Validity of maternal and infant outcomes within nationwide Medicaid data

Kristin Palmsten, **Krista F. Huybrechts**, **Mary K. Kowal**, **Helen Mogun**, and **Sonia Hernández-Díaz**

Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Kristin Palmsten, Sonia Hernández-Díaz); Department of Pediatrics, University of California, San Diego, La Jolla, California (Kristin Palmsten); Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Krista F. Huybrechts, Mary K. Kowal, Helen Mogun)

Abstract

Purpose—The aim of this study is to assess the validity of preeclampsia, congenital cardiac malformations, and persistent pulmonary hypertension of the newborn (PPHN) diagnoses in the US Medicaid Analytic eXtract (MAX), a nationwide healthcare utilization database that may be useful for perinatal research.

Methods—Using the 2000–2007 MAX, we identified more than 1 million pregnancies ending in live birth. We identified potential cases based on claims, reviewed their hospital medical records, and calculated the positive predictive values (PPV) and 95% confidence intervals using records as the reference.

Results—Among 183 women with any preeclampsia diagnoses, the PPV was 66.5% (53.6, 77.4%), but it increased to 94.5% (84.0, 98.3%) for inpatient preeclampsia diagnoses. The PPV for inpatient PPHN diagnoses (N=82) was 68.3% (57.6, 77.4%), but it increased to 89.6% (CI: 77.8, 95.5%) when restricting to infants not transferred to another facility shortly after birth (N=48). The PPV for cardiac malformations was 77.6% (65.7, 86.2%) when requiring inpatient codes on more than one date (N=63).

Conclusions—These PPVs are conservative, particularly when patients were transferred or received outpatient diagnoses, because we reviewed records from a single hospitalization only. PPVs improve with stringent identification criteria, at the cost of sensitivity, and can be used to correct for measurement error.

Keywords

congenital cardiac malformations; Medicaid; persistent pulmonary hypertension of the newborn; preeclampsia; pregnancy; validation study; pharmacoepidemiology

Introduction

Medicaid is the state and federal health insurance program for low income individuals in the United States, and Medicaid reimburses the medical expenses of over 40% of births in the US.[1] The Medicaid Analytic eXtract (MAX) contains beneficiary enrollment and healthcare utilization claims, including outpatient pharmacy dispensings and inpatient and outpatient diagnostic and procedure claims,[2] and may be a valuable resource for studies of medication use and safety in pregnancy.[3–6] We previously identified a cohort of over 1 million pregnant women and their live born infants from nationwide MAX data.[7]

Because healthcare utilization data are collected for administrative and payment purposes,[8] investigators using these databases for research should identify potential threats to study validity and implement strategies to address these limitations. In particular, outcome diagnoses recorded in MAX should be validated with medical records to inform the operational outcome definitions and to correct for measurement error through sensitivity analyses used in epidemiologic studies.[9]

The accuracy of diagnoses for pregnancy complications, delivery characteristics, and neonatal outcomes recorded in hospital discharge and healthcare claims databases, compared with information available in medical records, varies depending on the factor of interest and the data source.[10–22] Cooper et al described the validity of congenital malformation diagnoses among Medicaid beneficiaries in Tennessee.[22] Hennessy et al described the validation of sudden cardiac death and ventricular arrhythmia diagnoses in Medicaid and Medicare data from 5 states.[23–24] However, there are no previous studies that validate pregnancy-related factors recorded in nationwide Medicaid data.

We conducted studies of antidepressant use during pregnancy and risk for preeclampsia, persistent pulmonary hypertension of the newborn (PPHN), and congenital cardiac malformations (in particular ventricular septal defect (VSD), right ventricular outflow tract obstruction (RVOTO), and other cardiac malformations because of previously reported associations between antidepressants and VSD and RVOTO). Our primary goal was to assess the validity of these outcomes identified from MAX data using hospital medical records as the reference standard to inform our studies of antidepressant safety during pregnancy. We also wanted to assess the accuracy of outcomes for users and non-users of antidepressants and were able to do so for potential preeclampsia cases because of adequate numbers. Finally, we assessed additional obstetric factors using the records available from potential cases: multiparity, labor induction, cesarean delivery, and preterm delivery.

Methods

Study population

We conducted this validation study within a cohort of pregnancies ending in live birth that had previously been identified from 2000–2007 MAX data.[7] Briefly, women with delivery-related diagnoses and procedures were identified. Then, live-born infants were linked to these women by matching state, Medicaid Case Number, and maternal delivery dates with infant date of birth. Four major maternal eligibility criteria were required for cohort

inclusion: continuous enrollment in Medicaid, no private insurance, no restricted benefits, and appropriate enrollment type. The eligible subcohort size varies for each outcome of interest depending on the minimum eligibility period length and additional infant eligibility criteria required for each antidepressant safety study (Figure 1). This project was approved by the Brigham and Women's Hospital and Harvard School of Public Health Institutional Review Boards and a data use agreement was approved by the Centers for Medicare and Medicaid Services (CMS).

Step 1: Identification of potential cases—The criteria we used to identify pregnancies with the primary outcomes are listed in Table 1. We used both maternal and infant codes to identify PPHN and cardiac malformations because infant's claims may be recorded under the mother's ID for the first several months after birth.[25] Because a review of claims profiles suggested that just 1 diagnostic code for VSDs or RVOTOs may indicate a rule-out diagnosis, only individuals with diagnostic codes for VSDs or RVOTOs on at least two dates were classified as potential VSD and RVOTO cases.

Step 2: Exclusion of potential cases without hospital contact information— MAX data contain a state assigned hospital identifier, the Medicaid Billing Provider Number, but there was no centralized list of contact information for these identifiers.[26] Therefore, it was not possible to contact hospitals for claims validation using this identifier. Hospital contact information is available from the National Plan and Provider Enumeration System downloadable file[27] for the Medicare Provider Number available in Medicare inpatient data. Consequently, we created a database of hospital contact information by identifying the Medicare Provider Number that corresponded to the Medicaid Billing Provider Number. Among all pregnancies identified as having preeclampsia, PPHN, and cardiac malformations, we linked hospitalizations to the Medicare Provider Number, when available in the database, to obtain contact information. Pregnancies with no hospital contact information were excluded.

Step 3: Selection of a sample of potential cases for validation—Because of study feasibility and cost constraints, we selected for validation a sample of the potential cases with hospital information available. Our goal was to select 100 to 200 potential cases for each outcome group of interest. To assess preeclampsia validity by depression diagnoses and antidepressant use, we stratified potential preeclampsia cases according to depression-related diagnoses (International Classification of Diseases (ICD)-9 codes 296.x, 300.x, 309.x, 311.x) and antidepressant dispensings and sampled women from each stratum. Given the smaller pool of potential cases with available hospital information, all potential PPHN and cardiac malformation cases with hospital information available at the time of delivery were selected for validation. In addition, we selected a sample of potential cardiac malformation cases for which hospital information was not available at the time of delivery but was available for a subsequent hospitalization after the time of delivery (Figure 1).

Step 4: Social Security Number (SSN) linkage—MAX data do not contain direct personal identifiers such as names and addresses. However, SSNs can be requested from CMS for selected individuals. A CMS data vendor provided SSNs for potential cases to the

vendor that conducted the medical record abstraction. Investigators did not have access to SSNs or other personal identifiers.

Step 5: Medical record request—We requested that hospitals send the medical records of women identified as having preeclampsia and of infants identified as having PPHN or cardiac malformations based on claims. Of note, only SSN and date of birth were provided to the hospitals to identify the records because we did not have access to names, addresses, or other personal information. Because SSNs may not be assigned until after the time of birth, the mothers' SSNs and the infants' SSNs when available, were provided to locate the infants' records. Medical records were requested for 425 preeclampsia, 257 PPHN and 660 cardiac malformation potential cases (specifically 95 VSD, 24 RVOTO, and 541 other cardiac malformation potential cases). Written requests for records of interest were sent to 380 hospitals from 35 states. Hospitals that did not respond to the initial request were sent a second request and were contacted by phone.

We used information from the 2009 American Hospital Association Annual Survey Database to describe the characteristics of all hospitals in the US, hospitals that were sent a medical record request, and hospitals that fulfilled the request; we were unable to obtain hospital characteristics for all hospitals identified as having a potential case because of unavailable Medicare Provider Numbers. We described maternal characteristics available in claims data of potential cases that had medical records available and those of potential cases that did not have records available, either because we did not request a record or because the hospital did not send the requested record.

Step 6: Medical record abstraction—Medical records were abstracted by trained medical record reviewers using a standardized abstraction database designed by the study investigators. The first 50 record abstractions were re-abstracted, compared for quality control, and used to refine the abstraction database where necessary. The criteria used to confirm the outcomes with the abstracted data are listed in Table 1. We also assessed the validity of the definitions for nulliparity, labor induction, cesarean delivery, and preterm delivery (Appendix 1) among potential cases.

Step 7: Claims and redacted medical records review—To understand the sources of disagreement between claims and medical records, we reviewed the claims in MAX for the unconfirmed preeclampsia cases that did not have evidence of high blood pressure or a hypertension diagnosis in their delivery record and the claims for all unconfirmed PPHN and cardiac malformation cases. We also reviewed redacted medical records when the reasons for the inconsistency were unclear from claims alone.

Statistical analysis

We calculated the positive predictive value (PPV), i.e., the proportion of potential cases identified from MAX data that were confirmed by hospital medical record review, and 95% Wilson confidence intervals (CI). The PPVs for the potential preeclampsia cases were weighted to match the proportion of pregnancies in the depression-related diagnosis and antidepressant dispensing strata among all potential preeclampsia cases identified. Also, the

PPVs for the potential cardiac malformation cases were weighted to match the case mix distribution (i.e., VSD, RVOTO, and other cardiac malformation cases) among all potential cardiac malformation cases identified. In addition, to test and improve our algorithms for case identification in MAX, we applied alternative identification criteria. For preeclampsia, we restricted potential cases to those diagnosed during a hospitalization or with severe preeclampsia or eclampsia ICD-9 codes (642.5x or 6426x). For cardiac malformation, we additionally required procedure codes for cardiac surgery (Appendix 2). We also required diagnosis for other cardiac malformations on at least two dates. The reference standard medical records were from a single hospitalization; therefore, we did not have medical records from outpatient visits and from hospitalizations that occurred subsequent to a transfer from the original hospital. In secondary analyses, PPVs were estimated for potential cases likely to have complete medical record information, i.e., for women who were diagnosed with preeclampsia on or before the delivery date, and for infants who were not transferred to another hospital.

Results

Record requests were fulfilled by 168 (44%) different hospitals from 32 states. Hospitals were unable to fulfill the medical record request for various reasons including the following: the record was too old to locate, first and last names were necessary to locate the record, or the hospital required a signed letter of patient consent for record release, approval from their own institutional review board, or did not participate in research. Compared with all US hospitals, those that were sent a record request (i.e., had hospital information available and had at least one patient selected for validation) were less often from the Northeast, and were more often not-for-profit and accredited by the Joint Commission on Accreditation of Health Care Organizations (JCAHO) (Table 2). They also had a residency training program and a neonatal intensive care unit more often and had higher volume, as evidenced by higher median hospital beds, admissions, births, and personnel. Compared with all hospitals that were sent a record request, those that fulfilled the request less often had a residency training program and had lower volume.

Of potential cases selected for the record request, records were available for 183 (43%) preeclampsia, 82 (32%) PPHN, and 158 (24%) cardiac malformation (29% with records requested from the time of delivery, and 12% with records requested from after the time of delivery) potential cases. Maternal characteristics are listed in Appendix 3 according to whether or not the medical record was available. No consistent differences in maternal characteristics were observed across the different outcomes with the exception that medical records were available slightly more often for white women.

The PPV for any preeclampsia diagnosis was 66.5% (95% CI: 53.6, 77.4%). When restricting to potential cases with the first preeclampsia diagnosis on or before the delivery date, the PPV was 69.5% (95% CI: 56.0, 80.3%). The PPV for preeclampsia improved when using alternate identification criteria. The PPV was 94.5% (95% CI: 84.0, 98.3%) among potential cases with preeclampsia diagnoses recorded during a hospitalization (75% of potential cases), and it was 80.3% (95% CI: 53.3, 93.6%) among potential cases with severe preeclampsia or eclampsia diagnoses. There was no evidence of differential

misclassification by depression diagnosis or antidepressant dispensing status (Table 3). Of the 16 potential preeclampsia cases that were unconfirmed and had no evidence of high blood pressure or hypertension diagnosis in the delivery record, two had claims for preeclampsia diagnoses during the delivery hospitalization recorded in the inpatient claims file. The rest had preeclampsia diagnoses recorded in the outpatient claims file only; three women had their first preeclampsia diagnosis in the week to month following delivery and one had diagnoses both before and after the delivery hospitalization.

The PPV for PPHN was 68.3% (95% CI: 57.6, 77.4%), but when restricting to the potential cases that were not transferred to other hospitals (59% of potential cases), the PPV increased to 89.6% (95% CI: 77.8, 95.5%). Of the confirmed cases, 95% had evidence of severe respiratory distress; 70% had evidence of patent ductus arteriosus, patent foramen ovale, or other atrial septal defect, according to their medical record. According to claims profiles, unconfirmed PPHN cases typically had other respiratory problems or cardiac malformations related to preterm delivery and may have been identified as having PPHN because of a rule-out diagnosis.

The PPV for any cardiac malformation among potential cases with diagnostic codes on at least 2 dates was 77.6% (95% CI: 65.7, 86.2%) overall, and it was 76.2% (95% CI: 54.9, 89.4%) for VSD, although one of the cases only had patent ductus arteriosus recorded in the medical record at the time of hospital discharge (Table 4). The PPV would be 71.4% if that case was considered unconfirmed. There were only 3 potential RVOTO cases. The PPV was 79.5% (95% CI: 64.5, 89.2%) for potential cases with diagnostic codes for other malformations on at least 2 dates, and when requiring only 1 diagnosis date, the number of potential cases increased from 719 to 3,689, but the PPV decreased to 66.0% (95% CI: 56.1, 74.6%). From the claims review of the 44 potential cases that did not have evidence of a cardiac malformation in their medical record, we identified 7 (16% of the unconfirmed) pregnancies in which the mother had claims for the outcome, while the infant appeared to be healthy. We identified 9 (21% of the unconfirmed) pregnancies in which the code appeared to be a rule out diagnosis or a coding error (i.e., no other codes for cardiac malformations or interventions followed). We also identified 16 (36% of the unconfirmed) pregnancies that appeared to be cases because of claims for diagnoses on several dates and/or cardiac surgical procedures. These claims appeared after a likely hospital transfer or several days after the date of birth; therefore the medical records that were available did not cover these time periods of interest. The reason for the diagnosis in the remaining 27% was unclear.

Among potential preeclampsia cases, the PPVs for labor induction and cesarean delivery were nearly 100% (Table 5), while the PPV for preterm delivery was 75%.

Discussion

Conducting a validation study of MAX data was not straightforward because of the lack of personal and hospital identifiers. However, we have gained several important insights regarding the accuracy of data for research and the validation process itself. We observed that PPVs varied by outcome and identification criteria, ranging from 66–95%. More stringent criteria for outcome identification in claims resulted in lower sensitivity and fewer

potential cases but higher PPV. Based on the review of claims of unconfirmed cases, it seems that the PPVs are conservative estimates.

We only had medical record information available from one hospitalization, which was most often from the time of delivery. It is possible that we were unable to confirm preeclampsia cases with outpatient diagnoses only because we did not have outpatient medical records available. Furthermore, we suspect that we would have confirmed more PPHN and cardiac malformation cases if infants' records were available from hospital transfers and from outpatient visits. Many unconfirmed cases had multiple claims for the diagnoses and related surgeries in the weeks following delivery and were presumably transferred to a children's hospital with specialized units. The imperfect reference standard resulted in an underestimation of the PPV across outcomes, and we believe the PPV estimates would have increased if transfer-hospital and outpatient records been available.

Compared with any inpatient or outpatient preeclampsia diagnosis, restriction to preeclampsia diagnosis during a hospitalization resulted in higher PPV. If we had access to outpatient medical records, we may have confirmed some of the outpatient diagnoses. The PPV for a preeclampsia diagnosis made during a hospitalization was similar to or higher than findings from previous studies, which were from Sweden and Denmark.[10,12] The PPV we estimated for preeclampsia among those with severe preeclampsia diagnoses was similar to findings from previous US studies.[11,16]

The PPV for PPHN increased when restricting to cases that were not transferred, i.e., those with the most complete reference standard. The true PPV for PPHN is probably between 68%, the original estimate, and 90%, the estimate restricted to infants that were not transferred. The PPVs for cardiac malformations are in line with previous estimates and higher than the PPV from birth certificate data.[19,22] Cooper et al found that PPVs improved when both diagnostic and surgical procedure codes were required.[22] We found that requiring at least two diagnosis dates or a surgical procedure improved the PPV for cardiac malformations, but reduced the number of potential cases of which some were true positives.

The PPV for preterm delivery among potential preeclampsia cases (75%) was the same as that found for all women in one study.[18] However it was lower than two other studies that used more restrictive outcome identification criteria.[17,19] Restrictive identification criteria for preterm birth would likely improve validity while reducing the number of true cases identified; the optimal criteria may depend on the purpose of identifying preterm delivery, i.e., to study preterm delivery as an outcome or to estimate gestational length for exposure assessment. The validity of labor induction and cesarean delivery was excellent among women with preeclampsia diagnoses, and was similar to or higher than previous reports from hospital discharge data.[14–16] The value of the Medicaid eligibility variable on the estimated date of the last menstrual period can be used to identify multiparity in MAX data.

Our study had some additional limitations. First, the records were not randomly selected for validation (we had to select cases from those with available hospital contact information), and the proportion of requested records that were released for review was between 24–43%. However, we demonstrated that potential cases with available records and those without

available records were fairly similar with respect to measured maternal characteristics. We observed that hospitals that fulfilled record requests were slightly less often teaching hospitals and tended to be smaller compared to all hospitals that were sent a record request. Our results suggest that the unavailability of medical records for potential cases, primarily because of missing hospital identifiers, was random with respect to maternal characteristics, and we are assuming that the probability of being a true case does not depend on the availability of hospital contact information and the likelihood that the hospitals provided the records. Therefore, we expect our results to generalize to all potential cases in the cohort. Second, the PPVs for multiparity, labor induction, cesarean delivery, and preterm delivery were estimated among women identified as potential cases. The prevalence of these factors differs in the full cohort and among women identified as having preeclampsia; therefore, these PPVs may not generalize to the full cohort. Third, by including claims from both mothers and infants in our outcome definitions, we falsely classified infants as having outcomes when the diagnosis actually belonged to the mothers. Based on this information, we will modify our protocols accordingly for future research. The cohort that we identified consists of live births only, the woman-infant linkage method has not yet been validated, and to ensure complete follow up throughout pregnancy, we implemented restrictive eligibility criteria. These limitations should be considered when conducting studies of medication safety with these data.

While it is feasible to conduct a medical record review study using nationwide MAX data, there were many obstacles to obtaining records. The number of pregnancies available for medical record requests was limited by the lack of usable hospital identifiers in MAX. A centralized file of hospital contact information for Medicaid Billing Provider Numbers has since been made available by CMS.[28] The highest proportion of fulfilled record requests was for potential preeclampsia cases; hospitals had greater success locating women's records than locating infants' records based on the women's SSNs. The proportion of fulfilled record requests was higher for potential cardiac malformation cases in which we requested records from the time of delivery compared with after delivery. Future validation studies within MAX may focus on maternal outcomes and infant records from the time of delivery to increase the yield of records available for abstraction. Also, the number of records returned by hospitals could be improved if additional personal identifiers, e.g., first and last name, were available for record validation purposes.

This was the first validation study of maternal and neonatal outcomes identified within nationwide Medicaid data. For our studies of antidepressant safety during pregnancy, we identified outcome definitions that have high PPV, quantified the degree of outcome misclassification and can correct for it using sensitivity analysis,[29,30] and were able to rule out differential misclassification of preeclampsia by antidepressant and depression status. Although numbers were small, there was no evidence to suggest non-differential misclassification of PPHN and cardiac malformations. Furthermore, by reviewing both medical record and claims information, we learned that hospital medical records from the time of delivery may not be the gold standard and therefore provide conservative PPV estimates. We observed that requiring multiple diagnostic codes or both a diagnostic and a surgical procedure code to identify cardiac malformation cases and requiring preeclampsia diagnosis codes during a hospitalization to identify preeclampsia cases increased PPV but

lowered the number of cases identified. Although stricter outcome criteria decrease sensitivity, relative risks are unbiased in the presence of outcome misclassification with nondifferential sensitivity and perfect specificity.[31] In future studies, the use of strict outcome definitions with higher specificity is justified even at the cost of identifying fewer cases and lower sensitivity.

In conclusion, this study demonstrated that MAX data can be used to validly study several maternal and neonatal outcomes. There were barriers to obtaining hospital records for individuals identified from MAX data, but at least some of these obstacles could be removed to facilitate validation studies. The PPV for preeclampsia diagnoses made during a hospitalization was above 90%. Our best estimate of the PPV for PPHN is between 70–90% and the PPV for cardiac malformations on at least 2 dates is between 75–85%, had records been available from after the hospital transfer. These PPVs can be used to perform measurement error correction in epidemiologic studies.[29,30]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Garcia, G. [Accessed March 20, 2013] Maternal and child health (MCH) update: states increase eligibility for children's health in 2007. 2008. http://www.nga.org/files/live/sites/NGA/files/pdf/0811MCHUPDATE.PDF;jsessionid=7B47A647247DD4E5CB9B709C8F9797AE

2. Research Data Assistance Center. [Accessed September 1, 2013] Find a CMS Data File. http://www.resdac.org/cms-data/search?f[0]=im_field_program_type%3A2

3. Bateman BT, Hernandez-Diaz S, Huybrechts KF, et al. Patterns of outpatient antihypertensive medication use during pregnancy in a Medicaid population. Hypertension. 2012; 60(4):913–920. [PubMed: 22966012]

4. Huybrechts KF, Palmsten K, Mogun H, et al. National trends in antidepressant medication treatment among publicly insured pregnant women. Gen Hosp Psychiatry. 2013; 35(3):265–271. [PubMed: 23374897]

5. Palmsten K, Huybrechts KF, Michels KB, et al. Antidepressant Use and Risk for Preeclampsia. Epidemiology. 2013; 24(5):682–691. [PubMed: 23873072]

6. Palmsten K, Hernandez-Diaz S, Huybrechts KF, et al. Use of antidepressants near delivery and risk of postpartum hemorrhage: cohort study of low income women in the United States. BMJ. 2013; 347:f4877. [PubMed: 23965506]

7. Palmsten K, Huybrechts KF, Mogun H, et al. Harnessing the Medicaid Analytic eXtract (MAX) to Evaluate Medications in Pregnancy: Design Considerations. PLoS ONE. 2013; 8(6):e67405. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0067405. [PubMed: 23840692]

8. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005; 58(4):323–337. [PubMed: 15862718]

9. West, SL.; Ritchey, ME.; Poole, C. Validity of pharmacoepidemiologic drug and diagnosis data. In: Strom, BL.; Kimmel, SE.; Hennessy, S., editors. Pharmacoepidemiology. 5. Wiley-Blackwell; Chichester, West Sussex, UK: 2012. p. 757-794.

10. Ros HS, Cnattingius S, Lipworth L. Comparison of risk factors for preeclampsia and gestational hypertension in a population-based cohort study. Am J Epidemiol. 1998; 147(11):1062–1070. [PubMed: 9620050]

11. Geller SE, Ahmed S, Brown ML, Cox SM, Rosenberg D, Kilpatrick SJ. International Classification of Diseases-9th revision coding for preeclampsia: how accurate is it? Am J Obstet Gynecol. 2004; 190(6):1629–1633. [PubMed: 15284758]

12. Klemmensen AK, Olsen SF, Osterdal ML, Tabor A. Validity of preeclampsia-related diagnoses recorded in a national hospital registry and in a postpartum interview of the women. Am J Epidemiol. 2007; 166(2):117–124. [PubMed: 17556761]

13. Korst LM, Gregory KD, Gornbein JA. Elective primary caesarean delivery: accuracy of administrative data. Paediatr Perinat Epidemiol. 2004; 18(2):112–119. [PubMed: 14996250]

14. Romano PS, Yasmeen S, Schembri ME, Keyzer JM, Gilbert WM. Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. Obstet Gynecol. 2005; 106(4):717–725. [PubMed: 16199627]

15. Lydon-Rochelle MT, Holt VL, Nelson JC, et al. Accuracy of reporting maternal inhospital diagnoses and intrapartum procedures in Washington State linked birth records. Paediatr Perinat Epidemiol. 2005; 19(6):460–471. [PubMed: 16269074]

16. Yasmeen S, Romano PS, Schembri ME, Keyzer JM, Gilbert WM. Accuracy of obstetric diagnoses and procedures in hospital discharge data. Am J Obstet Gynecol. 2006; 194(4):992–1001. [PubMed: 16580288]

17. Eworuke E, Hampp C, Saidi A, Winterstein AG. An algorithm to identify preterm infants in administrative claims data. Pharmacoepidemiol Drug Saf. 2012; 21(6):640–650. [PubMed: 22504840]

18. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernández-Díaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. Pharmacoepidemiol Drug Saf. 2013; 22(1):16–24. [PubMed: 22550030]

19. Andrade SE, Scott PE, Davis RL, et al. Validity of health plan and birth certificate data for pregnancy research. Pharmacoepidemiol Drug Saf. 2013; 22(1):7–15. [PubMed: 22753079]

20. Hexter AC, Harris JA, Roeper P, Croen LA, Krueger P, Gant D. Evaluation of the hospital discharge diagnoses index and the birth certificate as sources of information on birth defects. Public Health Rep. 1990; 105(3):296–307. [PubMed: 2113690]

21. Frohnert BK, Lussky RC, Alms MA, Mendelsohn NJ, Symonik DM, Falken MC. Validity of hospital discharge data for identifying infants with cardiac defects. J Perinatol. 2005; 25(11):737–742. [PubMed: 16163368]

22. Cooper WO, Hernández-Díaz S, Gideon P, et al. Positive predictive value of computerized records for major congenital malformations. Pharmacoepidemiol Drug Saf. 2008; 17(5):455–460. [PubMed: 18081215]

23. Hennessy S, Leonard CE, Bilker WB. Researchers and HIPAA. Epidemiology. 2007; 18(4):518. [PubMed: 17568226]

24. Hennessy S, Leonard CE, Freeman CP, et al. Validation of diagnostic codes for outpatient-originating sudden cardiac death and ventricular arrhythmia in Medicaid and Medicare claims data. Pharmacoepidemiol Drug Saf. 2010; 19(6):555–562. [PubMed: 19844945]

25. Centers for Medicare & Medicaid Services. [Accessed March 20, 2013] Medicaid Analytic eXtract (MAX) General Information. MAX 1999–2005 state claims anomalies from the "2005 Files" zipped file within the "MAX Data 2005 to 2008 General Information, Data Dictionaries, Data Element Lists, Data Anomalies, Validation Table Measures and SAS Loads" zipped file. http://

www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/
MedicaidDataSourcesGenInfo/MAXGeneralInformation.html

26. Research Data Assistance Center. CMS: 102 Conducting Research with Medicaid Claims Data [workshop]; Minneapolis, MN. September 16–17, 2010;

27. Centers for Medicare and Medicaid Services. [Accessed May 6, 2013] Data Dissemination. http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProvIdentStand/DataDissemination.html

28. Bencio, D.; Sykes, J. [Accessed August 30, 2013] Medicaid Analytic Extract Provider Characteristics (MAXPC) Evaluation Report. 2009. http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/MAXPC_2009_Final_Eval_Rpt.pdf

29. 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. [PubMed: 16172102]

30. Lash, TL.; Fox, MP.; Fink, AK. Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer; 2009. Missclassification; p. 79-108.

31. Greenland, S.; Lash, TL. Bias Analysis. In: Rothman, KJ.; Greenland, S.; Lash, TL., editors. Modern Epidemiology. 3. Lippincott Williams & Wilkins; Philadelphia, PA: 2008. p. 358-359.

Key Points

- The accuracy of pregnancy-related outcomes in nationwide Medicaid data had not been established.

- By reviewing both medical record and claims information, we learned that hospital medical records from the time of delivery may not be the gold standard for the validation of perinatal diagnoses in claims data and therefore provide conservative positive predictive value (PPV) estimates.

- The PPV for preeclampsia diagnoses made during a hospitalization was above 90%. Our best estimate of the PPV for PPHN is between 70–90% and the PPV for cardiac malformations on at least 2 dates is between 75–85%, had records been available from after hospital transfers.

- These PPVs can be used to perform measurement error correction in epidemiologic studies.
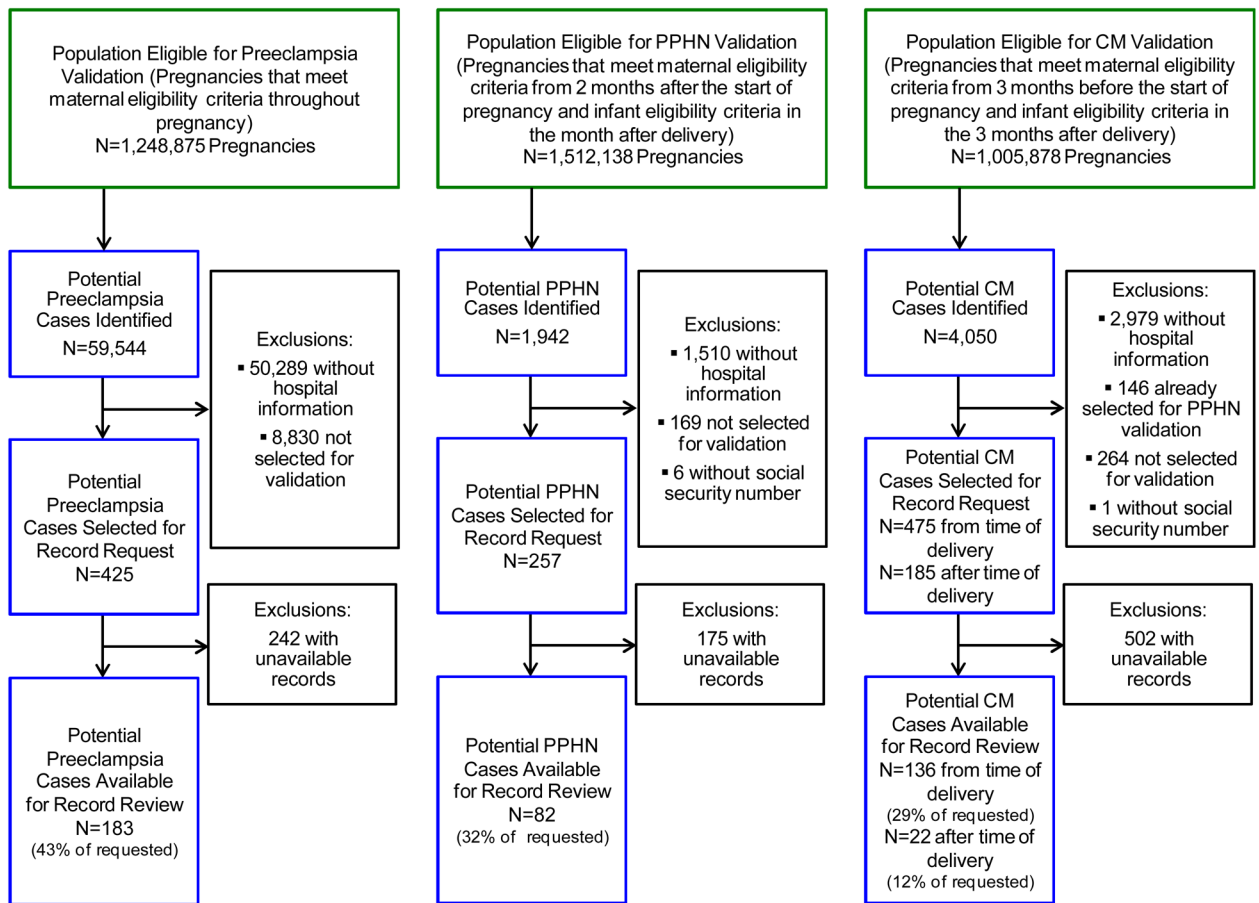
Figure 1.
Flow chart of pregnancies included in the validation study, Medicaid Analytic eXtract, 2000–2007.

Table 1

Identification Criteria Used to Identify Pregnancies With Potential Outcomes and Validation Criteria Used to Confirm the Outcomes.

Outcome	Identification Criteria	Type of Medical Record	Validation Criteria
Preeclampsia	Maternal ICD-9-CM diagnosis codes 642.4x, 642.5x, 642.6x, or 642.7x recorded in inpatient or outpatient records after 140 gestational days and within 30 days after the delivery date.	Maternal hospital record from the time of delivery.	Any type of preeclampsia diagnosis (mild, severe, superimposed, or eclampsia) recorded. Alternatively, chronic hypertension or gestational hypertension diagnosis or documentation of systolic blood pressure at least 140mm Hg or diastolic blood pressure at least 90 mm Hg, and proteinuria diagnosis or documentation of urine protein dipstick 2+ or urinary excretion 300mg for 24 hour specimen recorded.
Persistent Pulmonary Hypertension of the Newborn (PPHN)	Maternal or infant ICD-9-CM diagnosis codes 416.0x or 747.83 recorded in inpatient records within 30 days after the DOB.	Infant hospital record from the time of delivery.	Pulmonary hypertension or persistent fetal circulation recorded. Alternatively, severe respiratory failure (i.e., documentation of asphyxia, cyanotic congenital heart disease, cyanosis, respiratory distress syndrome, respiratory arrest, intubation, mechanical ventilation, oxygen therapy, nasal continuous positive airway pressure, high frequency ventilation, extracorporeal membrane oxygenation, or nitric oxide) and evidence of pulmonary hypertension (i.e., documentation from echocardiogram of right-to-left hemodynamic shunt or documentation from cardiac catheterization of 5% gradient between preductal and postductal oxygen).
Cardiac Malformations			
Ventricular Septal Defect (VSD)	Maternal or infant ICD-9-CM diagnosis codes for 745.4x recorded in inpatient records on at least 2 dates within the first 90 days after the DOB and no ICD-9-CM diagnosis codes for 758.x or 759.81–759.83 (chromosomal abnormalities) recorded in inpatient or outpatient records within 90 days after the DOB.	Infant hospital record from the time of delivery or after the delivery.	Cardiac malformation diagnosis recorded.
Right Ventricular Outflow Tract Obstruction (RVOTO)	Maternal or infant ICD-9-CM diagnosis codes 747.3x or 746.02 and no indication of preterm delivery, or 746.01, 746.09, or 746.83 recorded on at least 2 dates in inpatient records within the first 90 days after the DOB and no ICD-9-CM diagnosis codes for 758.x or 759.81–759.83 (chromosomal abnormalities) recorded in inpatient or outpatient records within 90 days after the DOB.	Infant hospital record from the time of delivery or after the delivery.	Cardiac malformation diagnosis recorded.
Other Cardiac Malformation	Maternal or infant ICD-9-CM diagnosis codes 745, 745.0x-745.3x, 745.6x-745.9x, 746, 746.00, 746.1x-746.3x, 746.5x, 746.7x, 746.8, 746.80–746.82, 746.84–746.89, 747, 747.1x–747.2x, 747.4x, 747.6, 747.60–747.68, or 747.8x in inpatient records and no ICD- 9-CM diagnosis codes for 758.x or 759.81–759.83 (chromosomal abnormalities) recorded in inpatient or outpatient records within 90 days after the DOB.	Infant hospital record from the time of delivery or after the delivery.	Cardiac malformation diagnosis recorded.

Abbreviations: DOB, date of birth; ICD-9-CM, International Classification of Diseases, 9th Revision, Clinical Modification.

Table 2

Hospital Characteristics Among All Hospitals, Those That Were Sent a Request, and Those That Fulfilled a Request.

Hospital Characteristics	All Hospitals* (n=6,342)		Hospitals That Were Sent a Record Request † (n=380)		Hospitals That Fulfilled a Record Request‡ (n=168)	
	N	%	N	%	N	%
Region						
Midwest	873	13.8	65	17.1	27	16.1
Northeast	2578	40.7	127	33.4	55	32.7
South	1728	27.3	139	36.6	58	34.5
West	1163	18.3	49	12.9	28	16.7
Hospital Policy Authority						
Nonfederal Government	1416	23.2	74	21.6	32	20.7
Not-For-Profit (Non-Government)	3148	51.6	214	70.5	111	71.6
For-Profit	1541	25.2	27	7.9	12	7.7
Primary Type of Service						
General Medical and Surgical	4822	89.1	329	96.2	155	100
Children's General	57	1.1	10	2.9	0	0
Other	536	9.9	3	0.9	0	0
Accreditation by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO)	4161	65.6	290	84.8	126	81.3
Residency Training Approved by the Accreditation Council for Graduate Medical Education	1094	17.3	100	29.2	36	23.2
Neonatal Intensive Care Facility	1436	29.7	161	50.6	66	46.5
Neonatal Intermediate Care Facility	1138	23.5	120	37.7	54	38.0
	Median	IQ Range	Median	IQ Range	Median	IQ Range
Total Hospital Beds	87	169	205.5	281	195	284
Obstetric Care Beds	3	16	17	24	17	24
Total Admissions	2285.5	7296	9732	14417	9364	14907
Total Births (Excluding Fetal Deaths)	36.5	784	1043	1766	1030	1725
Full Time Equivalent Total Personnel	362.5	830	1076.5	10461	947	1528

Abbreviation: IQ, interquartile.

*
Missing information is not noted.

†
38 hospitals were missing information on all hospital characteristics except region, additional missing information is not noted.

‡
13 hospitals were missing information on all hospital characteristics except region, additional missing information is not noted.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Proportions of Pregnancies Confirmed to Have Preeclampsia.

Identification Criteria	Number of Records Reviewed	Number Confirmed	Positive Predictive Value (95% Confidence Interval)
Any preeclampsia *	183	121	66.5 (53.6, 77.4)
No depression and no antidepressants	42	28	66.7 (51.6, 79.0)
No depression and antidepressants	55	40	72.7 (59.8, 82.7)
Depression and no antidepressants	49	29	59.2 (45.3, 71.8)
Depression and antidepressants	37	24	64.9 (48.8, 78.2)
Preeclampsia diagnosis before or on delivery date *	168	120	69.5 (56.0, 80.3)
Preeclampsia during a hospitalization *†	131	114	94.5 (84.0, 98.3)
No depression and no antidepressants	26	25	96.2 (81.1, 99.3)
No depression and antidepressants	43	39	90.7 (78.4, 96.3)
Depression and no antidepressants	36	27	75.0 (58.9, 86.3)
Depression and antidepressants	26	23	88.5 (71.0, 96.0)
Severe preeclampsia *	52	41	80.3 (53.3, 93.6)
Severe preeclampsia during a hospitalization *†	44	39	98.5 (89.3, 99.8)
Preeclampsia never during a hospitalization *†	52	7	17.5 (6.8, 38.3)

*
Indicates weighted positive predictive values and 95% confidence intervals

†
The diagnosis date occurred during a hospital admission and was recorded in the inpatient file or the other therapy file.

Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2015 June 01.

Table 4

Proportions of Pregnancies Confirmed to Have Persistent Pulmonary Hypertension of the Newborn (PPHN) and Cardiac Malformations.

Identification Criteria	Number of Records Reviewed	Number Confirmed	Positive Predictive Value (95% Confidence Interval)
Any PPHN	82	56	68.3 (57.6, 77.4)
PPHN, restricting to those who were not transferred to another facility	48	43	89.6 (77.8, 95.5)
Cardiac malformations, any diagnosis on >1 date *	63	49	77.6 (65.7, 86.2)
VSD diagnosis on >1 date†	21	16	76.2 (54.9, 89.4)
RVOTO diagnosis on >1 date	3	2	66.7 (20.8, 93.9)
Other cardiac malformation diagnosis on >1 date	39	31	79.5 (64.5, 89.2)
Other cardiac malformation diagnosis on 1 date only	97	64	66.0 (56.1, 74.6)
Other cardiac malformation diagnosis on 1 date only and cardiac surgery	22	15	68.2 (47.3, 83.6)
Cardiac malformations, any diagnosis on >1 date, or other malformation diagnosis on 1 date and cardiac surgery *	83	63	75.1 (64.4, 83.4)
Cardiac malformations, any diagnosis on >1 date and cardiac surgery *	27	23	76.6 (49.7, 91.7)

Abbreviations: RVOTO, right ventricular outflow tract obstruction; VSD, ventricular septal defect.

*
Indicates weighted positive predictive values and 95% confidence intervals

†
One of the cases had only patent ductus arteriosus recorded in the medical record at the time of hospital discharge. The positive predictive value would be 71.4% (95% CI: 50.0–86.2) if that case was considered unconfirmed.

Table 5

Proportions of Pregnancies Confirmed to Have Obstetric Factors Among Potential Preeclampsia Cases With Available Records.

Identification Criteria	Number of Records Reviewed	Number Confirmed	Positive Predictive Value (95% Confidence Interval)
Multiparity	129	112	86.6 (79.9, 91.6)
Induction	58	56	96.6 (88.3, 99.1)
Cesarean Delivery	75	74	98.7 (92.8, 99.8)
Preterm Delivery*	47	35	74.5 (60.5, 84.8)

*
The PPV for preterm delivery among potential PPHN cases with available records was 48.7% (95% CI: 33.9, 63.8), and it was 73.6% (95% CI: 62.4–82.4) among potential cardiac malformation cases.