Abstract Title: Determining an Association Between Family History of Cardiometabolic Disease and the Incidence of Preeclampsia in Pregnant Women of Different Racial and Age Groups using Artificial Intelligence and Machine Learning

Authors: Yara Yaghi[1], Zahra Izzi[1], Omar Barabandi[1], Ayham Elayan[1]
Institutions: 1. MedStar Health Research Institute

Background and Objective: Maternal cardiometabolic health encompasses the health of women pre, during, and post childbirth, and is important to monitor throughout the course of pregnancy. In order to pinpoint risk factors for various cardiometabolic diseases in mothers, we researched the correlation between mothers with a family history of cardiometabolic diseases, such as hypertension, cholesterol, diabetes, and heart failure, and looked for an association with mothers who develop preeclampsia. In order to conduct this, we used the Cardiometabolic Correlates Synthetic Dataset, specifically the Family History dataset and the New Event dataset.

Methods and Analysis: We used four key datasets: the Person dataset, the Visits dataset, the Diagnosis dataset, and the Family History dataset. After cleaning the data by removing duplicates and ensuring only unique variables were included, we merged these datasets. This data cleaning and merging process left us with a final cohort of 15,229 unique patients. For our analysis, we identified several predictor variables, including demographic factors such as race, age, ethnicity, language, and marital status, as well as clinical variables including admit type, facility type, insurance type, and encounter type. We also included specific family history conditions such as hypertension, diabetes, and cardiovascular disease, and the top 50 reasons for visits. Our primary outcomes of interest were preeclampsia, hypertension, and high blood pressure. We then split our data, using 70% to train our logistic regression model and reserving 30% for testing the model's performance. This approach ensured that the model we developed can generalize well to new data, allowing us to predict health outcomes effectively in the population.

Results: The logistic regression model was accurate at 71%, with good calibration and discrimination, particularly at high probabilities. By race, African American, Asian, and 'Other' racial groups carried a 4-5% higher risk for preeclampsia compared to White patients. Trends by age showed that younger women (18-24) were at highest risk of preeclampsia at approximately 59%, with declining risk up to approximately 55.5% in the 35-44 age group. For visit purpose, routine obstetric care was most common across all racial groups. However, the high cesarean

delivery rates noted among White women can contribute to their increasing risk of preeclampsia with older age.

Conclusion: Our analysis has highlighted significant variations in preeclampsia rates across different racial and age groups, particularly the decrease in rates for African American as they age, contrasted with an increase for White women. These findings emphasize the critical role that healthcare visit patterns, such as routine obstetric care and delivery-related visits, play in indicating health outcomes. For future studies we suggest consider exploring the underlying causes of the observed differences in preeclampsia rates across racial and age groups. Additionally, expanding the study to include a larger and more diverse sample could help to validate these findings and identify other factors that may contribute to preeclampsia risk.