

Timely Preterm-Birth Prediction Among Pregnant Women in Medicaid Without Preterm-Birth History

Xiao Wang, PhD; Luyu Xie, PharmD; Yolande M. Pengetnze, MD; Weiwei Ouyang, PhD; Hannah K. Ligon, MHA; Donna I. Persaud, MD; Sarah E. Messiah, PhD; and Barry Lachman, MD

Every year, more than half a million US babies (~1 in 10) are born prematurely,^{1,2} incurring annual societal costs of more than \$26 billion.³ Preterm birth (PTB), specifically early PTB (<35 weeks gestational age [GA]), is associated with lifelong medical and developmental comorbidities.⁴ US PTB rates decreased slightly from 2007 to 2013, then increased from 2014 to 2016. African Americans, Latinos, and socioeconomically disadvantaged populations are disproportionately affected.⁵

Population-level surveillance and interventions are needed to substantially curb PTB rates. Timely prediction of PTB could support scalable population-level interventions in large health systems and managed care or accountable care organizations. Published PTB prediction models, however, fail to leverage scalable data sources such as claims or publicly available community-level data sets. Current approaches to PTB prediction rely heavily on low-scale interventions during office visits, including clinical risk assessment of prior PTB, maternal age, pregnancy complications, and other obstetric history.⁶ Nontraditional risk factors, including educational achievement, body mass index, socioeconomic status, and mental and behavioral factors, are also known predictors⁷⁻¹¹ but are not well codified for routine incorporation into PTB prediction. Maternal stress, for instance, is associated with approximately 60% of PTBs, predominantly in socioeconomically disadvantaged populations,¹² but not routinely screened for. Many published PTB prediction models also rely on scarcely available biological data (eg, biomarkers, sonography).¹³⁻¹⁶

Although most PTBs occur among women with no history of PTB (including primigravid patients), published models fail to predict PTB in this subpopulation.¹⁷⁻¹⁹ Most models also do not report prospective or external validation, nor the timing of risk prediction, to gauge generalizability and clinical usefulness. In a systematic review by Meertens et al,¹⁷ none of the published PTB prediction models were accurate upon external validation. Similarly, Leneuve-Dorilas et al¹⁹ reported insufficient sensitivity for PTB prediction in external validation.

We therefore sought to build and prospectively validate a novel prediction model incorporating socioeconomic, clinical, and

ABSTRACT

OBJECTIVES: To develop and prospectively validate a novel model incorporating claims and community-level socioeconomic data to predict preterm birth at scale among pregnant Medicaid women with no history of preterm birth (PTB).

STUDY DESIGN: A longitudinal Texas Medicaid cohort study, with 2-year retrospective model building (October 2015–October 2017) and a 1-year prospective model validation phase (January 2018–December 2018).

METHODS: Inclusion criteria were females aged 11 to 55 years with at least 1 live singleton birth and no history of PTB. The primary outcome was live singleton birth earlier than 35 weeks. Covariates were medical/mental/behavioral comorbidities, obstetric history, sociodemographic characteristics, and health services utilization. Of multiple models built, the most parsimonious was selected to classify pregnancies as very high, high, medium, and low risk. Model performance was evaluated using positive predictive value (PPV), sensitivity, case identification ratio (1/PPV), and timing of prediction.

RESULTS: The model was built on 6689 pregnancies and validated on 7855 pregnancies. PTB rate earlier than 35 weeks was approximately 3.3%. Significant risk predictors included prenatal visit attendance, insurance gap days, and medical/obstetrical comorbidities. Model PPV was approximately 4-fold higher for very high-risk women (14.7% vs cohort [3.3%]) and so was the case identification ratio (1:7 vs 1:30, respectively). Sensitivity was good, with 57% of PTBs classified as medium risk or higher. Timing of prediction was clinically relevant, with more than 80% of PTBs risk stratified before 24 weeks.

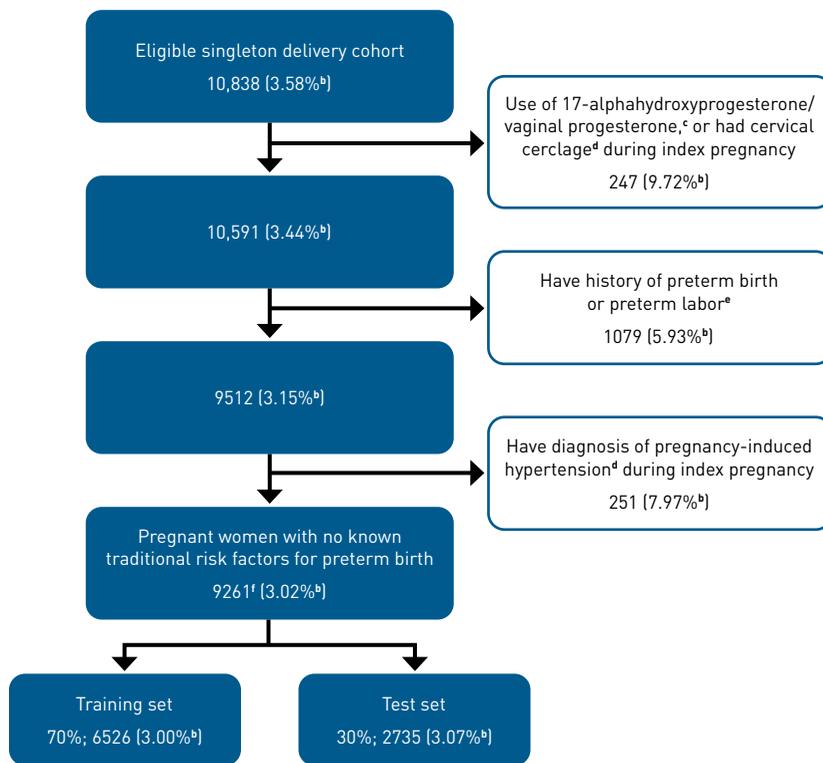
CONCLUSIONS: We report a novel PTB prediction model among pregnant Medicaid women without PTB history, which is timely, accurate, practical, and scalable. We leverage Medicaid and community data readily accessible by Medicaid plans to support population-level interventions to prevent PTBs.

Am J Manag Care. 2021;27(5):e145-e151. doi:10.37765/ajmc.2021.88636

TAKEAWAY POINTS

- A model is developed that accurately predicts preterm birth (PTB) among pregnant women in Medicaid without PTB history, to guide preventive interventions at scale.
- ▶ PTB is costly, and prediction is needed for preventive interventions among pregnant women without PTB history, representing approximately three-fourths of all PTBs.
- ▶ Published PTB prediction models fail to leverage readily available, scalable Medicaid and community data sets.
- ▶ A prediction model is developed and prospectively validated that incorporates socioeconomic and clinical data to accurately predict PTB at scale among pregnant women in Medicaid without PTB history.
- ▶ The model is accurate, timely, practical, and replicable at Medicaid managed care organizations for PTB risk-mitigating interventions at scale.

FIGURE 1. Retrospective Cohort Selection^a



CPT, Current Procedural Terminology; ICD-9, *International Classification of Diseases, Ninth Revision*; ICD-10, *ICD, Tenth Revision*; PCHP, Parkland Community Health Plan; PTB, preterm birth.

^aThe retrospective cohort was selected by eliminating pregnancies with traditional PTB risk factors or ongoing PTB prevention treatment. The final cohort included only pregnant women with no known traditional PTB risk factor and no ongoing treatment for PTB prevention at the time of prediction.

^bThe number in parentheses is the PTB rate earlier than 35 weeks' gestational age for the cohort subset.

^cUse of 17-alpha-hydroxyprogesterone and vaginal progesterone are defined as any filled prescription for 17-alpha-hydroxyprogesterone or vaginal progesterone as recorded in the PCHP pharmacy claims during the index pregnancy.

^dCervical cerclage and pregnancy-induced hypertension are defined by appropriate CPT and ICD-9/ICD-10 codes as recorded in the PCHP medical claims during the index pregnancy. [See eAppendix for CPT and ICD codes details.]

^eHistory of preterm labor or delivery is defined by appropriate ICD-9/ICD-10 codes as recorded in the PCHP medical claims during a previous pregnancy. [See eAppendix for CPT and ICD codes details.]

^fThe modeling sample of 9261 pregnancies includes duplicate pregnancies due to repeated sampling at different time periods, for a total of 6689 unique pregnancies.

health services utilization variables from large-scale data sets to predict PTB (<35 weeks) among pregnant women with no history of PTB, including primigravid women. We combined claims, membership, and readily available community data sets for a large Texas Medicaid population to develop a comprehensive set of clinical and community-level socioeconomic variables for PTB risk prediction.

METHODS

Study Design

This was a longitudinal cohort study of pregnant women members of the Parkland Community Health Plan (PCHP), a large, provider-owned Medicaid health maintenance organization in north Texas.

The study was conducted in 2 phases: a 2-year retrospective model building phase (October 2015-September 2017) (Figure 1) and a 1-year prospective model validation phase (January 2018-December 2018).

Cohort selection (Figure 1). Inclusion criteria were females aged 11 to 55 years with at least 1 live singleton birth documented in the PCHP claims data set during the designated study period. Exclusion criteria were (1) history of preterm labor or birth, (2) diagnosis of pregnancy-induced hypertension during the index pregnancy, (3) ongoing treatment with 17-alpha-hydroxyprogesterone or vaginal progesterone, and (4) documentation of cerclage during the index pregnancy.

Data sources were PCHP medical and pharmacy claims, membership and eligibility data sets, and the 2010 US Census data.²⁰

Key Variables Definition

Live singleton birth was defined by appropriate *International Classification of Diseases, Ninth Revision (ICD-9)* and *Tenth Revision (ICD-10)* codes (eAppendix [available at ajmc.com]) as documented in the PCHP claims data sets.

Index pregnancy was defined as the pregnancy for which a prediction was performed.

GA was defined as the number of completed weeks during pregnancy. GA at delivery was the number of completed weeks of gestation as defined by the delivery ICD-9/ICD-10 code. Where the delivery ICD-9/ICD-10 code had a precise number of completed weeks of gestation,

the precise number was used. Where the delivery ICD-9/ICD-10 code was a range of weeks, the median number of weeks was used. For example, for ICD-10 code P07.35 (32 completed weeks), the GA is 32 weeks; for ICD-9 code 765.26 (31-32 completed weeks), the GA is 31.5 weeks.

Presumed pregnancy start date was calculated retrospectively for every pregnancy by multiplying GA at delivery (in weeks) by 7 to approximate the number of gestational days, then subtracting the latter number from the delivery date to obtain the presumed pregnancy start date. For instance, a delivery on September 1, 2018, with ICD-10 code P07.35 (32 completed weeks) corresponded to 224 (32 × 7) gestational days, therefore the presumed pregnancy start date was January 20, 2018.

Primary Outcome

The primary outcome variable was PTB, defined as any live singleton birth with less than 35 completed weeks of gestation.

Predictors and Covariates

Obstetric history. Obstetric diagnoses and procedures during index and past pregnancies were identified using appropriate ICD-9/ICD-10 and Current Procedural Terminology (CPT) codes (eAppendix). Medication use during index and past pregnancies was derived by searching all or parts of medication names in the pharmacy claims data sets, then manually curating the search results to include only the desired medications.

Health services utilization. Outpatient visits and hospital admissions were identified by appropriate CPT codes (eAppendix). Prenatal visits were identified using Healthcare Effectiveness Data and Information Set criteria.²¹

Sociodemographic. Home address zip code was defined as the zip code of the patient's last home address recorded in the PCHP eligibility data set during the study period. For every patient, the number of distinct home addresses in the 12 months preceding prediction was computed as an indicator of housing instability.^{22,23} Socioeconomic variables from the 2010 US Census database²⁴ were aggregated at the zip code level and assigned to each patient's home address zip code. The zip code–level mean income ratio was calculated by dividing zip code–level mean income (2010 Census²⁴) by \$38,000, a reference value representing the upper limit of income for low-income families in the Dallas–Fort Worth area.²⁵ Insurance gap days were calculated from the PCHP membership data set as the number of days without insurance coverage during the 12 months preceding the time of prediction.

Modeling Approach

In the retrospective phase, multiple prediction models were built to predict PTB risk within 6 months of prediction and their accuracy evaluated.

Three cohorts of eligible pregnant women were sampled at 3 distinct time periods (October 2016, January 2017, and April 2017), spanning multiple seasons to account for seasonality in PTB risk.^{26,27}

For each pregnancy, a 12-month retrospective look-back period was scanned to determine the presumed pregnancy start date and collect data on candidate predictor variables. A 6-month look-ahead period was used for prediction. Some pregnancies were sampled more than once at different time periods.

Model Building and Testing

The modeling cohort (Figure 1) was split in a 70:30 ratio into training and test data sets. Resampling techniques were applied to balance the PTB prevalence rate in the 5% to 20% range in the training data set.

Using published literature and clinical expertise, 52 candidate predictors were preselected, including medical, mental, and behavioral comorbidities; obstetric history; current pregnancy-related diagnoses; socioeconomic indicators; demographic characteristics; medication use; and health services utilization (eAppendix Table 1). Additionally, patterns of variable missingness were assessed for inclusion into the model.

Modeling approaches. In univariable analysis, 26 predictors significantly associated with PTB ($P < .05$) were entered for multivariable model selection. Variables with P values less than .2 or that contributed to model performance were retained in the multivariable model. Logistic regression, multivariate adaptive regression spline, random forest, XGBoost, and neural networks were used for multivariable modeling. Models were evaluated using statistical measures of performance, including the receiver operating characteristic (ROC) area under the curve (AUC), precision-recall (PR) AUC, model parsimony, and model explanation. The most parsimonious model with the highest ROC-AUC and PR-AUC was selected, with a preference for explanatory models vs “black box” models, as the former provide meaningful insights to guide clinical interventions.

PTB risk categories. The selected model was applied to the test dataset to predict PTB risk for each pregnancy. The risk probabilities generated were grouped into categories: very high risk, defined as risk scores between the 97.5th and the 100th percentile; high risk, 90th to less than the 97.5th percentile; medium risk, 80th to less than the 90th percentile; and low risk, less than the 80th percentile. Clinically relevant model performance metrics were calculated for each risk category, including the positive predictive value (PPV), sensitivity, and case identification ratio, calculated as 1 divided by the PPV and representing the number of pregnant women who would need to be screened to identify 1 with true PTB risk. Risk categorization and clinical model performance metrics are more intuitive than statistical performance metrics for seamless integration into clinical and population health workflows.

Prospective Model Validation

From January to December 2018, the selected model was applied prospectively every month to eligible pregnancies (ie, validation cohort) and each pregnancy was assigned a risk category. For the purpose of this analysis, pregnancies were assigned the highest risk

TABLE 1. Study Population Characteristics²⁵

Characteristics	Retrospective cohort	Prospective/validation cohort
Number of unique pregnancies	6689	7855
Preterm birth rate (<35 weeks GA), n (%)	220 (3.29%)	262 (3.34%)
Maternal age at preterm delivery in years	25.45	27.08
Mean GA at preterm delivery in weeks	30.97	31.29
Race/ethnicity, %		
African American	0.76%	0.31%
White non-Hispanic	0.30%	0.1%
Hispanic	43.80%	46.02%
Asian or Pacific Islander	3.27%	2.69%
Other	51.59%	50.59%
Gestational diabetes during index pregnancy, n (%)	83 (1.24%)	124 (1.58%)
Threatened abortion during index pregnancy, n (%)	679 (10.15%)	859 (10.94%)
Mean number of addresses over the past 12 months	1.23	1.27
Zip code mean income ratio ^a	1.61	1.62
Mean number of inpatient admissions during index pregnancy	0.19	0.28
Mean number of prenatal/outpatient visits during index pregnancy	3.68	4.2
Mean number of "insurance gap" days in the past 12 months	267.18	293.46

GA, gestational age.

^aZip code mean income ratio was calculated as the ratio of the mean income for the patient's home address zip code to the income threshold for low-income households in Dallas-Fort Worth (\$38,000).²⁵

TABLE 2. Summary of Predictor Variables Retained in the Selected Model²⁵

Feature	Coefficient	P ^a
History of gestational diabetes	1.5726	.0060
Insurance gap days in past year	0.9355	.0384
Threatened abortion during index pregnancy	0.9091	.0002
Missing obstetric history	0.3843	.0545
Number of addresses in past year	-0.3492	.1019
Zip code mean income ratio ^b	-0.1285	.3716
Number of inpatient visits during index pregnancy	0.1014	.0962
Number of prenatal and/or outpatient visits during index pregnancy	-0.0640	.0319

^aModel retention threshold was $P < .2$.

^bZip code mean income ratio was calculated as the ratio of the mean income for the patient's home address zip code to the income threshold for low-income households in Dallas-Fort Worth (\$38,000).²⁵

level predicted during the pregnancy. Prospective PPV, sensitivity, and case identification ratio were calculated for each risk category.

For each pregnancy, timing of first risk prediction was assessed to determine if prediction occurred before 20 or before 24 weeks GA, 2 time frames relevant for timely clinical interventions and representing thresholds for guideline-recommended 17-alpha-hydroxyprogesterone administration and the fetal viability cutoff, respectively.^{6,28,29}

This study was approved by the University of Texas Southwestern Institutional Review Board.

RESULTS

About 3 in 4 (74%; 262/354) PTBs occurred among pregnant women with no PTB history, pregnancy-induced hypertension, or ongoing treatment with progesterone or cerclage. **Table 1**²⁵ summarizes the study population characteristics. The PTB rate was 3.29% in the retrospective cohort (ie, test data set) and 3.34% in the validation cohort. Most pregnant women were Hispanic or "other" race/ethnicity. The mean maternal age at preterm delivery was 27.1 years in the validation cohort. Median GA at preterm delivery was 31 completed weeks.

Model Selection and Components

Of 6689 unique pregnancies, 9261 samplings were made and randomly assigned, in a 70:30 ratio, to the training (6526) and test (2735) data sets. The PTB rate was comparable in both data sets, at approximately 3.3% (Figure 1).

Eight models were built and the most parsimonious with the highest ROC-AUC (0.679) and PR-AUC (0.064) was selected (see **eAppendix Table 2** for performance of all 8 models). The selected model was an explanatory multivariable logistic regression model developed using a bidirectional stepwise approach, retaining variables with a P value less than .2. Retained variables ranged from medical comorbidities and pregnancy-related conditions to prenatal care attendance and zip code-level median income ratio (**Table 2**²⁵). Prenatal visit attendance, number of home addresses in the past 12 months, and zip code median income were protective factors, with regression coefficients of -0.06, -0.35, and -0.13, respectively. The strongest risk factor was history of gestational diabetes during current or prior pregnancies, with a regression coefficient of 1.57. Other risk factors included number of insurance gap days (regression coefficient, 0.94), threatened abortion (regression coefficient, 0.91), missing obstetric history (regression coefficient, 0.38), and number of inpatient admissions (regression coefficient, 0.10). Although zip code-level median income was not significantly associated with PTB ($P = .37$), it was retained because it improved model performance.

Prospective Model Performance Validation

Prospective model PPV was 4-fold higher among very high-risk patients vs the entire cohort (14.7% vs 3.3%, respectively) (**Figure 2 [A]**), translating into a case identification ratio of 1:7

vs 1:30, respectively. The case identification ratio was 1:19 among high-risk patients, then substantially dropped among medium- and low-risk patients. Prospective model sensitivity was fair (Figure 2 [B]), with 42% of PTBs predicted as very high or high risk and a majority of pregnancies (57%) predicted as very high, high, or medium risk during pregnancy.

Timing of Pregnancy Risk Stratification

More than 80% of pregnancies with PTBs were first risk stratified before 24 weeks GA and 62% before 20 weeks GA (data not shown), a relevant timeliness for clinical interventions.

DISCUSSION

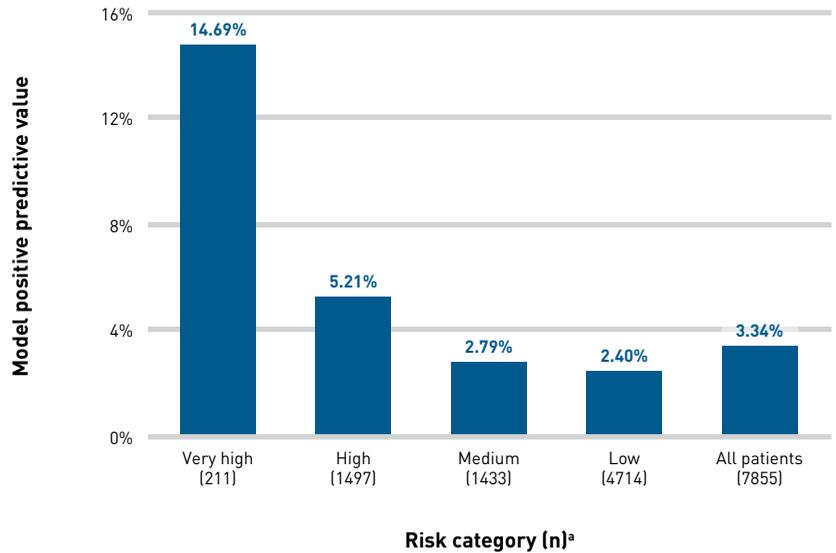
We report a prospectively validated prediction model that uniquely incorporates scalable data sources, such as claims and readily available socioeconomic data, to predict PTB (< 35 weeks) at scale among pregnant women with no history of PTB. PTB rates before 35 weeks in our cohorts are comparable with nationally reported rates.³ Preventing PTB among women with no history of PTB is crucial, as they represent 3 in 4 PTBs in our cohorts. Socioeconomic indicators are incorporated alongside clinical risk factors to create a holistic PTB risk profile for each woman. Prenatal visit attendance, gaps in insurance coverage, zip code-level income, home address changes, threatened abortion, and history of gestational diabetes are identified as independent predictors. Clinically relevant model performance metrics such as PPV and sensitivity are good. By quantifying the effect of known and novel clinical and social risk factors within the context of each other, our model provides an opportunity for personalized PTB prevention interventions.

Strengths and Limitations

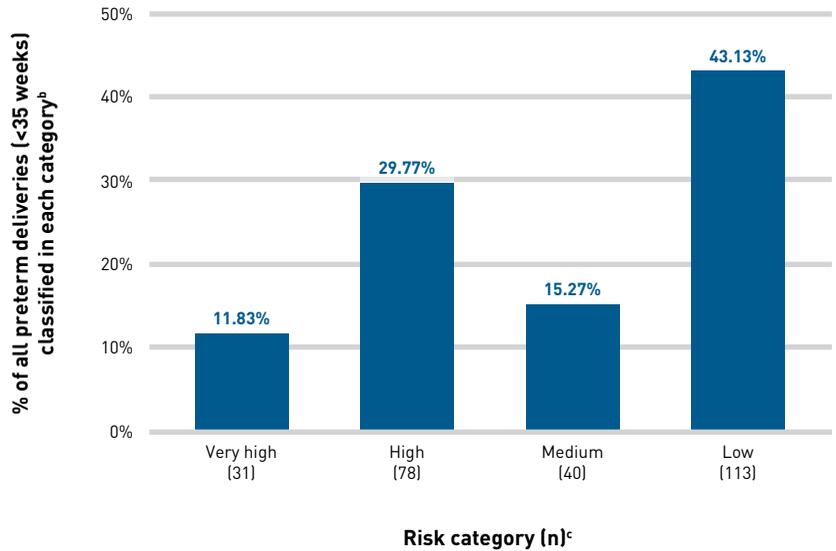
Our model is unique in that we use scalable data sources, such as claims and readily available community-level socioeconomic data, to provide timely and accurate PTB risk prediction among women with no history of PTB, accounting for socioeconomic context. This new model could be implemented seamlessly by payers and capitated health systems to support scalable and targeted, risk-tailored PTB prevention.

FIGURE 2. Prospective Validation of Model Positive Predictive Value and Sensitivity by Risk Category

A. Model Positive Predictive Value by Risk Category in the Prospective Validation Cohort^a



B. Model Sensitivity by Risk Category in the Prospective Validation Cohort^{b,c}



GA, gestational age; PTB, preterm birth.

^aThe model positive predictive value was calculated prospectively in the validation cohort for each PTB risk category. n is the number of pregnancies in the risk category.

^bPercentages were calculated as the number of PTBs before 35 weeks GA in each risk category divided by the total number of PTBs before 35 weeks GA in the entire validation cohort. All percentages add up to 100%.

^cThe model sensitivity was calculated prospectively in the validation cohort for each PTB risk category. The n is the number of PTBs before 35 weeks GA in each risk category.

Our model's high PPV among very high- and high-risk pregnancies translates into higher efficiency of at-risk case identification. For instance, using our risk model, an intervention targeting the very high-risk group would need to be applied to only 7 pregnant women for a chance to prevent 1 PTB (vs 30, if the model is not used), a 4-fold increase in case identification efficiency.

Actionable socioeconomic and clinical risk factors are quantified in the context of each other. For instance, the effect of prenatal care, a well-known protective factor for birth outcomes,³⁰ is quantified by our model within the context of other risk factors. This provides an opportunity for personalized interventions to eliminate barriers to prenatal care, including risk-driven prioritized scheduling, visit reminders, and transportation assistance. Our prediction model also focuses on a subpopulation of pregnant women for whom accurate PTB risk prediction has traditionally been elusive: women with no PTB history, including primigravid women.³¹ Most published risk prediction models fail to accurately predict PTB risk among primigravid women.^{17-19,32} Many studies also fail to incorporate socioeconomic indicators into PTB risk prediction, despite their importance as illustrated by Esplin et al,¹² who identified a distinct phenotype of PTB risk labeled "maternal stress"; for these women, predominant PTB risk factors were low socioeconomic status and mental/behavioral health conditions. Witt and colleagues also reported a 4-fold increase in PTB among women exposed to stress.^{10,11} The predominance of socioeconomic risk factors in this large subpopulation of pregnant women might partly explain the poor success of PTB prevention programs primarily focused on clinical interventions. Our model provides insights into the contribution of socioeconomic indicators (eg, zip code-level median income, insurance gap days, housing instability) to PTB risk, independent of clinical risk. We report that zip code-level median income is inversely associated with PTB risk, consistent with the findings of previous studies reporting an association between neighborhood socioeconomic conditions and health outcomes.³³ Identifying high-risk zip codes could guide geographical targeting for population-level interventions. In our study, housing instability, operationalized as home address changes within the preceding 12 months, is unexpectedly protective of PTB risk. Published studies identify housing instability as a risk factor for poor health outcomes, including low birth weight.³⁴⁻³⁶ It is possible that, in our study population, home address change is a marker of protective health behaviors, such as women moving into a new home during pregnancy to accommodate a growing family or to get closer to a support network prior to delivery. By incorporating socioeconomic indicators into PTB risk prediction, our model could guide targeted preventive interventions addressing social needs alongside clinical needs for at-risk pregnant women. Such a hybrid approach might be more optimal in preventing PTB among pregnant women with a predominance of socioeconomic risk factors through universal screening for social risk (eg, food insecurity, housing and transportation needs) and timely linkage to community resources.

Although not retained in the model, substance use was documented in 9.3% of study participants and associated with a PTB rate of 5.77% (vs 3.03% for no substance use; data not shown). With approximately 1 in 10 pregnant women affected and a 2-fold increase in PTB risk, universal substance use screening in early pregnancy, referrals for evidence-based treatment, and continuous monitoring throughout pregnancy should be considered in PTB prevention programs.

This study has several limitations. Claims data are typically associated with a lag time of 1 to 3 months,³⁷ which might compromise risk prediction timeliness. Moreover, claims data completeness might be affected by inaccurate provider reporting, gaps in insurance coverage, or the nature of data reported. For instance, the impact of prepregnancy maternal weight on PTB risk could not be accurately assessed because body mass index is not routinely reported in claims data.³⁸ Additionally, race and ethnicity were poorly reported in our claims data set, with high numbers of patients classified as "other" race/ethnicity whereas very few were positively identified as Black/African American. Moreover, mental and behavioral health data were missing in many patients due to the segregation of mental/behavioral claims data from medical claims data.³⁹ Abortion history, also, was underreported, with only 1.1% of the cohort affected (vs 19% in published studies⁴⁰), likely due to inadequate documentation, abortion occurring outside of Medicaid enrollment periods, or alternative payment methods with no associated claims. Abortion history underreporting likely affected our model's ability to identify some high-risk pregnancies, which could delay implementation of guideline-recommended routine cervical length screening⁴¹ and preventive cerclage, leading to missed opportunities for PTB prevention.

These data shortcomings could be addressed through improved claims data collection/reporting, access to electronic health records, and focused patient surveys, which would enhance the capacity for early identification of high-risk pregnancies and timely initiation of appropriate risk-tailored interventions using high-throughput population surveillance tools such as our model.

CONCLUSIONS

This study addresses an important gap in the literature by describing a new PTB prediction model targeting pregnant women in Medicaid without PTB history, including primigravid women, and incorporating both social and clinical risk into PTB risk prediction. This study also demonstrates how to leverage scalable Medicaid and neighborhood socioeconomic data sets, readily accessible to Medicaid managed care organizations, for PTB risk prediction. We quantify the effect of actionable socioeconomic and clinical factors to guide interventions tailored to each patient's unique risk profile and adaptive to changing risk. Our model can be used by Medicaid managed care organizations and health systems to drive population-level interventions addressing clinical and social risk to curb PTB rates and costs at scale. ■

Acknowledgments

Xiao Wang, PhD, and Luyu Xie, PharmD, contributed equally to this work and are listed as co-first authors.

Author Affiliations: Parkland Center for Clinical Innovation (XW, YMP, WO, HKL, DIP), Dallas, TX; now with United Health Group (WO), Minnetonka, MN; now with Parkland Health and Hospital System (DIP), Dallas, TX; Center for Pediatric Population Health (LX, SEM), Dallas, TX; Parkland Community Health Plan (BL), Dallas, TX.

Source of Funding: Parkland Community Health Plan and Parkland Center for Clinical Innovation.

Author Disclosures: Drs Wang and Pengetnze and Ms Ligon are employed by Parkland Center for Clinical Innovation, which has applied for a patent for all or part of the model discussed in the manuscript. The remaining authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (XW, LX, YMP, WO, DIP, BL); acquisition of data (YMP, WO, XW); analysis and interpretation of data (XW, YMP, WO, DIP, SEM, BL); drafting of the manuscript (XW, LX, YMP, HKL, DIP, SEM); critical revision of the manuscript for important intellectual content (XW, LX, YMP, HKL, DIP, SEM, BL); statistical analysis (XW, WO); provision of patients or study materials (YMP, BL); obtaining funding (YMP, BL); administrative, technical, or logistic support (YMP, HKL, DIP, SEM); and supervision (YMP, DIP, SEM).

Address Correspondence to: Yolande M. Pengetnze, MD, Parkland Center for Clinical Innovation, 8435 Stemmons Fwy, Ste 1150, Dallas, TX 75247. Email: yolande.pengetnze@pccinnovation.org.

REFERENCES

1. Preterm birth. World Health Organization. February 19, 2018. Accessed August 14, 2019. <https://www.who.int/news-room/fact-sheets/detail/preterm-birth>
2. Preterm birth. CDC. Updated October 30, 2020. Accessed April 6, 2021. <https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pretermbirth.htm>
3. Institute of Medicine. *Preterm Birth: Causes, Consequences, and Prevention*. The National Academies Press; 2007.
4. Frey HA, Klebanoff MA. The epidemiology, etiology, and costs of preterm birth. *Semin Fetal Neonatal Med*. 2016;21(2):68-73. doi:10.1016/j.siny.2015.12.011
5. Martin JA, Osterman MJK. Describing the increase in preterm births in the United States, 2014-2016. CDC. National Center for Health Statistics data brief 312. Updated June 13, 2018. Accessed August 14, 2019. <https://www.cdc.gov/nchs/products/databriefs/db312.htm>
6. Committee on Practice Bulletins—Obstetrics, The American College of Obstetricians and Gynecologists. Practice bulletin no. 130: prediction and prevention of preterm birth. *Obstet Gynecol*. 2012;120(4):964-973. doi:10.1097/AOG.0b013e3182723b1b
7. Baer RJ, Altman MR, Oltman SP, et al. Maternal factors influencing late entry into prenatal care: a stratified analysis by race or ethnicity and insurance status. *J Matern Fetal Neonatal Med*. 2019;32(20):3336-3342. doi:10.1080/14767058.2018.1463366
8. Margerison CE, Luo Z, Li Y. Economic conditions during pregnancy and preterm birth: a maternal fixed-effects analysis. *Paediatr Perinat Epidemiol*. 2019;33(2):154-161. doi:10.1111/pe.12534
9. Nuru-Jeter A, Dominguez TP, Hammond WP, et al. "It's the skin you're in": African-American women talk about their experiences of racism. an exploratory study to develop measures of racism for birth outcome studies. *Matern Child Health J*. 2009;13(1):29-39. doi:10.1007/s10995-008-0357-x
10. Witt WP, Cheng ER, Wisk LE, et al. Preterm birth in the United States: the impact of stressful life events prior to conception and maternal age. *Am J Public Health*. 2014;104(suppl 1):S73-S80. doi:10.2105/AJPH.2013.301688
11. Witt WP, Cheng ER, Wisk LE, et al. Maternal stressful life events prior to conception and the impact on infant birth weight in the United States. *Am J Public Health*. 2014;104(suppl 1):S81-S89. doi:10.2105/AJPH.2013.301544
12. Esplin MS, Manuck TA, Varner MW, et al. Cluster analysis of spontaneous preterm birth phenotypes identifies potential associations among preterm birth mechanisms. *Am J Obstet Gynecol*. 2015;213(3):429.e1-e9. doi:10.1016/j.ajog.2015.06.011
13. Souka AP, Papastefanou I, Michalitsi V, et al. Cervical length changes from the first to second trimester of pregnancy, and prediction of preterm birth by first-trimester sonographic cervical measurement. *J Ultrasound Med*. 2011;30(7):997-1002. doi:10.7863/jum.2011.30.7.997
14. Jelliffe-Pawlowski LL, Shaw GM, Currier RJ, et al. Association of early-preterm birth with abnormal levels of routinely collected first- and second-trimester biomarkers. *Am J Obstet Gynecol*. 2013;208(6):492.e1-e11. doi:10.1016/j.ajog.2013.02.012
15. Sananes N, Schuller E, Gaudineau A, et al. What is predictive of preterm delivery in the first trimester: isthmus or cervical length? *Prenat Diagn*. 2013;33(9):894-898. doi:10.1002/pd.4158
16. Jelliffe-Pawlowski LL, Baer RJ, Blumenfeld YJ, et al. Maternal characteristics and mid-pregnancy serum biomarkers as risk factors for subtypes of preterm birth. *BJOG*. 2015;122(11):1484-1493. doi:10.1111/1471-0528.13495
17. Meertens LJE, van Montfort P, Scheepers HCJ, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. *Acta Obstet Gynecol Scand*. 2018;97(8):907-920. doi:10.1111/aogs.13358
18. Luo W, Huning EYS, Tran T, Phung D, Venkatesh S. Screening for post 32-week preterm birth risk: how helpful is routine perinatal data collection? *Heliyon*. 2016;2(6):e00119. doi:10.1016/j.heliyon.2016.e00119
19. Leneuve-Dorilas M, Buekens P, Favre A, et al. Predictive factors of preterm delivery in French Guiana for singleton pregnancies: definition and validation of a predictive score. *J Matern Fetal Neonatal Med*. 2020;33(10):1709-1716. doi:10.1080/14767058.2018.1527898
20. Decennial Census of Population and Housing: datasets. United States Census Bureau. Accessed September 14, 2019. <https://www.census.gov/programs-surveys/decennial-census/data/datasets.2010.html>
21. HEDIS 2017 Technical Specifications for Health Plans, Volume 2. National Committee on Quality Assurance. October 3, 2016. Accessed April 6, 2021. https://www.ncqa.org/wp-content/uploads/2018/07/20171003_HEDIS_2017_Volume_2_Technical_Specifications_Update.pdf
22. Kushel MB, Gupta R, Gee L, Haas JS. Housing instability and food insecurity as barriers to health care among low-income Americans. *J Gen Intern Med*. 2006;21(1):71-77. doi:10.1111/j.1525-1497.2005.00278
23. Frederick TJ, Chwalek M, Hughes J, Karabanow J, Kidd S. How stable is stable? defining and measuring housing stability. *J Community Psychol*. 2014;42(8):964-979. doi:10.1002/jcop.21665
24. Income. United States Census Bureau. Accessed October 4, 2019. <https://www.census.gov/topics/income-poverty/income.html>
25. In Dallas-Fort Worth metro, 95% of mostly upper-income areas are dominated by White households. Pew Research Center. June 29, 2015. Accessed October 4, 2019. <https://www.pewresearch.org/fact-tank/2015/06/29/in-greater-dallas-area-segregation-by-income-and-race/f15-06-25-fair-housing2/>
26. Lee SJ, Steer PJ, Filippi V. Seasonal patterns and preterm birth: a systematic review of the literature and an analysis in a London-based cohort. *BJOG*. 2006;113(11):1280-1288. doi:10.1111/j.1471-0528.2006.01055.x
27. Darrow LA, Strickland MJ, Klein M, et al. Seasonality of birth and implications for temporal studies of preterm birth. *Epidemiology*. 2009;20(5):699-706. doi:10.1097/EDE.0b013e3181a66e96
28. Society for Maternal-Fetal Medicine Publications Committee. Progesterone and preterm birth prevention: translating clinical trials data into clinical practice. *Am J Obstet Gynecol*. 2012;206(5):376-386. doi:10.1016/j.ajog.2012.03.010
29. American College of Obstetricians and Gynecologists; Society for Maternal-Fetal Medicine. Obstetric Care Consensus No. 6: periviable birth. *Obstet Gynecol*. 2017;130(4):e187-e199. doi:10.1097/AOG.0000000000002352
30. Carter EB, Tuuli MG, Odibo AO, Macones GA, Cahill AG. Prenatal visit utilization and outcomes in pregnant women with type II and gestational diabetes. *J Perinatol*. 2017;37(2):122-126. doi:10.1038/jp.2016.175
31. Sananes N, Langer B, Gaudineau A, et al. Prediction of spontaneous preterm delivery in singleton pregnancies: where are we and where are we going? a review of literature. *J Obstet Gynaecol*. 2014;34(6):457-461. doi:10.3109/01443615.2014.896325
32. Morken NH, Källen K, Jacobsson B. Predicting risk of spontaneous preterm delivery in women with a singleton pregnancy. *Paediatr Perinat Epidemiol*. 2014;28(1):11-22. doi:10.1111/pe.12087
33. LeCounte ES, Swain GR. Life expectancy at birth in Milwaukee County: a zip code-level analysis. *J Patient Cent Res Rev*. 2017;4(4):213-220. doi:10.17294/2330-0698.1576
34. Taylor L. Housing and health: an overview of the literature. *Health Affairs*. June 7, 2018. Accessed October 3, 2019. <https://www.healthaffairs.org/doi/10.1377/hpb20180313.396577/tuul/>
35. Carrion BV, Earnshaw VA, Kershaw T, et al. Housing instability and birth weight among young urban mothers. *J Urban Health*. 2015;92(1):1-9. doi:10.1007/s11524-014-9913-4
36. Kumar NR, Raker CA, Ware CF, Phipps MG. Characterizing social determinants of health for adolescent mothers during the prenatal and postpartum periods. *Womens Health Issues*. 2017;27(5):565-572. doi:10.1016/j.whi.2017.03.009
37. Hartzema AG, Racoosin JA, MacCurdy TE, Gibbs JM, Kelman JA. Utilizing Medicare claims data for real-time drug safety evaluation: is it feasible? *Pharmacoepidemiol Drug Saf*. 2011;20(7):684-688. doi:10.1002/pds.2143
38. Ammann EM, Kalsekar I, Yoo A, Johnson SS. Validation of body mass index (BMI)-related ICD-9-CM and ICD-10-CM administrative diagnosis codes recorded in US claims data. *Pharmacoepidemiol Drug Saf*. 2018;27(10):1092-1100. doi:10.1002/pds.4617
39. Behavioral Health Data Exchange Consortium: ONC State Policy Consortium Project: final report. HealthIT.gov. June 2014. Accessed April 6, 2021. https://www.healthit.gov/sites/default/files/bhdeconsortiumfinalreport_06182014_508_compliant.pdf
40. Jones RK, Jerman J. Population group abortion rates and lifetime incidence of abortion: United States, 2008-2014. *Am J Public Health*. 2017;107(12):1904-1909. doi:10.2105/AJPH.2017.304042
41. McIntosh J, Feltovich H, Berghella V, Manuck T; Society for Maternal-Fetal Medicine. Role of routine cervical length screening for preterm birth prevention. *Am J Obstet Gynecol*. 2016;215(3):B2-B7. doi:10.1016/j.ajog.2016.04.027

Visit ajmc.com/link/88636 to download PDF and eAppendix

eAppendix. Variables Definition by International Classification of Disease Ninth and Tenth Revisions (ICD-9; ICD-10), and Current Procedural Terminology (CPT) Codes

1. Singleton Delivery:
 - ICD-9 codes: 765.0, 765.00 - 765.09, 765.1, 765.10 - 765.19, 765.2, 765.20 - 765.29;
 - ICD-10 codes: P07.00, P07.01 - P07.03, P07.10, P07.14 - P07.18, P07.20 -P07.26, P07.30 - P07.39, Z38.00, Z38.01, Z38.1, Z38.2;
2. Cervical Cerclage:
 - ICD-9 codes: 67.5, 67.51, 67.59;
 - ICD-10 codes: 0UVC0CZ, 0UVC0DZ, 0UVC0ZZ, 0UVC3CZ, 0UVC3ZZ, 0UVC4CZ, 0UVC4ZZ, 0UVC3DZ, 0UVC4DZ, 0UVC7DZ, 0UVC7ZZ, 0UVC8DZ, 0UVC8ZZ;
 - CPT code: 59320;
3. Preterm Birth or Labor – defined by mothers’ and/or babies’ diagnosis and/or procedure codes as follows:
 - a) Mothers’ diagnoses codes
 - ICD-9 codes: V13.21, V23.41, 644.0, 644.00, 644.03, 644.2, 644.20, 644.21;
 - ICD-10 codes: Z87.51; O09.211- O09.213, O09.219, O60.00, O60.02, O60.03, O60.10X0 - O60.14X0;
 - b) Babies’ diagnosis codes
 - ICD-9 codes: 765.0, 765.00 - 765.09, 765.1, 765.10 - 765.19, 765.2, 765.20 - 765.29;
 - ICD-10 codes: P07.00, P07.01 - P07.03, P07.10, P07.14 - P07.18, P07.20 -P07.26, P07.30 - P07.39;
4. Pregnancy-induced hypertension:
 - ICD-9 codes: 642.30 - 642.34, 642.40 - 642.44, 642.50 - 642.54, 642.60 - 642.64, 642.70 - 642.74, 642.90 - 642.94, 760.0;
 - ICD-10 codes: O11.1-O11.5, O11.9, O13, O13.1-O13.5, O13.9, O14.00-O14.05, O14.10-O14.15, O14.20-O14.25, O14.90-O14.95, O15.0, O15.02, O15.03, O15.1, O15.2, O15.9, O16.1-O16.5, O16.9, P00.0.
5. Outpatient visits CPT codes
 - CPT codes: 99201-99205, 99211-99216, 99241-99245, 99341-99345, 99347-99350, 99382-99386, 99392-99396, 99401-99404, 99411-99412, 99420, 99429;
6. Inpatient admissions CPT codes
 - CPT codes: 99217-99223, 99231-99236, 99238-99239, 99251-99255, 99291-99292;

eAppendix Table 1. Candidate Predictor Variables.

Category	Candidate Predictor Variables
Social and Demographic	Age, Home Address Zip-Code-Level Mean Income, Housing Instability, Nutritional Status, Insurance Gap in Past 12 Months
Behavioral	Smoking, Alcohol Use, Substance Use, Mental Health Diagnosis
Health Service Utilization	Number of Prenatal/Outpatient/Inpatient/Emergency Department Visit During Current Pregnancy
Medical Comorbidity	History of Hypertension, History of Diabetes, HIV, Sexually Transmitted Infection
Past and Recent Obstetric History	Gestational Diabetes in Index or Previous Pregnancies, Pregnancy-Induced Hypertension in Previous Pregnancies, Cervical Incompetence in Previous Pregnancies, Abortion in Previous Pregnancies, 17OH Progesterone Use in Previous Pregnancies, Threatened Abortion during Index Pregnancy
Missing Value	Whether a variable data is missing

eAppendix Table 2. Performance of Candidate Models Versus the Selected Model in The Training Dataset

Model (n*)	ROC[†] AUC[‡]	Precision- Recall AUC[‡]
Logistic regression LASSO (15)	0.676	0.064
XG Boosting (20)	0.665	0.059
Random Forest (20)	0.648	0.050
Neural Networks (15)	0.657	0.061
Fix1 (0) – No Resampling	0.669	0.060
Fix2 (15)	0.665	0.059
Fix3 (15) - Selected	0.679	0.064
Fix4 (15)	0.677	0.064

- *n is the preterm birth rate in the test dataset after resampling; [†]ROC is the receiver operating characteristic curve; [‡]AUC is the area under the curve.