U.S. Food and Drug Administration
Center for Biologics Evaluation and Research
Office of Biostatistics and Epidemiology

Biologics Effectiveness and Safety (BEST) Contract 1:
Data Tools and Infrastructure for Surveillance of Biologics Task Order 1

Study Protocol:
Validating Pregnancy Outcomes and Gestational Age in a Claims-EMR Linked Database
Version 2.6

June 2020

Authors: Moll K[1], Wong H-L[2], Fingar K[1], Hobbi S[3], Sheng M[1], Burrell T[1], Baer B[2], Eckert L[4], Munoz F[5], Obidi J[2], Sepulveda K[2], Shoaibi A[2], Anderson S[2]

[1]IBM Watson Health, Cambridge, MA, USA, [2]US Food and Drug Administration, Center for Biologics Evaluation and Research, 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

TABLE OF CONTENTS

History of Modifications

Version	Date	Modification
1.11	2/25/2019	• Clarified the objective of Aim 2 in the Introduction is to compare gestational age estimates and pregnancy outcomes from Aim 1 with "the clinician adjudication results based on the review of the structured data elements in the electronic medical record (EMR)." • Clarified the objective of Aim 3 in the Introduction, as well as the body of Aim 3. Added a reference to a similar study where the timing of first trimester in-utero exposure to letrozole was evaluated (Bird et al). • Added two procedure codes: S4037 to ET; S4035 to IUI
1.12	3/13/2019	• Added Aim 2a to describe activities related to identifying a study cohort with sufficient EMR data for outcomes validation
1.2	4/8/2019	• Incorporated AS comments; updated Aim 1 to reflect recent modifications
2	5/1/2019	• Added case selection strategy
2.1	6/29/2019	• Updated Reviewer Adjudication (Aim 2b) for clarity
2.2	7/15/2019	• Updated Background and Aims 1 and 2 sections for additional clarity, as well as additional content on the sampling strategy and post-validation analysis activities
2.3	8/19/2019	• Updated Aim 2b to capture additional methodology for validation analysis, including PPA calculation, and confidence interval calculation
2.4	9/12/2019	• Updated Aim 2b to add additional detail on PPA calculation, confidence interval calculation, and sensitivity analysis
2.5	10/23/2019	• Added special rule for full-term GA code O7582 to use 38 weeks as GA • Added section for Interreviewer Concordance Analysis
2.6	6/8/2020	• Minor stylistic edits for public posting

A Introduction

Background

- The Food and Drug Administration Center for Biologics Evaluation and Research is interested in analyzing administrative claims data to track pregnancy outcomes following maternal exposure to biologic products. Health outcomes of interest include live birth (LB), stillbirth (SB, 20+ weeks of gestation), and spontaneous abortion (SA, <20 weeks of gestation). LBs are further classified into preterm delivery of a live-born infant (PTB, <37 weeks of gestation) or full-term delivery of a live born infant (FTB, 37+ weeks of gestation).

- Each of these outcomes can be identified in claims data using International Classification of Diseases, Clinical Modification (ICD-CM) codes. All the outcomes are defined by gestational age (GA).

- On October 1, 2015, the United States transitioned to a modified version of the World Health Organization's International Classification of Diseases, Tenth Revision, Clinical Modification-Procedure Coding System (ICD-10-CM/PCS), replacing the ninth revision (ICD-9) diagnosis and procedure coding system with the ICD-10-CM diagnosis coding system for most inpatient and outpatient medical encounters and the ICD-10-PCS for inpatient hospital procedures. Although more granular detail is available on GA in ICD-10-CM than in ICD-9-CM, the coding quality is largely unknown.

- ICD-10-CM codes for PTB, FTB, SB, and SA, as well as for the GA in weeks, trimester, and preterm/full term status, should be examined before proceeding with future research. If the performance of claims-based algorithms is poor, predictive models may be explored.

- The Global Alignment of Immunization safety Assessment in pregnancy (GAIA) project was formed to establish a global, common understanding of outcomes and approaches to monitoring the safety of vaccines used in pregnancy. GAIA developed a set of case definitions for obstetric outcomes, including SA, SB, PTB, and GA. This study will use the GAIA case definitions as the reference method for case validation in electronic medical records (EMRs).

- The Pregnancy Outcomes validation study comprises (1) development of ICD-10-CM/PCS-based hierarchical algorithms to classify pregnancy outcomes and estimate GA in claims (Study Aim 1); (2) validation of the claims-based algorithms by physician adjudication of linked EMR charts using GAIA clinical definitions as the reference method through a semiautomated questionnaire-driven EMR chart review (Aim 2); and (3) calculation of the proportion of algorithm-determined outcomes or gestational age that were confirmed by EMR adjudication -the percentage positive agreement (PPA) (Aim 2). Exploratory aims include using the timing of fertility treatment and linked infant records in claims databases to validate GA estimates (Aim 3a) and predictive analytics using additional claims data elements to improve the algorithms (Aim 3b).

B Study Aims

Aim 1

- Using claims data, we will develop hierarchical algorithms to classify pregnancy outcomes (PTBs, FTBs, SBs, or SAs) and to estimate GA at the time of delivery or fetal loss.

Aim 2

- We will assess the performance of the claims-based algorithms developed in Aim 1 by comparing the results to physician adjudication of linked structured EMRs.

Aim 3 (Exploratory Aims)

- **Aim 3a.** Subject to time, budget, and study goals, we may conduct a concordance analysis using MarketScan data. The algorithms developed in Aim 1 will be evaluated among intrauterine insemination or embryo transfer procedure recipients with linked infant records for GA estimates and the classification of outcomes.

- **Aim 3b.** We may develop predictive models to better classify certain pregnancy outcomes.

C Methods

C1 Data Sources

- **For primary Aims 1 and 2, the IBM® MarketScan® Explorys® Claims-EMR Data Set (CED) will be used.**

  The CED is research-ready data that links the longitudinal treatment and economic claims records of patients from the MarketScan Commercial Claims and Encounters Database to the same patients' clinical records from Explorys EMR data. The IBM Explorys database is a large and living clinical data asset, built through direct connections to an expanding array of large health system partners, comprising of more than 30 health systems and spanning academic centers and community practices. The EMR available is limited to structured data elements that include diagnoses, procedures, immunizations, vital signs and biometrics, medical/surgical history, laboratory results, implantable devices, patient-reported outcomes, as well as inpatient drug administrations and ambulatory prescriptions. The combined data set deterministically links patients in both the MarketScan and Explorys data sets, and provides clinical, claims, and financial data to support health economics and outcomes research, market research, and epidemiologic analyses.

- **For exploratory Aim 3, the IBM MarketScan Commercial Database will be used, in addition to CED.**

  The IBM® MarketScan® Commercial Database includes health insurance claims across the continuum of care (e.g., inpatient, outpatient, outpatient pharmacy, carve-out behavioral healthcare) as well as enrollment data from large employers and health plans across the United States that provide private healthcare coverage for employees, their spouses, and dependents. This administrative claims database includes a variety of fee-for-service, preferred provider organization, and capitated health plans

C2 Study Population

We will include female patients aged 12–55 years at the time of the outcomes of interest who (1) were continuously enrolled with medical benefit during the pregnancy episode and (2) had the pregnancy outcome on or after August 1, 2016.

C3 Study Design

3.a Aim 1: Estimate Gestational Age (GA) and Classify Pregnancy Outcomes (LBs, SBs, SAs) Using Claims Data

We will adapt the pregnancy episode algorithm developed by Hornbrook et al.,[1] and later modified by Naleway et al.[2] and Matcho et al.,[3] to identify pregnancy episodes in this study.

i. Pregnancy Outcome Identification

Although the focus of this study is LBs, SBs, and SAs, any pregnancy endpoint can be relevant in determining which outcome events are clinically possible, given spacing between outcomes for the same woman (e.g., an abortion one week after a live birth is not clinically possible.), We will first identify all potential pregnancy endpoints based on the presence of ICD-10-CM/PCS and Healthcare Common Procedure Coding System (HCPCS) codes on the claim records. Events indicating the end of pregnancy include LBs, SBs, SAs, ectopic pregnancies (ECTs), elective abortions (ABs), trophoblastic and other abnormal products of conception (TROs), and deliveries with unknown outcomes (DELIVs). We will identify all events that indicate the end of a pregnancy

in all healthcare settings. Clinical codes[i] for identifying each outcome are available in Appendix D1: *Pregnancy Episodes Clinical Codes.xlsx.*

ii. Constructing Pregnancy Episodes

A *pregnancy episode* will be defined as the time frame between the estimated date of the last menstrual period and the date of the outcome.

a. Pregnancy Outcome Assignment

Pregnancy outcomes will be assigned sequentially, following a hierarchical order, starting with the outcomes considered most reliably coded in claims data. Within each type of outcome, we consider the date of first occurrence with the specific ICD-10-CM diagnosis or procedure codes as the date of the event, regardless of the care setting. Table 1 summarizes the minimum number of days required between two outcomes for the subsequent outcomes to be considered plausible.

- LBs will be placed on the timeline first. The first LB record on or after October 1, 2015, is assumed to be valid and placed on the timeline. Subsequent LB events will be evaluated in the order of occurrence and kept as valid only if the LB event is at least 182 days after the previous LB event.

- SBs will be the next outcome evaluated. Similarly, SBs are assessed sequentially in the order of occurrence. The date of the first SB event will be compared with the dates of any LBs that are already on the timeline, only to remain valid if the SB event occurred at least 168 days after a previous LB event and at least 182 days before a subsequent LB event. The remaining SB events will be assessed against all valid LB and SB events already on the timeline (SBs must be 168 days apart from each other).

- Next, deliveries with unknown outcomes (DELIVs) will be placed on the timeline through a similar process.

- The same process will continue in the following order of hierarchy.
  - Trophoblastic and other abnormal products of conception (TRO)
  - Ectopic pregnancy (ECT)
  - Elective abortion (AB)
  - Spontaneous abortion (SA)

Table 1. Minimum days/weeks between pregnancy outcomes

Preceding Outcome	Subsequent Outcome, Days (Weeks)						
	LB	**SB**	**DELIV**	**TRO**	**ECT**	**AB**	**SA**
LB	182 (26)	168 (24)	168 (24)	70 (10)	70 (10)	70 (10)	56 (8)
SB	182 (26)	168 (24)	168 (24)	70 (10)	70 (10)	70 (10)	56 (8)
DELIV	182 (26)	168 (24)	168 (24)	70 (10)	70 (10)	70 (10)	56 (8)
TRO	168 (24)	154 (22)	154 (22)	56 (8)	56 (8)	56 (8)	42 (6)
ECT	168 (24)	154 (22)	154 (22)	56 (8)	56 (8)	56 (8)	42 (6)
AB	168 (24)	154 (22)	154 (22)	56 (8)	56 (8)	56 (8)	42 (6)
SA	168 (24)	154 (22)	154 (22)	56 (8)	56 (8)	56 (8)	42 (6)

Abbreviations: AB, elective abortion; DELIV, delivery with unknown outcome; ECT, ectopic pregnancy; LB, live birth; SA, spontaneous abortion; SB, stillbirth; TRO, trophoblastic and other abnormal products of conception.

---

[i] Unless otherwise indicated, codes refer to ICD-10-CM diagnosis codes; for LBs, the codes in bold indicate there were multiple gestations, both live and stillborn. To determine pregnancy episodes, these will be put on the time scale but will eventually be excluded from the study.

After all valid pregnancy outcomes are placed on the timeline, we will select only the outcomes of interest (LBs, SBs, and SAs) occurring on or after August 1, 2016. This step ensures that the lookback periods for all outcomes are after October 1, 2015, when the ICD-10-CM/PCS coding system was implemented. We will set the maximum pregnancy term from an outcome date (Table 2) as the lookback window for the outcome. If a preceding outcome is present during the lookback window, the start date of the window will be adjusted according to a minimum allowable number of days before a subsequent pregnancy could start. For example, we will look back 301 days for an LB event to gather potentially relevant information for this LB event. However, if an SA is present 300 days prior, we will look back 300-14 days instead.

Table 2. Pregnancy term range and minimum spacing

Outcome	Spacing, Days (Weeks)		
	Maximum Pregnancy Term	Minimum Pregnancy Term	Minimum Days Before Subsequent Pregnancy Start
LB	301 (43)	154 (22)	28 (4)
SB	301 (43)	140 (20)	28 (4)
DELIV	301 (43)	140 (20)	28 (4)
TRO	112 (16)	42 (6)	14 (2)
ECT	84 (12)	42 (6)	14 (2)
AB	168 (24)	42 (6)	14 (2)
SA	133 (19)	28 (4)	14 (2)

Abbreviations: AB, elective abortion; DELIV, delivery with unknown outcome; ECT, ectopic pregnancy; LB, live birth; SA, spontaneous abortion; SB, stillbirth; TRO, trophoblastic and other abnormal products of conception.

b. Pregnancy Start Date and Gestational Age Determination

Several sources of information in claims data provide GA information:

- Procedures and timing of embryo transfer (ET) or intrauterine insemination (IUI)
- GA on delivery records indicating full-term/preterm status or trimester information
- Diagnoses with GA in weeks during prenatal or delivery encounters
- Pregnancy screening tests[ii] and timing

A pregnancy start date for each pregnancy episode will be assigned using the information within the lookback window, following a hierarchy:

1. ET or IUI date available. If multiple ET or IUI procedures are performed, we will select the one closest to the outcome date.
   a. Pregnancy start date = ET/IUI date – 7*2 weeks + 1
2. First-trimester ultrasonography with concurrent diagnosis codes indicating GA in weeks
   a. Pregnancy start date = date of scan – 7*GA weeks coded + 1
3. Nuchal translucency (NT) scan with concurrent diagnosis codes indicating GA in weeks
   a. Pregnancy start date = date of scan – 7*GA weeks coded + 1

[ii] Unless otherwise specified, this algorithm uses evidence of billed pregnancy screening tests using procedure codes. The results of such tests are not available in the IBM MarketScan claims data.

4. Anatomical ultrasound with concurrent diagnosis codes indicating GA in weeks
   a. Pregnancy start date = date of scan – 7*GA weeks coded + 1
5. Diagnosis codes indicating GA in weeks (Z3Ax codes)
   a. Pregnancy start date = date of diagnosis – 7*GA weeks coded + 1
   b. If multiple pregnancy start dates are calculated because there were multiple estimates of GA in weeks, we will select the mode (or median, if no mode exists) of potential pregnancy start dates as the pregnancy start date.
6. NT scan (guidelines recommend this scan to be performed before 13 5/7 weeks)
   a. Pregnancy start date = date of NT scan – 90 + 1 (NT timing of 90 days is based on Matcho et al.[3]).
7. Chorionic villus sampling (CVS) sampling (guidelines recommend done before 13 weeks)
   a. Pregnancy start date = date of CVS – 12 weeks*7 + 1 (CVS timing of 12 weeks was empirically estimated using MarketScan data).
8. Cell-free DNA prenatal test (usually done between 10 and 14 weeks)
   a. Pregnancy start date = date of cell-free DNA test – 12 weeks*7 + 1 (cell-free DNA test timing of 12 weeks was empirically estimated in MarketScan data)
9. Diagnosis codes indicating full-term status at delivery or fetal loss. Except for O471 (False labor at or after 37 completed weeks of gestation), full-term diagnosis codes must be within ±7 days of outcome dates to be considered related to the outcome. O471 must be less than 42 days before the outcome date to be considered as related to the outcome.
   a. For all full-term codes except for O7582 (Onset [spontaneous] of labor after 37 completed weeks of gestation but before 39 completed weeks of gestation, with delivery by [planned] cesarean section) and the outcome is either LB or SB, pregnancy start date = date of the outcome – 39*7 + 1 (Margulis et al.[4])
   b. For O7582 and the outcome is either LB or SB, pregnancy start date = date of the outcome – 38*7 + 1
10. ICD-10-CM diagnosis code indicating trimester at delivery or fetal loss (within ±7 days of the outcome date)
    a. First trimester (0–3 6/7 weeks): Pregnancy start date = date of the outcome – 70 + 1
    b. Second trimester (14 0/7–27 6/7 weeks): Pregnancy start date = date of the outcome – 147 + 1
    c. Third trimester (28 0/7–40 6/7 weeks): Pregnancy start date = date of the outcome – 241 + 1
    d. If multiple trimester codes are used and the mode exists, we will select the mode.
    e. If multiple trimester codes are used but different trimester codes are used equally, we will not assign pregnancy start date at this level and move on to the next level of the hierarchy.
11. Diagnosis codes indicating preterm status at delivery or fetal loss. Preterm diagnosis codes must be within ±7 days of outcome dates to be considered related to the outcome.
    a. If it is a preterm code and the outcome is either LB or SB, pregnancy start date = date of the outcome – 35*7 + 1 (Margulis et al.)
12. First gestational diabetes screening (usually done at GA between 24 and 28 weeks)
    a. Pregnancy start date = date of gestational diabetes screening – 26 weeks*7 + 1
13. ICD-10-CM diagnosis code indicating GA in trimester on services more than 7 days before the outcome date. Select the trimester code that is closest to the outcome date.
    a. First trimester (0–13 6/7 weeks): Pregnancy start date = service date with the trimester code – 70 + 1

      b.   Second trimester (14 0/7–27 6/7 weeks): Pregnancy start date = service date with the trimester code – 147 + 1

      c.   Third trimester (28 0/7–40 6/7 weeks): Pregnancy start date = service date with the trimester code – 241 + 1

At each level of the hierarchy, we will evaluate the estimated pregnancy start date and GA at the outcome date for reasonableness. The estimated pregnancy start date is considered plausible if it occurs between the following two time points. For Levels 1–4 of the hierarchy, if the estimated pregnancy start date is not plausible, the pregnancy episode is excluded. For Levels 5–13 of the hierarchy, if the estimated pregnancy start date is not plausible, the pregnancy start date will instead be estimated at the next level.

- Time point 1: The start date of the lookback window
- Time point 2: The start date of the minimum pregnancy term for the outcome, that is, GA of ≥22 weeks for LB, ≥20 weeks for SB, or ≥4 weeks for SA.

After each of the 13 steps, if there is no evidence to set the pregnancy start date, we will set the pregnancy start date as unknown. For pregnancy episodes with pregnancy start dates, GA in days at the time of the outcome is calculated as GA = outcome date – pregnancy start date + 1.

Note, a full list of codes related to GA is available in Appendix D1: *Pregnancy Episodes Clinical Codes.xlsx*.

c. Final Pregnancy Outcome Assignment

After GA at the time of outcomes is determined, pregnancy outcomes are assigned as follows:

1. LB: Separate LBs into PTBs and FTBs
    a. PTBs
        i.   If GA is <37 weeks at the date of the outcome, set an LB as a PTB.
    b. FTBs
        i.   If GA is ≥37 weeks at the date of the outcome, set an LB as a FTB.
2. SB and SA: With the built-in plausibility check in the GA hierarchical algorithm, SAs with an assigned GA will all have a GA between 4 and 19 weeks; SBs with an assigned GA will all have a GA between 20 and 43 weeks.

3.b   Aim 2a: Case Pool Selection and Assessment of an Infrastructure to Conduct Validation Studies Using a Claims-EMR Linked Database

No prior studies have attempted outcomes validation using structured components of EMR data. This study serves as a use case in building an infrastructure to facilitate validations of this type. In this aim, we examine the feasibility of this infrastructure in using structured components of EMR data to validate algorithms developed in claims-based data.

CED is a combination of MarketScan administrative claims and Explorys EMR data for a population that is present in both data sources. The population with at least one activity in EMR data and at least one claim in MarketScan claims data is linked deterministically. The advantage of this combined data is that patients' comprehensive longitudinal claims can be viewed along with snapshots of EMR activities and clinical results, potentially affording a more complete picture. Because of the "snapshot" nature of EMR data, we expect that only a subset of the cohort of interest identified in Aim 1 will have sufficient EMR information overlapping the gestational period to make the case validation possible. As detailed below, we will place cases determined as having sufficient EMR information in case pools. The final sample of cases will then be drawn from the case pools. In this aim, we will evaluate the impact of the attrition process from the full cohort to the case pools and then to the selected cases for adjudication. We will compare population characteristics between the cases that are selected and those that are not.

The reference method for the validation of outcomes and GA in this study will be the GAIA case definitions and guidelines developed by the Brighton Collaboration Working Groups. GAIA guidelines were designed to be applied in perspective studies, aiming to set a meaningful and standardized process for data collection. In the current retrospective study, we are limited to the structured data elements that were already recorded on patients' EMR records. We will make efforts to select cases with data elements that align with GAIA-specified measurements.

a. Tier 1 Case Pool Selection

A challenge of working with EMR data is that there is no equivalent to the enrollment information in claims data. Only patients' interactions with the providers that contribute their EMR data into CED are "visible" to us. As we select the cases for EMR chart review, we want to select those patients who are receiving regular obstetric care from providers contributing EMR data so that we have a relatively comprehensive view of patients' healthcare experiences during pregnancy.

The presence of GAIA Level 1 data elements for GA on the EMR charts will provide the most accurate information to determine the pregnancy start date and GA at the time of the outcome. GA is a central parameter underlying all four outcomes of interest. GAIA GA assessment guidelines consider that women with a certain menstrual date or who have undergone IUI or ET with a confirmatory first trimester scan (≤13 6/7 weeks) or a first-trimester ultrasound established date alone represent the "gold standard" of diagnostic certainty of GA determination.[5] According to a 2017 American College of Obstetricians and Gynecologists (ACOG) Committee Opinion,[6] ultrasound measurement of the embryo or fetus in the first trimester is the most accurate method to establish or confirm GA and has an accuracy of ±5–7 days.

Given the above considerations, cases that meet the following criteria will be placed in the Tier 1 case pool (relevant clinical concept lists are in Appendix D1: *EMR Concept List.xlsx* file):

1. During the gestational period, a patient has, on average, one encounter with an OB/GYN provider every 30 days (see "OB Visit" tab for provider specialties).
    a. The average is calculated as (total number of encounter records with OB/GYN from the estimated start of the pregnancy to the date of the pregnancy outcome recorded)/(estimated GA in months).
    AND

2. The following EMR data elements are documented on a patient's records:
    a. An ultrasound scan with documented GA that is in the first trimester (for potential first trimester ultrasound see tab "US SCAN SNOMED"; for GA concepts see "GA_LOINC" and "GA_SNOMED" tabs)
        OR
    b. Reported last menstrual period (LMP) date with an ultrasound scan performed between 4 weeks and 13 6/7 weeks after LMP date (see "LMP" tab)
        OR
    c. IUI or ET procedure with an ultrasound performed between 2 weeks and 11 6/7 weeks after the date of the IUI or ET procedure (see "IUI_ET_SNOMED" tab)
    AND
3. Pregnancy outcomes are documented on EMR records (see "Outcome" tab).

b. Tier 2 Case Pool Selection

Not all women seek obstetric care during the first trimester or have the GAIA gold standard measurements recorded to best determine GA. Among the cases that do not fall in the Tier 1 case pool, we will select Tier 2 cases on the basis of the availability of data elements that afford lower accuracy in GA assessment but remain specific. According to GAIA, a first trimester pelvic bimanual examination or a second trimester ultrasound established date are considered next-level measurements in GA estimate precision. ACOG reported that GA assessment by ultrasonography in the first part of the second trimester (between 14 0/7 and 21 6/7 weeks of gestation) has an accuracy of ±7–10 days, whereas gestational age estimation based on ultrasonographyperformed between 22 0/7 and 27 6/7 weeks of gestation has an accuracy of ±10–14 days. Cases that have Tier 2 elements reported on their EMR charts provide an opportunity to test the claims-derived algorithm where only an incomplete case history is available, though with lower accuracy.

Cases that are not in the Tier 1 pool but meet the following criteria will be placed in the Tier 2 case pool:

1. The following EMR data elements are present:
    a. A first-trimester obstetric encounter and a documented GA during pregnancy (first trimester OB encounter is based on provider specialty—see "OB visit" tab)
        OR
    b. A second-trimester ultrasonography with a documented GA (for potential second-trimester ultrasound, see tab "US SCAN SNOMED"; for GA concepts, see "GA_LOINC" and "GA_SNOMED" tabs)
        OR
    c. A second-trimester ultrasonography between 14 0/7 and 27 6/7 weeks after a reported LMP (for potential second trimester ultrasound, see tab "US SCAN SNOMED"; for LMP, see "LMP" tab)
    AND
2. Pregnancy outcomes are documented on EMR records (see "Outcome" tab)

c. Stillbirth Cases (Tier 3)

Stillbirth is the rarest of the outcomes of interest in this study. In 2013, 5.96 stillbirths were reported in the United States per 1,000 live births and fetal deaths.[7] For stillbirth cases that do not fall in the Tier 1 and Tier 2 case pools, we will select all stillbirth cases that meet the following minimum criteria for case validation:

1. Documented GA in EMR (for GA concepts see "GA_LOINC" and "GA_SNOMED" tabs)
    AND
2. Documented pregnancy outcomes in EMR (see "Outcome" tab)

d. Case Selection for Adjudication

We will describe demographic and clinical characteristics for the full cohort of pregnancy episodes and each case pool, stratified by outcome types. Counts and percentages will be provided for dichotomous and polychotomous variables. Continuous variables will be summarized by providing the mean and standard deviation (see table shells in Appendix D3: *Characteristics Comparison Table Shells.xlsx*). The distribution for some potential factors that may influence the PPA between the case pool and the target population in a claims database (i.e., IBM MarketScan Commercial Database) will be compared using standardized mean difference.

Based on how population characteristics differ between the full cohort and the case pools, we will examine potential selection bias. For outcomes with large sample sizes, such as FTBs, the selection bias may be addressed by stratified sampling. However, for rare outcomes such as SBs, we will include all available samples as validation cases. In the latter situation, we will attempt to address the selection bias by weighting the PPA of a stratum with the inverse selection probability of the corresponding stratum.

From each case pool, a sample of cases will be selected for each outcome of interest :

1. **Live birth – 200 cases**
   a. From Tier 1 case pool, a random sample of 50 FTBs and 50 PTBs
   b. From Tier 2 case pool, a random sample of 50 FTBs and 50 PTBs
2. **Spontaneous abortion – 100 cases**
   a. From Tier 1 case pool, a random sample of 50 SAs
   b. From Tier 2 case pool, a random sample of 50 SAs
3. **Stillbirth**
   a. All SBs in the Tiers 1, 2, and 3 case pools will be selected.
4. **Gestational age**
   a. The GA (at time of pregnancy outcome) for all sampled cases will be adjudicated by the clinician reviewers.

3.c   Aim 2b: Validation of Pregnancy Outcomes

a.   Operationalization of GAIA Case Definitions

Case definitions developed by the GAIA project (the GAIA definitions) will be used as the reference method for defining criteria necessary to adjudicate the outcomes based on EMR review. Table 3 summarizes the relevant GAIA definitions to be used for validation of the pregnancy outcomes of interest:

Table 3. Relevant GAIA definitions to be used for validation of pregnancy outcomes of interest

Pregnancy Outcome / GA	GAIA Case Definition for Validation	Link
Gestational age (at time of delivery)[a]	Preterm birth: Case definition and guidelines for data collection, analysis, and presentation of immunization safety data	http://dx.doi.org/10.1016/j.vaccine.2016.03.045
Live birth[a]	Preterm birth: Case definition and guidelines for data collection, analysis, and presentation of immunization safety data Live births will be adjudicated based on the presence of recorded live birth outcome, and the absence of other types of outcomes (e.g., SBs, SAs), and the GA recorded on the EMR, per GAIA case definition.[a]	N/A
Stillbirth	Stillbirth: Case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data	http://dx.doi.org/10.1016/j.vaccine.2016.03.044
Spontaneous abortion	Spontaneous abortion and ectopic pregnancy: Case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data	http://dx.doi.org/10.1016/j.vaccine.2017.01.047

Abbreviations: EMR, electronic medical record; GAIA, Global Alignment of Immunization safety Assessment in pregnancy; N/A, not available.

[a] Note that validated gestational age at time of delivery will also be used to distinguish between preterm live births and full-term live births.

Each GAIA case definition has defined criteria at different levels (between 1 and 4). Each level varies 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. Pregnancy outcomes validated with the Level 1 case definition can be considered to have met the gold standard.

As mentioned in previous sections, GAIA definitions were developed as guidelines for data collection for prospective studies focused on immunization safety in pregnancy, where the specified data elements can be collected during the study. The retrospective nature of this study limits the data used for validation to structured EMR data elements. To apply and interpret the

GAIA definitions, as currently written, structuring EMR data is not straightforward to the clinician reviewers and can also introduce interreviewer inconsistencies. There are several challenges to apply the GAIA definition to structured EMR data. Certain GAIA criteria require information inherently unavailable in structured EMR data (e.g., umbilical cord pulse); granular definitions to distinguish subsets of the same outcome are not feasible or the goals of this study (e.g., the distinction for antepartum and intrapartum SBs).

We will operationalize the case definitions to EMR language. Each of the GAIA level criteria will be mapped to specific clinical terms and concepts available in a structured EMR, while addressing the nuances noted above (see Appendix D4: *Operationalized GAIA Case Definitions.xlsx*). This process will be guided and verified by two original architects of the GAIA definitions, who serve as obstetrician clinical experts on the team. The operationalized GAIA definitions will serve as a standard that different clinician reviewers can reference during case adjudication and may be used in future retrospective studies to apply GAIA definition to structured EMR data.

b. Development of Patient EMR Charts in a Chart Abstraction Tool

For each selected case, all EMR records with encounter dates between 301 days before the estimated outcome date and 30 days after the estimated outcome date and medical histories recorded up till 30 days post outcome date will be gathered to create a patient's EMR chart. The EMR charts will be organized into different views in a Microsoft® Excel-based VBA Chart Abstraction Tool residing on a secure IBM Explorys Virtual Workbench. Clinician reviewers will be blinded from the claims data and the outcome determined by the claims-based algorithms. The reviewers will adjudicate cases solely based on the patient's EMR information in the views presented below.

1. **Summary:**
    a. Presents all diagnoses, procedures, and observations collected during pregnancy-related encounters, including
        i. All encounters with an OB/GYN (see the "OB Visit" tab in Appendix D2: *EMR Concept List.xslx*); and
        ii. Non-OB/GYN encounters with pregnancy-related coded elements (see the "Pregnancy SNOMED" and "GAIA LOINC" tabs in Appendix D2: *EMR Concept List.xslx*)

2. **Encounters:**
    a. Includes all encounters recorded during the pregnancy episode

3. **Diagnoses:**
    a. Shows all EMR records for diagnoses recorded, including encounter date, the ICD-10-CM code and description, and the Systematized Nomenclature of Medicine (SNOMED) concept and description

4. **Procedures:**
    a. Presents all procedures recorded, including encounter date, the Current Procedural Terminology (CPT®) code and description, and the SNOMED concept and description

5. **Observations:**
    a. Presents all EMRs for observations and lab results recorded, including encounter date, the Logical Observation Identifiers Names and Codes (LOINC®) code and description, as well as the observation value and standard unit of measurement

6. **Admissions:**

a. Provides a high-level view of all hospitalization encounters. This is a subset of the encounters.

7. **Problem list:**

a. Provides a clinician-recorded list of diagnoses and conditions recorded

8. **Medical history:**

a. Provides a patient-reported list of diagnoses and conditions

9. **Surgical history:**

a. Provides the patient-reported history of surgeries

10. **Health maintenance:**

a. Provides an aggregated view of preventive care that the patient received

11. **Drug:**

a. Includes records of all medications administered in an inpatient setting and medications prescribed in an outpatient setting

12. **Immunization:**

a. Includes records of immunizations administered

13. **Habit:**

a. Includes a list of patient-reported lifestyle habits such as smoking, drinking, and physical activity

The summary view of the Chart Abstraction Tool is intended to highlight the potentially most relevant information for the adjudication. For each case, we will populate the summary view with identified EMR clinical concepts and criteria according to the operationalized GAIA definitions. (see the "Pregnancy SNOMED" and "GAIA LOINC" tabs in Appendix D2: *EMR Concept List.xslx*).

All the other views are to provide comprehensive EMR information for each case. Within the Chart Abstraction Tool, the clinician adjudicators will be able to select a patient to view her EMR chart as described above and proceed to search, filter, and sort EMR data as needed to complete the clinical adjudication.

c. Clinician Review and Adjudication

Each pregnancy cases selected for adjudication will be randomly assigned to a clinician reviewer in the Chart Abstraction Tool, with the relevant information presented in a standard patient EMR chart. The operationalized GAIA case definitions (as well as the EMR data element mapping for the definitions) will be available in a different view for the clinicians to reference while reviewing a patient's EMR chart. The clinician adjudicators may navigate between the summary view and other detailed views of the patient EMR data to determine the pregnancy outcome and GA at the time of the pregnancy outcome according to the GAIA case definitions.

After reviewing the relevant information for each case, the clinician reviewer will be presented with a questionnaire to provide the following information:

1. Confirmation of pregnancy
2. The date of the pregnancy outcome on the EMR record
3. The pregnancy start date based on the earliest, most accurate GA recorded on the patient's record[iii]

[iii] If a GA recorded meets GAIA definition Level 1, this is the most accurate measure on which to base the pregnancy start date. If not, use the next best GA record to determine pregnancy start date (i.e., the earliest GA record that meets the next highest GAIA-defined level for GA).

a. And corresponding GAIA case definition level for GA (note that selecting the appropriate subcriteria to meet the definitions is required)
4. Identification of the pregnancy outcome
a. And corresponding GAIA case definition level (note that selecting the appropriate subcriteria to meet the definitions is required)
5. Whether there are insufficient data to answer any of the questions above
6. Whether they recommend another reviewer to review the case

The completed adjudication questionnaires will be submitted to a central repository for later analysis.

i. Gestational Age Validation Procedure

GA is adjudicated by determining the start date of a pregnancy episode (estimated LMP) and the date of the pregnancy outcome.

1. If there is sufficient evidence to determine the pregnancy start date using Levels 1, 2A, 2B, or 3A of the GAIA GA case definition, reviewers will indicate in the Reviewer Adjudication window (see Figure 1) the level of the GAIA definition (1–3A) at which they have adjudicated the pregnancy start date, based on the most accurate GA recorded on the patient's EMR chart.

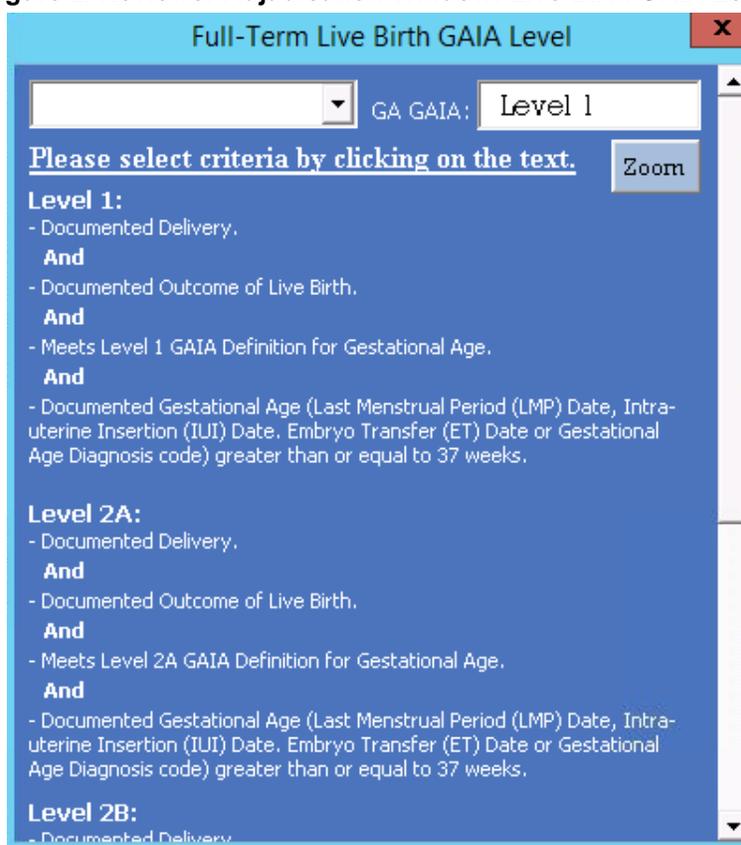**Figure 1. Reviewer Adjudication Window: Gestational Age GAIA Level**
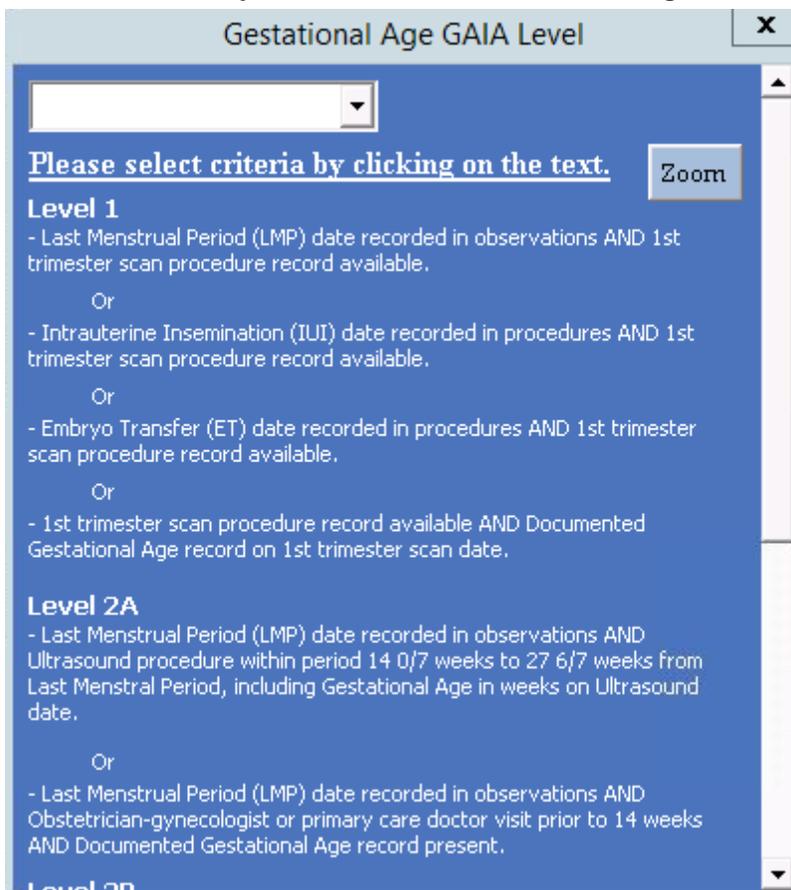
Abbreviation: Global Alignment of Immunization safety Assessment in pregnancy.

2. If there is insufficient evidence to determine the pregnancy start date, the GA for the pregnancy episode (and the outcome) cannot be satisfactorily adjudicated.

3. Reviewers may indicate whether they recommend another adjudicator to review the case for any reason.

4. Using the date of the pregnancy outcome recorded on EMR, the adjudicated GA at the time of the pregnancy outcome will then be calculated by the tool as (date of pregnancy outcome – start date of pregnancy episode + 1 day).

ii. Pregnancy Outcome Validation Procedure

1. If reviewers can determine the pregnancy outcome using evidence according to one of the levels defined in the GAIA case definition for the outcome, they will indicate the level of the GAIA definition (1–4) at which they believe the pregnancy outcome is adjudicated.

   a. Note that selecting the appropriate subcriteria to meet the definitions is required within the Reviewer Adjudication window (see Figure 2).

Figure 2. Reviewer Adjudication Window: Live Birth GAIA Level

Abbreviation: Global Alignment of Immunization safety Assessment in pregnancy.

2. If the reviewer finds insufficient evidence to adjudicate the pregnancy outcome at any level of the corresponding GAIA definition, the user will select "insufficient data" and describe the missing elements precluding an adjudication.

3. If the reviewer determines that the pregnancy outcome evaluated is not one of the four outcomes of interest (FTB, PTB, SA, or SB), the review will indicate the alternative

pregnancy outcome. In this case, the case should be flagged for a secondary reviewer to evaluate.

4. The reviewer can indicate whether they recommend another adjudicator to review the case for any reason, by checking the "Recommend for another review" box (see Figure 3). The reviewer can also enter notes at any time during the review by clicking the "Notepad" button on the Reviewer Adjudication window.

Figure 3. Reviewer Adjudication Window: Secondary Reviewer Request

d. Validation Analysis

i. Percentage Positive Agreement

In the sample selected for adjudication, we will calculate the percentage positive agreement (PPA), which is defined as the proportion of algorithm-determined outcomes or GA that are confirmed by EMR adjudication (Appendix D5: *Validation Analysis Table Shells.xslx*).

The cases not adjudicated due to the lack of EMR data are neither confirmed nor refuted and will be excluded from the calculation of the PPA.

General framework. For the sample of each outcome (n=50), the PPA is calculated by (1) summing the number of claims (test) cases with an algorithmically determined positive outcome that had a corresponding EMR-adjudicated positive outcome and (2) dividing the sum by the total number of positive test cases.

A statistical framework here is provided for conceptual illustration (see Table 4). Formally, each positive test case has three hypothetical outcomes from EMR adjudication: (1) positive, (2) negative, and (3) insufficient data. Each of these outcomes has a corresponding probability of occurrence. This data-generating process (whereby each test case contributes toward the sum of each type of outcome) can be formulated according to a multinomial distribution. Specifically, the three-outcome vector of counts $y = (y^+, y^-, y^\emptyset)$ is generated by a multinomial distribution with

probability vector $p = (p^+, p^-, p^\emptyset)$ and size n_{test}^+, which is the total number of test cases with a positive algorithmic outcome.

Table 4. Calculation of PPA for Pregnancy Outcomes

Claims Algorithm (Test)	EMR Adjudicated Cases (Reference Method)			*Row Total*
	Positive	**Negative**	**Insufficient Data**	
Positive	y^+	y^-	$y^\emptyset$	n_{test}^+
Negative	z^+	z^-	$z^\emptyset$	n_{test}^-
Column Total	n_{ref}^+	n_{ref}^-	$n_{ref}^\emptyset$	n

Abbreviation: CED, IBM MarketScan Explorys Claims-EMR Data Set.

Note that z^+, z^-, $z^\emptyset$, and n_{test}^- are included in Table 4 for conceptual purposes. As the algorithm is not finding negatives for the outcomes, these values are not to be calculated.

Table 5. Calculation of PPA for Gestational Age

Claims Algorithm (Test)	EMR Adjudicated Cases (Reference Method)			*Row Total*
	Adjudicated Gestational Age at Pregnancy Outcome (in days)		**Insufficient Data**	
	Positive = GA Agreement within ±7, ±14, ±21, or ±28 Days	**Negative = No GA Agreement within ±7, ±14, ±21, or ±28 Days**		
Estimated Gestational Age at Pregnancy Outcome (in days)	y^+	y^-	$y^\emptyset$	n_{test}
Column Total	n_{ref}^+	n_{ref}^-	$n_{ref}^\emptyset$	$n = n_{ref} = n_{test}$

Abbreviation: CED, IBM MarketScan Explorys Claims-EMR Data Set.

Note that for Gestational Age PPA calculation, for each agreement threshold (±7, ±14, ±21, and ±28 days):
- y^+ is calculated as the number of pregnancy episodes where $|Adjudicated\ GA - Estimated\ GA| \leq Agreement\ Threshold$.
- y^- is calculated as the number of pregnancy episodes where $|Adjudicated\ GA - Estimated\ GA| > Agreement\ Threshold$.
- $y^\emptyset$ is calculated as the number of pregnancy episodes where there is insufficient EMR data to adjudicate the gestational age.
- $\widehat{PPA}$ and $CI_{\widehat{PPA}}$ is calculated for each agreement threshold and stratified by each pregnancy outcome of interest (FTB, PTB, LB overall, SA, and SB).

PPA calculation. For conceptual clarity, as noted above, we exclude cases with insufficient data to calculate PPA. In other words, calculated in this way, the PPA characterizes the proportion of positive test cases that were positively determined by EMR adjudication and, importantly, renders unknown the true or false positive determination for cases with insufficient data. The PPA calculation directly excludes those cases from the denominator, leaving only true positive and false positive adjudications.

$$\widehat{PPA} = \frac{y^+}{n_{test}^+ - y^\emptyset} = \frac{y^+}{y^+ + y^-}$$

For gestational age and each pregnancy outcome of interest (FTB, PTB, LB overall, SA, and SB), the PPA estimate (and associated confidence intervals) will be calculated for 12 strata:

1. Cases sampled from the Tier 1 case pool
    a. PPA for cases adjudicated at GAIA outcome Level 1
    b. PPA for cases adjudicated at GAIA outcome Level 1 or 2
    c. PPA for cases adjudicated at GAIA outcome Level 1, 2, or 3
2. Cases sampled from the Tier 2 case pool
    d. PPA for cases adjudicated at GAIA outcome Level 1
    e. PPA for cases adjudicated at GAIA outcome Level 1 or 2
    f. PPA for cases adjudicated at GAIA outcome Level 1, 2, or 3
3. Cases sampled from the Tier 3 case pool
    g. PPA for cases adjudicated at GAIA outcome Level 1
    h. PPA for cases adjudicated at GAIA outcome Level 1 or 2
    i. PPA for cases adjudicated at GAIA outcome Level 1, 2, or 3
4. Cases sampled from all case pools (all tiers)
    j. PPA for cases adjudicated at GAIA outcome Level 1
    k. PPA for cases adjudicated at GAIA outcome Level 1 or 2
    l. PPA for cases adjudicated at GAIA outcome Level 1, 2, or 3

For each stratum, the denominator will be the sum of $y^+ + y^-$, representing the total number of cases within the specified validation sample tier with sufficient data to be adjudicated at the specified GAIA outcome level(s).

Extensions. By the general framework, the percentage of false positives and the percentage of cases with insufficient data can be calculated to estimate the multinomial probabilities p^- and $p^\emptyset$. Although the focus of the validation analysis is on the PPA, the other two hypothetical outcomes (in particular, $y^\emptyset$) provide information that can contribute to our understanding of the algorithm's success in identifying true positive outcomes (see Sensitivity Analysis section below).

ii. Confidence Intervals

To infer the PPA on the MarketScan target population, we will calculate confidence intervals. These confidence intervals capture statistical uncertainty induced by the validation sample selection.[iv] A set of confidence intervals for PPA will be calculated across all outcomes.

Citing literature about interval estimation for a binomial proportion,[8] we apply the recommended Agresti-Coull interval (a slight modification of the better-known Wilson interval) to address potential limitations in coverage probability for the confidence interval. Specifically, the Agresti-Coull interval is an adjustment to the standard Wald interval; at the nominal 0.05 level, the 95% interval is calculated as:

$$CI_{\widehat{PPA}} = \widehat{PPA} \pm 2\sqrt{\frac{\widehat{PPA}(1-\widehat{PPA})}{\tilde{n}}}$$

where $\widehat{PPA} = (y^+ + 2)/(y^+ + y^- + 4)$ and $\tilde{n} = y^+ + y^- + 4$. Colloquially, this is called the "add two successes and two failures" rule; by this rule, the method addresses the confidence interval's

---

[iv] Although the target population (i.e., enrollees in the MarketScan Database) can be construed as finite, the validation sample size is very small, so any finite population correction is negligible. Note that the target population for PPA calculation is actually a subpopulation—specifically, it is the pool of population cases with a positive outcome. Because the proportion of population cases with positive (or negative) outcomes is unknown, it must be estimated; the uncertainty from this estimation should be considered in standard error calculations for PPA. See Steinberg DM, Fine J, Chappell R. Sample size for positive and negative predictive value in diagnostic research using case–control designs. Biostatistics 2009;10:94-105.

potential lack of coverage at the nominal level (i.e., 0.05) for certain combinations of PPA and the binomial sample size $y^+ + y^-$.

iii. Sensitivity Analysis

The number of cases that presented insufficient data in the EMR, $y^\emptyset$, essentially represents those that could be neither positively nor negatively adjudicated. In one hypothetical scenario, all such cases could, in fact, have been positively adjudicated cases, thereby contributing to the count y^+; this scenario produces counts for an overestimate of PPA ($\widehat{PPA}^{over}$). On the other extreme, these cases could have been all negatively adjudicated and contribute to the count y^-, thereby producing an underestimate of PPA ($\widehat{PPA}^{under}$).

$$\widehat{PPA}^{over} = \frac{y^+ + y^\emptyset}{y^+ + \ y^- + \ y^\emptyset}$$
$$\widehat{PPA}^{under} = \frac{y^+}{y^+ + \ y^- + \ y^\emptyset}$$

These measures can be compared to assess the relative performance of the algorithm to identify true positive cases under the two assumptions about insufficient data: (1) whether all the positive test cases would have been positively determined by EMR adjudication and (2) whether they would have been negatively determined by EMR adjudication. In this way, the validation sample is preserved (i.e., by retaining all initially selected validation cases) while allowing for statistical inference on the PPA. For each stratum's PPA point estimate (detailed above), we will be presenting the $\widehat{PPA}^{over}$ and $\widehat{PPA}^{under}$ measures as a sensitivity analysis.

iv. Interreviewer Concordance Analysis

As a diagnostic measure to gauge the reliability of adjudication results between reviewers in this study, a concordance analysis will be conducted. For representativeness of each stratum, the proposed methodology is as follows:

1. From the cases with sufficient EMR for adjudications, randomly sample 20% (or 10 cases minimum) per outcome, per tier for FTB, PTB, and SA; select all SBs
2. Assign the cases to a reviewer different from the original reviewer.
3. Conduct the case adjudications following the same clinician adjudication process outlined in Section 3.c.c above for the selected cases.

Table 6. Interreviewer concordance sampling methodology

Outcome	Total		Tier 1-Relevant EMR Data		Tier 2-Relevant EMR Data	
	Adjudicated	*Sample for Rereview*	Adjudicated	*Sample for Rereview*	Adjudicated	*Sample for Rereview*
	No.	*No.*	No.	*No.*	No.	*No.*
LB	**185**	*40*	**90**	*20*	**95**	*20*
FTB	**92**	*20*	**46**	*10*	**46**	*10*
PTB	**93**	*20*	**44**	*10*	**49**	*10*
SA	**75**	*20*	**48**	*10*	**27**	*10*
SB	**24**	*24*	**5**	*5*	**9**	*9 (+10 SBs in Tier 3)*

Abbreviations: EMR, electronic medical record; FTB, full-term live birth; GA, gestational age; LB, live birth; PTB, preterm live birth; SA, spontaneous abortion; SB, stillbirth.

The results of the secondary adjudications will be compared against the results of the primary adjudication and then aggregated to calculate percent concordance for the outcome and percent concordance of the GA within ±7, ±14, ±21, and ±28 days.

3.d Aim 3: Exploratory Aims

a. Concordance Analysis Using MarketScan Data

According to the GAIA definition, fertility treatments such as IUI and ET are Level 1 evidence in determining GA. However, these procedures are not widely used. In 2015, the number of assisted reproductive technology procedures performed nationally per 1 million women aged 15–44 years was 2,832.[9] In CED data, only 1,200 enrollees received IUI or ET procedures after October 1, 2015. In the larger population of MarketScan data, approximately 38,000 enrollees received IUI or ET procedures after October 1, 2015, which resulted in more than 6,000 pregnancies. Additionally, maternal records can be linked to infant records in MarketScan data. For pregnancies initiated with IUI or ET, GA is considered known. With linked infant records, the status and gestational age of the infants also can be compared with the GA estimate for the mother. In a study evaluating the timing of first trimester in-utero exposure to letrozole, Bird et al.[10] selected the fertility procedure date as the point of reference when comparing the misclassification rates based on pregnancy start dates from different methods.

If deemed appropriate for a standalone follow-on study, in Aim 3a, the pregnancy outcomes and GA algorithm developed in Aim 1 will be applied to the large population in the MarketScan Commercial Database; thus, the obtained cohort will be further restricted to those pregnancies resulting from IUI/ET procedures by selecting the closest procedure to the pregnancy outcomes. Algorithm-based outcome classification and GA will be compared against outcomes based on infant records and GA based on the timing of IUI/ET procedures.

b. Predictive modeling

For a subset of pregnancy outcomes shown to have a poor PPA in Aim 2 (e.g., PPA of <80%), predictive models can be developed, through which a probability threshold can be determined to achieve an acceptable PPA. Depending on the scope and timeline of the current validation study, Aim 3b may be best suited as a standalone follow-on study.

- The dependent variable of the predictive model will be the outcome classification based on the gold standard determined from reviewing patient EMR charts. Potential predictors will include elements from claims data based on a literature review, consultation with clinical experts, and additional utilization measures. These may include—
  - Clinical codes indicating PTBs, FTBs, SBs, and SAs
  - Prenatal visits
  - Prenatal services and tests
  - Assisted reproductive services
  - Other complications and characteristics of pregnancy (e.g., gestational diabetes, severe maternal morbidities, delivery method)
  - Demographic information
  - Clinical history
  - Indicators of general health (e.g. Deyo-Charlson Comorbidity Index)
  - Medication use
  - Healthcare resource use
- A least absolute subset of predictors will be selected through logistic regression analysis while accounting for potentially correlated predictors.
- Using the values of the predictors in the model, patients' probability of being a true case will be estimated. A probability cutoff will be chosen to achieve a desirable PPA and sensitivity.

References

[1] Hornbrook MC, Whitlock EP, Berg CJ, et al. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. Health Serv Res 2007;42:908-27.

[2] Naleway AL, Golda R, Kuroskya S, et al. Identifying pregnancy episodes, outcomes, and mother–infant pairs in the Vaccine Safety Datalink. Vaccine 2013;31:2898-903.

[3] 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:e0192033.

[4] Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernández-Díaz S. Algorithms to estimate beginning of pregnancy in administrative databases. Parmacoepidemiol Drug Saf 2013;22:16-24.

[5] Quinn JA, Munoz FM, Gonik B, et al. Preterm birth: case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine 2016;34:6047-56.

[6] Committee on Obstetric Practice, the American Institute of Ultrasound in Medicine, and the Society for Maternal-Fetal Medicine. Committee Opinion No 700: Methods for Estimating Due Date. Obstet Gynecol 2017;129:e150-4.

[7] MacDorman MF, Gregory ECW. Fetal and Perinatal Mortality: United States, 2013. National Vital Staistics Reports. Volume 64, Number 8. Centers for Diseases Control and Prevention; July 23, 2015. (Accessed October, 2019, at https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_08.pdf.)

[8] Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Stat Sci 2001;16(2):101-33.

[9] Sunderam S, Kissin DM, et al. Assisted Reproductive Technology Surveillance — United States, 2015. Surveillance Summaries 2018 / 67(3);1–28

[10] Bird ST, Toh S, Sahin 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; 188:418-25.

D Appendix

D1 Pregnancy Episodes Clinical Codes

D2 EMR Concept List

D3 Characteristics Comparison Table Shells

D4 Operationalized GAIA Case Definitions

D5 Validation Analysis Table Shells