Supplementary Materials

Early Detection of High-Risk Pregnancies with Integrated Clinical and Social Data to Prevent Maternal and Infant Morbidity and Mortality

Sadiq Y. Patel PhD MS MSW, Chitra Akileswaran MD MBA, Sanjay Basu MD PhD

Corresponding author: sadiqyp@upenn.edu

The PDF file includes:

Materials and Methods

Study sample inclusion criteria

The final sample included 26 states and Washington, D.C., based on data quality reported by DQ Atlas. The analysis focused on women aged 15-49 who had their first pre-delivery pregnancy visit between October 2017 and December 2018 (a 15-month period). Eligibility criteria required members to have at least 6 out of 9 months of coverage during the "prediction measurement period" and at least 8 out of 12 months during the "outcome measurement period." Patients dually enrolled in Medicare and Medicaid were excluded.

Table 1: States to exclude and reason for exclusion

		2019	2018	2017
Eligibility File	Unstable enrollment benchmarks	RI	RI	RI
	Missing total Medicaid and CHIP enrollment	None	None	None
	Missing dual eligibility code	UT, AL	UT, AL	UT, AL
	Missing dually enrolled patients	KY	AZ, KY, MI, MA, RI	AZ, KY, MI, MA, RI
	Missing age	None	None	None
	Missing county FIPS code	Information not provided by DQ Atlas	Information not provided by DQ Atlas	Information not provided by DQ Atlas
Other Services File	Unstable claims volume	MN, MA, RI, NJ	MN, FL, MA, RI, NJ	MN, FL, MA, RI, NJ
	Claim type code	None		
	Missing diagnosis code	None	None	None
	Missing procedure code (professional claims)	UT	UT	UT
	Missing procedure code (institutional claims)	NY, TX	TX, NY, GA, PA	TX, NY, GA, PA

	Place of service code	None (present for each claim line)	UT	UT
	Date of service	Information not provided by DQ Atlas	Information not provided by DQ Atlas	Information not provided by DQ Atlas
Inpatient File	Unstable claims volume	OK, NY, MA, NH, RI, CT	OK, GA, NH, MA, RI, CT	OK, GA, NH, MA, RI, CT
	Claim type code	None (present for each claim line)	None (present for each claim line)	None (present for each claim line)
	Missing diagnosis code	TN	TN	TN
	Missing admission date	ID	GA	GA
	Missing discharge date	ID, MD	NE, IN, GA, MD	NE, IN, GA, MD
Rx File	Claims volume	NC	OH, NC, RI	OH, NC, RI

Identifying the first pre-delivery pregnancy first

We used the CMS algorithm to identify the first pre-delivery pregnancy outcome based on ICD-10-CM, revenue, and ICD-10-PCS codes. This analysis focused on ICD-10 diagnostic codes 1-12, admitting diagnosis codes, procedure codes 1-6, and revenue codes.

1. **Initial Identification**: We identified initial pregnancy visits occurring between October 2017 and December 2018.
2. **Exclusion of Prior Visits**: Patients with any prior pregnancy-related visits during this period were excluded from the sample.
3. **Exclusion of Delivery Outcomes**: Patients whose first pregnancy-related visit was classified as a delivery outcome—defined by CMS as live birth, miscarriage, stillbirth, termination, or unknown delivery outcome—were also removed.

Measurement period

The "prediction measurement period" refers to the 9 months preceding the first pre-delivery pregnancy visit. The "outcome measurement period" is defined as the 12 months immediately following the first pre-delivery pregnancy visit.

Eligibility criteria required members to have at least 6 out of 9 months of coverage during the prediction measurement period and at least 8 out of 12 months of coverage during the outcome measurement period.

Machine learning procedures

Model Training and Testing Process

We conducted repeated model training and testing 100 times using a Markov bootstrapping procedure. For each iteration, the model was trained and tested on randomly generated train-test splits, ensuring robust performance estimates.

Point Estimates and 95% Confidence Intervals

The point estimates and 95% confidence intervals for performance metrics were calculated based on the results of the 100 repeated train-test splits. All performance metrics noted in the prior results, including AUC, MCC, F1 Score, PPV, NPV, accuracy, sensitivity, and specificity, were measured to evaluate the model.

Hyperparameters

The following hyperparameters were used to find the optimal hyperparameters for each XGBoost model.

Hyperparameter GridSearch
Learning rate: 0.005, 0.01, 0.02, 0.05 Subsample: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 Min_child_weight: 0, 0.5, 1, 2, 3, 4, 5 Colsample_bytree: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 Max_depth: 10, 12, 14, 16, 18, 20, 22, 24, 26 Reg_alpha: 0.0, 0.25, 0.5, 0.75, 1.0 Reg_lambda: 0.0, 0.25, 0.5, 0.75, 1.0 Gamma: 0.0, 0.25, 0.5, 0.75, 1.0 Colsample_bylevel: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

Table S1.

Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement

Section/Topic			Checklist Item	Page / Section
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Title page (page 1)
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Abstract (page 2)
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Introduction (page 3)
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	Introduction (page 4)
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	Methods, study design and data sources (page 5)
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	Methods, cohort selection and outcome definition (page 5)
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centers.	Methods, cohort selection and outcome definition (page 5)
	5b	D;V	Describe eligibility criteria for participants.	Methods, cohort selection and outcome definition (page 5)
	5c	D;V	Give details of treatments received, if relevant.	N/A
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Methods, cohort selection and outcome definition (page 5)
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Methods, predictor variables (page 5-6)

	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	Figure 1, flowchart of sample inclusion/exclusion criteria
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Supplemental methods
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	Supplemental methods
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	Methods, statistical analysis (page 6)
	10c	V	For validation, describe how the predictions were calculated.	Methods, statistical analysis (page 6)
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	Methods, statistical analysis (page 6)
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	Supplemental methods
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Supplemental methods
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Table 1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Supplement Table X
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Results (page 7)
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Figures 2-4; Table 2
	15b	D	Explain how to use the prediction model.	Results (page 8)
Model	16	D;V	Report performance measures (with CIs) for the prediction model.	Table 2

performance				
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as non representative sample, few events per predictor, missing data).	Discussion (page 10)
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	Discussion (page 9-10)
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Discussion (page 9-10)
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	Discussion (page 9-10)
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Supplement
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	N/A, not funding received for the present study.

Table S2.
Outcome definitions and validation criteria

Each of the following outcomes required a Z37.XX ICD-10 code for delivery, along with one of the specified codes on the same hospitalization event, to be classified as an adverse outcome. To ensure external validity, we validated our sample estimate against a national estimate. Links to both the definitions and national estimate sources are provided.

Outcome	ICD-10 Definition criteria*	Definition source	Our estimate	National estimate	National estimate source
Severe maternal mortality and morbidity (SMM)	Z37.xx and one of the following complication codes: 1. I21.xx, I22.x 2. I71.xx, I79.0 3. N17.x, O90.4 4. J80, J95.1, J95.2, J95.3, J95.82x, J96.0x, J96.2x, J96.9x, R06.03, R09.2 5. O88.112, O88.113, O88.119, O88.12, O88.13 6. I46.x, I49.0x 7. 5A12012, 5A2204Z 8. D65, D68.8, D68.9, O45.002, O45.003, O45.009, O45.012, O45.013, O45.019, O45.022, O45.023, O45.029, O45.092, O45.093, O45.099, O46.002, O46.003, O46.009, O46.012, O46.013, O46.019, O46.022, O46.023, O46.029, O46.092, O46.093, O46.099, O67.0, O72.3 9. 30230H0, 30230K0, 30230L0, 30230M0, 30230N0, 30230P0, 30230R0, 30230T0, 30230H1, 30230K1, 30230L1, 30230M1, 30230N1, 30230P1, 30230R1, 30230T1, 30233H0, 30233K0, 30233L0, 30233M0, 30233N0, 30233P0, 30233R0, 30233T0, 30233H1, 30233K1, 30233L1, 30233M1, 30233N1, 30233P1, 30233R1, 30233T1, 30240H0, 30240K0, 30240L0, 30240M0, 30240N0, 30240P0, 30240R0, 30240T0, 30240H1, 30240K1, 30240L1, 30240M1, 30240N1, 30240P1, 30240R1, 30240T1, 30243H0, 30243K0, 30243L0, 30243M0, 30243N0, 30243P0, 30243R0, 30243T0, 30243H1, 30243K1, 30243L1, 30243M1, 30243N1, 30243P1, 30243R1, 30243T1 10. O15. X 11. I97.120, I97.121, I97.130, I97.131, I97.710, I97.711 12. A81.2, G45.x, G46.x, G93.49, H34.0x, I60.xx, I61.xx, I62.xx, I63.00, I63.01x, I63.1xx, I63.2xx, I63.3xx, I63.4xx, I63.5xx, I63.6, I63.8x, I63.9, I65.xx, I66.xx, I67.xx, I68.xx, O22.50, O22.52, 13. O22.53, I97.810, I97.811, I97.820, I97.821, O87.3 14. I50.1, I50.20, I50.21, I50.23, I50.30, I50.31, I50.33, I50.40, I50.41, I50.43, I50.810, I50.811, I50.813, I50.814, I50.82, I50.83, I50.84, I50.89, I50.9, J81.0 15. O29.112–O29.119, O29.122–O29.129, O29.192–O29.199, O29.212–O29.219, O29.292–O29.299, O74.0, O74.1, O74.2, O74.3, O89.0x, O89.1, O89.2, T88.2XXA, T88.3XXA	https://www.cdc.gov/maternal-infant-health/php/severe-maternal-morbidity/icd.html	2.36% of all deliveries were associated with an SMM diagnosis.	2.36%	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9375335/

	16. A32.7, A40.x, A41.x, I76, O85, O86.04, R65.20, R65.21, T81.12XA, T81.44XA 17. O75.1, R57.x, T78.2XXA, T81.10XA, T81.11XA, T81.19XA, T88.6XXA 18. D57.00, D57.01, D57.02, D57.211, D57.212, D57.219, D57.411, D57.412, D57.419, D57.811, D57.812, D57.819 19. I26.x, O88.012–O88.03, O88.212–O88.23, O88.312–O88.33, O88.812–O88.83, T80.0XXA 20. 0UT90ZL, 0UT90ZZ, 0UT97ZL, 0UT97ZZ 21. 0B110F4, 0B113F4, 0B114F4 22. 5A1935Z, 5A1945Z, 5A1955Z				
Adverse pregnancy outcome (APO)	Z37.xx and one of the following codes: 1. O42.xx 2. O11.xx, O14.xx, O15.xx 3. O13.xx, O16.xx 4. O99.81, O24.4x 5. O36.5x 6. O45.xx 7. Z37.1x, Z37.3x, Z37.4x, Z37.7x 8. O60.1x	https://pmc.ncbi.nlm.nih.gov/articles/PMC8520131/ https://pmc.ncbi.nlm.nih.gov/articles/PMC5266622/	35.7% of all deliveries were associated with an APO diagnosis.	42.6%	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8520131/
Neonatal intensive care (NICU) admissions	Z37.xx and one of the following codes: CPT codes: 1. P0720, P0721, P0722, P0723, P0724, P0725, P0726 2. P0730, P0731, P0732, P0733, P0734, P0735, P0736, P0737, P0738, P0739 3. P0701, P0702, P0703, P071, P0714, P0715, P0716, P0717, P0718 4. P050, P0500, P0502, P0503, P0504, P0505, P0506, P0507, P0508, P0509, P051, P0510, P0511, P0512, P0513, P0514, P0515, P0516, P0517, P0518 5. 99468, 99469, 99477, 99478, 99479, 99480 Revenue codes: 6. 0172, 0173, 0174, 0179	https://pmc.ncbi.nlm.nih.gov/articles/PMC10916318/#SD1	11.6% of all deliveries were associated with a NICU admission	9.6% 10.2%	https://pubmed.ncbi.nlm.nih.gov/38012436/ https://www.medicaid.gov/medicaid/quality-of-care/downloads/mih-beneficiary-profile.pdf

Table S3.
Predictor variables and definitions

Predictor Type	Measure	Definition
Demographic	State	Defined as a member's state of residence.
	Age	Defined as the age in 2017, measured in years
	Sex	Defined as male, female, or missing
	Race/ethnicity	Defined as Asian, White, Multiracial, Black, Hispanic, Native American, Hawaiian, or Missing
	Disabled	Defined as disabled or non-disabled. Members were classified as having a disability if either the Demographic and Eligibility file or the Disability and Need Supplement file indicated a disability.
Clinical history	Number of clinical conditions within each CCSR category, represented as separate columns	Clinical conditions are defined by using ICD-10 diagnostic codes, which are crosswalked to their corresponding Clinical Classification Software Refined (CCSR) categories. Each CCSR category is one-hot encoded into a unique column and summed to provide the total count of conditions for each category.
	Number of procedure codes within each BETOS category, represented as separate columns	Defined by crosswalking Current Procedural Terminology (CPT) codes to the Restructured Berenson-Eggers Type of Service (BETOS) System codes. Each BETOS category is one-hot encoded into a unique column and summed to provide the total count of procedures for each category.
	Number of NDC codes within each drug class category, represented as separate columns	Defined by crosswalking NDC codes to drug classes using the CMS Prescription Drug Data Collection. Each drug class category is one-hot encoded into a unique column and summed to provide the total count of medications for each category.
	Number of visits for each clinician type is represented as separate columns	Defined using the Centers for Medicare and Medicaid Services clinician specialty classification. Each clinician specialty category is one-hot encoded into a unique column and summed to calculate the total count for each specialty.
Measures of clinical risk	Number of all-cause ED visits and hospitalizations	Defined ED visits based on Current Procedural Terminology codes, revenue codes, and place-of-service codes. To count "episodes" of care, we linked ED and inpatient claim records for the same patient if dates of service were congruent or contiguous.
	Percentage of all-cause ED visits and hospitalizations for non-emergent conditions	We defined non-emergent ED visits as those meeting the New York University ED Patch Algorithm definition. Non-emergent inpatient admissions were defined using the Agency for Healthcare Research and Quality's Prevention Quality Indicator.
	Slope of all-cause ED visits and hospitalizations	Defined as the monthly rate of change for all-cause ED visits and hospitalizations across each month in 2017, representing the trend or trajectory over the year.
	Slope of non-emergent ED visits and hospitalizations	Defined as the monthly rate of change for non-emergent ED visits and hospitalizations across each month in 2017, representing the trend or trajectory over the year.

	Slope of number of medication prescriptions	Defined as the monthly rate of change for the number of medication prescriptions across each month in 2017, representing the trend or trajectory over the year.
	Number of hospitalization days	Defined as the total number of hospitalization days in 2017.
	Number of long-term care episodes	Defined as the total number of long-term care episodes in 2017.
	Days supply of medication	Defined as the total days supplied across all medications in 2017.
	Number of unique medications	Defined by the number of distinct National Drug Codes (NDC) present in a patient's medication history, representing each unique prescription.
	Total number of medication prescriptions	Defined by the total number of National Drug Codes (NDC) present in a patient's medication history, representing all prescriptions.
	Medication adherence	For each patient, we calculated the difference between the first and last prescription fill for each unique medication (based on the NDC code), divided by the total days supplied. We then took the mean of these values across all medications to provide an average measure for each patient.
Individual SDoH	Household size	Defined as single, 2-5 individuals, 6 or more individuals, or missing
	Household income	Defined as 0-100% or federal poverty level, 100-200%, 200% or more, or missing
	Proficient in English	Defined as yes, no, or missing
	Married	Defined as yes, no, or missing
	US citizen	Defined as yes, no, or missing
	Receipt of Supplement security income (SSI)	Defined as yes, no, or missing
	Receipt of social security disability insurance (SSDI)	Defined as yes, no, or missing
	Receipt of temporary assistance for needy families (TANF)	Defined as yes, no, or missing
County-level SDoH	Substance use services facility availability	Number of substance use services facilities accepting Medicaid per 1,000 population
	Substance use treatment facility availability	Number of substance use treatment facilities providing medication
	Mental health services facility availability	Number of mental health services facilities accepting Medicaid per 1,000 population
	Urgent care availability	Number of urgent care facilities per 1,000 population
	Physician assistant and nurse practitioner availability	Number of APRNs/PAs per 1,000 population
	Primary care physician availability	Defined as no shortage, partial county, or whole county, as defined by Health Professional Shortage Area (HPSA)
	Mental health provider availability	Defined as no shortage, partial county, or whole county, as defined by Health Professional Shortage Area (HPSA)
	Community mental health center availability	Number of community mental health centers per 1,000 population

	Federally quality health center availability	Number of federally qualified health centers per 1,000 population
	Rural health clinic availability	Number of rural health clinics per 1,000 population
	Hospitals with obstetrics department availability	Number of hospitals with obstetrics department per 1,000 population
	Percentage of private for-profit hospitals	Percentage of all hospitals that are private for-profit hospitals
	Proximity to urgent care	Median distance in miles to the nearest urgent care, calculated using population weighted tract centroids in the county
	Proximity to emergency department	Median distance in miles to the nearest emergency department, calculated using population weighted tract centroids in the county
	Proximity to hospital with obstetrics department	Median distance in miles to the nearest obstetrics department, calculated using population weighted tract centroids in the county
	Rurality	Population density per square mile
	Poverty rate	Percentage of population with income below poverty ratio under 0.50
	Public assistance rate	Percentage of households with public assistance income or receiving food stamps/SNAP benefits
	High school education rate	Percentage of population with less than high school education (ages 25 and over)
	Injury-related death rate	Number of deaths from injury per 100,000 population
	Opioid-related overdose rate	Number of drug overdose deaths involving any opioid per 100,000 population
	Air quality	Percentage of days with good air quality
	Heat index	Total number of days with daily maximum heat index, absolute threshold: 100°F

Table S4.
SDoH improvement sensitivity analysis

The baseline rate represents the mean rate across our patient sample (n=190,698). The adjusted rate reflects the mean rate after accounting for transformations in SDoH values. We hypothesized that patients living closer to healthcare resources or in areas with fewer drug overdose deaths or lower poverty rates would experience fewer adverse events. Additionally, areas with higher healthcare resources were expected to have fewer adverse events. Table 1 in the main text presents the county-level (n=1,325) mean, median, and interquartile range.

	Baseline rate mean (median, 25th, 75th)	Hypothesized adjustment to reduce adverse events	Adjusted rate mean (median, 25th, 75th)	Relative risk reduction estimates (% reduction in adverse events)	Cumulative population impact projection (% reduction in adverse events)
Number of APRNs/PAs per 1,000 population	2.3 (1.8, 1.4, 2.7)	75th percentile	3.0 (2.7, 2.7, 2.7)	3.21	3.21
Number of drug overdose deaths involving any opioid per 100,000 population	23.0 (20.8, 15.4, 27.4)	25th percentile	14.9 (15.4, 15.4, 15.4)	2.43	6.20
Percentage of population with income below poverty ratio under 0.50	6.5 (6.1, 5.0, 7.5)	25th percentile	4.7 (5.0, 5.0, 5.0)	1.63	8.27
Number of substance use services facilities accepting Medicaid per 1,000 population	0.034 (0.0, 0.0, 0.05)	75th percentile	0.04 (0.04, 0.04, 0.042)	1.47	11.37
Number of rural health clinics per 1,000 population	0.1 (0.0, 0.0, 0.0)	75th percentile	0.09 (0.05, 0.05, 0.05)	1.37	15.22
Median distance to hospital delivering obstetric care	4.5 (2.9, 2.2, 4.4)	25th percentile	2.1 (2.2, 2.2, 2.2)	1.35	17.61
Percentage of all hospitals that are private for-profit hospitals	61.1 (60.3, 54.1, 68.8)	25th percentile	52.2 (54.1, 54.1, 54.1)	1.27	21.56
Number of community mental health centers per 1,000 population	0.0 (0.0, 0.0, 0.0)	75th percentile	0.1 (0.1, 0.1, 0.1)	1.05	23.38
Number of federally qualified health centers per 1,000 population	0.04 (0.03, 0.01, 0.04)	75th percentile	0.1 (0.04, 0.04, 0.04)	0.72	24.96
Number of hospitals with obstetrics department per 1,000 population	0.019 (0.01, 0.01, 0.02)	75th percentile	0.02 (0.01, 0.01, 0.02)	0.68	25.68

Median distance in miles to the nearest urgent care, calculated using population weighted tract centroids in the county	8.1 (2.5, 1.8, 8.8)	25th percentile	1.7 (1.8, 1.8, 1.8)	0.36	26.37
Median distance in miles to the nearest emergency department, calculated using population weighted tract centroids in the county	3.5 (2.5, 1.9, 3.6)	25th percentile	1.8 (1.9, 1.9, 1.9)	0.33	30.58
Number of mental health services facilities accepting Medicaid per 1,000 population	0.03 (0.020, 0.013, 0.037)	75th percentile	0.05 (0.04, 0.04, 0.04)	0.11	31.60
Mental health provider availability as defined by HPSA as no shortage, partial county, or whole county	no shortage - 5.7% partial - 34.3% whole - 60.0%	Improve each county to the next highest category (eg, partial shortage to no shortage)	no shortage - 40.0% partial - 60.0% whole - 0%	0.09	31.74
Primary care availability as defined by HPSA as no shortage, partial county, or whole county	no shortage - 7.6% partial - 30.6% whole - 61.7%	Improve each county to the next highest category (eg, partial shortage to no shortage)	no shortage - 38.2% partial - 61.8% whole - 0%	0.04	31.84

Table S5.
Prevalence of SMM and APO submeasures

Submeasures do not add up to 100%, as an individual could have multiple diagnosis on the delivery episode

Severe maternal morbidity and mortality (n=4,154)	**Count**	**Percentage**
Blood Transfusion	1936	46.6
Acute Respiratory Distress Syndrome	524	12.6
Ventilation	520	12.5
Eclampsia	411	9.9
sepsis	402	9.7
Pulmonary Edema / Acute Heart Failure	240	5.8
Puerperal Cerebrovascular Disorders	147	3.5
Puerperal Cerebrovascular Disorders	147	3.5
Disseminated Intravascular Coagulation	138	3.3
shock	127	3.1
Air and Thrombotic Embolism	116	2.8
Hysterectomy	89	2.1
Conversion of Cardiac Rhythm	44	1.1
Amniotic Fluid Embolism	41	1
Sickle Cell Disease With Crisis	43	1
Cardiac Arrest / Ventricular Fibrillation	24	0.6
Acute Myocardial Infarction	20	0.5
Severe Anesthesia Complications	17	0.4
Aneurysm	7	0.2
Temporary Tracheostomy	4	0.1
Heart Failure / Arrest During Surgery or Procedure	0	0
Adverse pregnancy outcomes (n=68,922)	**Count**	**Percentage**
Premature rupture of membranes	19402	28.2

Gestational hypertension	16139	23.4
Gestational diabetes	14467	21.0
Preeclampsia	13844	20.1
Spontaneous preterm labor	12465	18.1
Poor fetal growth	9796	14.2
Placental abruption	2497	3.6
Stillbirth	1921	2.8

Table S6.
Gestational age during first pregnancy visit

Gestational age during first pregnancy visit among individuals with an adverse outcome.

Gestational age (weeks)	count	percentage
< 8	17124	23.78
8	5064	7.03
9	3749	5.21
10	3303	4.59
11	2945	4.09
12	3719	5.16
13	2562	3.56
14	1687	2.34
15	1527	2.12
16	1621	2.25
17	1323	1.84
18	1472	2.04
19	1753	2.43
20	2459	3.41
21	1365	1.9
22	1013	1.41
23	855	1.19
24	866	1.2
25	748	1.04
26	715	0.99
27	745	1.03
28	897	1.25

29	699	0.97
30	809	1.12
31	744	1.03
32	950	1.32
33	913	1.27
34	992	1.38
35	1132	1.57
29	699	0.97
30	809	1.12
31	744	1.03
32	950	1.32
33	913	1.27
34	992	1.38
35	1132	1.57
36	1503	2.09
37	1642	2.28
38	1715	2.38
39	1918	2.66
40	813	1.13
41	160	0.22
42	13	0.02
>48	1	0

Table S7.
Standardized mean differences across clinical and social risk factors

We measure the standardized mean differences between patients with and without an adverse event across all model features, considering a standardized mean difference of 0.10 or lower as indicating no meaningful differences.

The file is located: https://github.com/sadiqypatel/Pregnancy_Risk_Model

- ICD10 code groupers (for diagnoses) can be found:
  - https://hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp
- CPT code groupers (for types of care) can be found:
  - https://data.cms.gov/provider-summary-by-type-of-service/provider-service-classifications/restructured-betos-classification-system
- NDC code groupers (for medications) can be found:
  - https://www.cms.gov/cciio/programs-and-initiatives/other-insurance-protections/prescription-drug-data-collection
- Details regarding Area-level SDOH measures can be found in Materials and Methods Section

Table S8.
Model performance metrics

		AUC	Accuracy	F1	MCC	NPV	PPV	Sensitivity	Specificity
non-SDoH model Logistic regression	Composite	0.579 (0.578-0.580)	0.633 (0.632-0.634)	0.134 (0.133-0.135)	0.107 (0.106-0.108)	0.634 (0.633-0.635)	0.613 (0.610-0.616)	0.075 (0.074-0.076)	0.971 (0.970-0.972)
	NICU	0.861 (0.860-0.862)	0.965 (0.964-0.966)	0.047 (0.045-0.049)	0.116 (0.112-0.119)	0.966 (0.965-0.967)	0.593 (0.578-0.609)	0.025 (0.024-0.026)	0.999 (0.998-1.000)
	APO	0.583 (0.582-0.584)	0.651 (0.650-0.652)	0.114 (0.113-0.115)	0.101 (0.100-0.103)	0.653 (0.652-0.654)	0.603 (0.600-0.607)	0.063 (0.062-0.064)	0.977 (0.976-0.978)
	SMM	0.651 (0.649-0.653)	0.978 (0.977-0.979)	0.034 (0.031-0.036)	0.098 (0.092-0.103)	0.979 (0.978-0.980)	0.585 (0.557-0.612)	0.017 (0.016-0.019)	0.999 (0.998-1.000)
SDoH model Xgboost	Composite	0.954 (0.953,0.955)	0.894 (0.893, 0.895)	0.853 (0.852, 0.854)	0.773 (0.772, 0.774)	0.893 (0.892, 0.894)	0.896 (0.895, 0.897)	0.813 (0.812, 0.814)	0.943 (0.942, 0.944)
	NICU	0.971 (0.970-0.972)	0.989 (0.987-0.989)	0.816 (0.814-0.817)	0.822 (0.820-0.823)	0.989 (0.987-0.990)	0.702 (0.699-0.704)	0.702 (0.699-0.704)	0.999 (0.998-1.000)
	APO	0.917 (0.916-0.919)	0.900 (0.899-0.901)	0.852 0.851-0.853)	0.779 (0.778-0.780)	0.898 (0.897-0.899)	0.903 (0.902-0.904)	0.806 (0.805-0.807)	0.952 (0.951-0.953)
	SMM	0.953 (0.952-0.954)	0.994 (0.993-0.995)	0.831 (0.830-0.832)	0.839 (0.837-0.841)	0.994 (0.993-0.995)	0.991 (0.990-0.992)	0.716 (0.712-0.719)	0.999 (0.998-1.000)
non-SDoH model Xgboost	Composite	0.931 (0.930-0.932)	0.863 (0.862-0.864)	0.797 (0.796-0.798)	0.707 (0.706-0.708)	0.846 (0.845-0.847)	0.903 (0.902-0.904)	0.713 (0.712-0.714)	0.954 (0.953-0.955)
	NICU	0.965 (0.964-0.966)	0.987 (0.986-0.989)	0.778 (0.776-0.780)	0.790 (0.788-0.793)	0.987 (0.986-0.988)	0.985 (0.984-0.986)	0.643 (0.640-0.646)	0.999 (0.998-1.000)
	APO	0.930 (0.929-0.931)	0.869 (0.868-0.870)	0.793 (0.792-0.794)	0.712 (0.711-0.713)	0.854 (0.853-0.855)	0.909 (0.908-0.910)	0.704 (0.703-0.705)	0.961 (0.960-0.962)
	SMM	0.902 (0.901-0.904)	0.992 (0.991-0.993)	0.786 (0.784-0.789)	0.801 (0.799-0.804)	0.992 (0.991-0.993)	0.997 (0.996-0.998)	0.649 (0.646-0.653)	0.999 (0.998-1.000)

Table S9.
Constellation of social and clinical risk factors

The table is organized by measure type, with each row describing the percentage of members experiencing risk for each outcome measure.

Measure	Measure type	No adverse event	Any adverse event	APO	SMM	NICU
Percentage of members residing in counties in bottom quartile of air quality	Area SDoH	21.5	20.7	19.5	26.8	51.6
Percentage of members residing in counties in top quartile for distance to emergency department	Area SDoH	24.0	23.4	24.0	20.6	9.5
Percentage of members residing in counties in top quartile for distance to obstetric care distance	Area SDoH	25.2	24.4	25.0	21.6	9.1
Percentage of members residing in counties in bottom quartile of mental health facility rate	Area SDoH	25.1	24.5	23.5	28.7	48.5
Percentage of members residing in counties in bottom quartile of APRN rate	Area SDoH	33.4	32.0	30.8	35.1	65.0
Percentage of members residing in counties in bottom quartile of FQHC rate	Area SDoH	33.7	32.5	31.8	34.8	50.9
Percentage of members residing in counties in bottom quartile of rural health clinic rate	Area SDoH	54.8	56.7	56.8	55.1	58.1
Percentage of members residing in counties in bottom quartile of community mental health center rate	Area SDoH	69.2	67.6	67.6	67.6	82.7
Percentage of members residing in counties with whole county primary care shortage areas	Area SDoH	85.1	85.1	85.5	85.9	93.5
Percentage of members residing in counties in bottom quartile of urgent care facility rate	Area SDoH	85.8	85.8	85.6	87.1	92.0
Percentage of members residing in counties in bottom quartile of substance use facility rate	Area SDoH	100.0	100.0	100.0	100.0	100.0
Percentage of members with a bipolar diagnosis	Behavioral health	0.5	0.5	0.5	0.3	0.3
Percentage of members with a stress disorder diagnosis	Behavioral health	1.0	1.2	1.2	1.0	0.7

Percentage of members with an anxiety or depression diagnosis	Behavioral health	4.1	4.4	4.5	4.4	3.5
Percentage of members with at least 1 self harm diagnosis	Behavioral health	0.1	0.2	0.2	0.1	0.0
Percentage of members with at least 1 opioid use disorder diagnosis	Behavioral health	0.4	0.4	0.4	0.4	0.3
Percentage of members with at least 1 diabetes diagnosis	Clinical	0.3	0.7	0.7	0.7	1.1
Percentage of members with at least 1 obesity diagnosis	Clinical	1.6	2.3	2.3	1.9	2.9
Percentage of members with at least 1 hypertension diagnosis	Clinical	0.9	2.5	2.6	3.8	1.9
Percentage of members with at least 1 chest pain diagnosis	Clinical	2.8	3.2	3.2	3.2	3.6
Percentage of members with at least 1 medication prescription	Clinical	11.7	11.0	11.0	10.5	10.4
Percentage of members with at least 1 pre-diabetes diagnosis	Clinical	0.1	0.1	0.1	0.1	0.1
Percentage of members with at least 1 mild anemia diagnosis	Clinical	0.1	0.2	0.2	0.2	0.2
Percentage of members with at least 1 mild hypertension diagnosis	Clinical	0.2	0.4	0.4	0.5	0.3
Percentage of members with at least 1 nutrition deficiency diagnosis	Clinical	2.5	2.5	2.6	3.8	2.9
Percentage of members with a positive slope for medication prescriptions	Clinical	34.3	34.2	34.2	35.0	35.0
Percentage of members with at least 1 all-cause acute care visit	Clinical	19.7	20.5	20.5	22.6	19.6
Percentage of members with a positive slope for all-cause acute care visits	Clinical	18.8	20.5	20.5	22.5	20.0
Percentage of members with at least 1 injury-related diagnosis	Domestic violence	6.7	7.0	7.1	7.1	6.2
Percentage of members with at least 1 family history diagnosis	Family history	0.1	0.1	0.1	0.1	0.1
Percentage of members with at least 1 avoidable acute care visit	Primary care access	16.1	17.0	17.1	19.5	16.2

Percentage of members with a positive slope for avoidable acute care visits	Primary care access	12.6	13.9	13.9	15.0	13.4
Proportion of White patients (>1 means overrepresented relative to sample size; 1 means proportional)	Race/ethnicity	1.0	1.0	1.0	0.8	0.8
Proportion of Hispanic patients (>1 means overrepresented relative to sample size; 1 means proportional)	Race/ethnicity	1.0	1.0	0.9	1.0	1.6
Proportion of Black patients (>1 means overrepresented relative to sample size; 1 means proportional)	Race/ethnicity	0.9	1.1	1.1	1.2	0.5
Percentage of members with at least 1 SDoH-related diagnosis	Social risk	0.2	0.3	0.3	0.3	0.2

Table S10.
Model performance by race/ethnicity

The minority category included individuals identifying as Asian, Multiracial, Native American, or Hawaiian.

	SDoH model (xgboost)		non-SDoH model (xgboost)	
Race	**sensitivity**	**specificity**	**sensitivity**	**specificity**
minority	84.6 (83.8, 85.2)	93.3 (93.2, 94.0)	0.770 (0.765, 0.781)	0.932 (0.930, 0.938)
White	81.6 (81.0, 81.8)	95.3 (95.2, 95.4)	0.728 (0.725, 0.737)	0.942 (0.940, 0.944)
Hispanic	76.5 (75.9, 76.7)	95.0 (94.9, 95.2)	0.686 (0.682, 0.692)	0.964 (0.962, 0.965)
Black	81.1 (80.9, 81.8)	92.7 (92.4, 92.9)	0.710 (0.709, 0.715)	0.961 (0.960, 0.963)