

Development of Longitudinal, Linked Maternal-Infant Cohorts using the Epic Cosmos Electronic Health Record Dataset

Authors: Stephanie A. Leonard,^{1,2} Kevin Dysart,^{3,4} Alison Callahan,⁵ Sara Siadat,² Jiaqi Zhang,¹ Sara C. Handley,^{3,4,6} Krista F. Huybrechts,⁷ Iroque Igbinosa,¹ Brian T. Bateman²

Affiliations:

¹ Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, California

² Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, California

³ Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

⁴ Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania

⁵ Division of Computational Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, California

⁶ Leonard Davis Institute of Health Economics, The University of Pennsylvania, Philadelphia, Pennsylvania

⁷ Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

DATA AVAILABILITY STATEMENT

Access to the study dataset, Epic Cosmos, is available through participating healthcare organizations. Details are available at <https://cosmos.epic.com/>.

FUNDING STATEMENT

SAL is funded in part by a Career Development Award from the National Heart, Lung, and Blood Institute (K01HL171699); SCH and II are funded in part by Career Development Awards from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (K23HD109426; K12HD103084). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

CONFLICT OF INTEREST DISCLOSURE

KFH reports being an investigator on grants to her institution from UCB, Takeda and GSK for regulatory-mandated studies, unrelated to this work.

ETHICS APPROVAL STATEMENT

This study was determined to not be human subjects research by the Stanford Institutional Review Board.

SYNOPSIS

- **Study question:** The study sought to develop replicable steps for researchers to create longitudinal, linked maternal-infant cohorts using the Epic Cosmos electronic health record (EHR) dataset, assess completeness of key variables, evaluate potential selection bias with sequential restrictions for longitudinal healthcare encounters, and provide an example analysis.
- **What's already known:** Epic Cosmos is a relatively new dataset with high potential for perinatal epidemiologic research because of its large size and granular information from structured EHR data.
- **What this study adds:** This study provides a rigorous and reproducible approach to creating longitudinal, linked maternal-infant cohorts in Epic Cosmos and found overall high completeness of key variables, and high representativeness of the United States. Cohort characteristics did not meaningfully change with longitudinal cohort selection.

ABSTRACT

Background: Epic Cosmos is a relatively new centralized electronic health record dataset with high potential utility in perinatal epidemiologic research.

Objectives: The study objectives were to develop replicable steps to create longitudinal, linked maternal-infant cohorts in Cosmos, assess completeness of key variables, evaluate potential selection bias with restrictions for longitudinal healthcare encounters, and provide an example epidemiologic analysis.

Methods: We created maternal-infant cohorts by starting with live births during 2023-2024 recorded in the BirthFact data table and joining with additional data tables as needed. We selected and created variables for perinatal characteristics, common comorbidities, and routinely measured vital signs and laboratory values, and assessed variable completeness. We sequentially restricted the birth cohort for maternal-infant linkage and longitudinal healthcare from first-trimester prenatal care encounter through infant follow-up care within 12 weeks post-discharge from birth hospitalization. Finally, we conducted an example analysis of the association between high systolic blood pressure in the first trimester (≥ 140 mm Hg) and later onset of preeclampsia among those with chronic hypertension.

Results: The total linked birth cohort included 2,624,186 pregnancies. Completeness was $>90\%$ for most variables assessed but was 77% for racial and ethnic group and 76% for body mass index at delivery. Characteristics of the cohort were similar to those reported for the entire United States birth population based on birth certificate data, including similar regional and racial-ethnic composition. Longitudinal cohort restriction requiring linked records from first trimester prenatal care through infant follow-up care

reduced the cohort size to 509,148 pregnancies. However, restriction had minimal effects on cohort characteristics. In the example analysis, high systolic blood pressure was associated with increased risk of preeclampsia among those with chronic hypertension (aRR: 1.26; 95% CI: 1.22, 1.30).

Conclusions: This study provides a rigorous and reproducible approach to creating longitudinal, linked maternal-infant cohorts in Epic Cosmos and the analytical findings suggest high data quality and representativeness.

KEYWORDS: electronic health record data; medical informatics; epidemiologic methods; perinatal care; obstetrics; neonatology

BACKGROUND

Electronic health record (EHR) systems have been widely adopted in clinical practice to document patients' conditions, tests, and treatments¹ and contain granular, longitudinal health information collected in real-world clinical settings. These data are potentially a rich source for epidemiologic studies, but concerns about data quality and representativeness have limited their use.²⁻⁴ Data quality concerns include missing measurements and encounters. Representativeness concerns stem from the fact that EHR based studies typically rely on data from a small number of healthcare organizations. Combining EHR data from many institutions for research purposes has historically been a major challenge. Further, perinatal epidemiologic research frequently requires linkage of maternal and infant records, which is challenging in many EHR data sources.

Epic Cosmos is a relatively new dataset that centralizes EHR data from hundreds of healthcare organizations.⁵ Epic is the leading EHR software vendor in the United States, and healthcare organizations that use Epic Systems' EHR can now opt to contribute to Cosmos and access the Cosmos dataset. In Cosmos, information is pulled from all structured EHR fields, with high granularity of patient health information, including laboratory measurements, vital signs, and medications prescribed or administered. Organizations participating in Cosmos range from academic to community health systems, are located throughout the United States and select locations in Saudi Arabia, Lebanon, and Canada (currently), and continue to grow in number. The Cosmos dataset currently contains data from approximately 300 million patients seen at over 2,000 hospitals and 47,000 clinics.⁵

Interest in and use of Epic Cosmos for perinatal epidemiologic research are growing rapidly, creating a critical need to assess data quality and representativeness of maternal-infant cohorts in Cosmos.⁶⁻⁸ The goal of this study was to provide detailed, replicable steps to create longitudinal, linked maternal-infant cohorts in Cosmos, assess completeness of key variables, and evaluate potential selection bias via restrictions for longitudinal healthcare encounters. In addition, we sought to provide perinatal epidemiologic researchers with an applied example, by using the Cosmos dataset to estimate the association between systolic blood pressure (SBP) in the first trimester of pregnancy and later onset of preeclampsia among those with chronic hypertension. Prior studies in selective study samples have found that individuals with higher SBP in the first trimester were more likely to develop preeclampsia.⁹⁻¹¹ Screening for high SBP and modifying antihypertensive treatment early in pregnancy could be potential strategies to reduce the risk of preeclampsia, which affects approximately 30% of women with chronic hypertension and poses considerable risks to the pregnant woman and infant.^{12,13}

METHODS

Cohort Creation

Steps taken for cohort creation are described here and in greater detail in **Appendix 1**. We first selected all births recorded in the Cosmos BirthFact table during January 1, 2023 to December 31, 2024. The BirthFact table is populated for a patient when an EHR is created for an infant and documentation is made in the delivery summary. New EHRs are typically only created for liveborn infants and the existence of

a delivery summary indicates that the delivery occurred in the same healthcare organization that is contributing the data to Epic Cosmos. Therefore, pregnancies resulting in non-livebirths and those delivered at an organization that does not participate in Cosmos were not included. Such pregnancies would need to be identified using a different approach based on data tables other than BirthFact, including the PregnancyFact table. We selected 2023 to 2024 to have complete years of data and infant follow-up through 12 weeks post-discharge from birth hospitalization at the time of analysis, but the methods described here can be implemented for all available years in Cosmos when accounting for the necessary follow-up time in a given study. Data recorded in the BirthFact table are considered reliable as early as 2017 or when an organization started using Epic, whichever occurred later. Data are updated in Cosmos approximately every two weeks and we conducted analyses using data updated through 5/20/2026.

We then linked the birth records to pregnancy records by joining the BirthFact and PregnancyFact tables based on maternal records. The PregnancyFact table is populated when a patient receives healthcare during their pregnancy. The BirthFact table contains a record for each liveborn infant; for example, if an obstetric patient gives birth to liveborn twins, there will be two records in the BirthFact table and one record in the PregnancyFact table. Next, we linked to birth encounters during the same birth years (2023-2024) by joining to the EncounterFact table. The EncounterFact table contains a record for each encounter that a patient has with the healthcare organization that is documented in an EHR and encounters derived from billing data. We then removed unavailable birth encounters (i.e., missing, masked, deleted). BirthFact tables

could be populated retrospectively by a clinician at a later date and not be linked to a birth encounter. Next, we identified twins and higher-order births and, if multiple birth dates, assigned the earlier date as the delivery date of the pregnancy. In some instances, a stillborn infant will have an EHR created, which will be captured by the BirthFact table. We therefore excluded any pregnancies with an ICD code indicative of stillbirth. We note that it is recommended to use both ICD-9 and ICD-10 across all data years in Epic Cosmos due to the common data mapping process. Such recommendations are detailed in the Epic Cosmos data documentation.

Study Variables

We selected and created variables that are frequently used in perinatal epidemiologic research and surveillance. All variables are detailed in **Appendix 2**, including time windows used for diagnoses, procedures, or measurements. We additionally created variables that are important in obstetric or neonatal health but not typically available or reliable in administrative claims or vital records data. For example, parity, body mass index (BMI), infant birthweight, and Apgar score are not well-captured in administrative claims data; hypertensive disorders, diabetes, and depression are not well-captured in vital records data; and blood pressure measurements and hemoglobin values are not well-captured in either administrative claims or vital records data. For gestational age, if one pregnancy had multiple gestational ages, we assigned the lowest. If no gestational age was available, we calculated it as the delivery date from the BirthFact table minus the estimated start date of the pregnancy from the PregnancyFact

table. If that was missing, we set gestational age to missing. We calculated the last menstrual period as the delivery date minus the gestational age.

Longitudinal, Linked Maternal-Infant Cohorts

We then placed restrictions to identify maternal-infant dyads with continuity of healthcare capture in Epic beyond the birth hospitalization. Some patients give birth at hospitals that use Epic Systems for EHRs and participate in Cosmos, but receive their prenatal care in clinics that do not use Epic or do not participate in Cosmos such that these encounters are not captured in the Cosmos database. Excluding such patients is necessary for research questions with exposures, outcomes, or covariates that occur before or after the birth hospitalization, but reduce cohort size and therefore statistical power and may also induce selection bias. To examine this phenomenon, we started with the birth cohort and then added sequential restrictions based on encounter type, department of encounter, and timing of encounter relative to the start of pregnancy or date of delivery. First, we created a cohort with likely continuous healthcare documented in Epic throughout pregnancy and birth by requiring at least one prenatal care encounter in the first trimester (before gestational week 14) and at least one prenatal care encounter in the second trimester (from gestational week 14 to week 27). Second, we extended the restriction to require at least one postpartum care encounter within 12 weeks of delivery. Third, we added a requirement of at least one infant follow-up care encounter within 12 weeks of discharge from the birth hospitalization. These steps are further detailed in **Appendix 3**.

Statistical Analysis

We first calculated the frequencies and percentages of completeness for key perinatal variables. Next, we calculated the frequencies and percentages of maternal and neonatal characteristics in each of the four cohorts with increasingly restrictive longitudinal selection: (1) linked maternal-infant birth encounters, (2) birth and prenatal care encounters, (3) birth, prenatal, and postpartum care encounters, (4) birth, prenatal, postpartum, and infant follow-up care encounters. Finally, we conducted an example analysis, replicating the known association between high SBP in the first trimester of pregnancy and later onset of preeclampsia among pregnant individuals with chronic hypertension. For this analysis, we selected individuals with at least one prenatal care encounter in the first trimester from the total birth cohort. We then restricted to individuals with a diagnosis of chronic hypertension between three months prior to pregnancy and 20 weeks' gestation and a SBP recorded during a first trimester prenatal care encounter. We defined high SBP as ≥ 140 mm Hg following hypertension guidelines of the American College of Obstetricians and Gynecologists.¹⁴ We estimated risk ratios (RR) with 95% confidence intervals (CI) using multivariable modified Poisson regression models with robust standard errors. We used multiple imputation by chained equations (50 imputations) to account for missing values of the covariates: maternal age (0.04%), insurance status (6.4%), and race-ethnicity (18.7%). Analysis was conducted in R version 4.5.1.

RESULTS

The total cohort of births during 2023-2024 included 2,711,582 pregnancies to 2,652,556 mothers and 3,552,710 infants (**Figure 1**). After restricting to linked maternal-infant records, there was a 3% reduction in the number of pregnancies and mothers and a 25% reduction in the number of infants. Maternal-infant linkage in the birth cohort requires that the infant was delivered at a Cosmos-participating healthcare organization, has a mother documented in their family history, and the delivery episode was recorded in the mother's EHR. Therefore, an infant who received healthcare at a Cosmos-participating hospital or clinic but was born at a non-participating hospital could not be linked to a maternal record. Restricting the cohort to those with prenatal care encounters documented during the first and second trimesters, reduced the cohort size by an additional 62%. The excluded patients would be those who received prenatal care at a healthcare organization that does not participate in Epic Cosmos and those who received delayed or no prenatal care. Restricting the cohort to those with a postpartum visit, reduced the cohort size by 16%. Finally, requiring that there be a linked infant follow-up visit, reduced the cohort size by 41%. The cohort sizes in this most restrictive linked, longitudinal cohort were 501,178 pregnancies, 493,224 mothers, and 509,148 infants during the two-year study period. The cohort sizes differ for each group because a small proportion (~1.5%) of mothers had more than one pregnancy during the two years and a small proportion (~1.5%) of pregnancies were twins or higher-order multiples.

Data completeness varied among key perinatal variables (**Table 1**). Completeness was >97% for age, region of residence, rurality, parity, multiple gestation, gestational age, mode of birth, birthweight, and 5-minute Apgar score. It was lower (89-

92%) for hemoglobin and blood pressure measurements during delivery hospitalization and for payer status. Completeness was lowest for racial-ethnic group (77%) and BMI at delivery (76%).

Maternal health characteristics were similar across the four cohorts with increasingly restrictive longitudinal records (**Table 2**). The cohorts were diverse with respect to racial-ethnic group, payer status, and region of residence in the United States. The mean maternal age was 30 years old and most individuals lived in metropolitan areas, were multiparous, and had a BMI ≥ 30 kg/m² at delivery. In the total linked birth cohort, the prevalence of common morbidities was 11% for preeclampsia, 9% for gestational diabetes, 12% for depression, and 26% for anemia. At admission for delivery, mean hemoglobin was 11.8 g/dL (standard deviation (SD): 1.3) and mean blood pressure was 126/77 mm Hg (SD: 14.7 for systolic BP and 11.1 for diastolic BP).

Birth and infant characteristics were also similar across the four cohorts (**Table 3**). An exception was lower prevalence of extremely preterm birth (<28 wk), very low birthweight (<1500 g), and very low 5-minute Apgar score (0-3) in the cohort with an infant follow-up visit within 12 weeks of discharge from the birth hospitalization. In the total linked birth cohort, 3% of infants were twins, 33% were delivered by cesarean, 9% were born preterm (<37 wk), 9% had a low birthweight (<2500 g), and 2% had a low to moderate 5-minute Apgar score (<7).

The analysis of the association between high SBP in the first trimester and preeclampsia included 58,540 pregnancies. Of 2,624,186 pregnancies in the full cohort, 1,116,621 (43%) had a prenatal care encounter in the first trimester, 70,619 (6.3%) had a chronic hypertension diagnosis, and 58,540 (83%) had SBP recorded at a first-

trimester prenatal encounter. The prevalence of high SBP (≥ 140 mm Hg) was 28.6%. Among those with a high SBP, the prevalence of preeclampsia was 39.1%, compared with 30.4% among those without a high SBP (**Table 4**). The RR for the association between high SBP and preeclampsia was 1.29 (95% CI: 1.25, 1.32) in the unadjusted regression model and 1.26 (95% CI: 1.22, 1.30) in the adjusted regression model.

COMMENT

This study demonstrated replicable steps for longitudinal, linked maternal-infant cohort creation using the Epic Cosmos dataset, found low missingness of key variables, and did not find evidence of selection bias with longitudinal cohort selection. In an applied analytical example using the cohort to replicate a known association, we found that high SBP (≥ 140 mm Hg) in the first trimester was associated with approximately 20-30% higher risk of developing preeclampsia among individuals with chronic hypertension. This example shows the utility of Cosmos for perinatal epidemiologic research questions that benefit from a very large, generalizable U.S.-based sample and utilize information recorded in structured EHR.

The Epic Cosmos maternal-infant cohort was found to be highly representative of the United States. Comparing the Cosmos 2023-2024 linked birth cohort to national birth certificate data from 2023,¹⁵ 45% vs 41% lived in the South, 24% vs 20% in the Midwest, 17% vs 16% in the Northeast, and 14% vs 21% in the West. Racial-ethnic composition was also comparable in the Cosmos and birth certificate cohorts, with the largest groups being 46% vs 50% non-Hispanic White, 22% vs 26% Hispanic/Latino, and 15% vs 14% non-Hispanic Black.¹⁵ The prevalences of common conditions during

pregnancy, including gestational diabetes, anemia, and depression, were similar to those reported in other large studies in the United States.^{16–19} Diagnoses in EHR data may be prone to overreporting, such as coding for the purpose of ordering a diagnostic test, and improved phenotyping of perinatal conditions by Cosmos users for accurate identification of conditions and their onset is an important area of work that will be widely beneficial.² We also encourage researchers to apply filters that restrict time windows for the diagnosis during pregnancy and differentiate mutually exclusive conditions (e.g., preexisting diabetes and gestational diabetes.)

Data completeness in the Epic Cosmos maternal-infant cohort was found to be high overall. Missingness was less than 1% for key perinatal characteristics often derived from birth certificates and unavailable in claims datasets (e.g., parity, birthweight, Apgar score, rurality). Missingness was higher (~10%) for vital signs and laboratory results routinely measured at delivery admission (blood pressure and hemoglobin), and highest (~25%) for racial-ethnic group and BMI at delivery. As demonstrated in our example analysis, we would encourage researchers to use missing data techniques in their analyses with Cosmos to reduce the risk of bias.

Sequential restrictions for longitudinal cohort selection across the prenatal, perinatal, and neonatal periods reduced cohort size but did not indicate that selection bias was induced, which is reassuring. However, researchers should be aware that longitudinal cohort selection will exclude certain patient groups (e.g., patients with delayed or no prenatal care or those who switch healthcare organizations). Longitudinal restrictions did reduce cohort size considerably—from approximately 2.5 million pregnancies in the linked maternal-infant cohort to fewer than 500,000 pregnancies with

longitudinal care from the first trimester through infant follow-up care. This reduction was likely driven by individuals who had a delivery at a Cosmos-participating healthcare organization but received prenatal or infant follow-up care at a healthcare organization that does not participate in Cosmos. Further, the number of source data organizations in Cosmos increases over time as more and more organizations adopt Epic EHR systems. Researchers should be aware of this and take steps to mitigate induced bias, such as by restricting to source organizations with data across a full study period when conducting trend analyses.

Additional findings for researchers to note is that restriction to those with prenatal care and delivery encounters increased the prevalence of some comorbidities (e.g., antenatal depression from 12% to 17%). Further, restriction to those with infant follow-up care reduced the prevalence of extremely preterm birth, very low birthweight, and very low 5-minute Apgar score; these differences may reflect the exclusion of in-hospital neonatal deaths and transfer of high-risk newborns outside of Cosmos-participating organizations with this cohort restriction. We limited our study to births with over one year of follow-up data, which is an important consideration for studies examining outcomes of high-risk newborns, who will typically have long hospitalizations and may require hospital transfer.

This study has limitations. This initial study only included pregnancies that ended in a live birth and an area of future work is developing a robust approach to identify non-live births including stillbirths, spontaneous abortions and pregnancy terminations. This approach will be distinct and more complex as non-live births would not be captured in the BirthFact table and may not be recorded in the PregnancyFact table, depending on

the types of any healthcare services received during pregnancy and the loss. Another important consideration for our results is that we did not place restrictions on variables, such as removing outlier values or requiring multiple records to confirm a diagnosis. This was an intentional decision to make the cohort steps useful to a broad set of researchers, who would make such decisions in the analytical stage. For example, it would be possible in Cosmos for a researcher to require both a diagnosis of gestational diabetes and a positive 3-hour glucose tolerance test. Finally, it is important for researchers to know that Cosmos contains structured EHR data, including demographic information, vital signs, diagnoses, procedures, medications prescribed and administered, and laboratory test results. However, not all structured EHR data are currently available, including types of flowsheets and laboratory results that appear as a scanned document in a chart. Unstructured EHR data, such as clinical notes, are also not available.

In this study, we presented a rigorous and reproducible approach to creating longitudinal, linked maternal-infant cohorts in Epic Cosmos—a relatively new EHR dataset with a very large sample size and granular information from structured EHR data. Our findings demonstrate overall high completeness of key variables, high representativeness of the United States, and characteristics that do not meaningfully change with longitudinal cohort selection. We demonstrated the utility of the cohorts by analyzing the association between a vital sign measurement (systolic blood pressure) and a clinical outcome (preeclampsia superimposed on chronic hypertension). Overall, the study suggests that Epic Cosmos is a highly valuable data resource for perinatal epidemiologic research.

ACKNOWLEDGMENTS

Data used in this study came from Epic Cosmos, a dataset created in collaboration with a community of health systems using Epic representing more than 300 million patient records from over 2,000 hospitals and 47,000 clinics as of April 2026. The community represents patients from all 50 states, D.C., Canada, Lebanon, and Saudi Arabia.

REFERENCES

1. CDC. NEHRS Results and Publications. National Electronic Health Records Survey. December 15, 2025. Accessed April 21, 2026. <https://www.cdc.gov/nchs/nehrs/results/index.html>
2. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol*. 2021;21(1):234. doi:10.1186/s12874-021-01416-5
3. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013;20(1):144-151. doi:10.1136/amiajnl-2011-000681
4. Altman MR, Colorafi K, Daratha KB. The Reliability of Electronic Health Record Data Used for Obstetrical Research. *Appl Clin Inform*. 2018;09(1):156-162. doi:10.1055/s-0038-1627475
5. About | Epic Cosmos. Accessed April 22, 2026. <https://cosmos.epic.com/about/>
6. Son M, Gallagher K, Lo JY, et al. Coronavirus Disease 2019 (COVID-19) Pandemic and Pregnancy Outcomes in a U.S. Population. *Obstet Gynecol*. 2021;138(4):542. doi:10.1097/AOG.0000000000004547
7. Handley SC, Gallagher K, Lindgren E, et al. Postpartum Length of Stay and Hospital Readmission Before and During the Coronavirus Disease 2019 (COVID-19) Pandemic. *Obstet Gynecol*. 2022;139(3):381. doi:10.1097/AOG.0000000000004687
8. Scott K, Miller E, Culhane JF, et al. Trends in Vitamin K Administration Among Infants. *JAMA*. Published online December 8, 2025. doi:10.1001/jama.2025.21460

9. Nzelu D, Dumitrascu-Biris D, Nicolaidis KH, Kametas NA. Chronic hypertension: first-trimester blood pressure control and likelihood of severe hypertension, preeclampsia, and small for gestational age. *Am J Obstet Gynecol*. 2018;218(3):337.e1-337.e7. doi:10.1016/j.ajog.2017.12.235
10. Nuckols VR, Holwerda SW, Luehrs RE, et al. Beat-to-Beat Blood Pressure Variability in the First Trimester Is Associated With the Development of Preeclampsia in a Prospective Cohort. *Hypertension*. 2020;76(6):1800-1807. doi:10.1161/HYPERTENSIONAHA.120.15019
11. Baschat AA, Magder LS, Doyle LE, Atlas RO, Jenkins CB, Blitzer MG. Prediction of preeclampsia utilizing the first trimester screening examination. *Am J Obstet Gynecol*. 2014;211(5):514.e1-514.e7. doi:10.1016/j.ajog.2014.04.018
12. Leonard SA, Formanowski BL, Phibbs CS, et al. Chronic Hypertension in Pregnancy and Racial–Ethnic Disparities in Complications. *Obstet Gynecol*. 2023;142(4):862-871. doi:10.1097/AOG.0000000000005342
13. Rana S, Lemoine E, Granger JP, Karumanchi SA. Preeclampsia. *Circ Res*. 2019;124(7):1094-1112. doi:10.1161/CIRCRESAHA.118.313276
14. ACOG. ACOG Practice Bulletin No. 203: Chronic Hypertension in Pregnancy. *Obstet Gynecol*. 2019;133(1):e26. doi:10.1097/AOG.0000000000003020
15. Osterman MJK, Hamilton BE, Martin JA, Driscoll AK, Valenzuela CP. Births: Final Data for 2023. *Natl Vital Stat Rep*. 2025;74(1).
16. Liu EF, Ferrara A, Sridhar SB, Greenberg MB, Hedderson MM. Association Between Neighborhood Deprivation in Early Pregnancy and Gestational Diabetes Mellitus. *Obstet Gynecol*. 2024;144(5):670. doi:10.1097/AOG.0000000000005521
17. Gorsch LP, Wen T, Lonier JY, et al. Trends in delivery hospitalizations with pregestational and gestational diabetes mellitus and associated outcomes: 2000–2019. *Am J Obstet Gynecol*. 2023;229(1):63.e1-63.e14. doi:10.1016/j.ajog.2022.12.006
18. Igbinosa II, Booman A, Duverge-Corporan M, et al. Anemia During Pregnancy in a U.S. Nationwide Cohort, 2018-2023. *Obstet Gynecol*. Published online March 26, 2026. doi:10.1097/AOG.0000000000006270
19. Yin X, Sun N, Jiang N, et al. Prevalence and associated factors of antenatal depression: Systematic reviews and meta-analyses. *Clin Psychol Rev*. 2021;83:101932. doi:10.1016/j.cpr.2020.101932

Figure 1. Linked, longitudinal birth cohort selection.

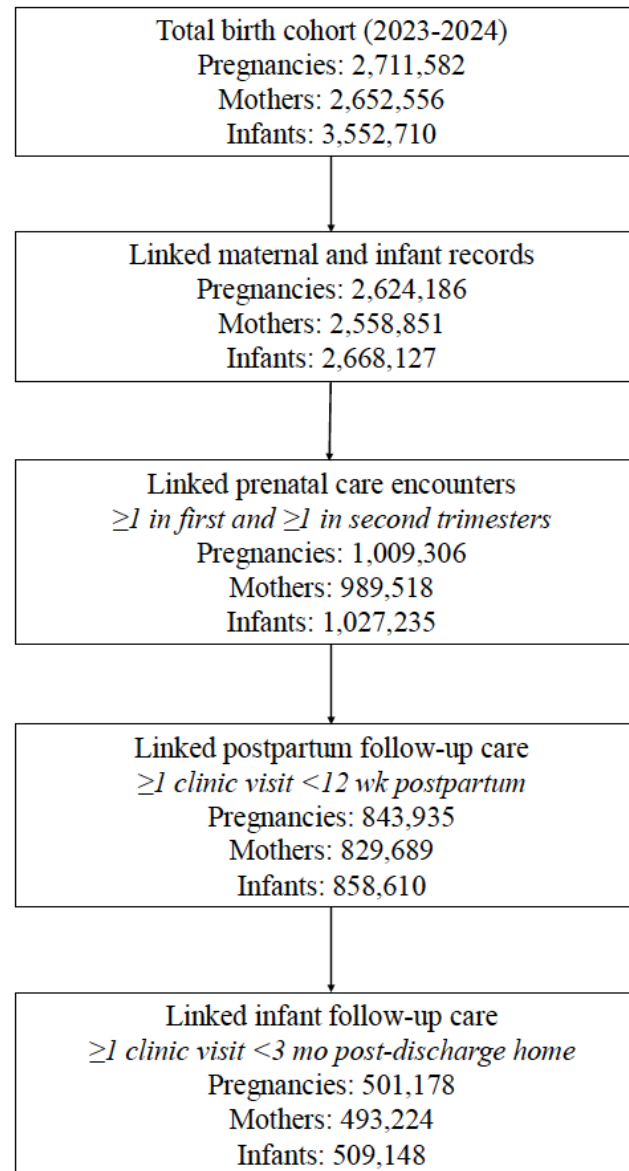


Table 1. Completeness of characteristics (N = 2,668,127 infant records in linked maternal-infant cohort)	
Characteristics	N (%)
Maternal age at delivery	2,668,126 (99.9)
Racial and ethnic group	2,048,540 (76.8)
Payer for delivery encounter	2,420,878 (90.7)
Region	2,641,016 (99.0)
Rurality	2,608,009 (97.8)
Delivery BMI	2,016,467 (75.6)
Parity	2,668,127 (100.0)
Hemoglobin at delivery admission	2,449,140 (91.8)
Blood pressure at delivery admission	2,361,543 (88.5)
Multiple gestation	2,668,127 (100.00)
Gestational age at birth	2,663,267 (99.8)
Mode of birth	2,647,368 (99.2)
Birthweight	2,665,081 (99.9)
Apgar score at 5 minutes	2,650,192 (99.3)

Table 2. Maternal health characteristics of pregnancies by linked maternal-infant cohort selection

	Birth encounters (n = 2,624,186)	Birth and prenatal care encounters (n = 1,009,306)	Birth, prenatal, & postpartum care encounters (n = 843,935)	Birth, prenatal, postpartum, & infant care encounters (n = 501,178)
Maternal age (y)	30 (6)	30 (6)	30 (6)	30 (6)
Racial and ethnic group ¹				
Hispanic/Latino	447,993 (22)	145,687 (19)	118,334 (18)	79,228 (21)
AI/AN	8,401 (0.4)	3,268 (0.4)	2,186 (0.3)	1,149 (0.3)
Asian	90,961 (5)	32,382 (4)	28,200 (4)	15,227 (4)
Black	307,511 (15)	107,725 (14)	84,914 (13)	50,557 (13)
NHOPI	6,832 (0.3)	1,546 (0.2)	1,135 (0.2)	659 (0.2)
White	922,745 (46)	392,485 (50)	333,305 (51)	182,022 (47)
Multiple races	201,709 (10)	95,949 (12)	82,157 (13)	52,888 (14)
Another race	28,413 (1)	8,819 (1)	7,401 (1)	4,300 (1)
Payer for delivery encounter ¹				
Commercial insurance	1,314,840 (55)	541,139 (58)	470,337 (60)	269,897 (58)
Medicaid	981,425 (41)	363,461 (39)	285,931 (37)	182,733 (39)
Military/Veterans	48,885 (2)	17,225 (2)	14,863 (2)	7,024 (2)
Other	35,599 (2)	14,666 (2)	12,713 (2)	8,260 (2)
Region ¹				

Northeast	432,161 (17)	190,995 (19)	157,227 (19)	91,305 (18)
Midwest	621,330 (24)	285,230 (29)	229,966 (28)	158,377 (32)
South	1,165,578 (45)	378,925 (38)	326,056 (39)	175,765 (35)
West	354,689 (14)	139,512 (14)	118,605 (14)	67,866 (14)
International ²	23,703 (0.9)	3,995 (0.4)	3,252 (0.4)	2,785 (0.6)
Rurality ¹				
Metropolitan	2,249,401 (88)	847,924 (85)	710,838 (86)	418,900 (85)
Non-metropolitan	315,662 (12)	145,850 (15)	120,286 (15)	74,151 (15)
Delivery BMI (kg/m ²) ¹				
<30	750,336 (38)	267,057 (35)	220,118 (35)	129,670 (34)
30-39.9	962,746 (48)	374,407 (49)	313,540 (50)	187,952 (50)
≥40	276,668 (14)	117,756 (16)	99,047 (16)	59,787 (16)
Parity				
Nulliparous	1,040,972 (40)	399,807 (40)	346,996 (41)	210,618 (42)
1-4 prior births	1,15,383 (58)	589,465 (58)	483,002 (57)	282,376 (56)
≥5 prior births	67,831 (3)	20,034 (2)	13,937 (2)	8,184 (2)
Chronic hypertension	101,477 (4)	65,052 (6)	55,189 (7)	62,091 (6)
Preeclampsia	286,136 (11)	119,321 (12)	103,079 (12)	62,091 (12)
Preexisting diabetes	98,348 (4)	49,191 (5)	41,350 (5)	24,868 (5)
Gestational diabetes	245,085 (9)	123,283 (12)	105,518 (13)	64,164 (13)
Depression in pregnancy	322,967 (12)	173,972 (17)	144,052 (17)	88,942 (18)
Anemia at	687,325 (26)	247,649 (25)	198,808 (24)	115,874 (23)

delivery				
Hemoglobin at delivery admission (g/dL)	11.8 (1.3)	11.8 (1.3)	11.9 (1.3)	11.9 (1.3)
Systolic blood pressure at delivery admission (mm Hg)	126 (14.7)	127 (14.4)	127 (14.5)	127 (14.4)
Diastolic blood pressure at delivery admission (mm Hg)	77 (11.1)	77.1 (10.9)	77.2 (10.9)	77.2 (10.8)

N (column) for categorical variables

Mean (SD) for continuous variables

¹ Cell frequencies do not sum to total sample size due to missing values

² Canada, Lebanon, Saudi Arabia

Table 3. Characteristics of infants by linked maternal-infant cohort selection

	Birth encounters (n = 2,668,127)	Birth and prenatal care encounters (n = 1,027,235)	Birth, prenatal, & postpartum care encounters (n = 858,610)	Birth, prenatal, postpartum, & infant care encounters (n = 509,148)
Plurality				
Singleton	2,580,963 (97)	991,730 (97)	829,528 (97)	493,031 (97)
Twin	85,068 (3)	34,480 (3)	28,306 (3)	15,791 (3)
Triplet or higher order	2,096 (<0.1)	1,025 (<0.1)	776 (<0.1)	326 (<0.1)
Gestational age at birth ¹ (wk)				
<28	15,168 (0.6)	5,168 (0.5)	4,219 (0.5)	275 (<0.1)
28 to <32	23,349 (0.9)	8,521 (0.8)	7,042 (0.8)	3,383 (0.7)
32 to <37	199,784 (8)	76,817 (8)	62,726 (7)	36,816 (7)
≥37	2,424,966 (91)	936,729 (91)	784,623 (91)	468,674 (92)
Mode of birth ¹				
Spontaneous vaginal	1,682,222 (64)	648,076 (64)	529,952 (62)	317,392 (63)
Operative vaginal	81,001 (3)	29,734 (3)	25,433 (3)	15,444 (3)
Cesarean	884,145 (33)	338,970 (33)	293,942 (35)	170,506 (34)
Birthweight ¹ (g)				
<1500	38,639 (1.4)	13,733 (1.3)	11,368 (1.3)	3,749 (0.7)
1500 to <2500	204,501 (8)	77,604 (8)	62,782 (7)	37,204 (7)
2500 to <4000	2,240,960 (84)	863,412 (84)	722,425 (84)	432,064 (85)
≥4000	180,981 (7)	71,734 (7)	61,440 (7)	35,914 (7)

Apgar score at 5 minutes ¹				
0-3	12,050 (0.5)	4,376 (0.4)	3,601 (0.4)	1,385 (0.3)
4-6	43,637 (1.6)	18,438 (1.8)	15,420 (1.8)	8,648 (1.7)
7-10	2,594,505 (98)	998,778 (98)	835,126 (98)	496,665 (98)

¹ Cell frequencies do not sum to total sample size due to missing values

Table 4. Association between high systolic blood pressure (SBP) (≥ 140 mm Hg) in the first trimester and the outcomes of preeclampsia among individuals with chronic hypertension (n = 58,540)	
Prevalence in those with high SBP (n = 16,746), n (%)	6,540 (39.1)
Prevalence in those without high SBP (n = 41,794), n (%)	12,695 (30.4)
Unadjusted RR (95% CI)	1.29 (1.25, 1.32)
Adjusted RR (95% CI)	1.26 (1.22, 1.30)

Adjusted for maternal age, parity, racial and ethnic group, and payer status.