PRELIMINARY & INCOMPLETE, PLEASE DO NOT CITE OR CIRCULATE

Applications of Predictive and Causal Machine Learning to Infant Health

Kandice A. Kapinos, Samuel Mann, and Kelly Hyde

February 2026

JEL Codes: I10, I18, C55

Keywords: artificial intelligence, health information technology, double machine learning, maternal and child health

* Kapinos: RAND, UT Southwestern, kkapinos@rand.org; Hyde: RAND, khyde@rand.org; Mann: RAND, mann@rand.org. This work was funded by the Richard King Mellon Foundation.

Abstract

Background: The United States spends more on maternity care than any other country yet loses nearly 20,000 infant and 700–1,000 mothers from pregnancy-related death annually. Machine learning (ML) offers promise for both identifying at-risk patients and evaluating preventive interventions, yet evidence on its feasibility in real-world perinatal care settings remains limited.

Objectives: We demonstrate two applications of ML using electronic health record (EHR) and administrative data from two large Pennsylvania healthcare systems: (1) prediction of adverse infant and maternal health outcomes, and (2) causal evaluation of a county-level nurse home visiting program on perinatal outcomes.

Data and Methods: We used data from a healthcare system serving Westmoreland County (WC; n = 5,037 singleton births, 2015–2021) and Allegheny County (AC; n = 80,953 singleton births, 2003–2016). For prediction, we trained gradient boosting models to predict preterm birth, low birth weight (LBW), and small for gestational age (SGA), evaluating performance using balanced accuracy, precision, recall, and AUROC. For causal inference, we applied double machine learning (DML) with lasso-based cross-fitting to estimate the effect of prenatal nurse home visiting on preterm birth and SGA in the AC sample, using both a Robinson partialing-out estimator and a doubly robust Augmented Inverse Probability Weighted (AIPW) estimator following Chernozhukov et al. (2018).

Results: In WC, gradient boosting models achieved balanced accuracy between 0.64 and 0.67 for infant outcomes, with recall ranging from 0.46 to 0.55 and precision from 0.13 to 0.20 suggesting meaningful but limited predictive performance given available data. Smoking during and prior to pregnancy was the strongest predictor of LBW and SGA; pre-pregnancy weight and hypertension were also consistently important across models. Using the AC data with DML, we found evidence that prenatal nurse home visiting was associated with a reduced risk of preterm delivery, which is consistent with published experimental studies.

Conclusions: ML methods show feasibility for perinatal risk stratification and program evaluation using routinely collected healthcare data, though predictive performance in smaller samples is constrained by data availability. DML offers a rigorous framework for evaluating public health interventions in observational settings. Further development and validation across more diverse populations and richer data environments is needed before clinical deployment.

I. Introduction

There are more than 20,000 infant deaths (Kochanek et al., 2024) and between 700 to 1,000 maternal deaths between pregnancy and 42 days postpartum out of about per 3.7 million live births per year (Hoyert 2024). For each maternal death, there are as many as 100 additional pregnant or postpartum women who experience other adverse health events (Harvey et al. 2018). Such events are considered “near misses” and labeled as “severe maternal morbidity (SMM).”

Our ability to identify risk of maternal and infant death and related morbidities using machine learning, a type of artificial intelligence (AI), is a growing area of research and practical clinical interest. In particular, because pregnant people typically have at least some interactions with the healthcare system prior to labor and delivery, AI can be used to leverage such large datasets both for the purposes of predicting risk but also for developing and testing interventions.

In this paper, we used data from two large healthcare systems in Pennsylvania to build and test the feasibility of applying machine learning methods for two critical applications in healthcare: identification of risk and evaluations of interventions.

II. Background

Infant Health Outcomes

Nationally, birth defects and preterm delivery are the leading causes of infant mortality (NICHD, 2023). Birth defects (structural or functional abnormalities that develop in utero) account for approximately 20% of infant deaths occurring within the

first year, even though only approximately 3% of infants are born with a congenital anomaly.

Around 10% of infants are delivered preterm (i.e., delivered before 37 weeks of gestation (CDC, 2025b)). However, two-thirds of infants that die within the first year of life were preterm deliveries, though the underlying causes of death tend to be attributed to related conditions that are significantly more common among preterm infants (e.g., low birth weight – defined by the World Health Organization as less than 2500 grams - and respiratory distress system) (Callaghan et al., 2006). In addition, about two-thirds of infant deaths – those occurring within the first year of life – occur within the first 28 days following delivery and are referred to as neonatal deaths (Ely and Driscoll, 2024).

Evidence suggests that there are several risk factors for preterm delivery related to maternal health both prior to and during pregnancy (e.g., maternal age, tobacco use, obesity, preeclampsia, and infections) (Martius et al., 1998; Goldenberg et al., 2008; McDonald et al., 2010; Räisänen et al., 2013). Estimates suggest between 65 and 70% of preterm births follow spontaneous preterm labor, which can occur with membranes intact or because they have ruptured prematurely. The remaining 30 to 35% of preterm births follow a medically indicated Cesarean delivery (Callaghan et al, 2006; Goldenberg et al., 2008). In addition, evidence suggests there are some effective obstetric interventions that can delay delivery (e.g. use of tocolytic drugs) (Haas et al., 2012).

The risk of adverse consequences on infant health outcomes declines as gestational age at delivery increases (Krans and Davis, 2012). Infants delivered preterm are at an elevated risk of not just mortality, but also immediate and longer-term negative health outcomes. Preterm infants are more likely to suffer from respiratory

problems, necrotizing enterocolitis, retinopathy, neonatal jaundice, and hypoxic-ischemic encephalopathy, which can cause longer-term neurodevelopmental issues (Platt, 2014). They also have higher incidence of cerebral palsy, epilepsy, cognitive impairments, and developmental coordination disorders (Platt, 2014).

Low birth weight occurs either because the infant was born preterm or due to intrauterine growth restriction (IUGR). Some of the same causes of preterm delivery can impact fetal growth but the biological pathway for IUGR is inadequate blood flow and nutrient delivery to the placenta (Sharma, Shastri, and Sharma, 2016). Regardless of cause, low birth weight is associated with increased risk of infant mortality (Richarson et al., 1993), and part of this driven by gestational age at delivery which also captures other aspects of fetal maturity that are related to mortality and morbidity (Committee to Study the Prevention of Low Birthweight, 1985).

Thus, while we might want to focus on using AI to identify risks of infant death, in some settings infant death may be rare complicating the performance of predictive models. Because prematurity and low birth weight are correlated with infant death as well as poor health outcomes among infants that survive, we primarily focus on these infant health outcomes in this analysis.

Finally, we note that pregnancies of multiples (i.e. non-singletons) are more likely to result in early delivery and low birth weight as well as several other pregnancy and delivery complications (Dongarwar et al., 2021). For this reason, we exclude multiples in our analysis but discuss how these methods could be extended to study multiples as well.

In this study, we first used an adaptive AI prediction model of to predict poor infant health outcomes using ensemble learning methods. Second, we used double machine learning (DML) to establish causal evidence on the effects of county-level programs and interventions developed to improve infant and maternal health.

III. Data

We used data from two large healthcare systems:

Westmoreland County (WC), PA: we received vital statistics, prenatal health care visits, and delivery outcomes for 7,060 mothers with 8,918 pregnancies delivered between 2015 and 2021 in a large integrated health care system. These data cover about 40% of all infants born in Westmoreland County. From this sample, we were able to obtain delivery outcomes for 6,651 infants (for 6,530 pregnancies across 5,314 unique mothers). We excluded pregnancies that ended due to miscarriage or abortion (n = 323), which yielded a sample of 6,328 infants (across 6,213 unique pregnancies and 5,080 unique mothers). We excluded births of multiples (n = 222) keeping only singleton births because multiples tend to be delivered earlier (Dongarwar et al., 2021), including them can impact analysis of both preterm delivery and low birth weight. We observed two pregnancies for 948 mothers (18.5%) and 3 or more for 75 mothers (1.5%).

Allegheny County (AC), PA: we received vital statistics records linked to prenatal care records, delivery hospitalizations, and use of county-level public health programs for 82,739 births occurring within a large integrated healthcare system in AC from 2003 to 2016. We included singleton births yielding a final analytic sample of 80,953. We also have integrated data on use of public social services programs, like Women, Infants, and

Children (WIC), housing assistance; Family Support Centers; substance use disorder services; and mental health services from the Allegheny County Department of Human Services Data Warehouse.

IV. Measures

Because there were so few deaths observable in our analytic samples, we primarily focused on 3 other infant health outcomes that have been linked to infant mortality: 1) preterm delivery (occurring prior to 37 weeks); 2) low birth weight (< 2500 grams); and 3) small for gestational age (weight at < 10th percentile adjusting for age).

The set of potential predictor variables available span multiple domains relating to maternal health risk factors, socio-demographics and health care utilization during the perinatal period. We included zip-code fixed effects as proxies for environmental measures due to little geographic variation in our sample. The full set of potential predictors included in our machine learning models derived from EHR information for each sample are shown in the Appendix. We highlight the key domains and example measures here:

1. Maternal demographics: age (categorical or continuous), marital status (categorical), race/ethnicity (categorical);
2. Maternal health care utilization: number of visits in each prenatal period and when prenatal care was initiated;
3. Maternal health conditions (current and history): diagnoses of 32 conditions, (including e.g., diabetes, anxiety, depression, and sexually transmitted infections);

4. Maternal health behaviors: current and previous tobacco use, current and previous alcohol use, and exposure to secondhand smoke;
5. Maternal health prenatal counseling: set of 46 indicators for whether particular types of prenatal counseling were received (e.g. nutrition, use of any medication, environmental hazards, signs of labor, etc. see Appendix for full list)

The AC sample includes details on use of and participation in a range county-level programs (for example, focused on mental health, parenting, infant care, and nutrition benefits). We focused on participation in a nurse home visiting program during pregnancy.

V. Statistical Approach

Prediction

We predict outcomes (infant health or intervention participation) using a machine learning model of the following form:

$$Y_{im} = f(X_{im}, \tau_i, \mu_i) + \varepsilon_{im} \quad (1)$$

In Equation (1), Y_{im} denotes the outcomes of interest for infant i born to mother m. The model incorporates a rich set of risk factors, X_{im}, capturing for example maternal demographics, pregnancy conditions, prenatal encounters, etc. To adjust for temporal changes in outcomes we include a series of birth year indicators, τ_i, and to account for geographic variation we include a series of 3-digit ZIP code level mothers' residence indicators, μ_i. The error term, ε_{im} represents the deviation between the model's prediction and the observed outcome. Although written compactly, the specification

encompasses many variables. Bold type interface signals vectors of multiple variables, rather than a single variable (e.g. X_{im}, τ_i, and μ_i).

The goal of the machine learning approach is to identify a function, $f(\bullet)$, that selects the most predictive variables in X_{im}, τ_i, and μ_i, and combines them in a way that minimizes the prediction error. In practical terms, this involves finding the most accurate possible model for the outcome under study.

Different algorithms arrive at the optimal function in different ways. There are several machine learning algorithms used for prediction, including penalized regression models (i.e. least absolute shrinkage and selection operator (LASSO), ridge regression, elastic net)(Zou and Hastie 2005, McNeish 2015), random forests (Hastie et al. 2009), and Bayesian networks (Lucas, Van der Gaag, and Abu-Hanna 2004) and generalized boosting models (Friedman, 2002; Kuhn, 2008). Penalized regression approaches "penalize" or reduce regression coefficients towards zero when a variable does not explain variation in the outcome. These approaches (ridge, elastic net, and adaptive LASSO), all employ alternative weighting schemes and/or calculate the "penalties" for each variable differently. Random forests extend classification and regression trees (CARTs), which essentially recursively partition the sample along explanatory variables with the goal of creating similar groups in each partition (Krzywinski and Altman 2017). Random forests collate numerous such random trees, averaging across trees. Another related type of "ensemble learning" method we used is gradient boosting, which is similar but in this case the random trees are built sequentially (to improve upon earlier trees) (Hastie et al. 2009).

Here, we implemented the gradient boosting method, which departs from traditional modeling approaches; in conventional models, the researcher chooses both which variables to include and the function form f(•). This approach can introduce bias if the selected variables or the assumed structure of the model (often linear) do not reflect the true data-generating process. In our framework the algorithm evaluates the variability in the data and identifies the most relevant predictors from X_{im}, τ_i, and μ_i, without relying on researcher pre-selection, thereby reducing model misspecification bias. Additionally, it systematically explores a wide range of functional forms (including linear and nonlinear) to determine the one that best fits the data, ensuring that the model structure id driven by predictive performance rather than prior assumptions.

Gradient boosting models begin with a simple baseline model, such as a linear function. They then examine the error term, ε_{im}, and uses this information to identify where the model falls shorts. Based on these errors, the algorithm fits a new function f(•), designed to reduce the error, compared to the prior function. This is done by leveraging the gradient of a loss function, which quantifies the discrepancy between predictions and actual outcomes. The process repeats, with each successive model targeting the remaining errors of the previous one, until no further improvement is possible (Friedman, Hastie and Tibshirani, 2000). This iterative procedure of building a sequence of weak learners, each improving upon the last, is known as boosting (Schapire, 1990). To prevent overfitting and improve out-of-sample performance, we

used three-fold cross-validation. The data were randomly divided into a training set (80%) and a holdout test set (20%).[1]

For model selection, we used the area under the precision-recall curve (AUPRC) as the optimization metric. This measure is appropriate for binary classification problems with highly imbalanced classes, such as the infant outcomes under study.[2] Once the model was trained, we assessed its performance on the holdout test set using balanced accuracy (which adjusts accuracy estimates as if the outcome were evenly distributed across observations), precision, recall, and the AUROC.[3]

Using the AC sample, we also aimed to evaluate the impact of a county-level nurse home visiting program used during the prenatal period on two infant health outcomes (preterm delivery and small for gestational age). To address the potential selection bias from systematic differences between individuals who use these programs and those who do not, we used a double machine learning (DML) or causal machine learning model that allows us to remove these potential biases (Feuerriegel et al. 2024). DML is a newer causal inference framework that can be used with a range of estimators (least absolute shrinkage and selection operator (lasso), boosted trees, and other

---

[1] While cross-validation inherently creates temporary holdout sets ("folds") during each iteration, a separate holdout test set can still be valuable when the dataset is large enough to permit it without compromising training. In such cases, evaluating the final model on an entirely unseen subset of data offers an additional and often more robust assessment of how well the model is likely to generalize to new observations.

[2] The AUPRC quantifies the balance between precision and recall, whereas the AUROC reflects the balance between specificity and sensitivity. Since recall and sensitivity both measure the true positive rate (the proportion of actual positive cases correctly identified) the key mathematical distinction between the two metrics comes from comparing precision to specificity. Precision is the share of the model's positive predictions that are correct ("true positives" out of all predicted positives). Specificity is the share of actual negative cases correctly identified as negative ("true negatives" out of all actual negatives). In situations where the positive class is rare, a model can achieve high specificity even if it performs poorly at detecting positive cases.

[3] Balanced accuracy has a baseline, or "no information rate," of 50%, i.e. the expected score for a naïve model that always predicts the majority class. A balanced accuracy below this threshold indicates that the model performs worse than simply guessing the most common outcome without using any training data.

ensemble methods). Traditionally, causal inference methods rely on the assumption of conditional exchangeability meaning that all confounders are included in the model (cite), which can be untenable with small samples or few measures. High dimensional data, however, minimizes the risk that this assumption does not hold. Again, we used cross-validation, repeatedly dividing a training dataset (80% of the full dataset) into subsets (called "folds"), and then validating each model on a fold excluded from the training process. These approaches improve generalizability and reduce overfitting. In the DML approach, the first step is to predict the main health outcome first using a ML model; we used lasso with all potential covariates standardized. The second step is to predict the probability of use of a DHS program or intervention using a similar ML model and set of covariates. The third step in the DML approach is to regress the residuals from the first estimation on the residuals from the second model to obtain the de-biased estimate of the relationship between use of one of these programs and infant and maternal health outcomes. RAND's Institutional Review Board (Project # 2025-NRNI-0034) approved this study.

VI. Results

Westmoreland County

In Table 1, we show basic descriptive statistics of the mothers in our final analytic sample for WC (n = 5037). Average maternal age at birth was 28.22 (standard deviation = 5.51). The sample is predominantly non-Hispanic White (96%). More than half of the mothers were married. The most common chronic conditions (present during or prior to pregnancy) were anxiety (29%) or an autoimmune condition (29%).

Table 2 presents the performance metrics for predicting preterm birth, LBW, and SGA models using all the available measures (see Appendix). We found balanced accuracy estimates between 0.64 and 0.67 which suggest that our models are performing better than a naïve "no information" model, which would have a balanced accuracy of 0.5. However, the models exhibit low precision and modest recall, confirming that the input data lack at least some meaningfully predictive information.

In a hospital context where a predictive model could influence decisions with life-or-death consequences, these models likely do not achieve a sufficient level of performance to be integrated into clinical processes in their current form, since they miss about half of positive cases (as evidenced by recall around or slightly above 0.5) and generate significantly more false positives than true positives (as evidenced by precision less than 0.2). However, the models performing significantly better in terms of recall than precision is a positive, since it is more costly to underestimate risk (and therefore not engage in potentially life-saving interventions) than to overestimate it in this context. In practical terms, a model performing better on recall than precision indicates that it tends to overestimate risk. Higher recall means that, among cases where the adverse outcome actually happened, it is more likely that the model predicted it would occur (true positives vs. false negatives); higher precision means that, among cases where the model predicted the adverse outcome would occur, it is more likely that it actually did occur (true positives vs. false positives). The model performance metrics in Table 2 indicate that between 46 and 55 percent of observed adverse outcomes were correctly predicted to occur by the model, whereas only 13 to 20 percent of predicted adverse outcomes actually occurred, i.e., between 80 and 87 percent of predicted

adverse outcomes did not actually occur. In a risk-averse healthcare context, particularly for interventions with low associated risks (e.g., checkups and counseling), recall likely receives higher weight. However, the very low precision may lead to excess interventions, which may be problematic if resources to support interventions (e.g., provider availability or funding) are scarce.

We also looked at feature importance, a measure of how significant each variable ("feature") included in the model's input training data was to its output predictions. More specifically, we calculate gain, a type of feature importance which approximates the average increase in predictive accuracy achieved by splits in the model decision tree using each respective input variable. In Figures 1 to 3, we show the top ten variables with the highest gain for each model. The gain (represented by the x-axis on the below figures) is normalized to 100, i.e., the raw gain is scaled so that the highest value equals 100 and all other values are adjusted proportionally. This means that a value of 50 can be interpreted as half the relative importance (gain) of the most important feature. Note that these values do not reflect causal relationships between the variables and the outcome, instead only reflecting relative strength of association. Additionally, while feature importance is useful to understand which variables contribute the most strongly to a model's predictions, it does not specify the directionality of the relationships between variables and outcomes (i.e., it is not clear whether an important variable is positively or negatively associated with the outcome). Because our models exhibit modest overall predictive performance, cannot causally identify these relationships, and require additional machine learning explanation methods to derive relationship directionality, we focus only on feature importance for this exercise.

Figures 1 to 3 show the top 10 feature importances across all 218 predictors included in the models.[4] The feature importance values are normalized so that the top feature has an importance of 100. Smoking (during and/or before pregnancy) is the number one predictor of both low birth weight and small for gestational age. Weight before pregnancy, healthcare utilization in certain specialties (OB-GYN, general surgery), and hypertension (either chronic or gestational) also rank highly for all three models. The other most important features, most of which have a relatively low estimated feature importance (less than one-tenth as important as the top predictor), vary across the models. However, especially for less significantly important variables, randomness at different levels of the model training process (the holdout test sample, randomly selected cross-validation folds, etc.) can result in randomness in estimated feature importances, meaning that subsequent retraining of the models could result in different feature importance estimates.

Allegheny County

In Table 3, we show basic descriptive statistics of the mothers in our final analytic sample for AC (n = 80,953). Average maternal age at birth was 28.66 (standard deviation = 5.98). The sample was more diverse than the WC sample but still predominantly non-Hispanic White (73%). More than half of the mothers were married. <ADD full descriptives to appendix>

The DML results suggest a decreased risk of preterm delivery but an increased risk of SGA (see Table 4). The adjusted odds ratio of preterm delivery was estimated at

[4] We present complete feature importance estimates for all predictors in the model in Appendix D.

0.88 (95%CI: 0.75, 1.00), whereas the adjusted odds ratio of an infant being small for gestational age was 1.46 (95% CI: 1.21, 1.77). <add performance metrics>

VII. Discussion

This study demonstrates the feasibility and value of applying machine learning to two distinct but complementary problems in perinatal health: outcome prediction and causal program evaluation, using data from two large integrated healthcare systems in Pennsylvania.

Prediction findings. Our gradient boosting models achieved moderate predictive performance for preterm birth, LBW, and SGA in the WC sample, with balanced accuracy estimates ranging from 0.64 to 0.67, but with relatively poor precision. Consistent with prior literature, smoking behavior and pre-pregnancy weight emerged as the strongest predictors across models, alongside hypertension and healthcare utilization patterns. While the models outperformed naïve baselines, precision remained low — between 0.13 to 0.20. In a clinical deployment context, these performance levels suggest the models are better suited to population-level risk stratification and resource allocation than to individual-level clinical decision-making in their current form. The asymmetry between recall and precision across models — where recall exceeded precision — is arguably appropriate for a risk-averse healthcare context, since failing to identify a high-risk pregnancy carries greater consequences than over-predicting risk. Nonetheless, very low precision raises legitimate concerns about resource strain from excess intervention, particularly where provider capacity or program funding is limited.

Causal evaluation findings. Using the AC sample, we estimated the causal effect of prenatal participation in a nurse home visiting program on preterm birth and SGA using double machine learning. The DML framework offers several advantages over traditional regression-based approaches in this setting: it accommodates a high-dimensional covariate space, makes no parametric assumptions about the functional form of confounding, and produces estimates that are robust to misspecification of either the outcome or treatment model under the doubly robust AIPW specification. We found that prenatal nurse home visiting was associated with a reduced risk of preterm birth, with results consistent across both the Robinson and Chernozhukov AIPW estimators (need to add). These findings are consistent with prior evidence on the effectiveness of home visiting programs, which have shown benefits across a range of child and maternal outcomes (CITE). However, our analysis shows the feasibility of using DML methods in cases where randomized controlled trials are not possible or ethical. Moreover, we provide causal evidence derived from ML-based methods applied to routine administrative data in a context that can easily be adapted and replicated across healthcare systems with integrated healthcare data.

Limitations. Several limitations warrant consideration. First, these samples are predominantly non-Hispanic White, limiting generalizability to more racially and socioeconomically diverse populations. Second, the outcome measures available — while clinically meaningful proxies — do not capture infant mortality directly, as deaths were too rare in both samples to support reliable model training. Third, the causal estimates rely on the assumption of conditional exchangeability; while our high-dimensional covariate set reduces the risk that important confounders are omitted, unmeasured

confounding (e.g., motivation to seek care, social support) cannot be ruled out. Fourth, the AC data, while rich, span 2003–2016, and both the composition of available programs and the healthcare landscape have changed meaningfully since that period. Finally, our prediction models were trained and tested within a single healthcare system; external validation in other settings is needed before broader deployment.

Implications and future directions. Together, these findings illustrate both the promise and the current boundaries of ML-based approaches to perinatal health. The prediction models perform well enough to identify high-risk subgroups and could potentially support population-level targeting of preventive services, even if individual-level precision remains limited. The DML findings suggest that routinely collected administrative data, combined with modern causal inference methods, can generate credible estimates of program effects without requiring randomized controlled trials — an important practical advantage given the cost and ethical constraints of experimental designs in this population. Future work should explore richer feature sets (including social determinants of health, environmental exposures, and wearable or remote monitoring data), validation in more diverse samples, and integration of these methods into real-time clinical decision support systems. Extending these methods to multiple gestation pregnancies, postpartum outcomes, and longer-term child development outcomes also represents a natural and important next step.

Table 1: Descriptive Statistics of WC Analytic Sample

Key Outcomes/	N	Percent
Preterm Delivery (37 weeks of gestation)	532	8.7%
Low Birth Weight (<2500 grams)	351	5.7%
Small for Gestational Age (<10th percentile)	538	8.8%
Age at First Birth Observed		
< 20	261	5%
20 – 24	1,091	22%
25 - 29	1,666	33%
30 – 34	1,321	26%
35 – 39	571	11%
40+	127	2%
Race/Ethnicity	Count	Percent
African American/Black	148	3%
Latino/Hispanic	56	1%
Other or Unknown	124	2%
White	4,706	94%
Marital Status	Count	Percent
Divorced	97	2%
Married	2,668	53%
Single	2,166	43%
Unknown/Missing/Widowed	106	2%
Prenatal Care	Count	Percent
Initiated in First Trimester		69%
Initiated in Second Trimester		24%
Initiated in Third Trimester		5%
Did Not Receive Any Prenatal Care		1%
Other Risk Factors Recorded in the Maternal Medical Chart	Count	Percent
Ever Used Alcohol Prior to Pregnancy[b]	1,000	20%
Ever Used Tobacco Use Prior to Pregnancy[b]	1,488	30%

Note: a. Measures created from EH generated flags (and not from actual diagnosis codes). b. Indicators created from "free text" fields and may not have been collected systematically from all patients. n = 5,037 mothers.

Table 2. Predictive Model Performance Metrics for WC Sample

Metric	Preterm Birth (<=37 weeks)	Low Birth Weight	Small for Gestational Age
AUPRC	0.18	0.14	0.18
Balanced Accuracy	0.64	0.64	0.67
Precision	0.20	0.13	0.20
Recall	0.46	0.51	0.55
AUROC	0.68	0.66	0.70

NOTE: The precision and recall statistics reported in the table are based on a raw probability cutoff of 0.5, i.e., any observation for which the model predicted a likelihood of greater than or equal to 0.5 was assigned the positive class and all others were assigned the negative class to dichotomize the raw probability output.

Table 3. Descriptive Statistics of AC Analytic Sample

Key Outcomes/	N	Percent
Preterm Delivery (37 weeks of gestation)	8,546	10.6%
Small for Gestational Age (<10th percentile)	3,367	8.8%
Age at First Birth Observed		
< 20	5,802	8%
20 – 24	15,315	20%
25 – 29	17,218	24%
30 – 34	23,461	31%
35 – 39	11,308	15%
40+	2,533	3%
Race/Ethnicity		
African American/Black	16,837	21%
Latino/Hispanic	709	1%
Other or Unknown	4,112	5%
White	59,295	73%
Marital Status		
Married	48,685	60%
Unmarried	32,268	40%
Prenatal Care		Mean (Std. Dev)
Number of Visits	60,999	12.58 (1.74)
Other Risk Factors Recorded in the Maternal Medical Chart	Count	Percent
Substance Use Disorder diagnosis (during pregnancy)[a]	1,000	20%
Smoked During First Trimester [b]	10,195	13%
Smoked During Second Trimester [b]	9,494	12%
Smoked During Third Trimester [b]	9,403	12%

Note: a. Based on diagnoses codes from encounters. b.Indicators created from EHR. n = 80,953 mothers.

Table 4. Double Machine Learning Results: Impact of Participation in Nurse Home Visits on Preterm Delivery and SGA in WC Analytic Sample

	Preterm Birth (<=37 weeks)
Variables Selected in Predicting Outcome (all measured prior to delivery)	Any Cardiovascular Condition, Any Respiratory Condition, Any Digestive Disorder, Any Infections during Pregnancy, Positive Toxicology Screen, Gestational Hypertension, Previous Preterm Delivery, Ever Smoked
Variables Selected in Predicting Nurse Home Visit during Pregnancy (all measured prior to delivery)	Maternal Age, Any Respiratory Condition, Any Mental, Behavioral or Neurodevelopment Condition, Any Digestive Disorder, Chlamydia Infection, Hypertension, Previous Preterm Delivery, Previous Cesarean Delivery, Smoking Prior to Conception, Smoking in First Trimester
OR: Any Nurse Home Visits During Pregnancy	0.88* (0.06)

Figure 1. Feature Importance for Preterm Birth (<=37 weeks) Model

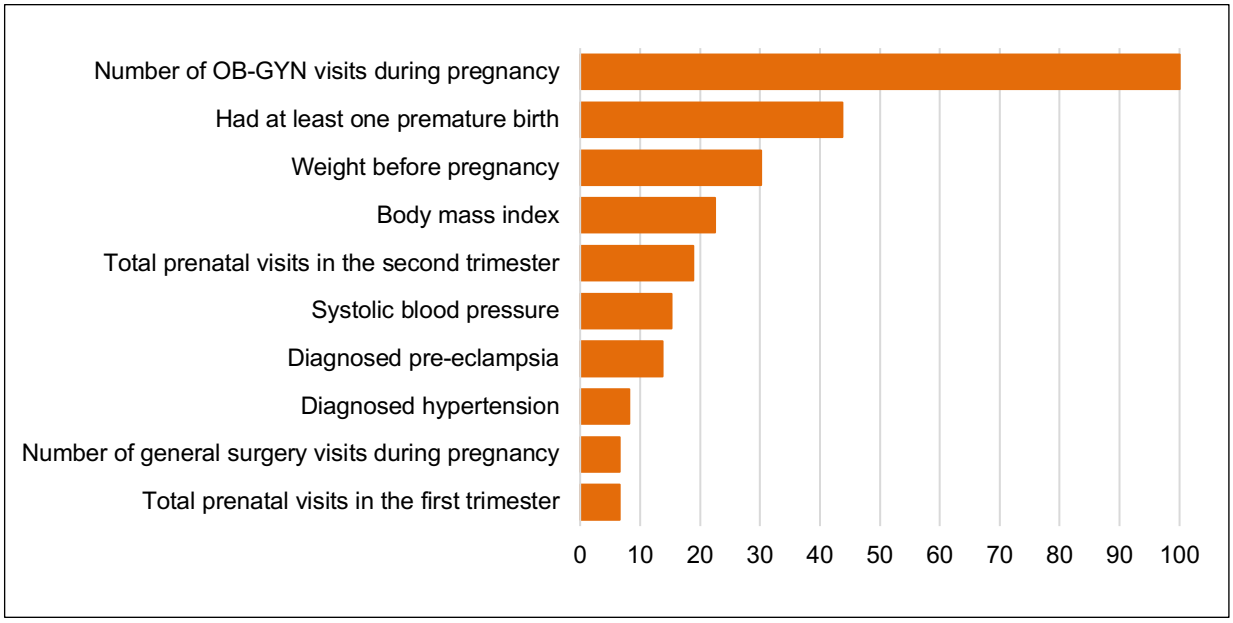


Figure 2. Feature Importance for Low-Birth-Weight Model

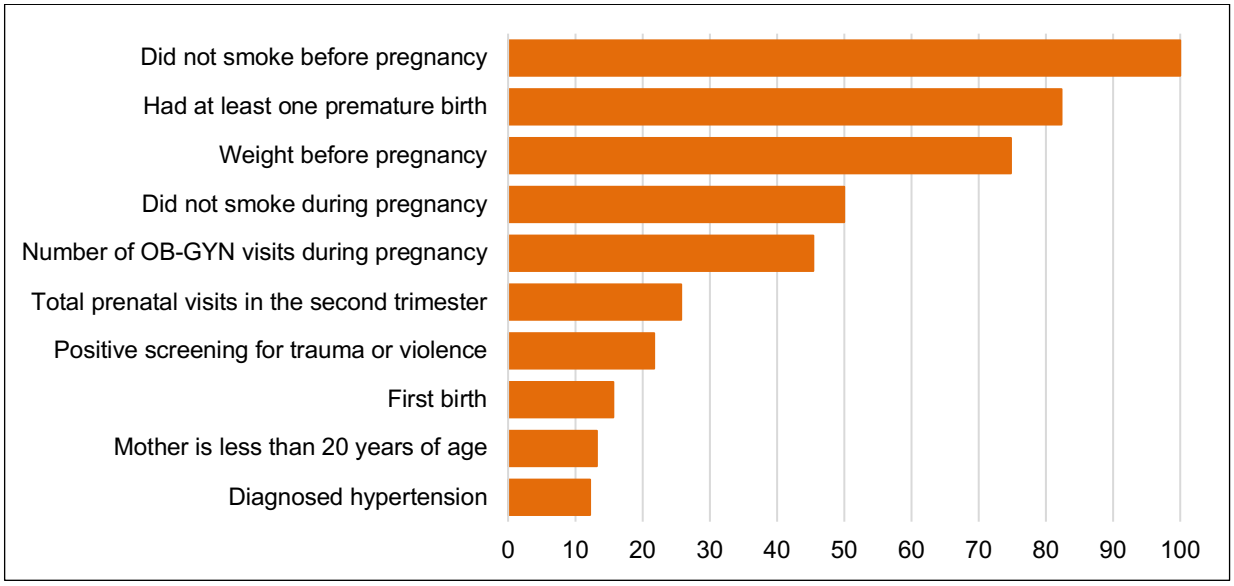
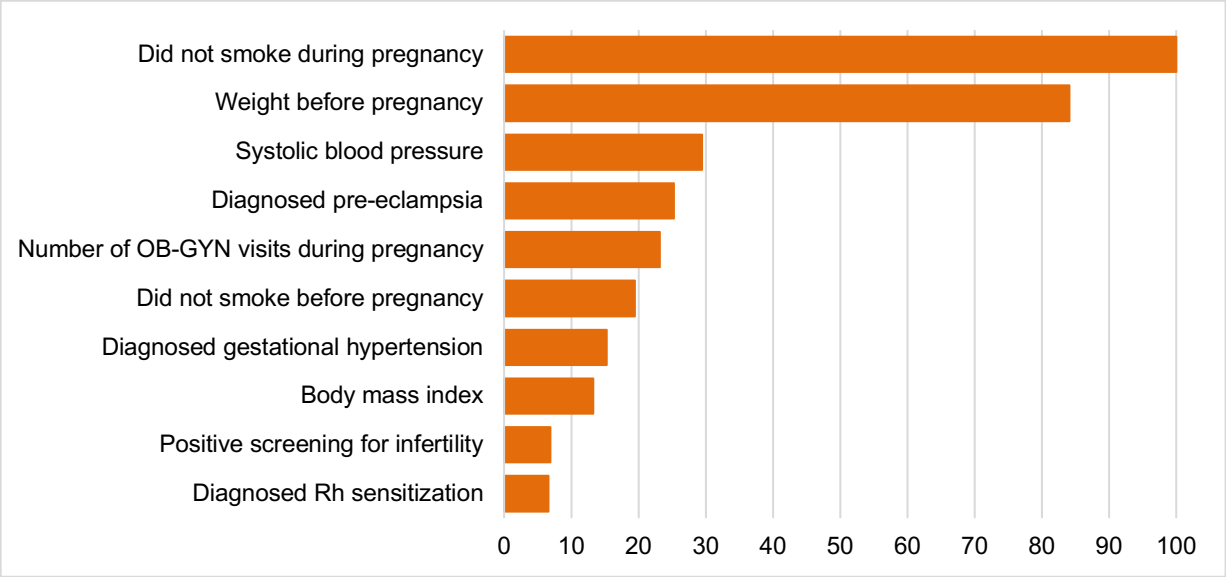


Figure 3. Feature Importance for Small for Gestational Age Model



REFERENCES

Bateman, B. T., J. M. Mhyre, S. Hernandez-Diaz, K. F. Huybrechts, M. A. Fischer, A. A. Creanga, W. M. Callaghan, and J. J. Gagne. 2013. "Development of a comorbidity index for use in obstetric patients." Obstet Gynecol no. 122 (5):957-965. doi: 10.1097/AOG.0b013e3182a603bb.

Bradley, A. P. 1997. "The use of the area under the roc curve in the evaluation of machine learning algorithms." Pattern Recognition no. 30 (7):1145-1159. doi: Doi 10.1016/S0031-3203(96)00142-2.

Centers for Disease Control and Prevention. 2024a. Identifying Severe Maternal Morbidity (SMM).

Centers for Disease Control and Prevention. Pregnancy Mortality Surveillance System 2024b [cited 12/12/2024. Available from https://www.cdc.gov/maternal-mortality/php/pregnancy-mortality-surveillance/index.html.

Centers for Disease Control and Prevention. Identifying Severe Maternal Morbidity (SMM) 2025 [cited 01/21/2025. Available from https://www.cdc.gov/maternal-infant-health/php/severe-maternal-morbidity/icd.html.

Feuerriegel, S., D. Frauen, V. Melnychuk, J. Schweisthal, K. Hess, A. Curth, S. Bauer, N. Kilbertus, I. S. Kohane, and M. van der Schaar. 2024. "Causal machine learning for predicting treatment outcomes." Nat Med no. 30 (4):958-968. doi: 10.1038/s41591-024-02902-1.

Harvey, E. M., S. Ahmed, S. E. Manning, H. Diop, C. Argani, and D. M. Strobino. 2018. "Severe Maternal Morbidity at Delivery and Risk of Hospital Encounters Within 6 Weeks and 1 Year Postpartum." J Womens Health (Larchmt) no. 27 (2):140-147. doi: 10.1089/jwh.2017.6437.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. "Boosting and additive trees." The elements of statistical learning: data mining, inference, and prediction:337-387.

Hoyert, DL. 2024. Maternal Mortality Rates in the United States, 2022. edited by National Center for Health Statistics.

Kern-Goldberger, A. R., E. A. Howell, S. K. Srinivas, and L. D. Levine. 2023. "What we talk about when we talk about severe maternal morbidity: a call to action to critically review severe maternal morbidity as an outcome measure." Am J Obstet Gynecol MFM no. 5 (5):100882. doi: 10.1016/j.ajogmf.2023.100882.

Krzywinski, M., and N. Altman. 2017. "Classification and regression trees." Nature Methods no. 14 (8):755-756. doi: 10.1038/nmeth.4370.

Lucas, Peter JF, Linda C Van der Gaag, and Ameen Abu-Hanna. 2004. "Bayesian networks in biomedicine and health-care." Artif Intell Med no. 30 (3):201-214. doi: 10.1016/j.artmed.2003.11.001.

McNeish, Daniel M. 2015. "Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences." Multivariate behavioral research no. 50 (5):471-484. doi: 10.1080/00273171.2015.1036965.

Payne, B. A., H. Ryan, J. Bone, L. A. Magee, A. B. Aarvold, J. Mark Ansermino, Z. A. Bhutta, M. Bowen, J. Guilherme Cecatti, C. Chazotte, T. Crozier, A. J. M. de Pont, O. Demirkiran, T. Duan, M. Kallen, W. Ganzevoort, M. Geary, D. Goffman, J. A. Hutcheon, K. S. Joseph, S. E. Lapinsky, I. Lataifeh, J. Li, S. Liskonova, E. M.

Hamel, F. M. McAuliffe, C. O'Herlihy, B. W. J. Mol, P. G. R. Seaward, R. Tadros, T. Togal, R. Qureshi, U. Vivian Ukah, D. Vasquez, E. Wallace, P. Yong, V. Zhou, K. R. Walley, P. von Dadelszen, and Cipher Group. 2018. "Development and internal validation of the multivariable CIPHER (Collaborative Integrated Pregnancy High-dependency Estimate of Risk) clinical risk prediction model." Crit Care no. 22 (1):278. doi: 10.1186/s13054-018-2215-6.

Snowden, J. M., A. Lyndon, P. Kan, A. El Ayadi, E. Main, and S. L. Carmichael. 2021. "Severe Maternal Morbidity: A Comparison of Definitions and Data Sources." Am J Epidemiol no. 190 (9):1890-1897. doi: 10.1093/aje/kwab077.

Steyerberg, E. W., A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan. 2010. "Assessing the performance of prediction models: a framework for traditional and novel measures." Epidemiology no. 21 (1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2.

Zou, Hui, and Trevor Hastie. 2005. "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society Series B: Statistical Methodology no. 67 (2):301-320.

APPENDIX A: Additional Westmoreland County Output

Appendix Table A1. Full List of WC Variables Included as Potential Predictors

Variable	Variable Type	Percent Missing	# Variables
Positive screening for heart disease	Binary	1.0%	1
Positive screening for kidney disease	Binary	1.1%	1
Positive screening for neurologic abnormality	Binary	1.1%	1
Positive screening for psychiatric condition	Binary	0.9%	1
Positive screening for hepatitis	Binary	1.1%	1
Positive screening for varicosities	Binary	1.1%	1
Positive screening for thyroid dysfunction	Binary	1.1%	1
Positive screening for trauma or violence	Binary	1.1%	1
History of blood transfusion	Binary	1.1%	1
Positive screening for seasonal allergies	Binary	1.1%	1
Positive screening for drug or latex allergy	Binary	1.0%	1
Positive screening for breast abnormality	Binary	1.1%	1
History of operations or hospitalization	Binary	0.9%	1
Positive screening for anesthetic complications	Binary	1.1%	1
History of abnormal Pap smears	Binary	1.1%	1
Positive screening for uterine abnormality	Binary	1.1%	1
Positive screening for infertility	Binary	1.1%	1
Positive screening for antiretroviral therapy	Binary	1.1%	1
Positive screening for relevant family history	Binary	1.2%	1
Other positive screening	Binary	1.3%	1
Head, eyes, ears, nose, and throat (HEENT) normal	Binary	25.8%	1
Fundi (uterus) normal	Binary	43.7%	1

Variable	Variable Type	Percent Missing	# Variables
Teeth normal	Binary	15.6%	1
Thyroid normal	Binary	10.0%	1
Breasts normal	Binary	12.3%	1
Lungs normal	Binary	7.9%	1
Heart normal	Binary	7.9%	1
Abdomen normal	Binary	8.0%	1
Extremities normal	Binary	8.4%	1
Skin normal	Binary	8.9%	1
Lymph nodes normal	Binary	15.1%	1
Vulva normal	Binary	8.4%	1
Vagina normal	Binary	8.4%	1
Cervix normal	Binary	8.4%	1
Adnexa normal	Binary	11.3%	1
Patient is 35 years of age or older as of estimated delivery date	Binary	1.1%	1
Mother was tested for fragile X syndrome	Binary	1.2%	1
Mother or father has at least one child	Binary	1.1%	1
Has had recurrent pregnancy loss	Binary	1.1%	1
New medications since last pregnancy	Binary	1.2%	1
Alcohol use before pregnancy	Binary	0.0%	1
Alcohol use during pregnancy	Binary	0.0%	1
Mother reported current substance or alcohol use	Binary	0.0%	1
Did not smoke before pregnancy	Binary	0.0%	1
Quit smoking before pregnancy	Binary	0.0%	1
Did not smoke during pregnancy	Binary	0.0%	1
Quit smoking during pregnancy	Binary	0.0%	1
Secondhand smoke exposure	Binary	0.0%	1
First pregnancy	Binary	0.7%	1
First birth	Binary	13.5%	1

Variable	Variable Type	Percent Missing	# Variables
Diagnosed chlamydia	Binary	0.0%	1
Diagnosed gonorrhea	Binary	0.0%	1
Diagnosed hepatitis B or C	Binary	0.0%	1
Diagnosed human immunodeficiency virus (HIV)	Binary	0.0%	1
Diagnosed human papillomavirus (HPV)	Binary	0.0%	1
Diagnosed other infections	Binary	0.0%	1
Diagnosed with syphilis	Binary	0.0%	1
Mother is less than 20 years of age	Binary	0.0%	1
Mother is 35 years of age or older	Binary	0.0%	1
Mother is married	Binary	0.0%	1
Mother is White	Binary	0.0%	1
Diagnosed anemia	Binary	0.4%	1
Diagnosed anxiety	Binary	0.4%	1
Diagnosed autoimmune disorder	Binary	0.0%	1
Diagnosed with any STD	Binary	0.0%	1
Diagnosed with depression	Binary	0.0%	1
Diagnosed with diabetes	Binary	0.0%	1
History of gynecological surgery	Binary	1.1%	1
Diagnosed hypertension	Binary	0.0%	1
Diagnosed obesity	Binary	0.4%	1
Diagnosed opioid use disorder	Binary	0.4%	1
Diagnosed pulmonary disease	Binary	1.0%	1
Diagnosed Rh sensitization	Binary	1.3%	1
Diagnosed substance use disorder	Binary	0.4%	1
Diagnosed gestational diabetes	Binary	0.4%	1
Diagnosed gestational hypertension	Binary	0.4%	1
Diagnosed pre-eclampsia	Binary	0.4%	1
Family reported substance or alcohol use	Binary	0.0%	1

Variable	Variable Type	Percent Missing	# Variables
Partner reported substance or alcohol use	Binary	0.0%	1
Mother reported past substance or alcohol use	Binary	0.0%	1
Total number of prenatal visits in the first trimester	Numeric	0.0%	1
Total number of prenatal visits in the second trimester	Numeric	0.0%	1
Total number of pregnancies	Numeric	0.7%	1
Has had at least one premature pregnancy	Binary	0.0%	1
Has had at least one ectopic pregnancy	Binary	0.0%	1
Has had at least one multiple birth pregnancy	Binary	0.0%	1
Body mass index	Numeric	19.2%	1
Systolic blood pressure	Numeric	20.3%	1
Diastolic blood pressure	Numeric	20.3%	1
Weight before pregnancy	Numeric	18.6%	1
Prenatal counseling indicators	Binary	0.0%	45
Specialist visits	Numeric	0.0%	25
When prenatal care started	Categorical, converted to 3 dummy variables	0.0%	3
Infrequent zip code (less than 10 births observed)	Binary	0.0%	1
Zip code dummies (at least 10 births observed)	Categorical, converted to 54 dummy variables	0.0%	54

Appendix Table A2. Appendix D. Complete Feature Importance Estimates by Model

Note that, for the table below, "--" signifies that the variable was not used by the model to predict the outcome, i.e., it has an importance of zero.

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Normal physical exam findings			
Abdomen normal	--	--	--
Adnexa normal	--	--	--
Breasts normal	--	--	1.99
Cervix normal	--	--	--
Extremities normal	--	--	--
Fundi (uterus) normal	--	--	--
Head, eyes, ears, nose, and throat (HEENT) normal	--	--	--
Heart normal	--	--	--
Lungs normal	--	--	--
Lymph nodes normal	--	--	--
Skin normal	--	--	--
Teeth normal	--	--	--
Thyroid normal	--	--	--
Vagina normal	--	--	--
Vulva normal	--	--	--
Positive screening for health conditions			
Other positive screening	3.35	--	--
Positive screening for anesthetic complications	--	--	--
Positive screening for antiretroviral therapy	--	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Positive screening for breast abnormality	--	--	--
Positive screening for drug or latex allergy	2.46	--	--
Positive screening for heart disease	--	--	--
Positive screening for hepatitis	--	--	--
Positive screening for infertility	1.04	--	6.89
Positive screening for kidney disease	1.71	--	--
Positive screening for neurologic abnormality	0.85	--	--
Positive screening for psychiatric condition	1.43	--	--
Positive screening for relevant family history	3.06	--	2.72
Positive screening for seasonal allergies	--	--	--
Positive screening for thyroid dysfunction	2.20	--	--
Positive screening for trauma or violence	1.12	21.73	2.96
Positive screening for uterine abnormality	2.17	--	--
Positive screening for varicosities	--	6.28	--
Diagnosed diseases			
Diagnosed anemia	--	--	--
Diagnosed anxiety	--	--	--
Diagnosed autoimmune disorder	--	--	--
Diagnosed chlamydia	1.26	--	1.69
Diagnosed depression	1.32	--	3.49
Diagnosed diabetes	--	--	2.72
Diagnosed gestational diabetes	2.10	--	--
Diagnosed gestational hypertension	--	11.99	15.23

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Diagnosed gonorrhea	--	--	--
Diagnosed hepatitis B or C	2.14	--	--
Diagnosed human immunodeficiency virus (HIV)	--	--	--
Diagnosed human papillomavirus (HPV)	--	--	--
Diagnosed hypertension	8.15	12.20	--
Diagnosed obesity	--	--	--
Diagnosed opioid use disorder	2.70	--	--
Diagnosed other infections	1.67	--	--
Diagnosed pre-eclampsia	13.80	--	25.31
Diagnosed pulmonary disease	2.93	--	--
Diagnosed Rh sensitization	--	--	6.56
Diagnosed substance use disorder	--	--	--
Diagnosed syphilis	--	--	--
Diagnosed with any STD	--	--	--
Mother, partner, and family behaviors and medical history			
Alcohol use before pregnancy	2.50	--	--
Alcohol use during pregnancy	0.79	--	--
Body mass index	22.50	6.02	13.27
Diastolic blood pressure	4.35	6.52	2.78
Did not smoke before pregnancy	2.95	100.00	19.47
Did not smoke during pregnancy	0.64	50.01	100.00
Family reported substance or alcohol use	--	--	--
First birth	3.65	15.65	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
First pregnancy	3.23	--	--
Had at least one prior ectopic pregnancy	1.22	--	--
Had at least one prior multiple birth pregnancy	0.76	--	--
Had at least one prior premature pregnancy	43.78	82.34	--
Has had recurrent pregnancy loss	--	--	--
History of abnormal Pap smears	1.77	--	--
History of blood transfusions	1.82	--	--
History of gynecological surgery	1.52	--	--
History of operations or hospitalization	0.49	--	--
Mother and/or father has at least one child	--	--	2.20
Mother is 35 years of age or older	1.20	--	--
Mother is less than 20 years of age	5.04	13.21	--
Mother is married	2.39	--	2.55
Mother is white	4.14	--	--
Mother reported current substance or alcohol use	--	--	4.20
Mother reported past substance or alcohol use	1.46	--	--
Mother was tested for fragile X syndrome	--	--	--
New medications since last pregnancy	--	--	--
Partner reported substance or alcohol use	--	--	--
Patient is 35 years of age or older as of estimated delivery date	0.52	--	--
Quit smoking before pregnancy	--	--	--
Quit smoking during pregnancy	--	--	--
Secondhand smoke exposure	1.19	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Systolic blood pressure	15.29	--	29.47
Total number of pregnancies	0.55	--	--
Weight before pregnancy	30.17	74.81	84.07
Healthcare utilization during pregnancy			
No prenatal care	--	--	--
Number of behavioral health visits during pregnancy	--	--	--
Number of cardiac surgery visits during pregnancy	--	--	--
Number of cardiology visits during pregnancy	5.41	--	--
Number of ear, nose, and throat (ENT) visits during pregnancy	--	--	4.76
Number of family medicine visits during pregnancy	5.09	5.43	1.91
Number of gastroenterology visits during pregnancy	--	--	1.69
Number of general surgery visits during pregnancy	6.65	--	--
Number of geriatrics visits during pregnancy	--	--	--
Number of internal medicine visits during pregnancy	2.64	--	--
Number of neurology visits during pregnancy	--	--	1.97
Number of neurosurgery visits during pregnancy	--	--	--
Number of OB-GYN visits during pregnancy	100.00	45.41	23.19
Number of orthopedics visits during pregnancy	3.46	5.91	--
Number of otolaryngology visits during pregnancy	--	--	--
Number of pain management visits during pregnancy	--	--	4.61
Number of pediatrics visits during pregnancy	--	--	--
Number of physical medicine and rehab visits during pregnancy	--	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Number of psychiatry visits during pregnancy	--	--	--
Number of psychology visits during pregnancy	--	--	--
Number of pulmonology visits during pregnancy	--	--	--
Number of rheumatology visits during pregnancy	--	--	--
Number of sports medicine visits during pregnancy	0.48	--	--
Number of thoracic surgery visits during pregnancy	--	--	--
Number of unknown specialty visits during pregnancy	--	--	--
Number of vascular surgery visits during pregnancy	--	--	--
Prenatal care started in second trimester	--	--	--
Prenatal care started in third trimester	--	--	--
Total number of prenatal visits in the first trimester	6.64	--	--
Total number of prenatal visits in the second trimester	18.87	25.75	1.99
Prenatal counseling			
Prenatal counseling on Abnormal Lab Values	--	--	--
Prenatal counseling on Alcohol	--	--	--
Prenatal counseling on Anesthesia/Analgesia Plans	2.71	--	--
Prenatal counseling on Anticipated Course of Prenatal Care	--	--	--
Prenatal counseling on Avoid Travel to Areas Where Risk of Congenitally Transmitted Infection(s) is High	5.12	--	--
Prenatal counseling on Avoidance of Saunas or Hot Tubs	--	--	--
Prenatal counseling on Breast or Bottle Feeding	0.71	--	--
Prenatal counseling on Breastfeeding	--	--	--
Prenatal counseling on Childbirth Classes/Hospital Facilities	4.58	--	2.14

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Prenatal counseling on Circumcision	--	--	--
Prenatal counseling on Dental Care	--	--	--
Prenatal counseling on Depression Screening (when indicated)	--	--	--
Prenatal counseling on Domestic Violence	--	--	--
Prenatal counseling on Donating or Banking Newborn Stem Cells	--	--	--
Prenatal counseling on Environmental/Work Hazards	--	--	--
Prenatal counseling on Exercise	--	--	--
Prenatal counseling on Family Medical Leave or Disability Forms	--	--	--
Prenatal counseling on Fetal Movement Monitoring	--	--	--
Prenatal counseling on HIV and Other Routine Prenatal Tests	--	--	--
Prenatal counseling on Illicit/Recreational Drugs	0.80	--	--
Prenatal counseling on Importance of Tdap Vaccination	--	--	--
Prenatal counseling on Indications for Ultrasound	--	--	--
Prenatal counseling on Influenza Vaccine	--	--	--
Prenatal counseling on Labor Signs	0.36	--	--
Prenatal counseling on Newborn Education (Newborn Screening, Jaundice, SIDS, Car Seat)	--	--	--
Prenatal counseling on Nutrition and Weight Gain Counseling: Special Diet	--	--	--
Prenatal counseling on Postpartum Depression	--	--	--
Prenatal counseling on Postpartum Family Planning/Tubal Sterilization	0.78	--	--
Prenatal counseling on Postterm Counseling	--	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Prenatal counseling on Risk Factors Identified By Prenatal History	--	--	--
Prenatal counseling on Screening for Aneuploidy	--	--	--
Prenatal counseling on Seat Belt Use	--	--	--
Prenatal counseling on Selecting a Newborn Care Provider	0.73	--	--
Prenatal counseling on Sexual Activity	--	--	--
Prenatal counseling on Signs and Symptoms of Pregnancy-Induced Hypertension	--	--	--
Prenatal counseling on Signs and Symptoms of Preterm Labor	--	--	--
Prenatal counseling on Smoking Counseling	--	--	--
Prenatal counseling on Teratogens	--	--	2.23
Prenatal counseling on Tobacco (Ask, Advise, Assess, Assist, and Arrange)	--	--	--
Prenatal counseling on Toxoplasmosis Precautions (Cats/Raw Meat)	--	--	--
Prenatal counseling on Travel	--	--	--
Prenatal counseling on Tubal Sterilization Consent	--	--	--
Prenatal counseling on Use of Any Meds (Including Supl, Vitamins, Herbs, or OTC Drugs)	--	--	--
Prenatal counseling on Use of Insect Repellent to Prevent Vector-borne Disease	--	--	--
Prenatal counseling on VBAC Counseling	--	--	--
Zip codes of mother's residence			
Zip code = 15012	--	--	--
Zip code = 15062	--	--	--
Zip code = 15085	--	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Zip code = 15089	--	--	--
Zip code = 15131	--	--	--
Zip code = 15146	--	--	--
Zip code = 15479	--	--	1.74
Zip code = 15501	--	--	--
Zip code = 15601	--	--	--
Zip code = 15610	0.84	--	--
Zip code = 15612	--	--	1.91
Zip code = 15613	--	--	--
Zip code = 15618	--	--	--
Zip code = 15619	--	--	--
Zip code = 15620	--	--	--
Zip code = 15622	--	--	--
Zip code = 15624	--	--	--
Zip code = 15626	--	5.48	--
Zip code = 15627	1.35	--	--
Zip code = 15628	--	--	--
Zip code = 15632	5.94	--	--
Zip code = 15634	--	--	--
Zip code = 15636	--	--	--
Zip code = 15637	--	--	--
Zip code = 15638	--	--	--
Zip code = 15639	--	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Zip code = 15642	6.32	--	--
Zip code = 15644	--	--	--
Zip code = 15650	--	--	--
Zip code = 15655	--	--	--
Zip code = 15658	--	--	--
Zip code = 15661	0.42	--	--
Zip code = 15665	--	--	--
Zip code = 15666	2.80	--	--
Zip code = 15670	--	--	--
Zip code = 15671	--	--	--
Zip code = 15672	--	--	--
Zip code = 15675	0.55	--	--
Zip code = 15676	--	--	--
Zip code = 15678	--	--	--
Zip code = 15679	0.53	--	--
Zip code = 15681	0.55	--	--
Zip code = 15683	--	--	--
Zip code = 15684	--	--	--
Zip code = 15687	--	--	--
Zip code = 15688	--	--	--
Zip code = 15690	--	--	--
Zip code = 15692	--	--	--
Zip code = 15696	--	--	--

Variable	Pre-term Birth (<=37 weeks) Model	LBW Model	SGA Model
Zip code = 15697	5.19	--	2.06
Zip code = 15698	--	--	--
Zip code = 15717	--	--	--
Zip code = 15923	--	--	--
Zip code = 15944	--	--	--
Infrequent zip code (less than 10 observed births)	5.27	--	--

APPENDIX B: Additional Alleghany County Output

<need to add>