



# Development and Validation of ICD-10-CM-based Algorithms for Date of Last Menstrual Period, Pregnancy Outcomes, and Infant Outcomes

Andrea K. Chomistek<sup>1</sup> · Kelesitse Phiri<sup>1</sup> · Michael C. Doherty<sup>1</sup> · Jenna F. Calderbank<sup>1</sup> · Stephanie E. Chiuve<sup>2</sup> · Brenda Hinman McIlroy<sup>2</sup> · Michael C. Snabes<sup>2</sup> · Cheryl Enger<sup>1</sup> · John D. Seeger<sup>1</sup>

Accepted: 22 November 2022 / Published online: 19 January 2023  
© The Author(s) 2023, corrected publication 2023

## Abstract

**Introduction and Objective** Validation studies of algorithms for pregnancy outcomes based on International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) codes are important for conducting drug safety research using administrative claims databases. To facilitate the conduct of pregnancy safety studies, this exploratory study aimed to develop and validate ICD-10-CM-based claims algorithms for date of last menstrual period (LMP) and pregnancy outcomes using medical records.

**Methods** Using a mother–infant–linked claims database, the study included women with a pregnancy between 2016–2017 and their infants. Claims-based algorithms for LMP date utilized codes for gestational age (Z3A codes). The primary outcomes were major congenital malformations (MCMs) and spontaneous abortion; additional secondary outcomes were also evaluated. Each pregnancy outcome was identified using a claims-based simple algorithm, defined as presence of  $\geq 1$  claim for the outcome. Positive predictive values (PPV) and 95% confidence intervals (CI) were calculated.

**Results** Overall, 586 medical records were sought and 365 (62.3%) were adjudicated, including 125 records each for MCMs and spontaneous abortion. Last menstrual period date was validated among maternal charts procured for pregnancy outcomes and fewer charts were adjudicated for the secondary outcomes. The median difference in days between LMP date based on Z3A codes and adjudicated LMP date was 4.0 (interquartile range: 2.0–10.0). The PPV of the simple algorithm for spontaneous abortion was 84.7% (95% CI 78.3, 91.2). The PPV for the MCM algorithm was  $< 70\%$ . The algorithms for the secondary outcomes pre-eclampsia, premature delivery, and low birthweight performed well, with PPVs  $> 70\%$ .

**Conclusions** The ICD-10-CM claims-based algorithm for spontaneous abortion performed well and may be used in pregnancy studies. Further algorithm refinement for MCMs is needed. The algorithms for LMP date and the secondary outcomes would benefit from additional validation in a larger sample.

## Key Points

The claims-based algorithm for last menstrual period that used ICD-10-CM Z3A codes accurately estimated the documented date of last menstrual period.

The ICD-10-CM claims-based algorithms for spontaneous abortion, pre-eclampsia, premature delivery, and low birthweight performed well, with positive predictive values exceeding 70%.

Algorithms for major congenital malformations, placenta previa, and small for gestational age did not perform well and require further refinement.

✉ Andrea K. Chomistek  
andrea.chomistek@optum.com

<sup>1</sup> Optum, 1325 Boylston Street, 11th Floor, Boston, MA 02215, USA

<sup>2</sup> AbbVie, Inc, North Chicago, IL, USA

## 1 Introduction

Administrative healthcare databases are increasingly used to evaluate medication safety during pregnancy [1–4]. These databases include claims submitted by healthcare providers for payment and records of patient encounters within healthcare systems, including pharmacy dispensing, inpatient and outpatient diagnoses, and procedures. As these databases are created primarily for administrative and billing purposes, rather than research, the validation of exposure and outcome variables defined by codes on claims against a gold standard, such as medical records, is essential to ensure the validity of research studies conducted using these data [5–8].

Algorithms to define pregnancy and infant outcomes based on claims using International Classification of Diseases, 9th Revision Clinical Modification (ICD-9-CM) codes have been developed and validated, including algorithms for pre-eclampsia [1, 7, 9], preterm birth [5], small for gestational age (SGA) [8, 9], and major congenital malformations (MCMs) [6, 10, 11]. However, previous validation studies of algorithms based on International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) codes have been limited to the outcomes of spontaneous abortion [12], preterm birth [12], stillbirth [12, 13], and a subset of MCMs [14]. Furthermore, the use of codes for gestational age (Z3A codes) introduced with ICD-10-CM may improve estimation of pregnancy start date, as prior algorithms have been based on the date of the observed pregnancy outcome [15–17].

The goal of this exploratory project was to develop and validate ICD-10-CM claims-based algorithms for key variables needed to conduct post-marketing pregnancy safety studies in claims databases. These variables included the estimated date of last menstrual period (LMP), which is necessary for establishing a pregnancy timeline, and multiple pregnancy outcomes that are used as key primary and secondary endpoints. The primary endpoints were MCMs and spontaneous abortion while the secondary endpoints consisted of placenta previa, pre-eclampsia, premature delivery, low birthweight, and SGA. The claims-based algorithms included a simple algorithm, defined as the presence of at least one claim for the outcome, and additional candidate algorithms based on patterns of services received with the goal of identifying a best-performing algorithm for each outcome.

## 2 Methods

### 2.1 Data Source

This study used data from the Optum Research Database (ORD), a claims database from a large US health insurer.

As early as 1993, medical and pharmacy claims data are available for 70 million individuals with both medical and pharmacy benefit coverage. The study population was identified using Optum's Dynamic Assessment of Pregnancies and Infants (DAPI), a process that includes a set of definitions and algorithms that are applied to claims data to identify pregnancies, outcomes, and link data from mothers and infants within the ORD [4, 18]. Due to the size of the ORD, there are approximately 200,000 new pregnancies identified each year within the database. All pregnancies are linked to infant(s) using a linkage algorithm that utilizes the infant's date of birth, the estimated delivery date, and a family member ID. Of pregnancies that result in live births, approximately 85% of mothers can be linked to an infant [19].

### 2.2 Study Population

Women aged 18–55 years with an estimated LMP date (i.e., pregnancy start date) and pregnancy end date between 01 January 2016 and 31 December 2017 were identified. This time period was chosen because this validation study was conducted as background for surveillance that began in 2018. The population was limited to women who had continuous medical and pharmacy benefit coverage for a minimum of 6 months prior to their estimated LMP date (i.e., the baseline period) through to the end of pregnancy. Within this study population, the infant study population was identified among pregnancies for which the mother and infant data could be linked.

The ORD contains data from health plans that contract for “administrative services only”; access to medical records was not allowed for patients enrolled in these health plans. As this study required medical record review, women and infants who were enrolled in “administrative services only” plans were excluded from the study population, and the study outcomes were identified among those remaining (Fig. 1a, b).

Each pregnancy was followed from the day after the estimated LMP date through to the first of the following: 60 days after the date of end of pregnancy, disenrollment from the health plan, or end of the study period. Infants who were linked to their mothers were followed from the estimated date of delivery through to the first of the following: disenrollment from the health plan, or end of the study period.

### 2.3 Protection of Human Subjects

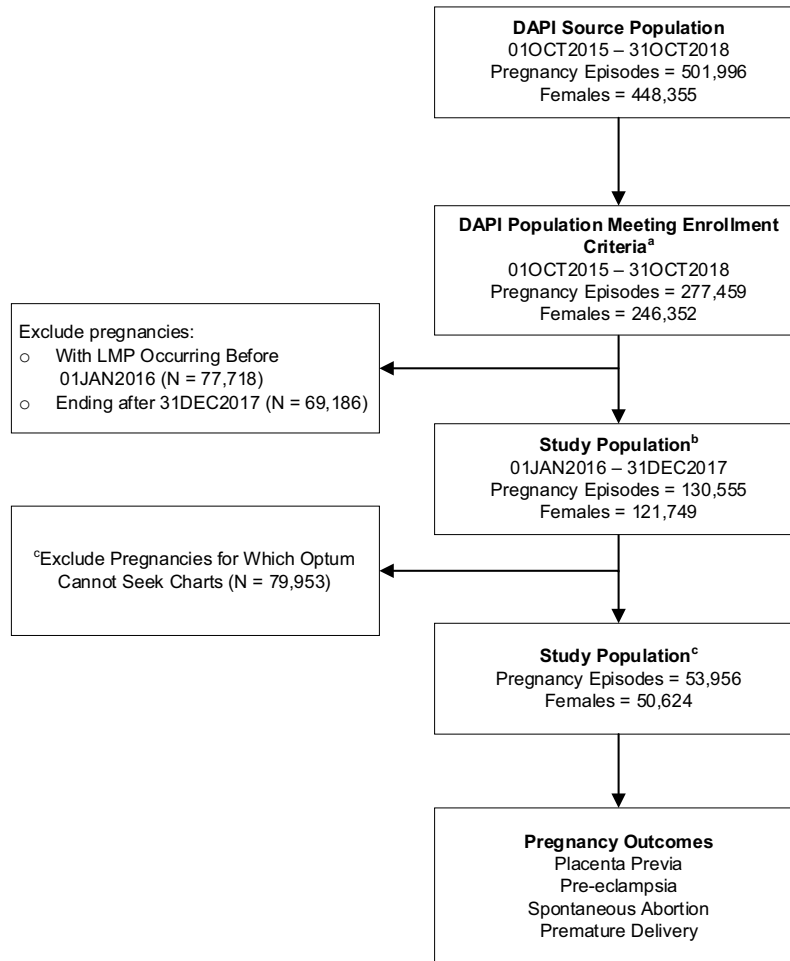
The study protocol was approved by the New England Institutional Review Board and all data access conformed to applicable Health Insurance Portability and Accountability Act policies.

## 2.4 Estimation of LMP Date

The algorithm to estimate LMP date utilized all available codes indicating weeks of gestation (Z3A.00 to Z3A.42, excluding Z3A.49). For each woman, the number of Z3A codes varied based on natural variability in number and timing of clinical visits. First, for each woman and each observed Z3A code, the LMP date was estimated by subtracting the weeks of gestation based on the Z3A code from

the date of service in the claim (e.g., if a Z3A.10 code [10 weeks gestation of pregnancy] was observed on 10 July 2019, 10 weeks was subtracted from the date, resulting in a LMP date of 01 May 2019). These LMP date estimations were repeated for each available Z3A code recorded on a claim during pregnancy, resulting in multiple estimated LMP dates for each woman, which were sequentially sorted. To identify pregnancy episodes, LMP clusters were created by grouping all LMP dates within 6 weeks of each other (from

(a) Cohort Creation Flow Diagram for Women/Pregnancies



**Fig. 1 a** Cohort creation flow diagram for women/pregnancies. *DAPI* Dynamic Assessment of Pregnancies and Infants, *LMP* last menstrual period. <sup>a</sup>Have continuous medical and pharmacy benefit coverage for a minimum of 6 months (182 days) prior to and including the estimated LMP through the end of pregnancy. <sup>b</sup>Earliest LMP occurring on or after 01JAN2016; end of pregnancy ending by 31DEC2017. <sup>c</sup>This step determines women for whom Optum can seek medical charts for the pregnancy and outcome, assessed at the pregnancy episode level. Pregnancies among women enrolled in administrative services only plans were excluded because access to medical records was not allowed for patients in these plans. The final study population consisted only of women with pregnancies for whom Optum could seek medical charts. **b** Cohort creation flow diagram for infants. <sup>a</sup>See

Figure 1a for details of the study population creation. <sup>b</sup>This is “multi-gestation” pregnancies that have livebirth(s) and stillbirth(s) (e.g. twins, one liveborn and one stillborn; quadruplets, some liveborn). <sup>c</sup>Includes stillbirths, ectopic, molar, and abortions (“spontaneous,” “elective,” and “other” per DAPI definitions). <sup>d</sup>Linked pregnancies are pregnancies for which the mother and infant data could be linked. Forty-one linked infants were from pregnancies ending in a non-live-birth, possibly representing misclassification of how these pregnancies ended. <sup>e</sup>This step determines infants for whom Optum can seek medical charts. Infants enrolled in administrative services only plans were excluded because access to medical records was not allowed for patients in these plans. The final study population consisted only of infants for whom Optum could seek medical charts

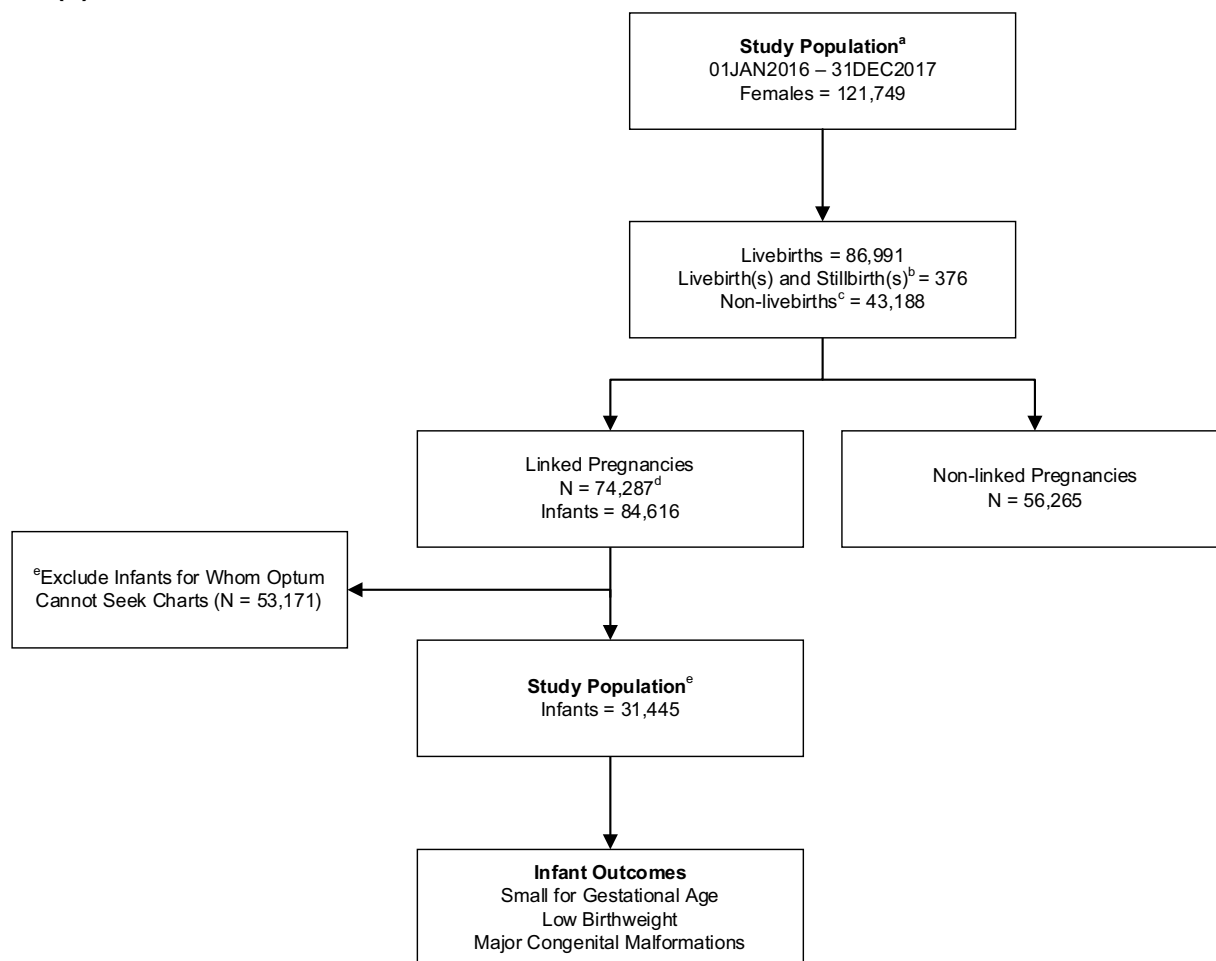
**(b) Cohort Creation Flow Diagram for Infants**

Fig. 1 (continued)

the earliest estimated LMP forward using up to a 6-week window, which was chosen based on previous publications [15, 17]). The LMP date for each pregnancy episode was estimated using two methods: (1) LMP date from the first observed Z3A code (i.e., earliest service date with a Z3A code) within the pregnancy episode, and (2) median LMP date based on all Z3A codes within the pregnancy episode.

For pregnancies where Z3A codes were not observed, algorithms informed by published literature and refined following obstetrician-gynecologist input were utilized to estimate the corresponding LMP [15–17, 20]. For these pregnancies, the estimated LMP was calculated based on algorithms that assume different lengths of gestation for full term singleton births (39 weeks), multiple births (36 weeks), stillbirths (28 weeks), abortions (10 weeks), trophoblastic diseases (8 weeks), and ectopic pregnancies (8 weeks).

## 2.5 Identification of Outcomes

Pregnancies and infants were classified according to the presence or absence of an outcome of interest. Outcome groups were not mutually exclusive. To maximize sensitivity, each study outcome was identified using a simple algorithm, defined as the presence of at least one claim for the outcome during follow-up, based on ICD-10-CM diagnostic codes and Current Procedural Terminology (CPT) codes<sup>1</sup> (Table 1, Supplemental Table 1). Spontaneous abortion, placenta previa, pre-eclampsia, and premature delivery were

<sup>1</sup> CPT © Copyright 2021 American Medical Association. All rights reserved. Fee schedules, relative value units, conversion factors and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein. CPT is a registered trademark of the American Medical Association.

**Table 1** International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) Diagnosis and CPT<sup>®a</sup> Codes to Estimate LMP Date and Identify Pregnancy and Infant Outcomes

<i>LMP date</i>			
<b>Code</b>	<b>Code description</b>		<b>Code type</b>
Z3A.**	Weeks of gestation		ICD-10-CM
<i>Pregnancy outcomes</i>			
<b>Outcome</b>	<b>Code</b>	<b>Code description</b>	<b>Code type</b>
Spontaneous abortion	O02.1	Missed abortion	ICD-10-CM
	O03.**	Spontaneous abortion	ICD-10-CM
	01965	Anesthesia for incomplete or missed abortion procedures	CPT <sup>®</sup> codes
	59800–59811	Treatment of spontaneous abortion	CPT <sup>®</sup> codes
Placenta previa	O44.**	Placenta previa	ICD-10-CM
Pre-eclampsia	O14.**	Pre-eclampsia	ICD-10-CM
	O11.*	Pre-existing hypertension with pre-eclampsia	ICD-10-CM
Premature delivery	O60.10**	Preterm labor with preterm delivery	ICD-10-CM
	O60.12**	Preterm labor second trimester with preterm delivery second trimester	ICD-10-CM
	O60.13**	Preterm labor second trimester with preterm delivery third trimester	ICD-10-CM
	O60.14**	Preterm labor third trimester with preterm delivery third trimester	ICD-10-CM
	P07.2*	Extreme immaturity of newborn	ICD-10-CM
	P07.3*	Preterm (premature) newborn (other)	ICD-10-CM
<i>Infant outcomes</i>			
<b>Outcome</b>	<b>Code</b>	<b>Code description</b>	<b>Code type</b>
Major congenital malformations <sup>b</sup>	Q00–Q07	Congenital malformations of the nervous system	ICD-10-CM
	Q10–Q18	Congenital malformations of eye, ear, face, and neck	ICD-10-CM
	Q20–Q28	Congenital malformations of the circulatory system	ICD-10-CM
	Q30–Q34	Congenital malformations of the respiratory system	ICD-10-CM
	Q35–Q37	Cleft lip and cleft palate	ICD-10-CM
	Q38–Q45	Other congenital malformations of the digestive system	ICD-10-CM
	Q50–Q56	Congenital malformations of genital organs	ICD-10-CM
	Q60–Q64	Congenital malformations of the urinary system	ICD-10-CM
	Q65–Q79	Congenital malformations and deformations of the musculoskeletal system	ICD-10-CM
	Q80–Q89	Other congenital malformations	ICD-10-CM
Low birthweight	P07.0*	Extremely low birth weight newborn	ICD-10-CM
	P07.1*	Other low birth weight newborn	ICD-10-CM
Small for gestational age	P05.**	Disorders of newborn related to slow fetal growth and fetal malnutrition	ICD-10-CM

<sup>a</sup>CPT<sup>®</sup> Copyright 2021 American Medical Association. All rights reserved

<sup>b</sup>The Metropolitan Atlanta Congenital Defects Program (MACDP) [22], European Surveillance of Congenital Anomalies (EUROCAT) [23] and New York State Department of Health Congenital Malformations Registry [24] references were used to create a list of minor anomalies to exclude from the code category Q00–Q89 in order to identify major congenital malformations

identified at any point during the pregnancy from pregnancy (maternal) claims; MCMs, low birthweight, and SGA were identified following delivery from infant claims.

## 2.6 Medical Record Procurement and Adjudication

From the outcomes identified using the simple algorithm, a subset of 750 charts based on an a priori number for each outcome (300 for MCMs, 200 for spontaneous abortion,

and 50 for each of the remaining 5 secondary outcomes) was selected randomly for medical record procurement and adjudication. Among the 300 potential cases of MCMs randomly selected for the validation sample, 92 cases were subsequently removed because they only had diagnosis codes for minor congenital malformations (Supplemental Fig. 1). As MCMs are typically the outcome of interest in pregnancy safety studies, minor malformations were excluded from the algorithm. Following this exclusion, 658 randomly selected outcomes were included in the validation sample.

Medical records for women were reviewed for LMP date, spontaneous abortion, placenta previa, pre-eclampsia, and premature delivery; medical records for infants were reviewed for MCMs, low birthweight, and SGA.

The date of LMP was adjudicated by an epidemiologist. Date of LMP, and gestational age (along with service date) when present, was recorded from each medical record. The final date of LMP was adjudicated based on all information in the record, informed by recommendations from The American College of Obstetrics and Gynecology [21].

Two geneticists with expertise in teratology (for MCMs) and 2 obstetrician-gynecologists (for all other pregnancy and infant outcomes) reviewed medical records for each potential case. The presence or absence of the diagnosis in the record was independently adjudicated by the 2 clinicians. Consensus was sought in the case of discrepant results. The clinical adjudicators were instructed to use their own clinical judgment and the guidance provided in the Appendix (Supplemental Material). For each outcome, a patient was classified as follows: (1) a definite case if the medical records included dated documentation that met the criteria for an outcome; (2) a probable case if all the criteria were not met, but sufficient information was present; (3) a non-case if the information in the record did not indicate presence of the outcome; and (4) insufficient information if the record did not contain the specific reports and notes needed to make an adjudication decision.

## 2.7 Statistical Analysis

Baseline characteristics of the source population overall and for the validation sample were examined. The frequency and percentage were calculated for categorical variables and the mean and standard deviation were calculated for continuous variables.

To assess the validity of estimated LMP date, the number of days between the claims-based estimates and the adjudicated LMP date was calculated by subtracting the adjudicated LMP from the estimated LMP. The claims-based estimates included:

1. LMP date from the first observed Z3A code within the pregnancy episode;
2. Median LMP date based on all Z3A codes within the pregnancy episode;
3. LMP date estimated using literature-based algorithms (for the subset of pregnancies where Z3A codes were not observed).

For each outcome, candidate algorithms that had been specified a priori, were used and performance metrics developed in order to identify the best-performing claims-based algorithm. Additional algorithms were developed by

reviewing claims profiles from the subset of chart-confirmed outcomes to identify patterns of claims associated with outcome confirmation, including type of service, clinician specialty, and temporality of claims [11]. The candidate algorithms that were evaluated included the following:

- Algorithm 1: at least 1 claim (simple algorithm);
- Algorithm 2: at least 2 claims on separate days;
- Algorithm 3: 2 claims separated by a specific number of days (e.g., 7, 14, 28 days);
- Algorithm 4: at least 1 claim from a specific provider specialty(ies) (e.g. hospital, obstetrics and gynecology);
- Algorithm 5: at least 1 claim from a specific site(s) of care (e.g. inpatient, outpatient visit, professional visit);
- Algorithm 6: additional algorithm for placenta previa based on timing of claim relative to delivery date.

The positive predictive value (PPV) and corresponding 95% confidence interval (CI) was calculated for the simple algorithm and each of the candidate algorithms for each outcome. The PPV was calculated as the sum of definite and probable cases divided by the number of potential cases reviewed, after excluding those whose charts lacked sufficient information to determine case status. The best-performing algorithm for each outcome was selected based on the PPV and the number of definite or probable cases identified by the algorithm, as a proxy for sensitivity. Sensitivity could not be determined since charts were only sought for claims-identified cases; potential cases without a claims-based diagnosis were not sought.

All analyses were conducted using SAS software, version 9.4 (SAS Institute Inc, Cary, NC).

## 3 Results

### 3.1 Study Population

Details of the cohort formation can be found in Fig. 1a, b. There were 53,956 pregnancy episodes among 50,624 women and 31,445 linked infants in the final study population. Descriptive characteristics of the study population are provided in Supplemental Table 2.

From the 53,956 pregnancy episodes, we identified 10,182 (18.9%) spontaneous abortion, 908 (1.7%) placenta previa, 2028 (3.8%) pre-eclampsia, and 1742 (3.2%) premature delivery outcomes using the simple algorithm (Table 2). Among 31,445 infants, 2600 (8.3%) had at least one MCM identified, 1711 (5.4%) were low birthweight, and 1273 (4.1%) were SGA based on the simple algorithm (Table 2).

Among the 658 randomly selected outcomes included in the validation sample, 72 cases could not be sent for chart procurement due to a priori provider refusal (i.e., the



**Table 2** Positive predictive values of claims-based simple algorithms (Algorithm 1) for pregnancy and infant outcomes based on adjudicated medical records

	Claims-identified outcomes	Randomly selected outcomes for the validation sample	Medical records sought <sup>a</sup>	Medical records procured	Medical records adjudicated <sup>b</sup>	Outcomes confirmed by adjudicator	PPV (%)	95% CI
<b>Outcomes</b>								
<i>Pregnancy outcomes</i>								
Spontaneous abortion	10,182	200	179	125	118	100	84.7	78.3–91.2
Placenta previa	908	50	47	26	23	3	13.0	0.0–26.8
Pre-eclampsia	2,028	50	48	34	23	18	78.3	61.4–95.1
Premature delivery	1742	50	44	30	26	24	92.3	82.1–100.0
<i>Infant outcomes</i>								
Major congenital malformations <sup>c</sup>	2600	208	174	130	125	55	44.0	35.3–52.7
Low birthweight	1,711	50	48	27	27	26	96.3	89.2–100.0
Small for gestational age	1,273	50	46	26	23	8	34.8	15.3–54.2
<b>Total</b>	<b>20,444</b>	<b>658</b>	<b>586</b>	<b>398</b>	<b>365</b>	<b>234</b>		

CI confidence interval, PPV positive predictive value

<sup>a</sup>Some cases identified by Optum could not be sent for chart procurement because the patients' providers have asked to not be contacted; hence, the decrease in the number of medical records sought for each outcome from the validation sample

<sup>b</sup>Records with the study outcome of interest for which medical records with sufficient information for adjudication were obtained

<sup>c</sup>300 potential cases were randomly selected for the validation sample; 92 cases were subsequently removed because they only had diagnosis codes for minor congenital malformations (see Supplemental Fig. 1 for details)

patients' providers were on the 'do not contact' list). Consequently, medical records were sought for 586 cases, of which 398 (67.9%) were procured and 365 (62.3%) were adjudicated (33 charts with insufficient information were excluded) (Table 2). Descriptive characteristics were similar for women whose charts could be adjudicated and women whose charts could not be adjudicated (Supplemental Table 3).

## 3.2 Validation of Claims-based Algorithms

### 3.2.1 Last Menstrual Period

Among the 215 records procured for the validation of pregnancy outcomes, 10 were excluded from the analysis for LMP because, within the claims data, the pregnancy episode overlapped with another episode within the same woman. As such, 205 records were used for LMP validation. Of these, 157 pregnancy episodes had at least one Z3A code. Table 3 compares the estimated median LMP date based on all Z3A codes within the pregnancy episode to the adjudicated LMP date. The median absolute difference in days was 4.0 (IQR: 2.0–10.0) overall and the median LMP date was  $\pm 7$  days from the adjudicated LMP date among 65.0% of pregnancies. According to pregnancy outcome, median LMP date was  $\pm 7$  days from adjudicated LMP date among 34.3% of pregnancies with spontaneous abortion, 89.7%

with premature delivery, 95.7% with placenta previa, and 90.6% with pre-eclampsia. The estimated median LMP date was later than the adjudicated LMP date for 126 pregnancies (80.3%).

Results for estimated LMP based on the first observed Z3A code were similar (Supplemental Table 4a). The 48 pregnancies for which a Z3A code was not observed were primarily spontaneous abortion (95.8%), and the difference in days between estimated and adjudicated LMP was 16.0 (IQR: 8.0–25.0) (Supplemental Table 4b).

### 3.2.2 Pregnancy Outcomes

For the pregnancy outcomes, 318 medical records were sought and 215 (67.6%) records were procured: 125 (69.8%) spontaneous abortion, 26 (55.3%) placenta previa, 34 (70.8%) pre-eclampsia, and 30 (68.1%) premature delivery (Table 2).

Among the 125 medical records reviewed for spontaneous abortion, 100 (80.0%) were adjudicated as definite or probable cases (Table 4). The PPV for the simple algorithm (Algorithm 1) was 84.7% (95% CI 78.3, 91.2). The additional candidate algorithms also performed well; the highest PPV was observed for Algorithm 3b which required 2 claims separated by at least 14 days (92.6%, 95% CI 82.7, 100).

The PPVs for the simple algorithm (Algorithm 1) for each of the secondary pregnancy outcomes were: 13.0%

**Table 3** Number of days between estimated median date of LMP based on all Z3A codes in DAPI and adjudicated LMP, overall and according to pregnancy outcome

	Overall ( <i>N</i> = 157)		Spontaneous abortion ( <i>N</i> = 73)		Premature delivery ( <i>N</i> = 29)		Placenta previa ( <i>N</i> = 23)		Pre-eclampsia ( <i>N</i> = 32)	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Difference in days*	4.0	2.0–10.0	10.0	4.5–26.0	2.5	1.8–4.0	3.0	2.0–4.0	2.5	1.5–4.5
<b>Stratified by age at LMP (years)</b>										
18–24	5.3	3.0–7.0	10.0	7.0–15.0	2.3	1.3–5.0	4.5	3.5–5.3	5.5	5.5–5.5
25–34	4.0	2.0–7.0	7.5	3.0–28.5	2.5	2.0–3.5	3.0	2.0–4.0	4.0	1.5–5.0
35–44	4.0	2.0–14.8	12.5	5.0–23.0	3.0	0.0–5.0	1.5	0.5–3.0	2.0	1.0–2.5
≥ 45	3.0	3.0–3.0	N/A	N/A	N/A	N/A	3.0	3.0–3.0	N/A	N/A
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<b>Days between estimated and adjudicated LMP</b>										
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
–29 or more days	2	1.3	2	2.7	0	0.0	0	0.0	0	0.0
–15 to –28	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
–8 to –14	1	0.6	1	1.4	0	0.0	0	0.0	0	0.0
–4 to –7	4	2.6	2	2.7	0	0.0	1	4.4	1	3.1
–1 to –3	1	1	1	1	0	0.0	0	0.0	0	0.0
0	9	6	4	5	2	6.9	1	4.4	2	6.3
+1 to +3	54	34.4	9	12.3	18	62.1	12	52.2	15	46.9
+4 to +7	34	21.7	9	12.3	6	20.7	8	34.8	11	34.4
+8 to +14	15	9.6	10	13.7	2	6.9	1	4.4	2	6.3
+15 to +28	13	8.3	13	17.8	0	0.0	0	0.0	0	0.0
+29 or more days	10	6.4	10	13.7	0	0.0	0	0.0	0	0.0
Insufficient information on LMP in chart	14	8.9	12	16.4	1	3.5	0	0.0	1	3.1

DAPI Dynamic Assessment of Pregnancies and Infants, IQR interquartile range, LMP last menstrual period, N/A not applicable

\*Based on absolute value of the number of days between estimated LMP and adjudicated LMP

(95% CI 0.0, 26.8) for placenta previa, 78.3% (95% CI 61.4, 95.1) for pre-eclampsia, and 92.3% (95% CI 82.1, 100.0) for premature delivery (Table 2). Positive predictive value estimates for all candidate claims-based algorithms developed for the secondary pregnancy outcomes are provided in Supplemental Table 5a–c. For placenta previa, the additional candidate algorithms also had low PPVs (best-performing PPV: 33.3%) (Supplemental Table 5a). Among 26 records for placenta previa, 15 had claims for complete placenta previa (6 with hemorrhage), one had a claim for partial placenta previa, and 10 had claims for low-lying placenta; all 3 confirmed cases had a claim for complete placenta previa with hemorrhage (Supplemental Table 6). For pre-eclampsia and premature delivery, the additional candidate algorithms performed as well or better than the simple algorithm, with some PPVs reaching 100%. (Supplemental Table 5b, c).

### 3.2.3 Infant Outcomes

Among the infant outcomes, 268 medical records were sought and 183 (68.3%) obtained: 130 (74.7%) MCMs, 27 (56.3%) low birthweight, and 26 (56.5%) SGA (Table 2).

Among the 130 medical records from infants identified as having a MCM by the simple algorithm (Algorithm 1), 54 (41.5%) were classified as definite cases, 1 (0.8%) as a probable case, 70 (53.8%) were non-cases, and 5 (3.8%) had insufficient information to determine case status (Table 5). Among the 70 non-cases, the adjudicators classified 51 (72.9%) as cases of a minor malformation. The PPV of the simple algorithm (Algorithm 1) was 44.0% (95% CI 35.3, 52.7), while Algorithm 3 which required 2 claims separated by at least 30 days had a PPV of 67.8% (95% CI 55.9, 79.7). The PPVs for MCMs by organ system



**Table 4** Positive predictive values for candidate claims-based algorithms: spontaneous abortion

Outcome algorithms	Algorithm-defined outcomes <sup>a</sup>	Records procured and adjudicated	Adjudicators' determination						PPV (%) <sup>b</sup>		95% CI	
			Definite cases		Probable cases		Non-cases		Insufficient Information			
			N	%	N	%	N	%	N	%		
Algorithm 1 (at least 1 claim)	200	125	82	65.6	18	14.4	18	14.4	7	5.6	84.7	78.3–91.2
Algorithm 2 (at least 2 claims on separate days)	132	85	61	71.8	9	10.6	10	11.8	5	5.9	87.5	80.3–94.7
Algorithm 3a (2 claims separated by at least 7 days)	85	54	39	72.2	5	9.3	6	11.1	4	7.4	88.0	79.0–97.0
Algorithm 3b (2 claims separated by at least 14 days)	47	30	22	73.3	3	10.0	2	6.7	3	10.0	92.6	82.7–100.0
Algorithm 4 (at least 1 claim from the hospital, OB/GYN or anesthesiology)	183	117	81	69.2	16	13.7	16	13.7	4	3.4	85.8	79.4–92.3
Algorithm 4a (at least 1 claim from the hospital)	96	62	48	77.4	6	9.7	7	11.3	1	1.6	88.5	80.5–96.5
Algorithm 4b (at least 1 claim from OB/GYN)	161	105	74	70.5	15	14.3	12	11.4	4	3.8	88.1	81.8–94.4
Algorithm 4c (at least 1 claim from anesthesiology)	73	51	41	80.4	5	9.8	5	9.8	0	0.0	90.2	82.0–98.4
Algorithm 5 (at least 1 claim from other or outpatient)	185	111	70	63.1	18	16.2	16	14.4	7	6.3	84.6	77.7–91.5

CI confidence interval, OB/GYN obstetrics and gynecology, PPV positive predictive value

<sup>a</sup>From the validation sample of randomly selected outcomes

<sup>b</sup>Charts with insufficient information for each algorithm were excluded from the denominator of the PPV; the numerator included definite and probable cases

**Table 5** Positive predictive values for candidate claims-based algorithms: major congenital malformations

Outcome algorithms	Algorithm-defined outcomes <sup>a</sup>	Records procured and adjudicated	Adjudicators' determination						PPV (%) <sup>b</sup>	95% CI		
			Definite cases		Probable Cases		Non-cases				Insufficient Information	
			N	%	N	%	N	%			N	%
Algorithm 1 (at least 1 claim)	208	130	54	41.5	1	0.8	70	53.8	5	3.8	44.0	35.3–52.7
Algorithm 2 (at least 2 claims on separate days)	134	89	49	55.1	1	1.1	36	40.4	3	3.4	58.1	47.7–68.6
Algorithm 3 (2 claims separated by at least 30 days)	86	62	39	62.9	1	1.6	19	30.6	3	4.8	67.8	55.9–79.7
Algorithm 4 (at least 1 claim from the hospital or pediatrician)	180	116	53	45.7	1	0.9	57	49.1	5	4.3	48.6	39.4–57.9
Algorithm 4a (at least 1 claim from the hospital)	138	87	47	54.0	1	1.1	36	41.4	3	3.4	57.1	46.6–67.7
Algorithm 4b (at least 1 claim from the pediatrician)	121	82	40	48.8	1	1.2	38	46.3	3	3.7	51.9	40.9–62.9
Algorithm 5 (at least 1 claim from an inpatient stay)	147	90	46	51.1	1	1.1	41	45.6	2	2.2	53.4	43.0–63.8

CI confidence interval, PPV positive predictive value

<sup>a</sup>From the validation sample of randomly selected outcomes

<sup>b</sup>Charts with insufficient information for each algorithm were excluded from the denominator of the PPV; the numerator included definite and probable cases

identified by the simple algorithm (Algorithm 1) are presented in Supplemental Table 7.

The PPV for the simple algorithm (Algorithm 1) for low birthweight was 96.3% (95% CI 89.2, 100.0) (Table 2) and the performance of additional candidate algorithms was similarly high (Supplemental Table 8a). For SGA, the PPV for the simple algorithm (Algorithm 1) was 34.8% (95% CI 15.3, 54.2) (Table 2); furthermore, the additional candidate algorithms all performed poorly (Supplemental Table 8b).

### 3.3 Best-performing Algorithms

Table 6 shows the proposed best-performing algorithms for each of the pregnancy and infant outcomes based on PPV  $\geq$  70.0%, and the number of potential cases and definite cases identified by the algorithm. The simple algorithm (Algorithm 1) performed best for spontaneous abortion, premature delivery, and low birthweight. For pre-eclampsia, the best-performing algorithm was the one that required at least one claim from an inpatient stay (Algorithm 5), which had a PPV of 85.7% (95% CI 70.7–100.0).

## 4 Discussion

In this exploratory study, several claims-based algorithms for pregnancy and infant outcomes were developed and validated through medical record review and adjudication. We also developed claims-based algorithms that accurately estimated the date of LMP among pregnancies resulting in a livebirth. The primary outcomes of interest were MCMs and spontaneous abortion, but algorithms for other pregnancy outcomes, infant outcomes, and LMP date were also evaluated. A simple algorithm based on a single claim was used to identify each outcome and a best-performing algorithm was determined based on the performance characteristics of all candidate algorithms. The algorithms performed well for spontaneous abortion, pre-eclampsia, premature delivery, and low birthweight and poorly for MCMs, placenta previa, and SGA.

Last menstrual period date was estimated accurately with Z3A codes, although it was observed that estimated LMP tended to be a few days later (median: 4.0, IQR: 2.0–10.0) than adjudicated LMP. This is likely because Z3A codes denote completed weeks of gestation; for example, a woman who had a doctor visit when her fetus was gestational age 10 weeks, 3 days would receive a Z3A.10 (10 weeks gestation of pregnancy) code on her claim. Additionally, estimated LMP date was less accurate for pregnancies with a claim for spontaneous abortion. The first specific Z3A code is Z3A.08 (8 weeks gestation of pregnancy), consistent with timing of first prenatal visit [29]. Prior to 8 weeks gestation,

**Table 6** Proposed claims-based *best-performing* algorithms for pregnancy and infant outcomes based on adjudicated medical record

Outcomes	Proposed claims-based <i>best-performing</i> algorithms					
	Algorithm	Algorithm-defined outcomes	Number of records adjudicated <sup>a</sup>	Confirmed by adjudicator	PPV (%)	95% CI
<b>Pregnancy outcomes</b>						
Spontaneous abortion	At least 1 claim (Algorithm 1)	200	118	100	84.7	78.3–91.2
Placenta previa	N/A <sup>b</sup>					
Pre-eclampsia	At least 1 claim from an inpatient stay (Algorithm 5)	43	21	18	85.7	70.7–100.0
Premature delivery	At least 1 claim (Algorithm 1)	44	26	24	92.3	82.1–100.0
<b>Infant outcomes</b>						
Major congenital malformations	N/A <sup>b</sup>					
Low birthweight	At least 1 claim (Algorithm 1)	48	27	26	96.3	89.2–100.0
Small for gestational age	N/A <sup>b</sup>					

CI confidence interval, N/A not applicable, PPV positive predictive value

<sup>a</sup>Records with the study outcome of interest for which medical records with sufficient information for adjudication were obtained

<sup>b</sup>Algorithms for placenta previa, small for gestational age and major malformations performed poorly, and therefore no best-performing claims-based algorithms are proposed

Z3A codes are non-specific; for these codes, we assigned 4 weeks gestation. As most miscarriages occur prior to the 12th week of pregnancy [30], a woman with spontaneous abortion may have had only non-specific Z3A codes in her claims or no Z3A codes at all if she had not yet sought clinical care, resulting in less accurate estimation of LMP date. Among pregnancies without a Z3A code in this study, 96% had spontaneous abortion as the outcome, for which we assigned 10 weeks gestation. Nonetheless, it was observed that the estimated LMP for these pregnancies was approximately 2 weeks different (median: 16.0 days, IQR: 8.0–25.0) than adjudicated LMP.

The simple algorithm for MCMs had a PPV of 44.0%. Among the 130 cases identified as having a MCM by the simple algorithm, 51 (39.2%) were adjudicated as minor malformations only. A previous study conducted by Carman et al within the ORD using ICD-9-CM codes also observed a PPV of 47.8% [11] for the simple algorithm, but the PPV for the final algorithm was 80.4%, which is higher than the PPV of 67.8% observed for the best-performing algorithm in this study. In the Carman et al study, the candidate algorithms were developed using a separate, iterative process for each body system category. The body system-specific algorithms were then applied to the infant study population, resulting in an overall PPV that was improved. Similarly, in a paper that was published after completion of the current study, Kharbanda et al proposed separate algorithms for each organ system when they converted previously validated ICD-9-CM

algorithms for MCMs to ICD-10-CM [14]. The algorithm PPVs were 80% or higher for most defects, although they only validated algorithms for seven targeted MCMs. In the current study, we developed and validated algorithms to identify any MCM. We subsequently examined the PPVs for the simple algorithm by body system, but there were small numbers of records adjudicated for several categories. The simple algorithm did not perform well for many of the body systems and additional candidate algorithms were not explored given the small number of records adjudicated within many body system categories. Nonetheless, given the promising results from Carman et al and Kharbanda et al, future work on algorithms for identifying MCMs should be directed towards developing and validating algorithms for additional MCMs, but doing so according to specific MCM categories. Additionally, it is necessary to refine the list of minor malformations for exclusion using available ICD-10-CM references [22–24] and clinical input.

For spontaneous abortion, the simple algorithm and candidate algorithms all performed well, with all PPVs approaching 85% or higher. Nonetheless, the PPVs observed in this study were slightly lower than the percent agreement between the claims-based algorithm for spontaneous abortion and physician adjudication of electronic medical records (EMRs) reported by Moll et al (100.0%, 95% CI 93.9, 100.0) [12]. One explanation for the better performance of the algorithm in Moll et al is the restriction of the validation sample to pregnancy episodes with a start date, which was estimated

by presence of at least one pregnancy-related code, not including the code for the outcome. This restriction resulted in 22% attrition for the spontaneous abortion outcome. In contrast, we estimated LMP date using outcome-based algorithms [15–17, 20] for pregnancies where Z3A codes were not observed, so zero pregnancies were excluded due to lack of pregnancy start date. Although this approach may have resulted in identifying some spontaneous abortions that were not true cases, it is less likely that spontaneous abortions that occurred very early in pregnancy, prior to the first prenatal visit, were excluded. Additionally, in the Moll et al study, the claims-based algorithms were validated using the structured components from linked (EMRs). In validation studies, however, the gold standard for diagnosis is based on a review of the complete medical record (i.e., structured and unstructured fields) and it is uncertain whether structured fields alone provide the same gold standard.

Although the other pregnancy and infant outcomes were investigated as secondary endpoints in this study, the results obtained will inform the next steps in the development of claims-based algorithms for these outcomes. The low PPV of the placenta previa algorithm may be due in part to revision of the clinical definition for the outcome after adjudication had begun. The simple claims-based algorithm used to identify cases included all ICD-10-CM codes under O44, including codes for low-lying placenta and codes for any trimester of diagnosis. However, placenta previa early in pregnancy often resolves as the uterus enlarges [25]. During adjudication, the definition was restricted to clinically relevant cases, including placenta previa that persisted into the second or third trimester, or for whom there was indication of a caesarian section delivery due to bleeding in the medical record. To account for this revised definition, an algorithm that required at least one claim within 2 weeks of the delivery date was developed. This algorithm also had a low PPV, potentially due to few charts meeting this definition. Future studies should start with a simple algorithm based on claims for complete placenta previa or previa with hemorrhage from an inpatient stay close to delivery or a claim for a caesarian section.

Algorithms for low birthweight performed well (all PPVs close to 100%) while algorithms for SGA performed poorly. This seeming contradiction likely stems from more standardized definitions of low birthweight. In some infants diagnosed as SGA by treating physicians, the birthweight and gestational age in the chart indicated that the infant was not below the 10th percentile of the growth curves [26]. This may occur when intrauterine growth restriction (IUGR) was indicated during pregnancy; however, IUGR and SGA may not be equivalent [27]. A recent study that validated an ICD-9 claims-based algorithm for SGA reported a higher PPV, but their validation criteria included birthweight below

the 20th percentile if accompanied by a diagnosis of SGA or IUGR [9]. Another potential explanation for the poor performance of the SGA algorithm is the inclusion of all ICD-10-CM P05 codes. Although this was done to improve sensitivity, it is possible that only a subset of codes may be relevant for the identification of true cases of SGA.

The PPV for the algorithm for premature delivery was 92%, which is higher than the percent agreement between the claims-based algorithm for preterm live birth and physician adjudication of EMRs reported by Moll et al (62.4%, 95% CI 52.0, 71.7) [12]. Nonetheless, the prevalence of premature delivery based on the algorithm in the current study (3.2%) was lower than the national estimate of 10% [28]. A likely explanation for the low prevalence is that we identified premature delivery from claims in the maternal record only; we did not examine claims for preterm birth in the infant record. Future studies of this outcome should consider using a combination of maternal and infant claims if possible.

This study had several strengths. The use of Optum's large DAPI population with access to source medical records enabled us to investigate the performance of claims-based algorithms for several outcomes relevant to pregnancy safety, including those that are relatively rare. The number of pregnancies in the study period accrued quickly due to the large population size, providing results rapidly to inform public health and regulatory decision making.

Nonetheless, this study had several limitations. To avoid missing any potential cases, the simple algorithm used to identify outcomes was based on a single diagnosis code, but this does not always reflect presence of disease. The diagnosis may be incorrectly coded, as observed for SGA, or the diagnosis code may reflect rule-out criteria or a minor rather than a major form of the condition, as observed for MCMs. Although more rigorous algorithms were developed, improvement in PPV may have been limited because all records were initially selected based on the simple algorithm. Additionally, in this study, sensitivity could not be calculated because charts were only sampled for claims-identified cases.

While the medical records served as the gold standard for validation, they may be incomplete. For example, 30% of medical records for pre-eclampsia had insufficient information to determine case status, mainly due to missing blood pressure and lab information. Further, a small number of charts was adjudicated for the secondary outcomes. Although this was largely by design as this was an exploratory study, only 62.3% of medical records sought for this study were procured and adjudicated, which is lower than historical procurement rates in the ORD (70–85%) [11]. Studies that seek medical records for pregnancies and infants have inherent challenges compared to other studies. For example, some relevant personally identifiable information

needed for requesting charts from providers, such as infant's first name or social security number, may be missing which may impact procurement for charts of outcomes identified soon after birth. Oversampling potential cases should be considered when conducting similar studies to try and overcome this issue.

For validation of LMP date, the analyses were restricted to pregnancies with a claim for an adverse pregnancy outcome, which may limit generalizability. Nonetheless, the pregnancies with premature delivery, placenta previa, and pre-eclampsia in this study often had multiple Z3A codes observed during the pregnancy which likely improved estimation of LMP. The accuracy of estimated LMP date among uncomplicated pregnancies is likely to be similar due to the high probability that multiple Z3A codes would be observed within a full-term pregnancy.

## 5 Conclusion

In conclusion, the ICD-10-CM claims-based algorithm for spontaneous abortion performed well and can be used in administrative databases. The algorithms for LMP date and the secondary outcomes pre-eclampsia, premature delivery, and low birthweight also performed well, but it would be beneficial to validate these algorithms in other study populations using a larger number of procured charts to ensure their generalizability. Furthermore, the value of applying the algorithm for premature delivery within both maternal and infant claims should be assessed. ICD-10-CM algorithms for MCMs, placenta previa, and SGA did not perform well; these algorithms are not recommended for use in research studies without further refinement. Future algorithm refinement for MCMs should build upon validation studies that have developed body system-specific algorithms that have performed well while also honing the list of minor malformations for exclusion. Additionally, the possible benefits of utilizing other data sources (e.g., electronic health record, national registries) to study MCMs as an outcome should be considered. For placenta previa, the outcome definition that is clinically relevant for medication safety studies should be determined based on trimester of diagnosis, clinical characteristics (e.g., hemorrhage), and presence of a caesarian section at delivery. Subsequently, the diagnosis and procedure codes included in the claims-based placenta previa algorithm can be adjusted accordingly. Future work on algorithms for SGA may consider alternative definitions that utilize codes for low birthweight when accompanied by gestational age. By building upon the findings from this exploratory study and similar studies, it is likely that improved ICD-10-CM algorithms for MCMs, placenta previa, and SGA can be developed.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40264-022-01261-5>.

**Acknowledgments** We would like to acknowledge the work of Ron Parambi, Senior Epidemiology Analyst, Optum; Nicole Brooks, Project Manager, Optum (formerly); Katherine Reed, Research Associate Team Lead, Optum; and Ryan Kilpatrick, Vice President, Global Epidemiology, AbbVie. This work was completed under a contract between Optum and AbbVie, Inc. and funded by AbbVie, Inc.

## Declarations

**Funding** This work was completed under a contract between Optum and AbbVie, Inc. and funded by AbbVie, Inc.

**Authors' Contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by AKC, KP, MCD, JFC, and CE. The first draft of the manuscript was written by AKC and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Competing Interests** AKC, MCD, JFC, JDS are employees of Optum and may own stock in UnitedHealth Group. KP and CE were employees of Optum. SEC, BHM, and MCS are employees of and may own stock/stock options in AbbVie, Inc.

**Ethics Approval** The study protocol was approved by the New England Institutional Review Board (IRB # 120180021) and all data access conformed to applicable Health Insurance Portability and Accountability Act policies.

**Consent to Participate** A waiver of consent was obtained for this study because the research involved no more than minimal risk to the subjects; the waiver or alteration would not adversely affect the rights and welfare of the subjects; the research could not practicably be conducted without the waiver or alteration. 45 CFR 46.116(d).

**Consent to Publish** Not applicable.

**Data Availability** The dataset analyzed during the current study is not publicly available but is available (with the exception of the adjudication results) from Optum through a data license agreement. More information can be found at the following website: <https://www.optum.com/business/solutions/life-sciences/real-world-data/claims-data.html>.

**Code Availability** Programming code is unavailable.

**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

- Palmsten K, Huybrechts KF, Mogun H, et al. Harnessing the medicaid analytic extract (MAX) to evaluate medications in pregnancy: design considerations. *PLoS ONE*. 2013;8(6): e67405. <https://doi.org/10.1371/journal.pone.0067405>.
- Margulis AV, Andrews EB. The safety of medications in pregnant women: an opportunity to use database studies. *Pediatrics*. 2017. <https://doi.org/10.1542/peds.2016-4194>.
- Andrade SE, Berard A, Nordeng HME, Wood ME, van Gelder M, 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>.
- Phiri K, Clifford RC, Gately RV, Doherty MC, Seeger JD. Development of a dynamic pregnancy database within the Optum Research Database. 2018:210-211.
- Eworuke E, Hampc C, Saidi A, Winterstein AG. An algorithm to identify preterm infants in administrative claims data. *Pharmacoepidemiol Drug Saf*. 2012;21(6):640–50. <https://doi.org/10.1002/pds.3264>.
- Cooper WO, Hernandez-Diaz S, Gideon P, et al. Positive predictive value of computerized records for major congenital malformations. *Pharmacoepidemiol Drug Saf*. 2008;17(5):455–60. <https://doi.org/10.1002/pds.1534>.
- Geller SE, Ahmed S, Brown ML, Cox SM, Rosenberg D, Kilpatrick SJ. International Classification of Diseases-9th revision coding for preeclampsia: how accurate is it? *Am J Obstet Gynecol*. 2004;190(6):1629–33. <https://doi.org/10.1016/j.ajog.2004.03.061>.
- Phiri K, Hernandez-Diaz S, Tsen LC, Puopolo KM, Seeger JD, Bateman BT. Accuracy of ICD-9-CM coding to identify small for gestational age newborns. *Pharmacoepidemiol Drug Saf*. 2015;24(4):381–8. <https://doi.org/10.1002/pds.3740>.
- He M, Huybrechts KF, Dejene SZ, et al. Validation of algorithms to identify adverse perinatal outcomes in the medicaid analytic extract database. *Pharmacoepidemiol Drug Saf*. 2020;29(4):419–26. <https://doi.org/10.1002/pds.4967>.
- Bateman BT, Hernandez-Diaz S, Straub L, et al. Association of first trimester prescription opioid use with congenital malformations in the offspring: population based cohort study. *BMJ*. 2021;372:102. <https://doi.org/10.1136/bmj.n102>.
- Carman WJ, Accortt NA, Anthony MS, Iles J, Enger C. Pregnancy and infant outcomes including major congenital malformations among women with chronic inflammatory arthritis or psoriasis, with and without etanercept use. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1109–18. <https://doi.org/10.1002/pds.4261>.
- Moll K, Wong HL, Fingar K, et al. Validating claims-based algorithms determining pregnancy outcomes and gestational age using a linked claims-electronic medical record database. *Drug Saf*. 2021;44(11):1151–64. <https://doi.org/10.1007/s40264-021-01113-8>.
- Andrade SE, Shinde M, Moore Simas TA, et al. Validation of an ICD-10-based algorithm to identify stillbirth in the sentinel system. *Pharmacoepidemiol Drug Saf*. 2021;30(9):1175–83. <https://doi.org/10.1002/pds.5300>.
- Kharbanda EO, Vazquez-Benitez G, DeSilva MB, et al. Developing algorithms for identifying major structural birth defects using automated electronic health data. *Pharmacoepidemiol Drug Saf*. 2021;30(2):266–74. <https://doi.org/10.1002/pds.5177>.
- 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(2):908–27. <https://doi.org/10.1111/j.1475-6773.2006.00635.x>.
- Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf*. 2013;22(1):16–24. <https://doi.org/10.1002/pds.3284>.
- 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>.
- Bertoia ML, Phiri K, Clifford CR, et al. Identification of pregnancies and infants within a US commercial healthcare administrative claims database. *Pharmacoepidemiol Drug Saf*. 2022;31(8):863–74. <https://doi.org/10.1002/pds.5483>.
- Seeger JD, M; Phiri, K; Bertoia, M; Seals, R; Wang, FT. Missing links in pregnancy safety studies: how different are mothers with non-linkable infants in claims databases? 2020.
- Margulis AV, Palmsten K, Andrade SE, et al. Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. *Pharmacoepidemiol Drug Saf*. 2015;24(4):335–42. <https://doi.org/10.1002/pds.3743>.
- Committee Opinion No 700: Methods for estimating the due date. *Obstet Gynecol*. 2017;129(5):e150–e154. doi: <https://doi.org/10.1097/aog.0000000000002046>
- Metropolitan Atlanta Congenital Defects Program (MACDP). Accessed July 16, 2019. <https://www.cdc.gov/ncbddd/birthdefects/macdp.html>
- European Surveillance of Congenital Anomalies (EUROCAT). Guide 1.4. Accessed March 23, 2020. [http://www.eurocat-network.eu/aboutus/datacollection/guidelinesforregistration/guide1\\_4](http://www.eurocat-network.eu/aboutus/datacollection/guidelinesforregistration/guide1_4)
- New York State Department of Health Congenital Malformations Registry. ICD-10 Codes List of Reportable Conditions. Accessed March 23, 2020. [https://www.health.ny.gov/statistics/environmental/public\\_health\\_tracking/health/birth\\_defects.htm](https://www.health.ny.gov/statistics/environmental/public_health_tracking/health/birth_defects.htm)
- Jansen C, Kleinrouweler CE, van Leeuwen L, Ruiter L, Mol BW, Pajkrt E. Which second trimester placenta previa remains a placenta previa in the third trimester: a prospective cohort study. *Eur J Obstet Gynecol Reprod Biol*. 2020;254:119–23. <https://doi.org/10.1016/j.ejogrb.2020.08.038>.
- Centers for Disease Control and Prevention, National Center for Health Statistics. Growth Charts. Accessed April 23, 2020. <https://www.cdc.gov/growthcharts/index.htm>
- Lee PA, Chernausk SD, Hokken-Koelega AC, Czernichow P. International small for gestational age advisory board consensus development conference statement: management of short children born small for gestational age, April 24-October 1, 2001. *Pediatrics*. 2003;111(6 Pt 1):1253–61. <https://doi.org/10.1542/peds.111.6.1253>.
- Centers for Disease Control and Prevention, National Center for Health Statistics. Birthweight and Gestation
- American Pregnancy Association. Your First Prenatal Visit. Accessed October 16, 2020. <https://americanpregnancy.org/healthy-pregnancy/planning/first-prenatal-visit-71023>
- March of Dimes. Miscarriage. Accessed July 28, 2020. <https://www.marchofdimes.org/complications/miscarriage.aspx>