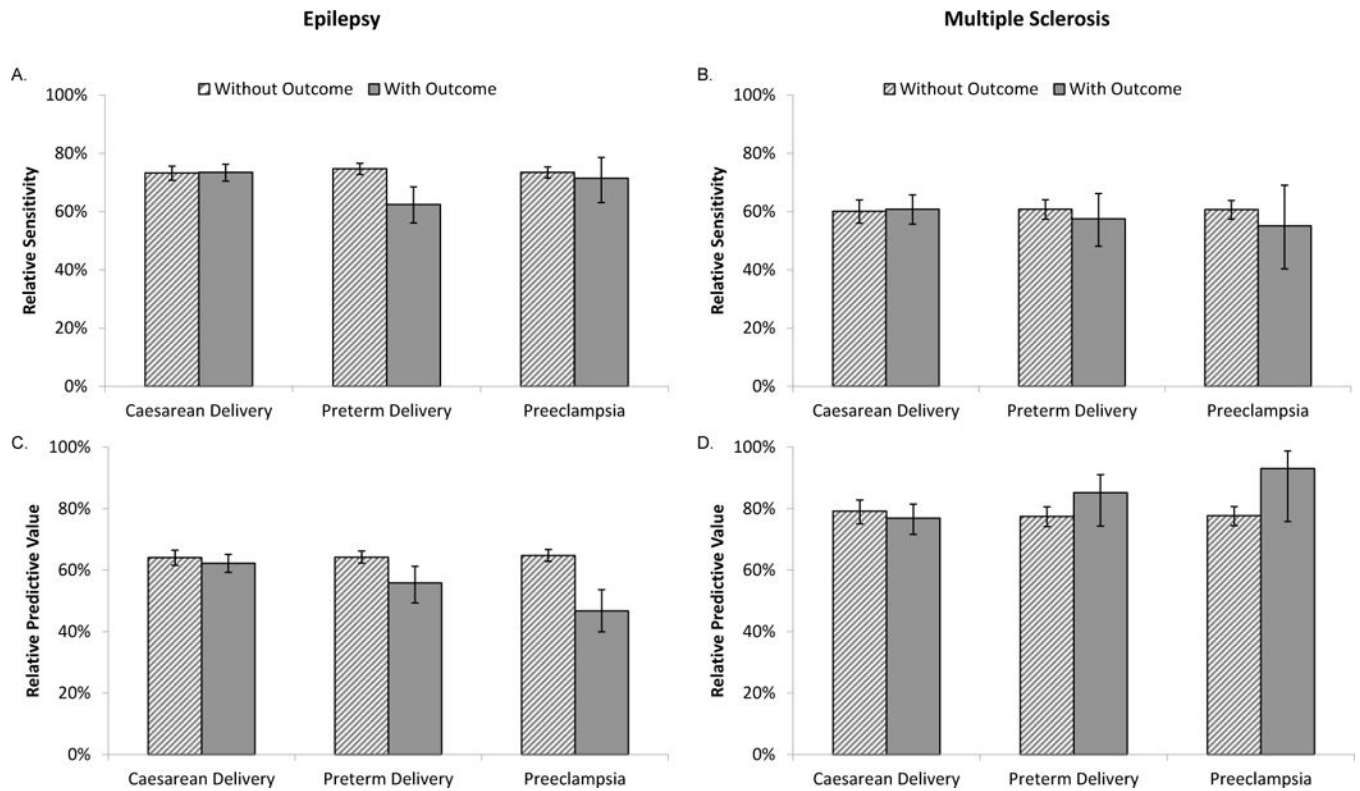


Figure 1. Relative Sensitivity (A-B) and Predictive Values (C-D) for Epilepsy (A, C) and Multiple Sclerosis (B, D) coded in the Delivery versus the Pre-Delivery Interval (Truven Health MarketScan® Commercial Claims and Encounters Database, 2011-2014 USA, N=909,065). Darker solid bars represent those with the delivery event; lighter hatched bars represent those without the delivery event. Error bars represent 95% confidence intervals.

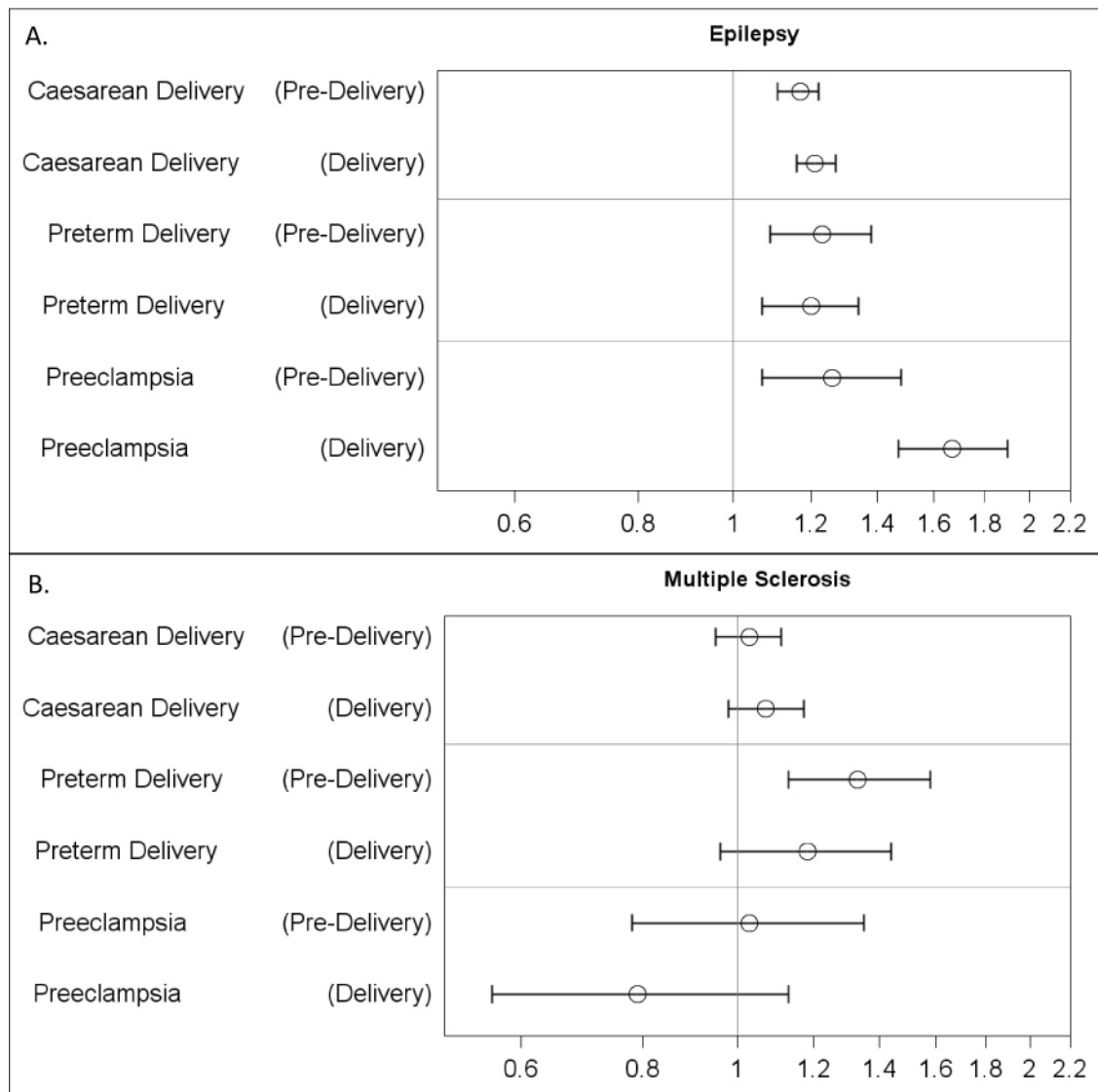


Figure 2.

Adjusted Risk Ratios of Obstetric Outcomes by (A) Epilepsy and (B) Multiple Sclerosis Defined in the Delivery and Pre-Delivery Interval (Truven Health MarketScan® Commercial Claims and Encounters Database, 2011-2014, USA, N=909,065). Risk ratios adjusted for maternal age, region (Northeast, Midwest/North Central, South, and West) and year of delivery. *Pre-Delivery* = Pre-existing condition recorded on at least two days in the 270 days before delivery until, and including, the delivery date (and at least one anti-convulsant prescription for epilepsy). *Delivery* = Pre-existing condition recorded on at least one day during the delivery hospitalization. *RR*= risk ratio, *LCI*= lower confidence interval, *UCI*= upper confidence interval.

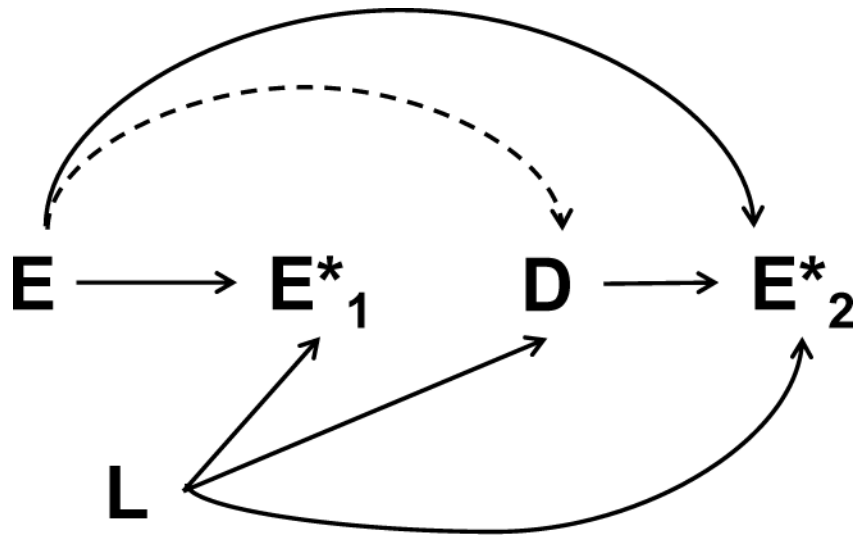


Figure 3. Directed acyclic graph indicating possible confounding. E = Pre-existing condition (e.g. epilepsy); D= Obstetric outcome (e.g. preeclampsia); E*₁ = Recording of pre-existing condition in pregnancy; E*₂ = Recording of pre-existing condition at delivery; L=Other covariates (e.g. Healthcare utilization). Dashed lines indicate the association of interest.