



Validating Claims-Based Algorithms Determining Pregnancy Outcomes and Gestational Age Using a Linked Claims-Electronic Medical Record Database

Keran Moll¹ · Hui Lee Wong² · Kathryn Fingar¹ · Shayan Hobbi³ · Minya Sheng¹ · Timothy A. Burrell¹ · Linda O. Eckert⁴ · Flor M. Munoz⁵ · Bethany Baer² · Azadeh Shoaibi² · Steven Anderson²

Accepted: 8 August 2021 / Published online: 30 September 2021
© The Author(s) 2021

Abstract

Introduction Pregnancy outcome identification and precise estimates of gestational age (GA) are critical in drug safety studies of pregnant women. Validated pregnancy outcome algorithms based on the International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System (ICD-10-CM/PCS) have not previously been published.

Methods We developed algorithms to classify pregnancy outcomes and estimate GA using ICD-10-CM/PCS and service codes on claims in the 2016–2018 IBM[®] MarketScan[®] Explorys[®] Claims-EMR Data Set and compared the results with ob-gyn adjudication of electronic medical records (EMRs). Obstetric services were grouped into episodes using hierarchical and spacing requirements. GA was based on evidence with the highest clinical accuracy. Among pregnancies with obstetric EMRs, 100 full-term live births (FTBs), 100 preterm live births (PTBs), 100 spontaneous abortions (SAs), and 24 stillbirths were selected for review. Physicians adjudicated cases using Global Alignment of Immunization safety Assessment in pregnancy (GAIA) definitions applied to structured EMRs.

Results The claims-based algorithms identified 34,204 pregnancies, of which 9.9% had obstetric EMRs. Of sampled pregnancies, 92 FTBs, 93 PTBs, 75 SAs, and 24 stillbirths were adjudicated. Among these pregnancies, the percent agreement was 97.8%, 62.4%, 100.0%, and 70.8% for FTBs, PTBs, SAs, and stillbirths, respectively. The percent agreement on GA within 7 and 28 days, respectively, was 85.9% and 100.0% for FTBs, 81.7% and 98.9% for PTBs, 61.3% and 94.7% for SAs, and 66.7% and 79.2% for stillbirths.

Conclusions The pregnancy outcome algorithms had high agreement with physician adjudication of EMRs and may inform post-market maternal safety surveillance.

1 Introduction

The safety of maternal exposure to vaccines and medications is rarely studied in clinical trials because pregnant women are usually excluded from pre-market research [1]. Most data on maternal vaccine and drug safety come from post-market

studies using billing codes on administrative claims and electronic medical records (EMRs) to evaluate exposures during pregnancy and their association with maternal and infant outcomes [1]. While the Food and Drug Administration (FDA) encourages inclusion of pregnant women and women of childbearing age in prelicensure clinical trials [2], maternal safety evidence is limited at the time of licensure. Thus, it is important to leverage real-world data, such as administrative claims, for post-market evaluation of the safety and effectiveness of vaccines and other drugs during pregnancy.

In addition, evidence suggests pregnant women with COVID-19 may be at an increased risk for severe illness compared with nonpregnant women [3]. To date, several COVID-19 vaccines of different platforms (novel mRNA or viral vector) have been authorized by the FDA on the US market and deployed broadly. However, safety evidence

✉ Keran Moll
kmoll@us.ibm.com

¹ IBM Watson Health, Cambridge, MA, USA

² Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

³ IBM Global Business Services, Bethesda, MD, USA

⁴ University of Washington School of Medicine, Seattle, WA, USA

⁵ Baylor College of Medicine, Houston, TX, USA

Key Points

We developed algorithms based on the International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System, to estimate pregnancy start dates and outcomes in claims. We validated the algorithms through ob-gyn adjudication of electronic medical records (EMRs) using the Global Alignment of Immunization safety Assessment in pregnancy (GAIA) framework.

The algorithms had a high level of agreement with ob-gyn adjudication of EMRs.

These algorithms may be used to evaluate maternal exposures to prescription drugs and vaccines, including COVID-19 vaccines, and their association with safety outcomes.

of maternal exposure is limited to small populations of pregnant women [4–6]. This highlights the importance of risk and benefit evaluation of maternal vaccination against COVID-19 in large longitudinal claims data using well-defined algorithms to inform clinicians, policymakers, and patients.

Precise pregnancy dating is needed to align acute exposures such as vaccines with key stages of embryologic development [7, 8]. Accurate estimates of gestational age (GA) and outcome identification in administrative data are important to the validity of pregnancy exposure studies [9]. Previous studies have developed and validated claims-based algorithms estimating pregnancy start dates and distinguishing between pregnancy outcomes [10–18], including preterm live births (PTBs, < 37 weeks of gestation), full-term live births (FTBs, 37 + weeks of gestation), stillbirths (20 + weeks of gestation), and spontaneous abortions (SAs, < 20 weeks of gestation). However, to our knowledge, all previous US-based studies have used data from before October 2015, when the nation transitioned to the International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System (ICD-10-CM/PCS), replacing the Ninth Revision (ICD-9-CM) for most inpatient and outpatient encounters. ICD-10-CM/PCS has greater diagnostic detail pertaining to GA at the time of the claim. More recent validation studies using ICD-10-CM/PCS are needed.

Relevant literature discusses the strengths and limitations of administrative claims data to evaluate medication safety in

pregnancy, highlighting the importance of rigorous assessment of claims-based algorithms [10]. Previous studies have compared claims with abstracted paper charts, linked birth records, and EMRs [11, 13–16, 18]. Official national statistics were also used as a comparison to evaluate accuracy of claims-based pregnancy outcome algorithms [12, 17]. The paper chart review is generally considered the gold standard but can be labor intensive and requires special access to medical charts. As EMRs have become more common, public and proprietary data sets now often include structured EMR components, although they lack certain chart detail (e.g., physician notes). This study used a claims-EMR linked database with an EMR repository incorporating data from more than 30 health systems. Although the data quality may vary across EMR systems, this linked database provides a large sample to evaluate whether structured EMR components contain enough information for physicians to validate claims-based algorithms through a clinical review, which has not been previously explored.

The Global Alignment of Immunization safety Assessment in pregnancy (GAIA) framework may improve the rigor of clinical review of EMRs but primarily has been used in clinical settings for prospective data collection. This framework provides standard, globally harmonized definitions of outcomes and data collection guidelines for monitoring the safety of vaccines used during pregnancy [19–23]. A collaborative network of experts developed case definitions for PTB, SA, and stillbirth as well as GA, which are categorized into levels of diagnostic certainty depending on the accuracy of available data (e.g., pregnancy dating in the first vs third trimester). Although it is unlikely that some GAIA criteria are detailed in retrospectively collected structured EMRs (e.g., umbilical cord pulse), other criteria may be recorded (e.g., GA on a first trimester scan). No prior studies have evaluated whether structured EMR components in a given EMR system can be operationalized using the GAIA framework, with what level of GAIA certainty physicians can conduct reviews, and whether EMR-based reviews can serve as a reference point that conceptually approximates a gold standard in the absence of manual unstructured chart reviews.

In this study, we used ICD-10-CM/PCS and Healthcare Common Procedure Coding System (HCPCS) codes on claims data to develop hierarchical algorithms for classifying pregnancy outcomes as PTBs, FTBs, SAs, or stillbirths and determining GA at the outcome. For a sample of pregnancy episodes, we calculated the agreement in the estimated GA and outcome types between the claims-based algorithms and physician adjudication of structured components from a repository of multiple EMR systems using the GAIA framework.

2 Methods

2.1 Data Source

The IBM[®] MarketScan[®] Explorys[®] Claims-EMR Data Set (CED) comprises reimbursement claims and EMR data. It deterministically links the longitudinal claims of patients from the IBM MarketScan Commercial Database to the same patients' clinical records from the IBM Explorys EHR Database. The Explorys EHR Database is a large clinical data asset built through direct connections to large health system partners, including more than 30 health systems and spanning academic and community practices. The EMRs available are limited to structured elements—diagnoses, procedures, immunizations, vital signs and biometrics, medical/surgical history, laboratory results, implantable devices, patient-reported outcomes, and inpatient drug administrations and ambulatory prescriptions. The combined data set provides clinical, claims, and financial data to support research and analysis.

2.2 Study Population

We included pregnancy episodes ending in PTB, FTB, SA, or stillbirth from August 1, 2016, through October 31, 2018, for women aged 12–55 years at the time of the outcome who were continuously enrolled during the prenatal period, allowing for an enrollment gap of up to 45 days. We chose August 1, 2016, to evaluate the prenatal period for claims spanning only the ICD-10-CM/PCS system. Figure 1 depicts the steps to select claims and episodes to include and the sample drawn for clinician adjudication using EMRs.

2.3 Definition of Pregnancy-Related Endpoints

Although we focused on live births, stillbirths, and SAs, any pregnancy endpoint can be relevant in determining clinically possible outcomes, given spacing between different outcomes for the same woman (e.g., an abortion 1 week after a live birth is not clinically possible). Based on a focused literature review [10–18, 24], we first identified all potential pregnancy endpoints using ICD-10-CM/PCS, HCPCS, and diagnosis-related groups on service records from all healthcare settings. Events indicating a pregnancy's end included live births, stillbirths, SAs, ectopic pregnancies, elective abortions, trophoblastic and other abnormal products of conception, and deliveries with unknown outcomes (DELIV). Supplemental Table 1 (Online Resource 1, see electronic supplementary material [ESM]) presents clinical codes for identifying each outcome. Online Resource 2 (see ESM) is a Microsoft[®] Excel file that provides clinical codes and additional descriptive statistics.

2.4 Definition of Pregnancy Episodes

Pregnancy endpoints were first assigned by service day. When clinical codes indicating different endpoint types (other than DELIV) occurred on the same day, the service records for that day were excluded, because the conflicting information was likely due to miscoding. This exclusion rarely removed an outcome entirely because true outcomes were usually coded on multiple service dates. DELIV was assigned only if no other types of endpoint codes were present on the same date (see Supplemental Table 2, Online Resource 1, ESM).

Next, we evaluated pregnancy endpoints on the timeline of the same woman. Some endpoints are implausible because they occur too close together. Additionally, pregnancy-related clinical codes on multiple services may represent the same pregnancy (e.g., a visit to both the doctor's office and an ambulatory surgery center following an SA). Thus, we adapted pregnancy algorithms developed by Hornbrook et al. [13] and modified by Matcho et al. [16] and Naleway et al. [18] to group service records into pregnancy episodes.

Pregnancy episodes were constructed following a hierarchy, starting with the outcomes considered most reliably coded with respect to outcome identification and timing in claims data [16] and based on consultation with physicians advising this study (authors LE and FM). We started with the earliest service record with the most reliable outcome and then evaluated subsequent service records classified as the same outcome type. We repeated this process for service records classified as outcomes of lower hierarchy. Live births were placed on the timeline first, followed by stillbirths, DELIV, trophoblastic and other abnormal products of conception, ectopic pregnancies, elective abortions, and SAs [13, 16]. Subsequent outcomes of the same type and outcomes further down the hierarchy were placed on the timeline only if they met a minimum allowable time between the date of that outcome and the date of other outcomes already on the timeline (Supplemental Table 3, Online Resource 1, see ESM). Thus, this process sequentially placed outcomes on the timeline, starting with those highest on the hierarchy and ending with those lowest on the hierarchy (Supplemental Fig. 1, Online Resource 1, see ESM).

From plausible outcomes on the timeline, we selected only those of interest (live births, stillbirths, and SAs) occurring on or after August 1, 2016. Fig. 2 displays the study flowchart, starting from outcomes of interest and detailing each attrition step with reasons for exclusion and the number of episodes excluded. For each outcome on the timeline representing a pregnancy's end, we assembled all service records within a prenatal window. The *prenatal window* is the time period between the outcome date and the earliest possible pregnancy start date based on a maximum pregnancy term, which varied by outcome type (Supplemental

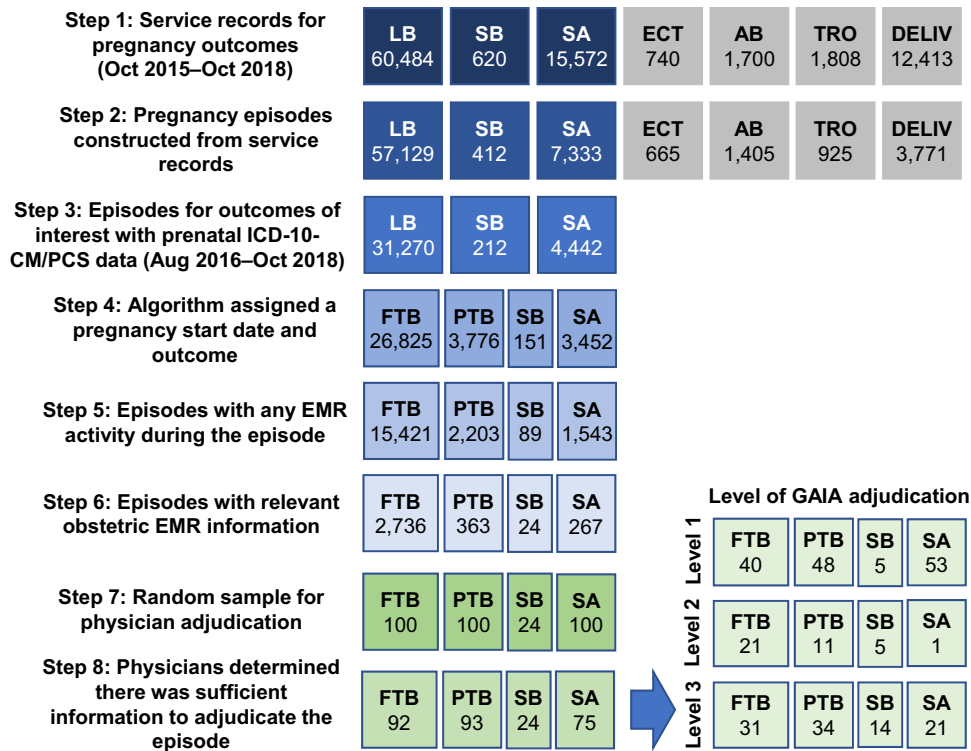


Fig. 1 Sample selection and attrition. Relevant obstetric EMRs only mapped to GAIA levels 1–3; thus, no pregnancy episodes were adjudicated at GAIA level 4. A total of 1663 pregnancy episodes could not be assigned a pregnancy start date based on information in the claims; an additional 20 live births, 33 stillbirths, and 4 spontaneous abortions were not assigned a pregnancy start date because the estimated gestational age was implausible (stillbirths were required to be 20+ weeks, SAs < 20 weeks, and live births 22 + weeks). *AB*

elective abortion, *DELIV* deliveries with unknown outcomes, *ECT* ectopic pregnancy, *EMR* electronic medical record, *FTB* full-term live birth, *GAIA* Global Alignment of Immunization safety Assessment in pregnancy, *ICD-10-CM/PCS* International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System, *LB* live birth, *PTB* preterm live birth, *SA* spontaneous abortion, *SB* stillbirth, *TRO* trophoblastic and other abnormal products of conception

Table 4, Online Resource 1, see ESM). If we encountered a previous outcome during the prenatal window, we adjusted the window’s start date according to a minimum allowable number of days before a subsequent pregnancy could start (Fig. 2).

2.5 Algorithms Determining Pregnancy Outcome and Start Date

Certain services in the prenatal window may provide more reliable estimates for a pregnancy start date (e.g., the date of embryo transfer likely resulting in the pregnancy is reliable in determining pregnancy start date). Additionally, information from multiple services may indicate pregnancy start dates that are inconsistent with one another. Therefore, we developed hierarchical algorithms to assign pregnancy start dates, calculate GA at the outcome date, and confirm the outcome (i.e., that stillbirths occur at ≥ 20 weeks of gestation, that SAs occur at < 20 weeks, and to differentiate PTBs from FTBs at 37 weeks).

Several sources in claims data provide GA information: the timing of embryo transfer or intrauterine insemination; ICD-10-CM codes indicating GA in weeks, trimester, or preterm status during prenatal or delivery encounters; and prenatal screening tests (Supplemental Table 5, Online Resource 1, see ESM). We consulted prior literature and two practicing physicians (one ob-gyn, one pediatrician) to assess the reliability of GA in claims data and developed the following hierarchy steps 1–13 (see overview in Table 1). We moved through each step sequentially until a pregnancy start date was assigned.

Step 1 First, pregnancy start dates were assigned based on the date of intrauterine insemination or embryo transfer procedure during the prenatal window if those procedures were recorded. The start date was calculated as procedure date – 14 days + 1. If these procedures were recorded on multiple dates, we selected the service record closest to the outcome.

Steps 2–5 Next, we looked for diagnoses indicating GA in weeks on the same date of first trimester ultrasounds (step 2), nuchal translucency scans (step 3), anatomic ultrasounds

(step 4), and then on any other type of service record (step 5). We calculated the pregnancy start date as service date – GA in weeks * 7 + 1. If there were multiple service records for first trimester ultrasounds, nuchal translucency scans, or anatomic ultrasounds that documented GA, we selected the first. If there were multiple service records of other types that documented GA in weeks, multiple pregnancy start dates were calculated based on each documented GA and the mode or, if no mode exists, the median was selected as the assigned start date.

Steps 6–8 Next, we looked for HCPCS codes for nuchal translucency scans (usually before 13 5/7 weeks according to guidelines, step 6), chorionic villus sampling (usually before 13 weeks, step 7), and prenatal cell-free DNA screening (usually between 10 and 14 weeks, step 8) without a diagnosis indicating GA on the service record. We calculated the pregnancy start date as service date – 90 days + 1 for nuchal translucency scans and as service date – 12 weeks * 7 + 1 for the latter two tests. These tests and the values of 90 days and 12 weeks were selected in consultation with ob-gyn experts, published studies, and empirical estimates of screening timing from the MarketScan Commercial Database. If multiple service records existed for nuchal translucency scans, chorionic villus sampling, or cell-free DNA screenings, we selected the first.

Steps 9–11 Next, within 7 days of the outcome, we looked for records with diagnoses indicating full-term status (step 9), diagnoses or HCPCS documenting that the pregnancy had reached a certain trimester (step 10), and diagnoses indicating preterm status (step 11). For claims for live birth and stillbirth episodes with full-term codes, we calculated the pregnancy start date as outcome date – 39 weeks * 7 + 1. For services with diagnoses or HCPCS codes indicating trimester, we calculated the pregnancy start date as outcome date – 70, 147, or 241 days + 1, for the first, second, and third trimester, respectively. These days represent the midpoints of each trimester. There were no instances where multiple records contained conflicting information on the trimester of the outcome. For records indicating the delivery was preterm, we calculated the pregnancy start date as outcome date – 35 weeks * 7 + 1. Margulis et al. [15] reported that assigning a GA of 35 weeks to preterm birth and 39 weeks to term births maximized positive agreement on GA within 1 week, comparing birth discharge records and a chart review.

Step 12 Next, we looked for the first glucose tolerance test. We calculated the pregnancy start date as screening date – 26 weeks * 7 + 1. We chose 26 weeks because it is the midpoint of the guideline-recommended range for this test (24–28 weeks).

Step 13 Finally, we looked for prenatal services > 7 days before the outcome with ICD-10-CM codes indicating a trimester of pregnancy at the time of the service. We selected

the trimester code on the service date closest to the outcome and assigned the pregnancy start date as service date – 70, 147, or 241 days + 1, for the first, second, and third trimester, respectively.

At each step, we evaluated the estimated pregnancy start date and GA at the outcome for plausibility. The estimated pregnancy start date was considered plausible if it occurred between the maximum pregnancy term (as defined above) and the minimum pregnancy term (Supplemental Table 4, see ESM). For steps 1–4, if the estimated pregnancy start date was not plausible, we excluded the pregnancy episode without evaluating subsequent steps. For steps 5–13, if the estimated pregnancy start date was implausible, we attempted to estimate it in subsequent steps of the hierarchy. If at step 13 the start date was still not assigned or was implausible, we set it to missing and excluded the pregnancy episode.

We retained only stillbirths assigned a GA of 20 + weeks, SAs assigned a GA of < 20 weeks, and live births assigned a GA of 22 + weeks. Live births were further classified as PTB if they had a GA at delivery of < 37 weeks and FTB if GA was 37 + weeks. GA in days at the outcome was calculated as outcome date – pregnancy start date + 1.

As a reference, we calculated GA estimates from all GA-related data elements as described above but without imposing the stepwise selection. Supplemental Table 6 (Online Resource 1, see ESM) provides the differences in GA calculated from multiple service records when there was more than one estimate for a pregnancy episode.

2.6 Validation Sample

Like paper chart reviews, EMR charts consisting of sufficient pregnancy-related information are required for physician adjudication. Not all pregnancy episodes identified in the claims data had linked EMR data overlapping the same period. We further limited our study population to patients who received routine obstetric care from providers contributing EMRs to the CED during the prenatal window so that EMR charts resembling paper charts with a comprehensive view of patients' healthcare experiences during pregnancy could be constructed. Live births and SAs that met either of the following two scenarios and had a documented pregnancy outcome on the EMR were included in the case pool. For a rare outcome such as stillbirth, we selected any stillbirth with a documented GA on any prenatal EMR and a documented pregnancy outcome. We searched EMRs for the following information between the maximum pregnancy term, which was adjusted as described above if previous episodes occurred during that period, and 7 days after the outcome.

Scenario 1 During the gestational period, the patient had, on average, at least one encounter recorded in EMRs with

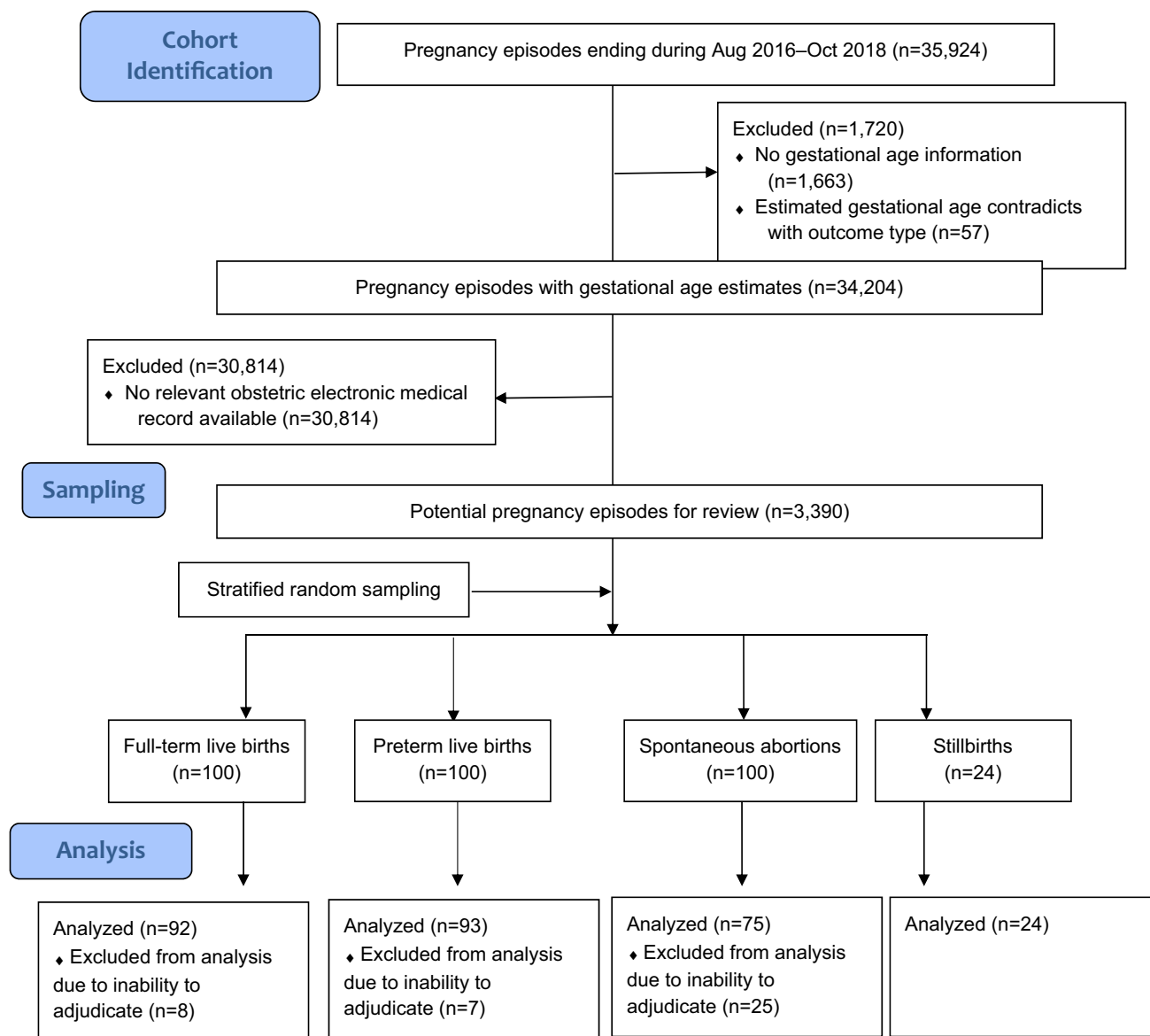


Fig. 2 Study flowchart. Note: Cohort Identification step includes only pregnancy episodes of interest (live births, stillbirths, and spontaneous abortions).

an ob-gyn every 30 days, and one of the following EMR elements was documented: an ultrasonography in the first trimester with documented GA, reported last menstrual period (LMP) with an ultrasound scan performed 4–13 6/7 weeks after the LMP date, or intrauterine insemination or embryo transfer with an ultrasound performed 2–11 6/7 weeks after the procedure date.

Scenario 2 The patient did not meet the scenario 1 criteria, but one of the following EMR elements was documented: a first-trimester obstetric encounter and a documented GA on any prenatal EMR, a second-trimester ultrasonography with a documented GA, or a second trimester ultrasonography 14 0/7–27 6/7 weeks after reported LMP.

Scenarios 1 and 2 represent women with different types of prenatal EMRs, potentially indicating different prenatal healthcare-seeking behaviors or access. For FTBs, PTBs, and SAs, we conducted stratified sampling and randomly selected 50 episodes from each scenario described above for a total of 100 episodes of each outcome so that women with different types of EMRs would be represented in the adjudication sample. We selected all stillbirths with a documented outcome and GA. Supplemental Table 7 (Online Resource 1, see ESM) contains the Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine (SNOMED) Clinical Terms used to identify

Table 1 Algorithms to assign pregnancy episodes a start date and an outcome

Step at (and method by) which the final pregnancy start date was assigned	Full-term live birth		Preterm live birth		Stillbirth		Spontaneous abortion	
	No.	%	No.	%	No.	%	No.	%
Total pregnancy episodes assigned a start date and an outcome	26,825	100.0	3776	100.0	151	100.0	3452	100.0
1. Intrauterine insemination/embryo transfer (procedure date – 14 days + 1)	610	2.3	191	5.1	9	6.0	217	6.3
2. Z3A code on 1st trimester ultrasound (service date – GA in weeks * 7 + 1)	6178	23.0	1058	28.0	41	27.2	732	21.2
3. Z3A code on NT scan (service date – GA in weeks * 7 + 1)	1628	6.1	224	5.9	5	3.3	15	0.4
4. Z3A code on anatomic ultrasound (service date – GA in weeks * 7 + 1)	3397	12.7	655	17.3	33	21.9	1	0.0
5. Z3A code on another type of service (service date – GA in weeks * 7 + 1)								
5a. Mode of multiple Z3A codes	7224	26.9	853	22.6	21	13.9	79	2.3
5b. Median of multiple Z3A codes	4385	16.3	527	14.0	16	10.6	649	18.8
5c. Single Z3A code	2973	11.1	177	4.7	19	12.6	744	21.6
6. NT scan without Z3A code (service date – 90 days + 1)	148	0.6	8	0.2	1	0.7	5	0.1
7. CVS (service date – 12 weeks * 7 + 1)	0	0.0	0	0.0	0	0.0	0	0.0
8. Cell-free DNA screening (service date – 12 weeks * 7 + 1)	35	0.1	11	0.3	0	0.0	7	0.2
9. Full-term code for live birth or stillbirth +/- 7 days of outcome date (outcome date – 39 weeks * 7 + 1)	227	0.8	0	0.0	2	1.3	0	0.0
10. Trimester codes +/- 7 days of outcome date								
10a. 1st trimester code (outcome date – 70 days + 1)	0	0.0	0	0.0	0	0.0	828	24.0
10b. 2nd trimester code (outcome date – 147 days + 1)	0	0.0	0	0.0	4	2.6	0	0.0
10c. 3rd trimester code (outcome date – 241 days + 1)	0	0.0	44	1.2	0	0.0	0	0.0
11. Preterm code +/- 7 days of outcome date (outcome date – 35 weeks * 7 + 1)	0	0.0	0	0.0	0	0.0	0	0.0
12. Glucose screening (screening date – 26 weeks * 7 + 1)	15	0.1	9	0.2	0	0.0	0	0.0
13. Prenatal service > 7 days before the outcome date with a trimester code (service date – 70, 147, or 241 days + 1, for the 1st, 2nd, and 3rd trimester, respectively)	5	0.0	19	0.5	0	0.0	175	5.1

Z3A are the first three digits of a series of ICD-10-CM codes specifying the week of gestation in the fourth and fifth digit (e.g., Z3A29 indicates 29 weeks of gestation). For a list of ICD-10-CM codes that are not specific with respect to GA but that indicate full-term or preterm status, or trimester of pregnancy, see Supplemental Table 5 of Online Resource 1 (ESM), which also contains the codes used to identify the types of procedures, screenings, and tests in this table

CVS chorionic villus sampling, GA gestational age, ICD-10-CM International Classification of Diseases, Tenth Revision, Clinical Modification, NT nuchal translucency

ultrasounds, intrauterine insemination, embryo transfer, GA, and pregnancy outcomes in EMRs.

2.7 Chart Review by Ob-Gyns

Ob-gyn adjudicators were provided a detailed guide on reviewing EMR data using an Excel-based online chart abstraction tool [25] and asked to assign a GAIA certainty level at the end of their review. GAIA case definitions for determining GA at delivery or fetal loss and identifying PTBs, SAs, and stillbirths were developed by the Brighton Collaboration Working Groups [19–23]. The GAIA framework was chosen because it is the result of a call from the World Health Organization for a globally harmonized approach to define universally recognized reference standards for vaccine safety studies during pregnancy. The case definitions established were based on feedback from experts from 13 organizations, consisting of more than 200 volunteers and 25 working groups. Although motivated primarily

by systematic evaluation of immunization in pregnancy, GAIA standard case definitions may have broader applications in other medication safety studies during pregnancy. Each GAIA case definition has defined criteria at levels 1–4, which vary in sensitivity and specificity for case ascertainment. Level 1 is the highest level with maximum sensitivity and specificity, level 2 remains sensitive and specific despite missing certain diagnostic parameters, level 3 is less sensitive but with specificity, and level 4 affords the lowest level of certainty, relying primarily on self-reported evidence, and thus is not discernable in EMRs. Therefore, EMRs with relevant obstetric information only mapped to GAIA levels 1–3, and these levels were used for physician adjudication.

2.8 Analysis

2.8.1 Assessment of Selection Bias

To evaluate potential selection bias of our results, we compared demographic and clinical characteristics among pregnancy episodes for which an outcome and start date were assigned in the CED (target population) with those included in the pool (study population) from which we randomly selected cases for clinical review and adjudication. We evaluated potential imbalances by calculating the standardized mean difference (SMD) for characteristics between these groups. In discussing the results, we use an absolute SMD value of > 0.20 to distinguish potentially meaningful differences between groups [26, 27]; however, we present exact values in the tables.

Additionally, we conducted a similar comparison of characteristics between cases excluded from adjudication (due to insufficient EMR information) and the sample pool.

2.8.2 Percent Agreement

We evaluated the claims-based algorithms by calculating the agreement in pregnancy outcome type and GA determined by the algorithms and the adjudicator-identified results. The *percent agreement* is the proportion of algorithm-determined outcomes or GAs confirmed by EMR adjudication. Despite efforts to select cases with sufficient obstetric EMRs, the reviewers determined that 40 cases had insufficient information for adjudication, so these were excluded from percent agreement calculations.

We quantified the range of potential bias introduced by excluding cases from adjudication in a sensitivity analysis. We computed percent agreement under two hypothetical scenarios: (1) assuming all 40 excluded cases had negative agreement with physician adjudication and (2) assuming all 40 excluded cases had positive agreement with physician adjudication.

3 Results

3.1 Sample Attrition

We identified 35,924 pregnancy episodes in the CED: 31,270 live births, 212 stillbirths, and 4442 SAs (Fig. 1). Of these episodes, the claims-based algorithms assigned a pregnancy start date and confirmed the outcome for 95.2% (34,204, including 26,825 FTBs, 3776 PTBs, 151 stillbirths, and 3452 SAs). Of these, 3390 (9.9%) had obstetric EMRs during the prenatal window and made up the case pool from which the sample was drawn for adjudication (100 FTBs, 100 PTBs, 24 stillbirths, and 100 SAs). Of sampled

episodes, 284 were adjudicated (92 FTBs, 93 PTBs, 24 stillbirths, and 75 SAs): 146 (51.4%) were adjudicated at GAIA level 1 certainty (40 FTBs, 48 PTBs, 5 stillbirths, and 53 SAs), 38 (13.4%) were adjudicated at level 2 certainty (21 FTBs, 11 PTBs, 5 stillbirths, and 1 SA), and 100 (35.2%) were adjudicated at level 3 certainty (31 FTBs, 34 PTBs, 14 stillbirths, and 21 SAs). Overall, 40 (12.3%) sampled episodes could not be adjudicated and were excluded from the validation analysis (8 FTBs, 7 PTBs, 0 stillbirths, and 25 SAs).

Additional information on attrition of service records for endpoint assignment is displayed in Supplemental Table 2 (see ESM). Stillbirths were most likely to have conflicting information on the same day (only 625/1222 stillbirth records did *not* have conflicting information).

3.2 Algorithms

Table 1 shows the step in the algorithm hierarchy at which the pregnancy start date was assigned by outcome type, before we excluded pregnancy episodes without enough obstetric EMRs. The algorithms produced a preterm birth rate of 12.3% (3776 PTBs out of 30,601 live births).

Of FTBs, 44.0% were assigned a pregnancy start date in steps 1–4 of the algorithms. Another 54.4% of FTBs were assigned a start date in step 5. Of PTBs and stillbirths, 56.4% and 58.3%, respectively, were assigned a pregnancy start date in steps 1–4. Another 41.2% and 37.1%, respectively, were assigned a start date in step 5. Thus, nearly all live births and stillbirths were assigned a pregnancy start date based on information deemed fairly reliable, in steps 1–5.

Of SAs, 28.0% were assigned a pregnancy start date in steps 1–4 of the algorithms and another 42.6% were assigned a start date in step 5. Nearly 30% of SAs were assigned a start date based on a code indicating a pregnancy trimester, either on a service record within 7 days of the outcome (24.0%, step 10) or > 7 days before the outcome (5.1%, step 13).

Table 2 shows the distribution of GA at the time of the outcome assigned by the claims-based algorithms. The median GA was 10 weeks for SAs (minimum 4 weeks, maximum 19 weeks), 27 weeks for stillbirths (minimum 20 weeks, maximum 42 weeks), and 38 weeks for live births (minimum 22 weeks, maximum 42 weeks).

3.3 Assessment of Selection Bias

Table 3 presents statistics to evaluate potential selection bias introduced through sample attrition, comparing pregnancy episodes identified in claims of the CED with the subset of episodes included in the case pool for adjudication. The episodes eligible to be sampled for adjudication were more likely to be for white women than were those in the CED

Table 2 Distribution of gestational age at the time of birth or fetal loss across outcomes

Outcome	No.	Distribution of gestational age at birth or fetal loss, weeks							
		Minimum	10th percentile	25th percentile	50th percentile (median)	75th percentile	90th percentile	Maximum	Mean
Live birth	30,601	22	36	38	38	39	40	42	38
Stillbirth	151	20	21	22	27	34	38	42	28
Spontaneous abortion	3452	4	7	8	10	10	12	19	10

pregnancy population overall; otherwise, the populations generally were similar. There were additional imbalances on other variables for stillbirths, likely due to small sample size (24 stillbirths in case pool).

Supplemental Table 8 (Online Resource 1, see ESM) compares the 40 cases excluded from adjudication with the overall sample pool. The excluded cases were older than the sample pool for FTB (30.5 vs 28.7 years) and SA (32.2 vs 30.4 years) but younger for PTB (25.7 vs 29.6 years). The excluded SA cases had a shorter average GA than the SA sample pool (8.6 vs 10.1 weeks).

3.4 Adjudication Results

3.4.1 Outcomes

Table 4 displays the results comparing assignment of GA and outcomes from the claims-based algorithms with those determined through ob-gyn adjudication. Of pregnancy episodes adjudicated with GAIA certainty levels 1–3, the percent agreement confirming the outcome was 100.0% for live births overall (95% confidence interval [CI] 97.5–100.0) and for SAs (95% CI 93.9–100.0), 70.8% for stillbirths (95% CI 50.2–85.5), and 97.8% for FTBs (95% CI 91.8–99.9). PTBs had the lowest percent agreement of 62.4% (95% CI 52.0–71.7).

3.4.2 Gestational Age

We measured the percentage of pregnancy episodes where the algorithm-determined and adjudicator-assigned GA agreed within 7, 14, 21, and 28 days of each other. For all live births, the percent agreement was 83.8% within 7 days (95% CI 77.6–88.5) and ranged up to 99.5% within 28 days (95% CI 96.6–100.0). The percent agreement was highest for FTBs, ranging from 85.9% within 7 days (95% CI 77.0–91.8) to 100.0% within 28 days (95% CI 95.0–100.0). The percent agreement within 7 days and 28 days, respectively, ranged from 81.7% (95% CI 72.4–88.5) to 98.9% (95% CI 93.4–100.0) for PTBs, from 61.3% (95% CI 49.8–71.7) to 94.7% (95% CI 86.4–98.4) for SAs, and from 66.7% (95% CI 46.2–82.4) to 79.2% (95% CI 58.6–91.4) for stillbirths.

3.4.3 Sensitivity Analysis

Supplemental Table 9 (Online Resource 1, see ESM) shows the range of percent agreement under the two extreme hypothetical scenarios in the sensitivity analysis: 90.0–98.0% for FTB, 58.0–65.0% for PTB, and 75.0–100.0% for SA. No cases were excluded from adjudication for stillbirth.

4 Discussion

This study leveraged a linked claims-EMR database to develop algorithms for identifying pregnancy outcomes (FTB, PTB, SA, and stillbirth) and estimating GA in claims and to adjudicate cases in the linked EMRs. Physicians used structured EMR components and the GAIA framework as the reference method to adjudicate the cases, along with a semiautomated chart abstraction tool. The ICD-10-CM/PCS claims-based algorithms had high agreement with the physician adjudication of EMRs confirming the outcomes of live births overall (100.0%), SAs (100.0%), and stillbirths (70.8%). These agreement metrics generally are consistent with published performance metrics for ICD-9-CM algorithms. Naleway et al. [18] found a percent agreement of 96% for live births and 92% for SAs. Hornbrook et al. [13] had a percent agreement of 100% for live births and 97% for SAs. Although both groups reported stillbirth agreement substantially higher than ours (92% [Naleway] and 88% [Hornbrook] vs 70.8% [our study]), stillbirth samples are small in all studies (11 [Naleway] and 24 [Hornbrook] vs 24 [our study]).

Our methods further classified live births as FTBs and PTBs based on estimated GA at the outcome. The percent agreement was high for FTBs (97.8%) but lower for PTBs (62.4%). Small differences in GA could produce larger discrepancies in the dichotomous outcome of PTB. Of the 93 PTBs reviewed, reviewers classified 35 cases to FTBs, 23 of which had a difference between the algorithm-estimated and the adjudicator-determined GA of 1 week (36 vs 37 weeks) (not shown); that is, 23/35 misclassifications of PTB by the algorithms were due to the underestimation of GA by only 1 week. When all live births were evaluated together, the claims-based algorithms assigned a GA that

Table 3 Comparison of pregnancy episodes for which an outcome and start date were assigned across target and study populations

Characteristic	Full-term live birth			Preterm live birth			Stillbirth			Spontaneous abortion		
	CED	Sample pool	SMD	CED	Sample pool	SMD	CED	Sample pool	SMD	CED	Sample pool	SMD
Total, no.	26,825	2736	N/A	3776	363	N/A	151	24	N/A	3452	267	N/A
Age, years, %												
< 15	0.1	0.1	0.02	0.1	0.3	0.03	0.0	0.0	N/A	0.0	0.0	-0.02
15-17	1.4	1.6	0.01	1.2	0.6	-0.07	2.6	0.0	-0.23	1.0	0.0	-0.14
18-24	16.9	16.7	-0.01	16.2	14.3	-0.05	17.9	29.2	0.27	11.6	17.6	0.17
25-34	68.8	70.5	0.04	65.8	68.0	0.05	57.0	33.3	-0.49	59.8	58.4	-0.03
35-44	12.6	11.1	-0.05	16.3	16.0	-0.01	22.5	37.5	0.33	27.3	24.0	-0.08
45 +	0.1	0.0	-0.04	0.4	0.8	0.06	0.0	0.0	N/A	0.3	0.0	-0.08
Age, years, mean	28.9	28.7	-0.03	29.4	29.6	0.03	29.6	30.3	0.11	30.9	30.4	-0.09
Race/ethnicity, %												
Asian	3.1	4.2	0.06	3.4	3.9	0.02	0.7	0.0	-0.12	3.1	5.2	0.11
Black	9.4	8.9	-0.02	12.8	10.5	-0.07	17.2	16.7	-0.01	13.2	12.7	-0.01
Hispanic	1.1	1.7	0.05	1.4	1.9	0.05	1.3	0.0	-0.16	1.2	2.2	0.08
White	72.4	80.8	0.20	69.9	78.0	0.19	67.5	75.0	0.17	69.4	78.7	0.21
Other	6.6	7.4	0.03	6.3	7.7	0.06	2.6	4.2	0.08	7.1	10.5	0.12
Multiracial	0.8	1.3	0.05	0.8	0.8	0.01	1.3	0.0	-0.16	0.7	1.1	0.05
Prenatal ED visit, %	29.6	31.1	0.03	37.4	40.5	0.06	31.1	20.8	-0.24	23.2	25.5	0.05
GDM, %	17.5	18.5	0.03	26.0	30.9	0.11	10.6	4.2	-0.25	1.9	3.4	0.10
Preeclampsia, %	4.1	3.9	-0.01	23.6	24.0	0.01	6.0	4.2	-0.08	0.1	0.0	-0.03
C-section, %												
Primary	13.8	14.0	0.00	27.2	26.2	-0.02	9.3	12.5	0.10	0.0	0.0	N/A
Repeat	14.8	14.4	-0.01	18.7	19.0	0.01	4.6	4.2	-0.02	0.0	0.0	N/A
GA at outcome, weeks, mean	39.0	39.0	-0.05	34.6	34.7	0.02	28.6	30.1	0.25	9.8	10.1	0.12

Age is the age at time of the outcome. ED visits, gestational diabetes, and preeclampsia were measured as any-listed diagnosis on any claim in the prenatal window
 CED IBM MarketScan Explorers Claims-EMR Data Set, ED emergency department, GA gestational age, GDM gestational diabetes mellitus, N/A not available, SMD standardized mean differ-
 ence

was within 28 days of the adjudicator-assigned GA 99.5% of the time. These ICD-10-CM/PCS results also are consistent with reported agreement for ICD-9-CM-defined GA for live births within 30 days (98%) [18].

The algorithms produced a preterm delivery rate of 12.3% in the claims data, which is higher than the rate of preterm birth reported in the vital statistics birth file (9.9% in 2016) [28]. This was likely due to the underestimation of GA described above. The rate of stillbirth per 1000 live births and stillbirths in our study was 4.9—slightly lower than the national rate reported using vital statistics data (5.9) [29] but similar to estimates in another validation study of a privately insured population using ICD-9-CM (5.4) [13]. Thus, the algorithms may overestimate PTBs and underestimate stillbirths.

Our study supports the application of data from an EMR repository consisting of more than 30 health systems and more than 5 million patients with both EMR and claims data for validation studies. Approximately 16–18% of pregnancy episodes had enough obstetric EMRs to be included in the sample pool. Except for stillbirths, this provided a large enough sample for physician adjudication. However, attrition raises questions of selection bias and generalizability, which may also be present in prior paper-based and EMR-based chart review validation studies. Use of EMR data, which can be queried, in place of paper charts, which must be manually abstracted, for validation significantly reduces labor, cost, and time, making validation studies more feasible. These methods may be applied to other EMR systems that use SNOMED and LOINC clinical concepts for structured components.

Like most validation studies, the cases selected for adjudication were a convenience sample for which EMR charts with sufficient obstetric information were available. We evaluated potential selection bias by comparing characteristics among the pregnancy episodes in the database for which an outcome and start date were assigned with those with sufficient obstetric EMR information to be included in the sample pool from which we randomly selected cases for clinical review and adjudication. The sample pool generally resembled the broader population of pregnancy episodes. However, the proportion of white women was higher in the sample pool. This difference could be due to racial disparities in care-seeking behavior or access to care, with white women having more prenatal encounters than women of other races/ethnicities; therefore, they were more likely to have enough obstetric information to be included in the sample pool. The generalizability of our validated algorithms to other privately insured populations should be informed by this potential selection bias. Additionally, the privately insured pregnant population identified using the algorithms may not represent pregnant women in the United States;

Table 4 Comparison of the claims-based algorithms with the results from physician adjudication

Outcome	Adjudicated with GAIA levels 1–3 certainty No.	Percent agreement	
		%	95% CI
Outcomes			
Live birth	185	100.0	97.5–100.0
Full-term live birth	92	97.8	91.8–99.9
Preterm live birth	93	62.4	52.0–71.7
Stillbirth	24	70.8	50.2–85.5
Spontaneous abortion	75	100.0	93.9–100.0
Gestational age			
Live birth			
± 7 days	185	83.8	77.6–88.5
± 14 days	185	95.7	91.5–98.0
± 21 days	185	97.3	93.5–99.0
± 28 days	185	99.5	96.6–100.0
Full-term live birth			
± 7 days	92	85.9	77.0–91.8
± 14 days	92	98.9	93.3–100.0
± 21 days	92	98.9	93.3–100.0
± 28 days	92	100.0	95.0–100.0
Preterm live birth			
± 7 days	93	81.7	72.4–88.5
± 14 days	93	92.5	84.8–96.6
± 21 days	93	95.7	88.9–98.7
± 28 days	93	98.9	93.4–100.0
Stillbirth			
± 7 days	24	66.7	46.2–82.4
± 14 days	24	79.2	58.6–91.4
± 21 days	24	79.2	58.6–91.4
± 28 days	24	79.2	58.6–91.4
Spontaneous abortion			
± 7 days	75	61.3	49.8–71.7
± 14 days	75	81.3	70.7–88.8
± 21 days	75	88.0	78.3–93.9
± 28 days	75	94.7	86.4–98.4

CI confidence interval, GAIA Global Alignment of Immunization safety Assessment in pregnancy

thus, our findings may have limited generalizability to other populations, such as Medicaid-insured individuals.

Finally, our study supports application of the GAIA framework for validation studies. GAIA guidelines were designed for case ascertainment in clinical trials, aiming to set a meaningful and standardized process for data collection. We were limited to retrospectively collected structured data elements recorded on EMRs and needed to operationalize the GAIA framework for our purposes. This work can be a resource for mapping SNOMED and LOINC codes to GAIA concepts in future studies. In translating GAIA-based

case definitions to EMR language, standardization and automation were gained with some unavoidable loss of clinical granularity. Because the GAIA framework comprises standard, globally harmonized outcome definitions and data collection guidelines for monitoring the safety of vaccines during pregnancy, it is informative, valuable, and feasible to use the framework in observational studies as well.

This study has several limitations. The algorithms are based on administrative claims, for which the quality of coding may vary over time, settings, and regions. Claims used in this study included the period right after ICD-10-CM/PCS implementation, and the quality of coding may improve over time. Our findings may have limited generalizability to other claims databases that differ greatly from the care settings and regions represented in our data.

This study focuses on four pregnancy outcomes of interest (FTB, PTB, SA, and stillbirth). The algorithms do not define some other pregnancy outcomes (therapeutic abortion, elective abortion, and ectopic pregnancy) that may be important to certain research questions related to drug safety. The algorithms classified several clinical codes into a single outcome; however, misclassification could occur because these codes lack specificity (e.g., O08 [Complications following ectopic and molar pregnancy] is classified as ectopic pregnancy but can also indicate molar pregnancies). Supplemental Table 1 highlights these codes (see ESM).

Pregnancy episodes with a start and end date for the four outcomes of interest, and hence retained in this study, are a subset of pregnancies in the claims; thus, our methods are not intended for estimating the incidence of pregnancy outcomes within a population. A defined pregnancy episode with a start and end date is critical to determine medication exposure during pregnancy for safety studies. GA is critical to the assignment of pregnancy start date and the specificity of the outcomes because it distinguishes between SAs and stillbirths and between PTBs and FTBs. Some pregnancy episodes were excluded because they lacked a pregnancy start date estimate. This attrition varied by outcome: 22.3% (990/4442) for SAs and 28.8% (61/212) for stillbirths, versus 2.1% for live births (669/31,270). For SAs, the high attrition was primarily due to a lack of GA information. SAs have shorter gestational periods with fewer encounters to record GA, and the ICD-10-CM coding rules specify that Z3A codes (weeks of GA) should not be assigned for pregnancies with abortive outcomes. For stillbirths, the high attrition was primarily due to the exclusion of a subset that had an estimated GA of < 20 weeks. Although these cases could be SAs miscoded as stillbirths, we did not reclassify them to SAs because of the uncertainty and the low number. Despite the strict criteria in classifying stillbirths, the percent agreement of stillbirths (70.8%) was lower than that of live births and SAs. Of the 24 stillbirths reviewed, 7 were determined to be misclassifications (5 adjudicated as SAs

and 2 as FTBs) (not shown). At the service record level, stillbirths were also most likely to have conflicting information, including concurrent diagnosis of SAs or live births (not shown). This indicates that the ICD-10-CM/PCS coding of stillbirths is less reliable than coding of other outcomes during the study period.

Although our use of EMRs was innovative, using structured EMR components alone as a reference method has not been evaluated in other validation studies. Future studies may advance our work by comparing structured EMRs to EMRs with clinical notes and to full medical charts (the gold standard) for pregnancy outcome ascertainment. Furthermore, pregnancies identified in claims do not always have EMRs available during pregnancy. Pregnancies with relevant EMRs for validation are a convenience sample. Those with shorter gestational periods, such as SAs, may be less likely than longer pregnancies to be selected into the validation sample because they have a shorter window for healthcare encounters producing relevant EMRs.

Additionally, of the validation sample, 40 cases were excluded because of insufficient EMRs for adjudication. Although data were sparse, there were some noted differences between the excluded cases and sample pool. However, exclusion of the 40 cases is unlikely to be associated with coding accuracy in claims and, therefore, with the algorithms' performance estimate. The range of percent agreement, if the 40 cases could have been adjudicated, was similar to the estimates from the main analysis.

5 Conclusions

We found high levels of agreement between ICD-10-CM/PCS claims-based algorithms and physician adjudication of EMRs regarding GA and the outcomes of live births, stillbirths, and SAs. The study suggests the feasibility of using structured components of EMRs in a clinical review process for pregnancy outcome adjudication guided by the GAIA framework. The algorithms presented may be applied to administrative claims data in studies evaluating maternal exposures to prescription drugs and vaccines and their association with maternal and infant outcomes. They may be a timely addition to existing algorithms in assessing new interventions during pregnancy in the COVID-19 pandemic.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40264-021-01113-8>.

Acknowledgements We thank our physician adjudicators: Dr Dana Lee (University of Maryland Medical System), Dr Anne Foster (Presbyterian Medical Group), and Dr Michael Padula (Children's Hospital of Philadelphia). We appreciate the following colleagues' valuable comments to the manuscript: Dr Jeff Roberts and Dr Cindy Zhou from the US Food and Drug Administration, Center for Biologics Evaluation

and Research; Dr Julia Simard from Stanford University; and Michael Lu and Jiemin Liao from Acumen, LLC. The manuscript was reviewed and approved through the clearance process of the US Food and Drug Administration, Center for Biologics Evaluation and Research.

Declarations

Funding This work was supported by the US Food and Drug Administration through the Department of Health and Human Services (HHS) Contract Number HHSF223201810022I Task Order 75F40119F19002.

Conflicts of interest/Competing interests Keran Moll, Hui Lee Wong, Kathryn Fingar, Shayan Hobbi, Minya Sheng, Timothy A. Burrell, Linda O. Eckert, Flor M. Munoz, Bethany Baer, Azadeh Shoaibi, and Steven Anderson declare that they have no conflict of interest.

Availability of data and material All data supporting the findings of this study are available within the article and its supplementary information files.

Code availability Programming codes are available upon request.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Author contributions Keran Moll, Hui Lee Wong, Kathryn Fingar, Shayan Hobbi, Timothy A. Burrell, Linda O. Eckert, Flor M. Munoz, and Azadeh Shoaibi contributed to the study conception and design. Acquisition of data was performed by Keran Moll, Shayan Hobbi, and Minya Sheng. Analysis and interpretation of data were conducted by Keran Moll, Hui Lee Wong, Kathryn Fingar, Shayan Hobbi, Timothy A. Burrell, Bethany Baer, Azadeh Shoaibi, and Steven Anderson. All authors contributed to drafting or revising the article. All authors read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial 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-nc/4.0/>.

References

1. Feibus K, Goldkind SF. Pregnant women and clinical trials: scientific, regulatory, and ethical considerations. 18 Oct 2010. <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/Women/HealthResearch/UCM243540.ppt>. Accessed 20 Sep 2019
2. Food and Drug Administration, Center for Biologics Evaluation and Research. Development and licensure of vaccines to prevent COVID-19: guidance for industry. Jun 2020. <https://www.fda.gov/media/139638/download>. Accessed 23 Nov 2020
3. Zambrano LD, Ellington S, Strid P, Galang RR, Oduyebo T, Ton VT, et al. Update: characteristics of symptomatic women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status—United States, January 22–October 3, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(44):1641–7. <https://doi.org/10.15585/mmwr.mm6944e3>.
4. Shimabukuro TT, Kim SY, Myers TR, Moro PL, Oduyebo T, Panagiotakopoulos L, et al. Preliminary findings of mRNA COVID-19 vaccine safety in pregnant persons. *N Engl J Med.* 2021;384(24):2273–82. <https://doi.org/10.1056/NEJMoa2104983>.
5. Collier AY, McMahan K, Yu J, Tostanoski LH, Aguayo R, Ansel J, et al. Immunogenicity of COVID-19 mRNA vaccines in pregnant and lactating women. *JAMA.* 2021;325(23):2370–80. <https://doi.org/10.1001/jama.2021.7563>.
6. Shanes ED, Otero S, Mithal LB, Mupanomunda CA, Miller ES, Goldstein JA. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccination in pregnancy: measures of immunity and placental histopathology. *Obstet Gynecol.* 2021;138(2):281–3. <https://doi.org/10.1097/AOG.0000000000004457>.
7. Embryogenesis and fetal morphological development. In: Cunningham F, Leveno KJ, Bloom SL, Spong CY, Dashe JS, Hoffman BL, et al., editors. *Williams obstetrics*. 24th ed. New York: McGraw Hill; 2014. pp. 127–153.
8. Ailes EC, Simeone RM, Dawson AL, Petersen EE, Gilboa SM. Using insurance claims data to identify and estimate critical periods in pregnancy: an application to antidepressants. *Birth Defects Res A Clin Mol Teratol.* 2016;106(11):927–34. <https://doi.org/10.1002/bdra.23573>.
9. Bird ST, Toh S, Sahin L, Andrade SE, Gelperin K, Taylor L, et al. Misclassification in assessment of first trimester in-utero exposure to drugs used proximally to conception: the example of letrozole utilization for infertility treatment. *Am J Epidemiol.* 2019;188(2):418–25. <https://doi.org/10.1093/aje/kwy237>.
10. Andrade SE, Bérard A, Nordeng HME, Wood ME, van Gelder MMHJ, Toh S. Administrative claims data versus augmented pregnancy data for the study of pharmaceutical treatments in pregnancy. *Curr Epidemiol Rep.* 2017;4(2):106–16. <https://doi.org/10.1007/s40471-017-0104-1>.
11. Andrade SE, Scott PE, Davis RL, Li D-K, Getahun D, Cheetham TC, 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>.
12. Blotière P-O, Weill A, Dalichamp M, Billionnet C, Mezzarobba M, Raguideau F, et al. Development of an algorithm to identify pregnancy episodes and related outcomes in health care claims databases: an application to antiepileptic drug use in 4.9 million pregnant women in France. *Pharmacoepidemiol Drug Saf.* 2018;27(7):763–70. <https://doi.org/10.1002/pds.4556>.
13. Hornbrook MC, Whitlock EP, Berg CJ, Callaghan WM, Bachman DJ, Gold R, et al. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. *Health Serv Res.* 2007;42(2):908–27. <https://doi.org/10.1111/j.1475-6773.2006.00635.x>.
14. Li Q, Andrade SE, Cooper WO, Davis RL, Dublin S, Hammad TA, et al. Validation of an algorithm to estimate gestational age in electronic health plan databases. *Pharmacoepidemiol Drug Saf.* 2013;22(5):524–32. <https://doi.org/10.1002/pds.3407>.
15. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernández-Díaz S. Algorithms to estimate beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf.* 2013;22(5):16–24. <https://doi.org/10.1002/pds.3284>.
16. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. *PLoS ONE.* 2018;13(2): e0192033. <https://doi.org/10.1371/journal.pone.0192033>.

17. Mikolajczyk RT, Kraut AA, Garbe E. Evaluation of pregnancy outcome records in the German Pharmacoepidemiological Research Database (GePaRD). *Pharmacoepidemiol Drug Saf*. 2013;22(8):873–80. <https://doi.org/10.1002/pds.3467>.
18. Naleway AL, Gold R, Kurosky S, Riedlinger K, Henninger ML, Nordin JD, et al. Identifying pregnancy episodes, outcomes, and mother–infant pairs in the Vaccine Safety Datalink. *Vaccine*. 2013;31(27):2898–903. <https://doi.org/10.1016/j.vaccine.2013.03.069>.
19. Jones CE, Munoz FM, Kochhar S, Vergnano S, Cutland CL, Steinhoff M, et al. Guidance for the collection of case report form variables to assess safety in clinical trials of vaccines in pregnancy. *Vaccine*. 2016;34(49):6007–14. <https://doi.org/10.1016/j.vaccine.2016.07.007>.
20. Jones CE, Munoz FM, Spiegel HML, Heininger U, Zuber PLF, Edwards KM, et al. Guideline for collection, analysis and presentation of safety data in clinical trials of vaccines in pregnant women. *Vaccine*. 2016;34(49):5998–6006. <https://doi.org/10.1016/j.vaccine.2016.07.032>.
21. Quinn JA, Munoz FM, Gonik B, Frau L, Cutland C, Mallett-Moore T, et al. Preterm birth: case definition and guidelines for data collection, analysis, and presentation of immunisation safety data. *Vaccine*. 2016;34(49):6047–56. <https://doi.org/10.1016/j.vaccine.2016.03.045>.
22. Rouse CE, Eckert LO, Babarinsa I, Fay E, Gupta M, Harrison MS, et al. Spontaneous abortion and ectopic pregnancy: case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. *Vaccine*. 2017;35(48 Pt A):6563–74. <https://doi.org/10.1016/j.vaccine.2017.01.047>.
23. Da Silva FT, Gonik B, McMillan M, Keech C, Dellicour S, Bhange S, et al. Stillbirth: case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. *Vaccine*. 2016;34(49):6057–68. <https://doi.org/10.1016/j.vaccine.2016.03.044>.
24. Kuklina EV, Whiteman MK, Hillis SD, Jamieson DJ, Meikle SF, Posner SF, et al. An enhanced method for identifying obstetric deliveries: implications for estimating maternal morbidity. *Matern Child Health J*. 2008;12(4):469–77. <https://doi.org/10.1007/s10995-007-0256-6>.
25. Moll K, Wong H-L, Fingar K, Hobbi S, Sheng M, Burrell T, et al. Study protocol: validating pregnancy outcomes and gestational age in a claims-EMR linked database version 2.6. Food and Drug Administration. Jun 2020. https://www.bestinitiative.org/wp-content/uploads/2020/08/Validating_Pregnancy_Outcomes_Linked_Database_Protocol_2020-1.pdf. Accessed 11 Jun 2021
26. Murad MH, Wang Z, Chu H, Lin L. When continuous outcomes are measured using different scales: guide for meta-analysis and interpretation. *BMJ*. 2019;364: k4817. <https://doi.org/10.1136/bmj.k4817>.
27. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates; 1988.
28. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Drake P. Births: final data for 2016. *National Vital Statistics Reports*. Volume 67, Number 1. National Center for Health Statistics. 2018. https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_01.pdf. Accessed 14 Oct 2020
29. Centers for Disease Control and Prevention. Data and statistics. Page last reviewed 13 Aug 2020. <https://www.cdc.gov/ncbddd/stillbirth/data.html>. Accessed 11 Jun 2021