Innovation in Applied Nursing Informatics
G. Strudwick et al. (Eds.)

doi:10.3233/SHTI240150

Black-White Differences in Chronic Stress Exposures to Predict Preterm Birth: Interpretable, Race/Ethnicity-Specific Machine Learning Models

Sangmi KIM,[a,1] Melinda K. HIGGINS,[a] Patricia BRENNAN,[b] George M. SLAVICH,[c] Li XIONG,[d] Vicki HERTZBERG,[a] Ursula KELLY,[a] and Anne L. DUNLOP[e]
[a] *Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, USA*
[b] *Department of Psychology, Emory University, Atlanta, GA*
[c] *Department of Psychiatry and Biobehavioral Sciences, UCLA, Los Angeles, CA, USA*
[d] *Department of Computer Science, Emory University, Atlanta, GA*
[e] *Department of Gynecology and Obstetrics, School of Medicine, Emory University, Atlanta, GA, USA*

Abstract. We developed Multivariate Adaptive Regression Splines (MARS) machine learning models of chronic stressors using the Pregnancy Risk Assessment Monitoring System data (2012-2017) to predict preterm birth (PTB) more accurately and identify chronic stressors driving PTB among non-Hispanic (N-H) Black and N-H White pregnant women in the U.S. We trained the MARS models using 5-fold cross-validation, whose performance was evaluated with AUC. We computed variable importance for PTB prediction. Our models showed high accuracy (AUC: 0.754-0.765). The number of prenatal care visits, premature rupture of membrane, and medical conditions were the most important variables in predicting PTB across the populations. Chronic stressors (e.g., low maternal education and violence) and their correlates were pivotal for PTB prediction only for N-H Black women. Interpretable, race/ethnicity-specific MARS models can predict PTB accurately and explain the most impactful life stressors and their magnitude of effect on PTB risk among N-H Black and N-H White women.

Keywords. Chronic stress, preterm birth, disparity, machine learning, PRAMS

1. Introduction

In the U.S., the Black-White inequalities in preterm birth (PTB; < 37 weeks' gestation) have persisted over the years, such that non-Hispanic (N-H) Black (14.8%) women are 55% more likely than N-H White women (9.5%) to experience PTB [1]. Nevertheless, the underlying causes of this Black-White difference are not fully understood. While well-established maternal risk factors explain only about half of the PTB risk [2], growing evidence attributes the remaining unexplained risk to chronic stress [3]. However, study findings are mixed regarding the effects of chronic stress exposures

[1] Corresponding Author: Sangmi Kim, Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Road NE, Atlanta, GA 30322, sangmi.kim@emory.edu, +1 (484) 821-7337

(i.e., chronic stressors) on PTB. One contributor to this inconsistency is limitations in study design and modeling to capture the complexities around women's chronic stressors, which have been conceptualized variably across studies and out of racial/ethnic contexts. Also, the existing statistical models that included stress as variables assumed linear and independent associations between stressors and between stressors and outcomes. However, such models are less effective in capturing the dynamics of chronic stressors that are synergistic, accumulate over time, and vary in types and effects by race/ethnicity.

To address these evidence gaps, the current study employed a more flexible and sophisticated method, such as machine learning (ML), for more accurate computation of chronic stressors and prediction of PTB among N-H Black and N-H White women in the U.S. An increasing number of studies have used various ML models to predict PTB and different data types, mostly electronic health records data collected from local hospitals [4]. Only a handful of previous studies used large population data (e.g., national survey data) that contained variables rich in socioeconomic, psychological, or behavioral factors beyond biomedical factors and were more representative of the population of pregnant women in the U.S. [5–7]. Furthermore, previous studies using ML paid less attention to understanding the implications of predictions, making the developed ML models opaque and, in turn, less trustworthy.

Therefore, this study aimed to (1) develop ML models of chronic stress exposures to predict PTB risk among N-H Black women, N-H White women, and racial/ethnic groups combined; (2) evaluate the models' prediction accuracy; (3) identify and compare important features for PTB prediction among the three models. To the best of our knowledge, this study is the first to develop interpretable ML models with various chronic stressors as predictors, along with sociodemographic, medical, and behavioral factors, to predict PTB among N-H Black and N-H White pregnant women in the U.S. by analyzing a large national survey data.

2. Methods

2.1. Data Source and Study Population

This secondary data analysis used data from the Pregnancy Risk Assessment Monitoring System (PRAMS) linked with birth certificate data (2012-2017). PRAMS is an ongoing, population-based surveillance project established by the Centers for Disease Control and Prevention (CDC) to monitor maternal attitudes and experiences before, during, and shortly after pregnancy [8]. Each state's PRAMS questionnaire is unique, even though most items are shared in common across states. The study sample consisted of first-time mothers who: (1) were aged younger than 50 years at the time of childbirth; (2) delivered live singleton births without birth defects; (3) identified themselves as N-H Black or N-H White. After removing missing data, the final sample size was 78,356 (representing 359,855 women with the sampling weight taken into account).

2.2. Measures

44 out of 669 variables were selected and modeled. The outcome variable, gestational age, was dichotomized into PTB and term birth. The chronic stressors included health

insurance coverage during pregnancy, yearly total household income, maternal educational attainment, receiving the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), physical abuse by a husband/partner before and during pregnancy, and 11 stressful life events (SLEs). As enhancers of stress, we also included depression before pregnancy. The sociodemographic factors included maternal race/ethnicity, maternal age, and marital status. The medical factors included the number of pregnancy terminations in the past, presence of pre-pregnancy health conditions (e.g., diabetes mellitus), pre-pregnancy BMI, gestational diabetes, pregnancy complications (e.g., fever), and other medical risk factors. Lastly, the behavioral factors encompassed multivitamin intake, pregnancy intention, the receipt of prenatal care (PNC), initiation of the PNC in the first trimester, and smoking before and during pregnancy. A majority of these variables were categorical. In addition, survey years and U.S. states were modeled to factor in potential temporal and spatial variations.

2.3. Analysis Procedures

We used a Multivariate Adaptive Regression Splines (MARS) model to predict PTB among three groups: N-H Black women, N-H White women, and both. MARS is a nonparametric, multivariate regression method that can estimate complex non-linear relations by a series of spline (i.e., piecewise curve) functions of the predictor variables. As a nonparametric approach, MARS does not make any underlying assumptions about the distribution of the predictor variables [9]. Importantly, MARS can estimate the relative feature importance via generalized cross-validation (GCV) [9]. We split the data into training (70%) and test sets (30%). Given the unbalanced data, we partitioned the data in a way that each of the training and test sets contained the same PTB:term birth ratio. We built three models: 1) a baseline model without feature interactions; 2) a second-degree interaction model; and 3) a third-degree interaction model. We implemented 5-fold cross-validation to select the model with the smallest residual, which was evaluated on the test set later. We also analyzed both original and weighted data to develop ML models for comparison. The original data were the ones collected by the CDC in a way that mothers of low-birth-weight infants, those living in high-risk geographic areas, and racial/ethnic minority groups were oversampled [10]. The weighted data were the ones that the sampling weight was applied to represent the population of pregnant women who birthed in certain states and survey years. However, how to include sampling weight in ML models is not clearly documented, and developing weighted ML models requires extensive computing resources [11]. Thus, we approximated weighting ML models by replicating each observation by the highest integer number of the assigned sampling weight and training and testing ML models on those replicated data. Finally, we calibrated our best-performing model for each population using logistic, isotonic, or beta calibration methods. We evaluated the model performance via the Area Under the Receiver Operating Characteristic Curve (AUC). To make the ML models interpretable, we originally created a white-box model, like MARS, and simultaneously interpreted already trained MARS models post hoc by assessing the feature importance, partial dependence plot, and individual conditional expectation curve. All data analysis was conducted using R version 4.0.2 (2020-06-22).

3. Results

We compared the unweighted and weighted model performance according to the study population, interaction, and dataset (**Table 1**). While both unweighted and weighted models showed fairly high accuracy, the weighted models showed lower accuracy than the unweighted models across the board (AUC_w = 0.754-0.765, AUC_{uw} = 0.799-0.814). The prediction accuracy stayed the same after calibration. Our final models identified the number of PNC visits, premature rupture of membrane (PROM), and medical conditions as the top three important features across the racial/ethnic groups. Unlike N-H White women, the important features of N-H Black women included a range of chronic stressors, such as maternal education, household income, SLEs (e.g., move to a new address and imprisonment of husband/partner/self), and physical abuse during pregnancy. Moreover, hypertension before pregnancy, states (i.e., GA, CO, and LA), history of pregnancy termination, BMI before pregnancy, multivitamin intake, and smoking were identified as important among N-H Black women, most of which are closely associated with chronic stress. On the other hand, the initiation of PNC in the first trimester and BMI before pregnancy were important features of N-H White women.

Table 1. Preterm risk prediction accuracy by model characteristic

	Training			Testing
	No Interaction	2-way Interaction	3-way Interaction	
	Unweighted			
Pooled	0.814 (0.809-0.819)	**0.815 (0.810-0.820)**	0.815 (0.810-0.820)	0.813 (0.805-0.820)
N-H Black	0.816 (0.806-0.820)	**0.820 (0.810-0.830)**	0.817 (0.807-0.827)	0.799 (0.783-0.815)
N-H White	0.814 (0.809-0.820)	**0.815 (0.809-0.821)**	0.815 (0.809-0.821)	0.814 (0.805-0.823)
	Weighted			
Pooled	0.755 (0.754-0.756)	0.757 (0.756-0.758)	**0.758 (0.757-0.759)**	0.757 (0.756-0.759)
N-H Black	0.737 (0.734-0.739)	0.751 (0.748-0.753)	**0.757 (0.754-0.759)**	0.754 (0.750-0.757)
N-H White	**0.765 (0.764-0.767)**	0.759 (0.758-0.760)	0.759 (0.758-0.760)	0.765 (0.763-0.766)

4. Discussion

This study built interpretable and race/ethnicity-specific MARS models to predict PTB among N-H Black and N-H White pregnant women in the U.S. using large national survey data. The study compared the prediction accuracy between the models with different specifications, as well as with different datasets: original (unweighted) data that oversampled high-risk pregnant women and weighted data more representative of the pregnant women in the U.S. Overall, the MARS models predicted PTB with high accuracy, where the weighted models, despite their generalizability, slightly underperformed the unweighted models. More importantly, there were commonalities and differences in the important features for PTB prediction between N-H Black and N-H White women. The number of PNC visits, PROM, and medical conditions were the most important features for both racial/ethnic groups. Only the N-H Black model identified several chronic stressors and their medical and behavioral correlates as important features for PTB prediction.

Our MARS models performed better than linear models in previous studies [7]. Moreover, their prediction accuracy was higher than previous studies using different ML models and national data, including the PRAMS data [5–7]. While the important features for PTB prediction in this study were evidenced by the existing literature, those unique to N-H Black and N-H White women could be found only because we stratified the data and developed race/ethnicity-specific models. This study has some limitations. Our study inherited limitations of the self-reported measures (e.g., response bias). It also included only individual-level factors to predict PTB due to data limitations. Our weighted models did not directly factor in the survey's sampling weight for modeling due to the extensive computing power required. Lastly, the input features in the models did not include stress biomarkers, which could have increased the prediction accuracy.

5. Conclusions

Our study established the role of interpretable, race/ethnicity-specific ML models as a useful tool to generate risk prediction systems that could inform key factors behind the PTB prediction unique to N-H Black and N-H White pregnant women for targeted prevention and intervention. That multiple chronic stressors and their correlates made a significant and unique contribution to PTB prediction among N-H Black but not N-H White women called for more efforts to tackle the identified chronic stressors to alleviate the Black-White inequalities in PTB. Lastly, incorporating such risk prediction systems into care may support a diagnosis of PTB and assist healthcare providers in preparing expectant families and managing at-risk PTB infants.

Acknowledgements

This study was supported by the NIH/NINR under Award Number K01NR019651.

References

[1] M. Osterman, B. Hamilton E., J. Martin, A. Driscoll, and C. Valenzuela, Births: Final Data for 2021, Atlanta, Georgia, 2023. doi:10.15620/cdc:122047.
[2] American College of Obstetricians and Gynecologists Committee on Health Care for Undeserved Women, ACOG Committee Opinion No. 343: Psychosocial Risk Factors: Perinatal Screening and Intervention, *Obstetrics & Gynecology*. **108** (2006) 469. doi:10.1097/00006250-200608000-00046.
[3] R.E. Behrman, and A.S. Butler, Committee on Understanding Premature Birth and Assuring Healthy Outcomes IoMotNA. Preterm Birth: Causes, Consequences, and Prevention, (2007).
[4] T. Włodarczyk, S. Płotka, T. Szczepański, P. Rokita, N. Sochacki-Wojcicka, J. Wojcicki, M. Lipa, and T. Trzciński, Machine learning methods for preterm birth prediction: a review, *Electronics (Basel)*. **10** (2021) 586.
[5] A. Esty, M. Frize, J. Gilchrist, and E. Bariciak, Applying data preprocessing methods to predict premature birth, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018: pp. 6096–6099.
[6] M. Frize, N. Yu, and S. Weyand, Effectiveness of a hybrid pattern classifier for medical applications, *Int J Hybrid Intell Syst*. **8** (2011) 71–79.
[7] A. Koivu, and M. Sairanen, Predicting risk of stillbirth and preterm pregnancies with machine learning, *Health Inf Sci Syst*. **8** (2020) 14. doi:10.1007/s13755-020-00105-9.
[8] Centers for Disease Control and Prevention, About PRAMS, (2022). https://cdc.gov/prams/about/prams-faq.htm (accessed December 5, 2023).
[9] R. Menon, G. Bhat, G.R. Saade, and H. Spratt, Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth, *Acta Obstet Gynecol Scand*. **93** (2014) 382–391.

[10] H.B. Shulman, D. V. D'Angelo, L. Harrison, R.A. Smith, and L. Warner, The Pregnancy Risk Assessment Monitoring System (PRAMS): Overview of Design and Methodology, *Am J Public Health*. **108** (2018) 1305–1313. doi:10.2105/AJPH.2018.304563.

[11] N. MacNell, L. Feinstein, J. Wilkerson, P.M. Salo, S.A. Molsberry, M.B. Fessler, P.S. Thorne, A.A. Motsinger-Reif, and D.C. Zeldin, Implementing machine learning methods with complex survey data: Lessons learned on the impacts of accounting sampling weights in gradient boosting, *PLoS One*. **18** (2023) e0280387.