

RESEARCH

Open Access



# Concordance of neonatal critical condition data between secondary databases: Florida and Texas birth certificate Linkage with medicaid analytic extract

Yasser Albogami<sup>1,2</sup>, Yanmin Zhu<sup>1</sup>, Xi Wang<sup>1</sup> and Almut G Winterstein<sup>1,3\*</sup>

## Abstract

**Background** Limited information is available about neonates' critical conditions data quality. The study aim was to measure the agreement regarding presence of neonatal critical conditions between Medicaid Analytic eXtract claims data and Birth Certificate (BC) records.

**Methods** Claims data files of neonates born between 1999–2010 and their mothers were linked to birth certificates in the states of Texas and Florida. In claims data, neonatal critical conditions were identified using medical encounter claims records within the first 30 days postpartum, while in birth certificates, the conditions were identified based on predetermined variables. We calculated the prevalence of cases within each data source that were identified by its comparator, in addition to calculating overall agreement and kappa statistics.

**Results** The sample included 558,224 and 981,120 neonates in Florida and Texas, respectively. Kappa values show poor agreement (< 20%) for all critical conditions except neonatal intensive care unit (NICU) admission, which showed moderate (> 50%) and substantial (> 60%) agreement in Florida and Texas, respectively. claims data resulted in higher prevalences and capture of a larger proportion of cases than the BC, except for assisted ventilation.

**Conclusions** Claims data and BC showed low agreement on neonatal critical conditions except for NICU admission. Each data source identified cases most of which the comparator failed to capture, with higher prevalences estimated within claims data except for assisted ventilation.

**Keywords** Medicaid, Claims data, Birth certificates, Concordance, Neonatal complications, Measurement, Agreement, Sensitivity, Specificity, False positive

## Introduction

Randomized clinical trials tend to exclude vulnerable populations such as pregnant women and neonates, children under 30 days of age, for ethical reasons [1, 2], necessitating observational studies, using prospective cohorts or real-world data to be the cornerstone for perinatal drug safety and effectiveness assessments. Studies that assess maternal and/or neonatal health outcomes rely commonly on administrative data sources including Birth Certificates (BCs) and claims data [3–9].

\*Correspondence:

Almut G Winterstein  
almut@ufl.edu

<sup>1</sup> Department of Pharmaceutical Outcomes and Policy, College of Pharmacy, University of Florida, Gainesville, Florida, USA

<sup>2</sup> Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

<sup>3</sup> Center for Drug Evaluation and Safety, University of Florida College of Pharmacy, Gainesville, Florida, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

BCs have been utilized widely for public health surveillance in maternal and neonatal epidemiology because they collect detailed information on delivery and the neonate's condition in addition to information pertaining to pregnancy course and maternal characteristics. Claims data, on the other hand, provide rich longitudinal information on healthcare utilization that allows inferences about maternal clinical history and drug exposure before and during pregnancy and outcomes in mothers and infants after delivery. Because both data sources are not collected for research purposes, data quality is of concern. For studies that aim to make causal inferences about drug exposure and outcomes, misclassification bias, especially low specificity of outcome definitions [10], can bias the estimated exposure-outcome association. Accordingly, some studies have evaluated the quality of a number of variables in both BCs and claims data for multiple states [11–15], while others have used variables in BCs and/or claims data with unknown sensitivity (the ability to identify true cases) or specificity (the ability to identify true non-cases) [16–18].

Studies that have evaluated the validity of either BCs or claims data, usually against medical records, have centered around maternal variables including method of delivery, diabetes, hypertension, gestational age [19, 20], and major birth defects such as cleft palate or heart malformations [13, 15, 21]. Limited information is available about neonates' critical conditions data quality such as birth injury or respiratory distress syndrome. This study aimed to evaluate the concordance between BCs and claims data on several neonatal critical conditions and quantified the extent of possible false positive cases in each data source.

## Methods

### Data Sources

Texas and Florida BC data from 1999–2010 were harvested and deterministically linked to Medicaid claims data. BC data contain information on neonates' demographics, date of birth, gestational age, weight, complications during labor, conditions and maternal data. The Medicaid Analytic eXtract (MAX) files include in- and outpatient encounter claims for neonates and their mothers in addition to pharmacy claims of dispensed prescriptions. We required that both neonates and their mothers be continuously enrolled in the insurance for at least 30 days postpartum to capture neonatal health information in claims data.

### BC and MAX linkage

Mothers and neonates in MAX and BCs were linked using a two-step deterministic linkage procedure [22]. In brief, if both the mother's and neonate's Social Security

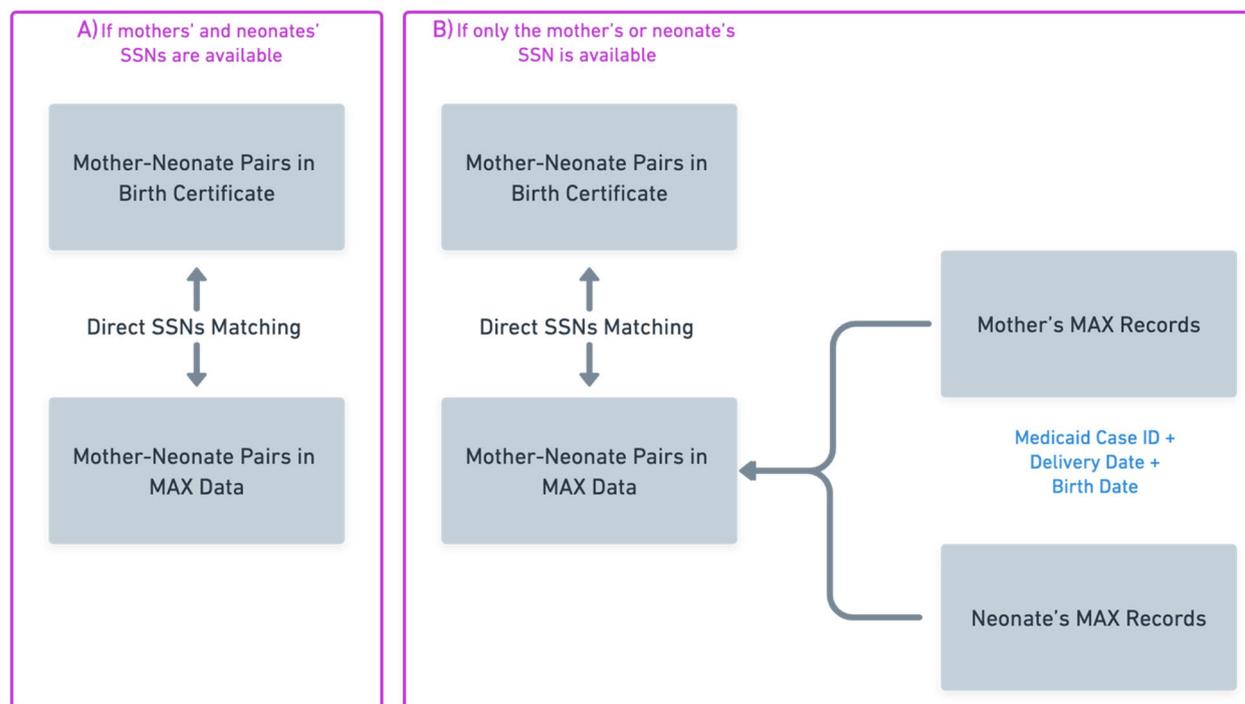
Numbers (SSNs) were available for a mother-neonate pair in MAX and BC, both mothers and neonates were linked directly using exact SSN matching. If only the mother's or neonate's SSN was available, we employed an established linkage algorithm for both Texas and Florida [23]. We first linked neonates and the mothers in MAX using a Medicaid family identifier (i.e., the case ID) and delivery claim dates. Neonates' birth date was then matched to delivery. The established algorithm was assessed for Texas and Florida showing >99% of the positive predictive value (the proportion of true cases among total identified cases) [24]. After creating the mother-neonate pairs in MAX, the available SSN was then used to link the mother-neonate pairs in MAX to the BC (Fig. 1).

### Measurement of variables in MAX/BC

We measured the occurrence of birth injury, assisted ventilation (AV), seizure, respiratory distress syndrome (RDS) and neonatal intensive care unit (NICU) admission for each eligible neonate in both databases. In MAX data, the cases were ascertained based on the presence of at least one inpatient or outpatient encounter with condition-specific diagnosis codes in any diagnoses field within 30 days of birth on either neonates or mothers' records using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) or Current Procedural Terminology (CPT) codes (Table 1). The claims codes to identify NICU admission, RDS, AV and seizures have been previously validated against medical records and have shown high sensitivity and specificity or positive predictive value [14–16, 25, 26]. In BCs, clinical measurements are recorded manually in an official form which is subsequently entered in the birth certificate database. A certified registrars, midwives and healthcare providers obtain the clinical information, according to the institution's policy [27, 28]. Therefore, a condition was flagged for a neonate if it was checked on the BC form, i.e., given value 1 in the database.

### Statistical analysis

Characteristics of the linked mother-neonate pairs in both states were examined in addition to the prevalence of each condition in each data source. We determined the level of agreement between data sources by calculating crude Kappa statistic. Unlike MAX where a missing diagnosis code is interpreted as absence of a condition, the BCs explicitly specify whether a condition was present or not, i.e., the BC forces selection of yes or no for a condition. Thus, we excluded BC records with missing determination of neonatal critical conditions, which occurred for less than 2% of records. Following usual convention, Kappa was categorized into high (>80%), substantial (61%–80%), moderate (41%–60%), fair (21%–40%) and



**Fig. 1** MAX and BC Linkage Flowchart

poor ( $\leq 20\%$ ) [29]. We also calculated the sensitivity of each data source as if the comparator were considered the gold standard in an attempt to quantify the ability of each source to capture cases identified by the other. For NICU admission, the analysis was conducted from 2004–2010 because this data field was added after 2003 as a result of a new BC form [30]. All analyses were performed using SAS 9.4 statistical software (SAS, Cary, NC). The study was approved by the University of Florida, Centers for Medicare and Medicaid Services, and Florida and Texas Departments of Health Institutional Review and Privacy Boards.

## Results

The study sample was 1,539,344 mother-neonate pairs after successful linkage of MAX and BCs. The sample included 558,224 and 981,120 mother-neonate pairs for Florida and Texas, respectively. The median age of the mothers was 23.7 in Florida and 22.9 in Texas (Table 2). Half of mothers who gave birth in Texas had at most graduated from secondary school. In contrast, mothers in Florida who had at least a high school diploma comprised almost three quarters. About 97% of deliveries occurred in medical facilities in Florida which was lower than in Texas (99.9%). The median number of prenatal visits was eleven in Florida which was slightly higher when compared to Texas. Preterm deliveries

comprised 10.2% of linked pairs in Florida and 11.4% in Texas. Concordantly, we found higher prevalence of NICU admissions, RDS and AV in Texas compared to Florida regardless of data source (Table 3).

The prevalence of neonatal critical conditions was consistently higher in MAX except for AV where the BC captured more than twice as many cases as MAX (Table 3). RDS showed a more than fourfold higher prevalence in MAX than in the BC and seizures were hardly ever captured in the BC. The highest absolute differences in prevalence between MAX and BC was for RDS (4%) in Texas. Figure 2 depicts the dissimilarity between MAX and BCs indicating that the majority of cases was predominantly identified by one data source but not the other. Except for NICU admission, more than 80% of all ascertained cases were captured by either MAX or BC but not both.

The agreement was moderate for NICU admission in Florida (Kappa = 56%) and reached the substantial level in Texas (63%). For respiratory distress syndrome and assisted ventilation, Kappa showed poor agreement between BC and MAX in Texas (19% and 16%) and Florida (16%–15%). There was 8%–10% agreement between MAX and BC for seizures in the two states (Table 4). Birth injury showed extremely low agreement (Kappa 1.5%). Kappa values were generally slightly lower for Florida than Texas with MAX

**Table 1** ICD-9 & CPT Codes Used in MAX.

| Variables   | Type     | Codes  | Description  | Validation                         |
|---|----------|--------|--|------------------------------------|
| Neonatal Intensive Care Unit Admission (NICU Admission) | CPT      | 99,295 | Initial inpatient neonatal critical care ( <i>Deleted 01 Jan 2009</i> )  | PPV = 92%<br>(Medicaid Data [15])  |
|   | CPT      | 99,296 | Subsequent inpatient neonatal critical care ( <i>Deleted 01 Jan 2009</i> )   |                                    |
|   | CPT      | 99,297 | Subsequent inpatient neonatal critical care ( <i>Deleted 01 Jan 2009</i> )   |                                    |
|   | CPT      | 99,298 | Subsequent intensive care ( <i>Deleted 01 Jan 2009</i> )   |                                    |
|   | CPT      | 99,299 | Subsequent intensive care ( <i>Deleted 01 Jan 2009</i> )   |                                    |
|   | CPT      | 99,300 | Subsequent intensive care ( <i>Deleted 01 Jan 2009</i> )   |                                    |
|   | CPT      | 99,468 | Initial inpatient neonatal critical care, per day, for the evaluation and management of a critically ill neonate, 28 days of age or younger  |                                    |
|   | CPT      | 99,469 | Subsequent inpatient neonatal critical care, per day, for the evaluation and management of a critically ill neonate, 28 days of age or younger   |                                    |
|   | CPT      | 99,477 | Initial hospital care, per day, for the evaluation and management of the neonate, 28 days of age or younger, who requires intensive observation, frequent interventions, and other intensive care services |                                    |
|   | CPT      | 99,478 | Subsequent intensive care, per day, for the evaluation and management of the recovering very low birth weight infant (present body weight less than 1500 g)  |                                    |
| Respiratory Distress Syndrome (RDS)                     | CPT      | 99,479 | Subsequent intensive care, per day, for the evaluation and management of the recovering low birth weight infant (present body weight of 1500–2500 g)   | PPV = 97%<br>(Discharge Data [25]) |
|   | CPT      | 99,480 | Subsequent intensive care, per day, for the evaluation and management of the recovering infant (present body weight of 2501–5000 g)  |                                    |
| Seizure   | ICD-9 dx | 769.xx | Respiratory distress syndrome in newborn   | PPV = 86%<br>(Medicaid data [37])  |
| Assisted Ventilation (AV)                               | ICD-9 pc | 96.7   | Other Continuous Invasive Mechanical Ventilation   | PPV = 93%<br>(Discharge data [26]) |
|   | ICD-9 pc | 96.70  | Continuous Invasive Mechanical Ventilation of Unspecified Duration   |                                    |
|   | ICD-9 pc | 96.71  | Continuous Invasive Mechanical Ventilation for Less Than 96 Consecutive Hours  |                                    |
|   | ICD-9 pc | 96.72  | Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours or More  |                                    |
|   | ICD-9 pc | 93.90  | Non-Invasive Mechanical Ventilation  |                                    |
| Birth Injury  | ICD-9 dx | 767.xx | Birth trauma   | NA                                 |

ICD-9 PC International Classification of Diseases, Ninth Revision, Clinical Modification (Procedure Codes)

ICD-9 DX International Classification of Diseases, Ninth Revision, Clinical Modification (Diagnosis Codes)

CPT Current Procedural Terminology

identifying lesser cases captured by the BC, but the BCs of both states showing similar sensitivity to capture cases in MAX.

Assuming that neither data source captured false positives, we found generally better sensitivity for MAX to capture BC cases than vice versa: MAX captured > 80% of NICU admissions that were recorded in Texas and 64% in Florida, while the BC only captured about half of all NICU admissions in MAX (Table 4). The only exception was AV, with about 70% of all identified assisted ventilation cases recorded on the BC and better capture of MAX cases in the BC (about 30%) than vice versa (about 10%).

## Discussion

Our comparison of birth certificate and Medicaid claims data in measuring neonatal critical conditions in two large states found large variation in concordance between the two data sources. Capture of NICU admission showed moderate to substantial agreement between BCs and MAX, while variables such as birth injury and seizures had extremely low agreement. Importantly, except for NICU admission, 70% or more percent of cases in MAX were not captured by BCs in either state. MAX found higher prevalences for all examined conditions except for need for assisted ventilation with more than twice as many cases identified in the BC.

**Table 2** Characteristics of mother and deliveries in the birth Certificate and Medicaid Analytic eExtract linked cohort

| Characteristics                           | FLORIDA<br>(N = 558,224) | TEXAS<br>(N = 981,120) |
|---|--------------------------|------------------------|
| Mother age (median, SD) <sup>a</sup>      | 23.7 (5.7)               | 22.9 (5.3)             |
| Neonate's gender <sup>a</sup>             |                          |                        |
| Male                                      | 284,821 (51.0)           | 499,071 (50.9)         |
| Female                                    | 272,798 (48.9)           | 482,049 (49.1)         |
| Missing                                   | 605 (0.1)                | 0                      |
| Mother race (%) <sup>a</sup>              |                          |                        |
| White                                     | 231,583 (41.5)           | 261,530 (26.7)         |
| Black                                     | 177,530 (31.8)           | 151,157 (15.4)         |
| Hispanic                                  | 121,740 (21.8)           | 518,202 (52.8)         |
| Others                                    | 6579 (1.2)               | 10,557 (1.1)           |
| Missing                                   | 20,792 (3.7)             | 39,674 (4.0)           |
| Mother education (%) <sup>b</sup>         |                          |                        |
| Secondary school or less                  | 149,570 (26.8)           | 488,092 (49.7)         |
| High school                               | 245,514 (44.0)           | 255,125 (26.0)         |
| Undergraduate                             | 147,612 (26.4)           | 210,553 (21.5)         |
| Graduate                                  | 4879 (0.9)               | 23,264 (2.4)           |
| Missing                                   | 10,649 (1.9)             | 3636 (0.4)             |
| Gestational Age (median, SD) <sup>b</sup> | 39 (2.0)                 | 39 (2.0)               |
| Mother alcohol use (%) <sup>b</sup>       |                          |                        |
| Yes                                       | 2385 (0.4)               | 3533 (0.9)             |
| No  | 549,170 (98.4)           | 372,618 (98.8)         |
| Missing                                   | 6669 (1.2)               | 949 (0.3)              |
| Mother tobacco use (%) <sup>b</sup>       |                          |                        |
| Smoking                                   | 76,689 (13.7)            | 45,397 (12.0)          |
| Quitted                                   | 7158 (1.3)               | NA <sup>*</sup>        |
| Not smoking                               | 472,943 (84.7)           | 330,816 (87.7)         |
| Missing                                   | 1434 (0.3)               | 887 (0.3)              |
| Birthplace (%) <sup>b</sup>               |                          |                        |
| Medical facility                          | 540,183 (96.8)           | 980,267 (99.9)         |
| Home                                      | 4450 (0.8)               | 727 (0.07)             |
| Missing                                   | 13,591 (2.4)             | 124 (0.0)              |
| Prenatal visit (median, SD) <sup>b</sup>  | 11 (3.8)                 | 10 (4.0)               |
| Preterm (%) <sup>b</sup>                  |                          |                        |
| Yes                                       | 56,934 (10.2)            | 112,482 (11.4)         |
| No  | 492,731 (88.3)           | 863,893 (88.1)         |
| Missing                                   | 8559 (1.5)               | 4745 (0.5)             |
| Birthweight (%) <sup>b</sup>              |                          |                        |
| Very low (< 1500 g)                       | 6686 (1.2)               | 12,651 (1.3)           |
| Low (1500 – 2499 g)                       | 41,197 (7.4)             | 77,423 (7.9)           |
| Normal (2500 – 3999 g)                    | 472,673 (84.7)           | 841,045 (85.7)         |
| High (> = 4000 g)                         | 37,065 (6.6)             | 50,001 (5.1)           |
| Missing                                   | 603 (0.1)                | 0                      |

<sup>a</sup> data obtained from the Medicaid Analytic eExtract data

<sup>b</sup> data obtained from the Birth Certificate data

**Table 3** Prevalence<sup>a</sup> of neonatal critical conditions

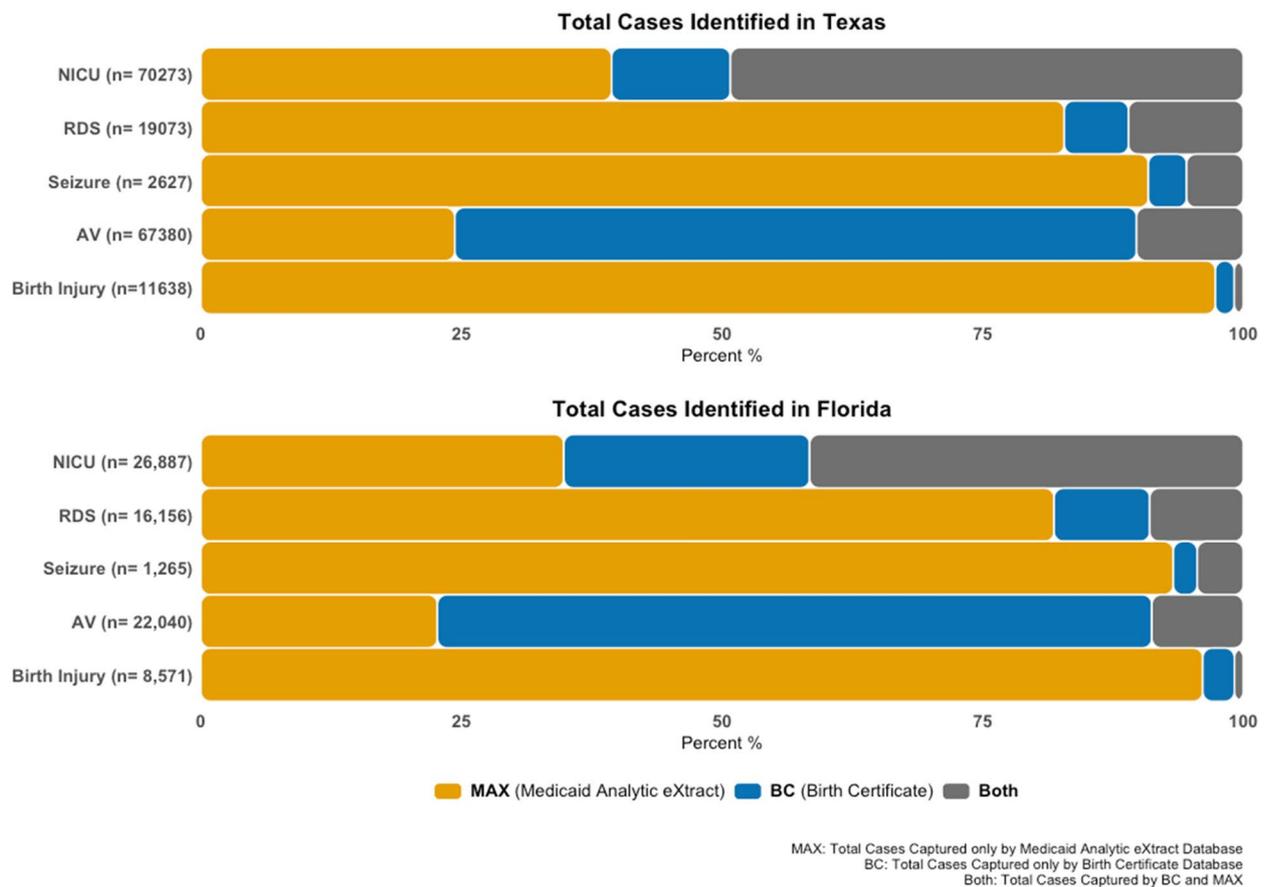
| Neonatal critical condition                             | FLORIDA |      | TEXAS |      |
|---|---------|------|-------|------|
|   | MAX     | BC   | MAX   | BC   |
| Neonatal intensive care unit admission (NICU admission) | 7.40    | 6.31 | 10.00 | 7.05 |
| Respiratory distress syndrome (RDS)                     | 2.72    | 0.54 | 4.74  | 0.87 |
| Seizure   | 0.24    | 0.02 | 0.25  | 0.02 |
| Assisted ventilation (AV)                               | 1.27    | 3.14 | 2.37  | 5.19 |
| Birth injury  | 1.54    | 0.06 | 1.89  | 0.05 |

MAX Medicaid Analytic eExtract, BC Birth Certificate

<sup>a</sup> Prevalence was calculated as the number of identified cases per 100 neonate/mother pairs

The overall higher prevalences captured by MAX may be explained by either lower sensitivity of data capture in BCs or false positive cases captured in MAX. Considering our use of previously validated claims-based algorithms, suggesting PPV > = 86% for all evaluated conditions, low specificity of MAX-derived cases is unlikely to explain the lower prevalences reported in BCs. This conclusion is supported by a recent evaluation of BC data by the Centers for Disease Control and Prevention (CDC), which found quality to vary across states and hospitals [31]. The study, limited to two unidentified states and only 8 hospitals, reported agreement for NICU admissions and need for AV to be substantial or high for one state and low to extremely low for the other. The study confirmed that especially the sensitivity of BC is low, which reiterates previous findings that highlight underreporting of medical conditions as a key limitation of BCs [28, 32–34]. Specific to seizures, one research group studied neonatal seizure by evaluating capture on the BC [14] against hospital discharge and Medicaid claims [35]. Consistent with our finding, the authors reported low Kappa estimates (9–12%), which they attributed to BC underreporting. Of note, the timing of BC completion relative to the occurrence or discovery of neonatal conditions may further reduce BC sensitivity.

Failure of MAX data to capture cases detected by the BC on the other hand, may be explained by erroneous data on the BC or limited sensitivity of MAX. The previously quoted CDC study found in one state that the proportion of false positives on the BC varied between 15–30% for AV (depending on definition) and 11% for NICU admission. A more recent report from five hospitals in New York city shows a similar pattern of false-positive AV cases [36], suggesting that BCs may not only lack sensitivity but also have some specificity issues in correctly identifying neonatal conditions [26]. Similarly, Zollinger et al [11] compared Indiana BC data to medical records for a variety of variables



**Fig. 2** All cases distributed by data sources in Texas and Florida

**Table 4** Estimated Kappa and sensitivity

| State | Neonatal Critical Condition                             | MAX    | BC     | Both   | Kappa | MAX Sensitivity <sup>a</sup> | BC Sensitivity <sup>b</sup> |
|-------|---|--------|--------|--------|-------|------------------------------|-----------------------------|
| FL    | Neonatal intensive care unit admission (NICU admission) | 20,548 | 17,524 | 11,185 | 55.7  | 0.64                         | 0.54                        |
|       | Respiratory distress syndrome (RDS)                     | 14,673 | 2,933  | 1,450  | 15.7  | 0.49                         | 0.10                        |
|       | Seizure   | 1,236  | 85     | 56     | 8.4   | 0.66                         | 0.05                        |
|       | Assisted ventilation (AV)                               | 6,933  | 17,039 | 1,932  | 14.5  | 0.11                         | 0.28                        |
|       | Birth injury  | 8,311  | 333    | 73     | 1.5   | 0.22                         | 0.01                        |
| TX    | Neonatal intensive care unit admission (NICU admission) | 62,258 | 42,592 | 34,577 | 62.8  | 0.81                         | 0.56                        |
|       | Respiratory distress syndrome (RDS)                     | 17,897 | 3,276  | 2,100  | 18.6  | 0.64                         | 0.12                        |
|       | Seizure   | 2,531  | 239    | 143    | 10.2  | 0.60                         | 0.06                        |
|       | Assisted ventilation (AV)                               | 23,311 | 50,969 | 6,900  | 15.8  | 0.14                         | 0.30                        |
|       | Birth injury  | 11,427 | 312    | 101    | 1.6   | 0.32                         | 0.01                        |

BC Birth Certificate, FL Florida, TX Texas, MAX: Medicaid Analytic eXtract;

<sup>a</sup> Ability of MAX to capture all cases identified by BC (Cases identified by both / all cases in BC)

<sup>b</sup> Ability of BC to capture all cases identified by MAX (Cases identified by both / all cases in MAX)

including RDS, AV, seizure and birth injury. The corresponding results showed extremely low specificity of BCs in identifying RDS, AV, seizure and birth injury

cases. Administrative claims databases, on the contrary, have been compared to medical records for some variables and concluded overall high specificity. Two studies

assessed ICD-9-CM codes for neonatal mechanical ventilation against medical records resulting in specificity of 99.7% and 97.1% in Canadian and Australian discharge databases, respectively [25, 26]. The same Australian discharge database showed 92.4% specificity for RDS. In addition, Bateman et al [37] validated neonatal seizure codes in MAX data and found a high positive predicative value of 86%. Although some of these validation studies occurred outside the U.S. and may not be generalizable to local coding conventions, the similarity in findings regarding specificity issues in BCs raises concerns about bias when used in inferential analyses. It should be noted that claims data's focus on billing of medical services may result in varying sensitivity by prioritizing those diagnoses that are relevant for reimbursement decisions.

Our findings, taken together, support the assumption that MAX data may produce superior sensitivity and specificity in capturing the evaluated neonatal conditions. An exception of this observation appears to be in the measurement of need for assisted ventilation, which may be inherent in coding practices for billing purposes. For example, short-term need for assisted ventilation immediately after delivery, which is captured on the BC, may be incorporated in capitated billing arrangements and thus not appear itemized in billing data.

Motivation or incentives as well as training may influence the quality of data and discrepancy between BC and MAX. Both CDC and the American College of Obstetricians and Gynecologists have also pointed to lack of standardization of obstetric clinical data definitions [31, 38]. Ongoing efforts to improve BC documentation via enhanced training and standardization warrant future studies as more recent MAX and BC data are made available.

Although our results only cover two states and are therefore not nationally representative, Texas and Florida supply a large portion of infant information to the national surveillance systems. Thus, national trends of neonatal critical conditions such as seizures that are estimated from BCs ought to be interpreted with caution. Moreover, causal inference research that use BC neonatal critical conditions should interpret the results with extreme caution given the high false positive rates of the BC data. Claims data sources appear to be more suited for causal inference research studying the neonatal critical conditions.

In conclusion, we compared the extent of agreement between BCs and claims data regarding capture of neonatal critical conditions in two large states and found low agreement. Both data sources captured cases that the other one did not, presumably due to underreporting or capture of false positives. Future research ought to

examine reasons for discrepancies between the two data sources.

#### Acknowledgements

We would like to thank the National Center for Health Statistics, Florida Department of Health, and the Texas Department of State Health Services for providing the birth certificate data. The study conclusions are the authors' own and do not necessarily reflect the opinion of the State Departments of Health.

#### Authors' contribution

Contributors YA, YZ and AW conceptualized the study. YA and YZ performed data analyses. All authors contributed to study design and analytical approach. YA prepared the draft of the manuscript and all the coauthors revised the draft critically. The author(s) read and approved the final manuscript.

#### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Availability of data and materials

The data that support the findings of this study are available from the Texas State Department of Health and the Florida State Department of Health but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the State Departments of Health in Texas and Florida.

#### Declarations

##### Ethics approval and consent to participate

This research does not involve human subjects, and there is no intervention or interaction with individuals and thus was approved and exempt from issuing informed consent by the Institutional Review Boards and Privacy Boards at the University of Florida, as well as the Florida and Texas State Departments of Health. All methods were carried out in accordance with relevant guidelines and regulations.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 25 July 2022 Accepted: 3 February 2023

Published online: 20 February 2023

#### References

- Welch MJ, Lally R, Miller JE, et al. The ethics and regulatory landscape of including vulnerable populations in pragmatic clinical trials. *Clin Trials*. 2015;12(5):503–10. <https://doi.org/10.1177/1740774515597701>.
- Blehar MC, Spong C, Grady C, Goldkind SF, Sahin L, Clayton JA. Enrolling pregnant women: issues in clinical research. *Womens Health Issues*. 2013;23(1):e39–45. <https://doi.org/10.1016/j.whi.2012.10.003>.
- Huybrechts KF, Palmsten K, Avorn J, et al. Antidepressant use in pregnancy and the risk of cardiac defects. *N Engl J Med*. 2014;370(25):2397–407. <https://doi.org/10.1056/NEJMoa1312828>.
- Stewart DL, Ryan KJ, Seare JG, Pinsky B, Becker L, Frogel M. Association of RSV-related hospitalization and non-compliance with palivizumab among commercially insured infants: a retrospective claims analysis. *BMC Infect Dis*. 2013;13:334. <https://doi.org/10.1186/1471-2334-13-334>.
- Deputy NP, Sharma AJ, Kim SY, Hinkle SN. Prevalence and characteristics associated with gestational weight gain adequacy. *Obstet Gynecol*. 2015;125(4):773–81. <https://doi.org/10.1097/AOG.0000000000000739>.
- Avalos LA, Chen H, Yang C, et al. The prevalence and trends of antiviral medication use during pregnancy in the US: a population-based study of

- 664,297 deliveries in 2001–2007. *Matern Child Health J.* 2014;18(1):64–72. <https://doi.org/10.1007/s10995-013-1234-9>.
7. Savitz DA, Danilack VA, Engel SM, Elston B, Lipkind HS. Descriptive epidemiology of chronic hypertension, gestational hypertension, and preeclampsia in New York State, 1995–2004. *Matern Child Health J.* 2014;18(4):829–38. <https://doi.org/10.1007/s10995-013-1307-9>.
  8. DeSisto CL, Kim SY, Sharma AJ. Prevalence estimates of gestational diabetes mellitus in the United States, Pregnancy Risk Assessment Monitoring System (PRAMS), 2007–2010. *Prev Chronic Dis.* 2014;11:E104. <https://doi.org/10.5888/pcd11.130415>.
  9. McLaughlin CC, Baptiste MS, Schymura MJ, Nasca PC, Zdeb MS. Maternal and infant birth characteristics and hepatoblastoma. *Am J Epidemiol.* 2006;163(9):818–28. <https://doi.org/10.1093/aje/kwj104>.
  10. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol.* 1977;105(5):488–95. <https://doi.org/10.1093/oxfordjournals.aje.a112408>.
  11. Zollinger TW, Przybylski MJ, Gamache RE. Reliability of Indiana birth certificate data compared to medical records. *Ann Epidemiol.* 2006;16(1):1–10. <https://doi.org/10.1016/j.annepidem.2005.03.005>.
  12. Lain SJ, Hadfield RM, Raynes-Greenow CH, et al. Quality of data in perinatal population health databases: a systematic review. *Med Care.* 2012;50(4):e7–20. <https://doi.org/10.1097/MLR.0b013e31821d2b1d>.
  13. Palmsten K, Huybrechts KF, Kowal MK, Mogun H, Hernández-Díaz S. Validity of maternal and infant outcomes within nationwide Medicaid data. *Pharmacoepidemiol Drug Saf.* 2014;23(6):646–55. <https://doi.org/10.1002/pds.3627>.
  14. Li Q, Lenski M, Copeland G, et al. Recording of Neonatal Seizures in Birth Certificates, Maternal Interviews, and Hospital Discharge Abstracts in a Cerebral Palsy Case-Control Study in Michigan. *J Child Neurol.* 2016;31(7):817–23. <https://doi.org/10.1177/0883073815620678>.
  15. 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. <https://doi.org/10.1002/pds.3319>.
  16. Bird TM, Bronstein JM, Hall RW, Lowery CL, Nugent R, Mays GP. Late preterm infants: birth outcomes and health care utilization in the first year. *Pediatrics.* 2010;126(2):e311–9. <https://doi.org/10.1542/peds.2009-2869>.
  17. Cheng YW, Snowden JM, King TL, Caughey AB. Selected perinatal outcomes associated with planned home births in the United States. *Am J Obstet Gynecol.* 2013;209(4):325.e1–8. <https://doi.org/10.1016/j.ajog.2013.06.022>.
  18. Amendah DD, Grosse SD, Bertrand J. Medical expenditures of children in the United States with fetal alcohol syndrome. *Neurotoxicol Teratol.* 2011;33(2):322–4. <https://doi.org/10.1016/j.ntt.2010.10.008>.
  19. 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. <https://doi.org/10.1002/pds.3284>.
  20. Berry K, Pesko MF, Hesdorffer DC, Shellhaas RA, Seirup JK, Grinspan ZM. An evaluation of national birth certificate data for neonatal seizure epidemiology. *Epilepsia.* 2017;58(3):446–55. <https://doi.org/10.1111/epi.13665>.
  21. Boulet SL, Shin M, Kirby RS, Goodman D, Correa A. Sensitivity of birth certificate reports of birth defects in Atlanta, 1995–2005: effects of maternal, infant, and hospital characteristics. *Public Health Rep.* 2011;126(2):186–94. <https://doi.org/10.1177/003335491112600209>.
  22. Zhu Y, Hampp C, Wang X, et al. Validation of algorithms to estimate gestational age at birth in the Medicaid Analytic eXtract—Quantifying the misclassification of maternal drug exposure during pregnancy. *Pharmacoepidemiol Drug Saf.* 2020. <https://doi.org/10.1002/pds.5126>.
  23. 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. <https://doi.org/10.1371/journal.pone.0067405>.
  24. Knox CA, Hampp C, Palmsten K, et al. Validation of mother-infant linkage using Medicaid Case ID variable within the Medicaid Analytic eXtract (MAX) database. *Pharmacoepidemiol Drug Saf.* 2019. <https://doi.org/10.1002/pds.4843>.
  25. Ford JB, Roberts CL, Algert CS, et al. Using hospital discharge data for determining neonatal morbidity and mortality: a validation study. *BMC Health Serv Res.* 2007;7:188. <https://doi.org/10.1186/1472-6963-7-188>.
  26. Quan H, Parsons GA, Ghali WA. Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. *Med Care.* 2004;42(8):801–9. <https://doi.org/10.1097/01.mlr.0000132391.59713.0d>.
  27. Cohen B, Bernson D, Sappenfield W, Kirby RS, Kissin D, Zhang Y, Copeland G, Zhang Z, Macaluso M; States Monitoring Assisted Reproductive Technology Collaborative. Accuracy of assisted reproductive technology information on birth certificates: Florida and Massachusetts, 2004–06. *Paediatr Perinat Epidemiol.* 2014;28(3):181–90. <https://doi.org/10.1111/ppe.12110>. Epub 2014 Feb 17.
  28. Northam S, Knapp TR. The reliability and validity of birth certificates. *J Obstet Gynecol Neonatal Nurs.* 2006;35(1):3–12. <https://doi.org/10.1111/j.1552-6909.2006.00016.x>.
  29. Zingg T, Uldry E, Omoumi P, et al. Interobserver reliability of the Tile classification system for pelvic fractures among radiologists and surgeons. *Eur Radiol.* 2020. <https://doi.org/10.1007/s00330-020-07247-0>.
  30. Brumberg HL, Dozor D, Golombek SG. History of the birth certificate: from inception to the future of electronic data. *J Perinatol.* 2012;32(6):407–11. <https://doi.org/10.1038/jp.2012.3>.
  31. Martin JA, Wilson EC, Osterman MJK, Saadi EW, Sutton SR, Hamilton BE. Assessing the quality of medical and health data from the 2003 birth certificate revision: results from two states. *Natl Vital Stat Rep.* 2013;62(2):1–19.
  32. DiGiuseppe DL, Aron DC, Ranbom L, Harper DL, Rosenthal GE. Reliability of birth certificate data: a multi-hospital comparison to medical records information. *Matern Child Health J.* 2002;6(3):169–79. <https://doi.org/10.1023/a:1019726112597>.
  33. Roohan PJ, Josberger RE, Acar J, Dabir P, Feder HM, Gagliano PJ. Validation of birth certificate data in New York State. *J Community Health.* 2003;28(5):335–46. <https://doi.org/10.1023/a:1025492512915>.
  34. Piper JM, Mitchel EF, Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *Am J Epidemiol.* 1993;137(7):758–68. <https://doi.org/10.1093/oxfordjournals.aje.a116736>.
  35. Li Q, Jenkins DD, Kinsman SL. Birth Settings and the Validation of Neonatal Seizures Recorded in Birth Certificates Compared to Medicaid Claims and Hospital Discharge Abstracts Among Live Births in South Carolina, 1996–2013. *Matern Child Health J.* 2017;21(5):1047–54. <https://doi.org/10.1007/s10995-016-2200-0>.
  36. Gregory ECW, Martin JA, Argov EL, Osterman MJK. Assessing the quality of medical and health data from the 2003 birth certificate revision: results from new york city. *Natl Vital Stat Rep.* 2019;68(8):1–20.
  37. Bateman BT, Huybrechts KF, Maeda A, et al. Calcium channel blocker exposure in late pregnancy and the risk of neonatal seizures. *Obstet Gynecol.* 2015;126(2):271–8. <https://doi.org/10.1097/AOG.0000000000000908>.
  38. ACOG Committee Opinion No. 748 Summary: The Importance of Vital Records and Statistics for the Obstetrician-Gynecologist. *Obstet Gynecol.* 2018;132(2):537–8. <https://doi.org/10.1097/AOG.0000000000002760>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

