Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort

Jonathan S. Schor, PhD; Adesh Kadambi, BEng; Isabel Fulcher, PhD; Kartik K. Venkatesh, MD, PhD; Mark A. Clapp, MD, MPH; Senan Ebrahim, MD, PhD; Ali Ebrahim, PhD; Timothy Wen, MD, MPH

BACKGROUND: Hypertensive disorders of pregnancy (HDP) are significant drivers of maternal and neonatal morbidity and mortality. Current management strategies include early identification and initiation of risk mitigating interventions facilitated by a rules-based checklist. Advanced analytic techniques, such as machine learning, can potentially offer improved and refined predictive capabilities.

OBJECTIVE: To develop and internally validate a machine learning prediction model for hypertensive disorders of pregnancy (HDP) when initiating prenatal care.

STUDY DESIGN: We developed a prediction model using data from the prospective multisite cohort Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) among low-risk individuals without a prior history of aspirin utilization for preeclampsia prevention. The primary outcome was the development of HDP. Random forest modeling was utilized to develop predictive models. Recursive feature elimination (RFE) was employed to create a reduced model for each outcome. Area under the curve (AUC), 95% confidence intervals (CI), and calibration curves were utilized to assess discrimination and accuracy. Sensitivity analyses were conducted to compare the sensitivity and specificity of the reduced model compared to existing risk factor-based algorithms.

RESULTS: Of 9,124 assessed low risk nulliparous individuals, 21% (n=1,927) developed HDP. The prediction model for HDP had satisfactory discrimination with an AUC of 0.73 (95% CI: 0.70, 0.75). After RFE, a parsimonious reduced model with 30 features was created with an AUC of 0.71 (95% CI: 0.68, 0.74). Variables included in the model after RFE included body mass index at the first study visit, pre-pregnancy weight, first trimester complete blood count results, and maximum systolic blood pressure at the first visit. Calibration curves for all models revealed relatively stable agreement between predicted and observed probabilities. Sensitivity analysis noted superior sensitivity (AUC 0.80 vs 0.65) and specificity (0.65 vs 0.53) of the model compared to traditional risk factor-based algorithms.

CONCLUSION: In cohort of low-risk nulliparous pregnant individuals, a prediction model may accurately predict HDP diagnosis at the time of initiating prenatal care and aid employment of close interval monitoring and prophylactic measures earlier in pregnancy.

Key words: Hypertensive disorders of pregnancy, Machine learning, Risk prediction

Introduction

Hypertensive disorders of pregnancy (HDP), which includes gestational hypertension, preeclampsia with and without severe features, and eclampsia, affect up to 10% of all pregnancies and are a significant contributor to maternal and neonatal morbidity and mortality.[1–3] Short-term complications include renal and hepatic dysfunction, stroke, and seizures, and long-term complications include cardiovascular disease and chronic hypertension in later life.[4,5] In addition to maternal complications, HDP has significant impacts on fetal outcomes, including an increased risk of fetal death, preterm delivery, and the subsequent clinical sequelae of prematurity.[3,4] The

From the Delfina Care Inc, San Francisco, CA, USA (Schor, Kadambi, Fulcher, Venkatesh, Clapp, Ebrahim, Ebrahim and Wen); University of California, San Francisco (UCSF) Medical Scientist Training Program, San Francisco, CA, USA (Schor); Department of Biomedical Engineering, University of Toronto, Toronto, Canada (Kadambi); Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts, USA (Fulcher); Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, Ohio State University Medical Center, Columbus, OH, USA (Venkatesh); Division of Maternal Fetal Medicine, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco (UCSF), San Francisco, CA USA (Wen)

The parent study was supported by grant funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD): U10 HD063036; U10 HD063072; U10 HD063047; U10 HD063037; U10 HD063041; U10 HD063020; U10 HD063046; U10 HD063048; and U10 HD063053. In addition, support was provided by Clinical and Translational Science Institutes: UL1TR001108 and UL1TR000153. We acknowledge NICHD DASH for providing the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-be data that was used for this research.

Disclosure: Drs. Wen, Venkatesh, and Clapp serve as consultants on the Medical Advisory Board for Delfina, Inc. Dr. Schor is a data scientist for Delfina. Dr. Wen also serves as the Associate Chief Medical Officer for Delfina. Mr. Adesh Kadambi is a data scientist for Delfina. Dr. Isabel Fulcher is the Vice President of Data Science for Delfina. Dr. A. Ebrahim is the Chief Technology Officer for Delfina. Dr. S. Ebrahim is the Chief Executive Officer for Delfina.

Corresponding author: Timothy Wen, MD, MPH. timothy.wen2@ucsf.edu

2666-5778/$36.00

© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

http://dx.doi.org/10.1016/j.xagr.2024.100386

AJOG Global Reports at a Glance

Why was this study conducted?

This study was conducted to develop and internally validate a machine learning prediction model for predicting hypertensive disorders of pregnancy (HDP) in the first trimester using features typically ascertained by the first prenatal care visit found in a publicly available data set.

Key findings

- In a low-risk nulliparous pregnancy cohort, a prediction model for hypertensive disorders of pregnancy may accurately predict HDP diagnosis at the time of initiating prenatal care with satisfactory discrimination, an area under the receiver operator curve (AUC) of 0.73 (95% CI: 0.70, 0.75).
- A reduced parsimonious model developed using recursive feature elimination exhibits similar discriminatory capability (AUC: 0.71, 95% CI: 0.68, 0.74).
- Sensitivity analyses noted an improved sensitivity and specificity in predicting HDP when utilizing this model over traditional risk factors analysis.

What does this add to what is known?

- The prediction of hypertensive disorders of pregnancy early in pregnancy using a machine learning approach derived from publicly available data is feasible with satisfactory discrimination and superior sensitivity and specificity when compared to current risk-based algorithms.
- Implementation of this algorithm could potentially identify more patients at risk for HDP and by extension, could benefit from preeclampsia prevention strategies.

frequency of HDP has increased from 6.0% to 12.0% of all delivery hospitalizations from 2000 to 2018.[6–8]

Because of the increasing prevalence of HDP and the severity of both maternal and fetal complications, studies in the past have attempted to develop therapeutics and risk prediction algorithms to identify individuals who are at highest risk of developing HDP.[9–11] Earlier identification of high-risk patients may benefit from risk mitigating interventions such as the prophylactic use of aspirin, closer interval surveillance, and timed delivery; all of which could potentially reduce the incidence of disease and by extension, the associated short- and long-term complications.[11–14] The current risk stratification model, developed by the American College of Obstetricians and Gynecologists and supported by the United States Preventative Services Task Force, employed involves a risk factor screening rules-based algorithm to identify individuals who are at higher risk of developing hypertensive disorders of pregnancy.[15] Additionally, previously published predictive models for HDP rely on a combination of serum analytes or maternal sociodemographic and clinical characteristics that are not a part of routine prenatal care, limiting their utility in clinical practice.[16–18]

Advanced analytic techniques, such as machine learning, offer the ability to leverage the computational power of computers to identifying relationships and nuances in data that traditional risk factor-based stratification algorithms may not account for.[19] Given this gap, the objective of this study was to develop a prediction model to identify individuals at high risk of developing HDP among low-risk nulliparous individuals at the time of initiating prenatal care. We utilized a machine learning based approach using data available in the electronic health record as part of routine prenatal care.

Methods

Parent study

This study was a secondary analysis utilizing data from the Nulliparous Pregnancy Outcomes Study: Monitoring mothers-to-be (nuMoM2b). This was a prospective cohort that enrolled 10,038 nulliparous pregnant individuals between 6 weeks 0 days and 13 weeks 6 days from 2010 to 2013. Individuals were excluded if they were less than 13 years old, had 3 or more prior pregnancy losses, planned termination of pregnancy, had fetal malformations concerning for aneuploidy, had donor oocyte pregnancy, and were unable to provide full consent.[20] Individuals participated in 3 study visits during their pregnancy, and data including maternal characteristics, clinical assessments, medical record abstraction, and standardized questionnaires were abstracted.[20] Additional details of study procedures have been well described in the literature.[20]

Outcome

The primary outcome was a diagnosis with any HDP between 20 weeks gestation and 2 weeks postpartum, an outcome utilized by prior nuMoM2b secondary analyses.[21–23] HDP included diagnoses of gestational hypertension, preeclampsia with/without severe features, superimposed preeclampsia with/without severe features, and eclampsia. The outcome window included the postpartum period as postpartum preeclampsia diagnoses have significant burdens on patient and health systems like antepartum and intrapartum diagnoses, thus warranting inclusion into our outcome.

Selection of predictors

For the primary outcome, we only considered variables that were collected at the first study visit (between 6 weeks 0 days and 13 weeks 6 days estimated gestational age) and that are part of the electronic medical record by the first prenatal visit. This time cutoff was selected based on the United States Preventive Services Task Force (USPSTF) recommendation that individuals at risk of HDP be started on low-dose aspirin after 12 weeks of pregnancy.[10] Following assessment of the available variables by healthcare providers (TW, KKV, MAC) for the criteria above, we selected 143 predictors for further assessment, including demographic and health factors, laboratory results, medical conditions and diagnosis, and family health history. Any predictors in either set that had missing values were specifically

encoded as "-10000" for individuals for which the value was unknown, allowing the subsequent models to learn even in the absence of a predictor. See Appendix 1 for a full list of features utilized in the initial model building. We did not include information on maternal race and ethnicity in the training of our prediction model, but we used this variable to evaluate model performance by race and ethnicity in stratified analyses.

Model development

For our primary outcome of HDP diagnosis, we developed a random forest model to differentiate between individuals with and without HDP with the full set of predictors. A random forest model uses random subsets of training data, sampled with replacement, to generate a collection of decision trees to predict the outcome.[24] This model was selected as it was able to flexibly handle numerous predictors that may contain non-linear effects, to incorporate variable selection methods to identify the top predictors, is a commonly used model in medical literature with consistent performance in classification tasks.[25,26] The overall outcome of the random forest is selected by majority consensus among all trees, allowing for strengths of one tree to compensate for weaknesses or gaps in others.[24] To generate the random forest, we first randomly allocated individuals to a training set (80% of patients) or a holdout testing set (20% of patients); the model was trained on the training set and evaluated on the testing set. To find optimal hyperparameters for the model, such as the number of decision trees to include, we used gridsearch with 3-fold cross validation with sensitivity as the scoring metric. This method chooses the hyperparameter values that will yield the highest sensitivity by evaluating various combinations of hyperparameter values using training and testing sets with 3:1 allocation. This yielded a maximum tree depth of 50, a minimum number of samples per leaf node of 1, and a minimum number of samples required to split, and an internal node of 2. We utilized balanced class weights to account for imbalance in HDP diagnosis, which ensures that misclassification in the minority class (HDP diagnosis) will be penalized similarly to the majority class (no HDP diagnosis). To improve model interpretability, we created a "reduced model" with 30 predictors using recursive feature elimination, which iteratively removed features to create a parsimonious model that optimized model predictability.

Model evaluation

All random forest models were evaluated on the holdout testing sets. Our evaluation metrics include plots of the receiver operator characteristic (ROC) curves and the area under the curve (AUC). The ROC curve is threshold-agnostic, meaning that the performance of our model is displayed without a specific sensitivity or specificity threshold in mind − allowing one to choose the desired cutoffs based on a clinical objective. AUC is often used as the primary performance metric for prediction models as it provides a numeric summary for the ROC curve. We report ROC curves and AUC for the full and reduced models for both outcomes. The 95% confidence intervals (CI) for each AUC was calculated using the fast DeLong method.[27] Statistical comparison of the AUC's between the full and reduced models was accomplished using the DeLong's test.[27,28] Calibration curves were calculated using the sklearn "calibration_curve" function in Python to assess how close predicted probabilities were to the truth.[29]

We conducted several sensitivity analyses. We first evaluated the primary random forest model performance stratified by self-reported maternal race and ethnicity (Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, and Other). We subsequently compared the performance of the primary model with and without the inclusion of maternal race and ethnicity. A modified DeLong's test for AUC comparison of independent samples using unpaired t-tests with unequal sample size and variance was utilized for comparison between maternal race/ethnicity groups with non-Hispanic white as the reference group.[27,28] Second, we assessed whether the prediction model could be used to better assign individuals to receipt of aspirin prophylaxis to prevent preeclampsia compared to current guidelines. Currently, the American College of Obstetricians and Gynecologists (ACOG) recommends a standard of low dose aspirin for prevention of HDP in patients who are at high risk.[30] We calculated the number of patients who would be prescribed aspirin using ACOG criteria compared to the number of patients who would be recommended aspirin using the current model at different sensitivity thresholds.[30] We evaluated moderate and high risk factors including: history of kidney disease, pregestational diabetes, hypertension, antiphospholipid syndrome or acquired thrombophilia, systemic lupus erythematosus, family history of preeclampsia, maternal age $\geq$ 35 years, BMI $>$ 30, income under the poverty line, and self-reported non-Hispanic Black race and ethnicity in concordance with current ACOG guidelines.[30]

Results

Of the 10,038 total patients enrolled in the parent study, 9,289 consented to having their anonymized data released into shared National Institutes of Health databases. From this subset, we excluded 165 patients who were on aspirin at any time during pregnancy. The final sample size was 9,124 patients, of whom 1,927 (21%) were diagnosed with HDP between 20 weeks gestation and 2 weeks postpartum. Mean maternal age was not statistically different between individuals who developed HDP and did not develop HDP (Table 1). The HDP cohort had a higher BMI at the first visit (28.4% vs. 25.8%, $p<.001$), higher proportion of pregestational diabetes (3.2% vs. 1.0%, $p<.001$), and higher proportion of individuals with histories of hypertension (15.8% vs. 2.6%, $p<.001$) (Table 1). Among patients who developed an HDP, 1585 (82%) had gestational hypertension, 186 (10%) had mild preeclampsia, 48 (2%) had superimposed preeclampsia, 103 (5%) had severe preeclampsia and 5 (<1%) had eclampsia. Additional

TABLE 1
Distribution of all reduced model variables, race and ethnicity, and hypertensive disorders of pregnancy (HDP) type by HDP diagnosis (N=9124)

Variable	Overall N=9124	HDP N=1927	No HDP N=7197	*P*-value[a]
Race and Hispanic Ethnicity				
Hispanic	1,591 (17.4%)	247 (12.8%)	1,344 (18.7%)	
Non-Hispanic White	5,457 (59.8%)	1,155 (59.9%)	4,302 (59.8%)	
Non-Hispanic Black	1,244 (13.6%)	358 (18.6%)	886 (12.3%)	
Asian	357 (3.9%)	59 (3.1%)	298 (4.1%)	
Other/not classified	475 (5.2%)	108 (5.6%)	367 (5.1%)	
Type of HDP				
None	7,197 (78.9%)			
Gestational hypertension	1,585 (17.4%)			
Mild preeclampsia	186 (2.0%)			
Superimposed preeclampsia	48 (0.5%)			
Severe preeclampsia	103 (1.1%)			
Eclampsia	5 (0.05%)			
Reduced Features, *mean (std dev) [missing N]* or *total N (%) [missing N]*				
Hemoglobin g/dL	12.9 (2.4) [610]	12.9 (2.8) [39]	12.9 (2.3) [571]	0.99
Hematocrit %	38.0 (3.0) [629]	38.0 (3.00) [44]	38.0 (3.0) [585]	0.72
MCV fL/cell	89.5 (11.3) [1348]	88.9 (5.7) [237]	89.6 (12.4) [1111]	<0.001
Platelet count x10 3 /mm 3 (x10^3/μL)	253.4 (58.0) [1116]	262.5 (60.6) [171]	250.9 (57.0) [945]	<0.001
Hemoglobin electrophoresis test result				
Normal (AA)	2,319 (25.4%)	481 (25.0%)	1,838 (25.5%)	0.99
Elevated A2	31 (0.3%)	6 (0.3%)	25 (0.3%)	
Hemoglobin S/C/F	133 (1.5%)	26 (1.3%)	107 (1.5%)	
Other hemoglobin	43 (0.5%)	9 (0.5%)	34 (0.5%)	
Unknown	6,598 (72.3%)	1,405 (72.9%)	5,193 (72.2%)	
Blood type				
A	3,107 (34.1%)	693 (36.0%)	2,414 (33.5%)	0.72
B	1,208 (13.2%)	261 (13.5%)	947 (13.2%)	
O	4,039 (44.3%)	869 (45.1%)	3,170 (44.0%)	
AB	364 (4.0%)	86 (4.5%)	278 (3.9%)	
Unknown	406 (4.4%)	18 (0.9%)	388 (5.4%)	
Chlamydial screen performed?	146 (1.6%) [1826]	29 (1.5%) [284]	117 (1.6%) [1542]	0.50
Urine culture taken?	1,201 (13.2%) [3074]	296 (15.4%) [585]	905 (12.6%) [2489]	0.02
Urine culture - Organism code[b]	[7940]	[1635]	[6305]	0.020
Nuchal translucency measured?	4,179 (45.8%) [315]	1,005 (52.2%) [2]	3,174 (44.1%) [313]	<0.001
Nuchal translucency - Measurement (mm)	1.5 (1.0) [4991]	1.5 (0.4) [936]	1.5 (1.2) [4055]	0.41
Treated for depression prior to pregnancy	1,011 (11.1%) [495]	255 (13.2%) [1]	756 (10.5%) [494]	0.020

Schor et al. Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort. AJOG Glob Rep 2024.

(continued)

TABLE 1
Distribution of all reduced model variables, race and ethnicity, and hypertensive disorders of pregnancy (HDP) type by HDP diagnosis (N=9124) *(continued)*

Variable	Overall N=9124	HDP N=1927	No HDP N=7197	*P*-value[a]
Treated for anxiety prior to pregnancy	730 (8.0%) [495]	197 (10.2%) [1]	533 (7.4%) [494]	<0.001
Age (years) at visit 1 calculated from DOB	26.8 (5.7) [10]	26.9 (5.8) [0]	26.8 (5.6) [10]	0.53
BMI (kg/m^2) at visit 1	26.4 (6.3) [208]	28.4 (7.0) [35]	25.8 (6.0) [173]	<0.001
Ever used tobacco	3,803 (41.7%) [19]	907 (47.1%) [1]	2,896 (40.2%) [18]	<0.001
Pre-pregnancy weight — lbs	152.0 (39.4) [195]	164.4 (44.4) [27]	148.6 (37.2) [168]	<0.001
Diabetes ever diagnosed	136 (1.5%) [502]	61 (3.2%) [1]	75 (1.0%) [501]	<0.001
Has ever used illegal drugs or drugs not prescribed	3,158 (34.6%) [22]	737 (38.2%) [3]	2,421 (33.6%) [19]	<0.001
History of hypertension	488 (5.3%) [266]	304 (15.8%) [2]	184 (2.6%) [264]	<0.001
History of asthma	1,176 (12.9%) [926]	270 (14.0%) [173]	906 (12.6%) [753]	0.17
History of migraines	1,652 (18.1%) [508]	396 (20.6%) [34]	1,256 (17.5%) [474]	0.03
History of UTIs	3,883 (42.6%) [508]	840 (43.6%) [34]	3,043 (42.3%) [474]	0.51
History of yeast infection	3,245 (35.6%) [508]	728 (37.8%) [34]	2,517 (35.0%) [474]	0.43
Family history of birth at <37 weeks	1,723 (18.9%) [891]	400 (20.8%) [127]	1,323 (18.4%) [764]	0.14
Family history of birth weight <2500 grams	1,242 (13.6%) [900]	295 (15.3%) [129]	947 (13.2%) [771]	0.09
Family history of preeclampsia, eclampsia, toxemia, or pregnancy-induced hypertension	888 (9.7%) [1005]	210 (10.9%) [150]	678 (9.4%) [855]	0.19
Family history of hypertension	2,333 (25.6%) [3948]	583 (30.3%) [766]	1,750 (24.3%) [3182]	<0.001
Max systolic BP at first visit	109.2 (11.0) [198]	112.7 (11.2) [29]	108.3 (10.7) [169]	<0.001
Max diastolic BP at first visit	67.0 (8.4) [198]	69.2 (8.8) [29]	66.4 (8.2) [169]	<0.001

[a] *P*-values calculated using Chi-squared test for categorical variables and independent t-test for continuous variables.; [b] Full details of all 27 organism categories omitted for brevity, missing N displayed for each group.

Schor et al. Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort. AJOG Glob Rep 2024.

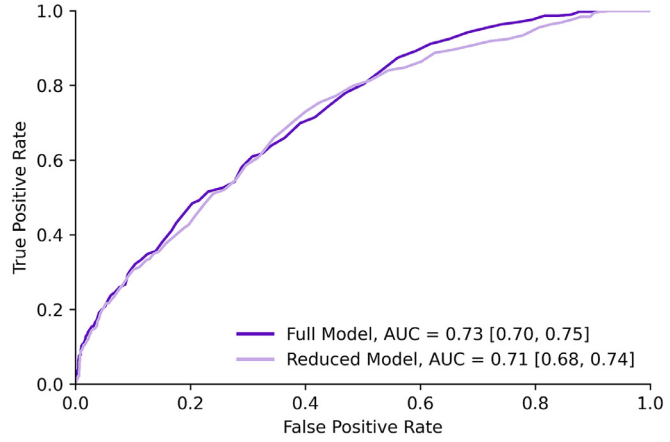
characteristics of this study population have been previously described in the literature.[20]

First, we constructed 2 random forest models, one with the full set of covariates predicting the occurrence of HDP and the other with a subset. The first model included 143 features and achieved an AUC of 0.73 (95% CI: 0.70, 0.75; Figure 1, See Supplemental Figure 1 for the associated calibration plot). The second model was a reduced model of 30 features and achieved an AUC of 0.71 (95% CI: 0.68, 0.74; Figure 1, See Supplemental Figure 2 for the associated calibration plot). There was no difference between the AUC's of the full and reduced models (p=0.12). The distribution of the 30 features and race and ethnicity stratified by HDP diagnosis are given in Table 1 and included features such as complete blood count components, blood type, urine culture, and historical features such as chronic hypertension, pregestational diabetes, and depression among others. Laboratory results referenced in the model reflected the laboratory tests collected at the initial visit. Calibration plots for both random forest models noted near perfect calibration (Supplemental Figures 1 and 2). The calibration plot noted some deviation at the higher deciles of predicted probability. Individual variable importance in the reduced model is depicted in Supplemental Figure 3, with the features most important to the model after RFE including body mass index at the first study visit, pre-pregnancy weight, complete blood count results in the first trimester, and maximum systolic blood pressure at the first study visit.

In our first sensitivity analyses, we assessed the performance of the primary reduced model when stratified by maternal race and ethnicity. Figure 2 demonstrates the AUROC curves of the primary reduced random forest model predicting the occurrence of HDP with an AUC range of 0.66 (95% CI: 0.58, 0.74) and 0.78 (95% CI: 0.63, 0.94), performing lowest among Hispanic patients and highest among Asian patients (Figure 2). Model performance did not vary by race and ethnicity (p>.05 for all). When maternal race and ethnicity was included in the primary reduced model to assess the potential benefits of inclusion, there was no statistically significant difference in performance (Reduced Model AUC with maternal race/ethnicity: 0.72, 95% CI: 0.69, 0.75) (Supplemental Figure 4).

In our second sensitivity analyses, we compared our reduced primary model performance in predicting HDP to existing ACOG aspirin recommendation guidelines in predicting HDP.

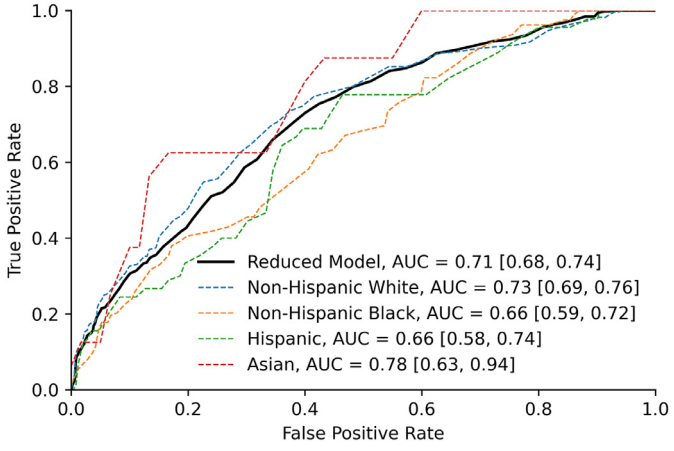
FIGURE 1
Receiver operating characteristics curve for predicting any hypertensive disorders of pregnancy among all pregnant persons for the full and reduced model



Schor et al. Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort. AJOG Glob Rep 2024.

When compared to ACOG aspirin guidelines, the current model performed better than current guidelines. Existing ACOG guidelines had a sensitivity of 0.65 and specificity of 0.53 in nuMoM2b (Figure 3). When set at the same specificity level, the current model had a sensitivity of 0.80 (Table 2). This translates to 15 additional HDP patients for every 100 HDP patients initiated on aspirin compared to ACOG guidelines. When set at the same level of sensitivity, our reduced model had a specificity of 0.65 (Table 2). This translates to 12 additional non-HDP patients for every 100 non-HDP patients who would not be initiated on aspirin using the current model compared to guidelines.

FIGURE 2
Receiver operating characteristics curves for predicting any hypertensive disorders of pregnancy by maternal race and ethnicity in the reduced model



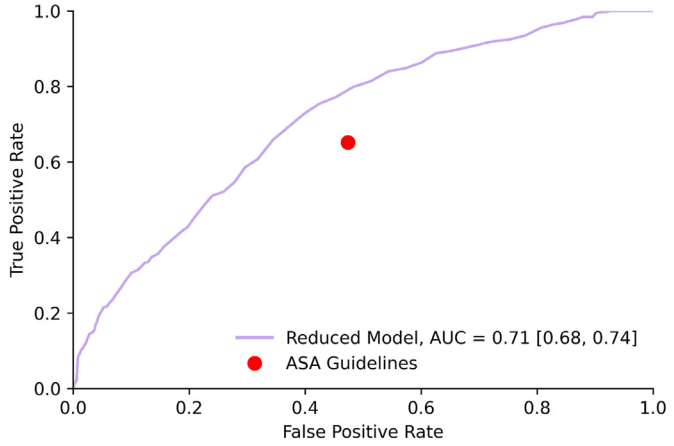
Schor et al. Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort. AJOG Glob Rep 2024.

Discussion

We found that a prediction model can accurately predict HDP diagnosis at the time of initiating prenatal care in a cohort of low-risk nulliparous pregnant individuals. We constructed both a full (inclusive 207 predictors) and reduced model to predict the development of HDP with moderate discrimination. Our findings noted that even after simplifying to a reduced model with features traditionally available at the first prenatal care visit, our predictive capability for HDP did not diminish. Features that contributed to the reduced model included early pregnancy body mass index, pre-pregnancy weight, initial complete blood count results, and maximum systolic blood pressure (Supplemental Figure 3). Prior studies have utilized advanced analytic techniques to model this outcome; however, these studies were primarily conducted at single institutions or foreign countries or utilized features that are not captured during routine prenatal care in the United States.[31–36]

The findings of our model have potential significant clinical implications. Accurate identification of patients who are at high risk for HDP, especially at the first visit, allows for not only earlier recognition of the risk of HDP, but the potential for earlier and standardized implementation of preventive measures.[37] Currently, the USPSTF and ACOG recommend the initiation of low dose aspirin as the primary preventative measure for mitigating the risk of HDP development.[9,10,30] The implementation of these guidelines is made on a risk stratification algorithm based on maternal and prior obstetric history expert opinion, and does not consider more granular, patient-specific factors such as vital signs, laboratory data, or other current pregnancy attributes. At the same level of specificity, the current prediction model had greater sensitivity for identifying individuals for aspirin prophylaxis to prevent preeclampsia compared to current ACOG guidelines (Table 2). This could potentially allow for expanded initiation of aspirin to patients who would not have qualified previously based on prior ACOG

FIGURE 3
Comparison of American College of Obstetricians and Gynecologists aspirin guidelines (ASA) and prediction of any hypertensive disorders of pregnancy among all pregnant persons in the reduced model



Schor et al. Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort. AJOG Glob Rep 2024.

guidelines but still are at high risk of HDP, thus potentially further decreasing the overall population risk of HDP.[38]

Furthermore, our model allows for improved risk stratification that would assist in increased monitoring at later stages of pregnancy, enabling earlier intervention including hospital admission and induction of labor. Prior studies have noted significant associations between long term maternal hypertensive disease relating to antepartum and intrapartum diagnoses of HDP.[5,21,39] Given this association, delivery prior to diagnosis of HDP could decrease the cardiovascular disease burden later in life. Future iterations of our model could be utilized to aid in the selection of patients for induction and serve as additional evidentiary support for risk reducing inductions of labor.[40] Additionally, given the need for therapeutic interventions to treat HDP, accurate risk stratification would assist in identifying individuals for potential randomized controlled trials or advanced therapeutics in the future.

Our sensitivity analyses found no significant difference in model performance by maternal race and ethnicity cohorts. Further, inclusion of maternal race and ethnicity into the model did not improve predictive accuracy significantly. This finding highlights the potential of advanced analytic models to reduce bias in the current healthcare setting. Physician bias and structural racism have been cited in the literature as contributors to adverse maternal and neonatal outcomes in general and can be extrapolated to management of HDP.[41] The use of an automated machine learning model to risk stratify patients can objectively identify higher risk patients for early recognition and prompt management of HDP.

Our study is one of the first to utilize data from a national prospective cohort study in a data-driven effort to predict HDP using features that are available in the electronic health record and through the course of routine prenatal care. The data were prospectively collected at multiple institutions with standardized methods and validated collection tools which provided assurances regarding data quality, validity, and capture.[20] Additionally, the data collected was real-world data that is typically captured during routine prenatal care, allowing for ease of external validation and integration. While many factors were included in the full and even reduced models, the intended utilization of this model would be an interplay between the electronic health record and an application programming interface for automated risk calculation. Furthermore, given the large sample size, we were able to internally validate the findings of our initial models. Our

TABLE 2
Early initiation of aspirin comparing prediction model for any hypertensive disorders of pregnancy to American College of Obstetricians and Gynecologists aspirin guidelines

Sens of proposed model	Spec of proposed model	Number of HDP patients correctly started on aspirin *(out of 100 HDP patients)*			Number of non-HDP patients incorrectly started on aspirin *(out of 100 non-HDP patients)*		
		Proposed	ACOG	Diff	Proposed	ACOG	Diff
0.80	0.53	80	65	15	47	47	0
0.65	0.65	65	65	0	35	47	12

Schor et al. Using machine learning to predict the risk of developing hypertensive disorders of pregnancy using a contemporary nulliparous cohort. AJOG Glob Rep 2024.

sample also contained information prior to the adoption of the ACOG aspirin guidelines, which allowed us to assess HDP prediction prior to the adoption of aspirin in standard practice. As with other risk prediction algorithms, this model will require external validation and refinement before utilization in clinical settings. While nuMoM2b was inclusive of many features of routine pregnancy, there are likely additional features collected during prenatal care that could aid in refining the precision of this model. Finally, nuMoM2b only contained data from nulliparous patients, leaving questions regarding generalizability to the general population including multiparous patients until further external validation studies are conducted. Based on established correlations between HDP risk in sequential pregnancies, we hypothesize that generating predictive models in multiparous patients including data from patients' prior pregnancies will yield similar or higher performance characteristics as observed in this study.[42]

Conclusion

Using results from a prospective cohort study, we constructed 2 random forest models that utilize prenatal factors to predict the development of hypertensive disorders of pregnancy. These models outperform current guidelines in recommending aspirin prophylaxis. Following additional external validation and refinement, the deployment of these automated models in clinical practice can aid in identification of hypertensive disorders of pregnancy and standardized implementation of increased monitoring and prophylactic measures.

Patient consent statement

This study utilized deidentified data and is considered exempt from IRB review given the deidentified nature of the data. ■

CRediT authorship contribution statement

Jonathan S. Schor: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Adesh Kadambi:** Writing – review & editing, Formal analysis, Conceptualization. **Isabel Fulcher:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Kartik K. Venkatesh:** Writing – review & editing, Supervision, Conceptualization. **Mark A. Clapp:** Writing – review & editing, Formal analysis, Conceptualization. **Senan Ebrahim:** Writing – review & editing, Writing – original draft, Data curation. **Ali Ebrahim:** Writing – review & editing, Formal analysis, Data curation, Conceptualization. **Timothy Wen:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. ■

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.xagr.2024.100386.

REFERENCES

1. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol 2009;33(3):130–7.
2. Harmon QE, Huang L, Umbach DM, et al. Risk of fetal death with preeclampsia. Obstetr Gynecol 2015;125(3):628.
3. Steegers EA, Von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. The Lancet 2010;376(9741):631–44.
4. Magee LA, Nicolaides KH, Von Dadelszen P. Preeclampsia. N Engl J Med 2022;386(19):1817–32.
5. Levine LD, Ky B, Chirinos JA, et al. Prospective evaluation of cardiovascular risk 10 years after a hypertensive disorder of pregnancy. J Am Coll Cardiol 2022;79(24):2401–11. https://doi.org/10.1016/j.jacc.2022.03.383.
6. Wen T, Schmidt CN, Sobhani NC, et al. Trends and outcomes for deliveries with hypertensive disorders of pregnancy from 2000 to 2018: a repeated cross-sectional study. BJOG: Int J Obstetr Gynaecol 2022;129(7):1050–60.
7. Wallis AB, Saftlas AF, Hsia J, Atrash HK. Secular trends in the rates of preeclampsia, eclampsia, and gestational hypertension, United States, 1987–2004. Am J Hypertens 2008;21(5):521–6.
8. Gibson KS, Hameed AB. Society for maternal-fetal medicine special statement: checklist for postpartum discharge of women with hypertensive disorders. Am J Obstetr Gynecol 2020;223(4):B18–21.
9. Espinoza J, Vidaeff A, Pettker CM, Simhan H. Gestational hypertension and preeclampsia. Obstetr Gynecol 2019;133(1):E1–E25.
10. LeFevre ML. Force* UPST. Low-dose aspirin use for the prevention of morbidity and mortality from preeclampsia: US Preventive Services Task Force recommendation statement. Ann Internal Med 2014;161(11):819–26.
11. Ma'ayeh M, Constantine MM. Prevention of preeclampsia. Semin Fetal Neonatal Med 2020;25(5):111123.
12. Rolnik DL, Wright D, Poon LC, et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. New Engl J. Med 2017;377(7):613–22.
13. Roberge S, Bujold E, Nicolaides KH. Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. Am J Obstetr Gynecol 2018;218(3):287–293.e1.
14. Koopmans CM, Bijlenga D, Groen H, et al. Induction of labour versus expectant monitoring for gestational hypertension or mild preeclampsia after 36 weeks' gestation (HYPITAT): a multicentre, open-label randomised controlled trial. The Lancet 2009;374(9694):979–88.
15. Obstetricians ACOG, Gynecologists. Low-dose aspirin use for the prevention of preeclampsia and related morbidity and mortality. ACOG; 2021.
16. Tsiakkas A, Saiid Y, Wright A, Wright D, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 30–34 weeks' gestation. Am J Obstetr Gynecol 2016;215(1):87.e1–87.e17.
17. Gallo DM, Wright D, Casanova C, Campanero M, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 19–24 weeks' gestation. Am J Obstetr Gynecol 2016;214(5):619.e1–619.e17.
18. Hoffman MK, Ma N, Roberts A. A machine learning algorithm for predicting maternal readmission for hypertensive disorders of pregnancy. Am J Obstetr Gynecol MFM 2021;3(1):100250.
19. Clapp MA, McCoy TH. The potential of big data for obstetrics discovery. Curr Opin Endocrinol Diabetes Obes 2021;28(6):553–7. https://doi.org/10.1097/med.0000000000000679.
20. Haas DM, Parker CB, Wing DA, et al. A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). Am J Obstetr Gynecol 2015;212(4):539.e1–539.e24.
21. Haas DM, Parker CB, Marsh DJ, et al. Association of adverse pregnancy outcomes with hypertension 2 to 7 years postpartum. J Am Heart Assoc 2019;8(19):e013092.
22. Facco FL, Parker CB, Reddy UM, et al. Association between sleep-disordered breathing and hypertensive disorders of pregnancy and gestational diabetes mellitus. Obstetr Gynecol 2017;129(1):31.

23. Goretsky A, Dmitrienko A, Tang I, et al. Data preparation of the nuMoM2b dataset. medRxiv. 2021.
24. Parmar A, Katariya R, Patel V. A review on random forest: An ensemble classifier. In International conference on intelligent data communication technologies and internet of things (ICICI) 2018. Springer International Publishing; 2019. p. 758–63.
25. Loef B, Wong A, Janssen NA, et al. Using random forest to identify longitudinal predictors of health in a 30-year cohort study. Sci Rep 2022;12(1):10372.
26. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? J Machine Learning Res 2014;15(1):3133–81.
27. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Signal Proc Lett 2014;21 (11):1389–93.
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988: 837–45.
29. Van RG, Drake F. Python 3 reference manual. Scotts Valley, CA: CreateSpace 2009;10:1593511.
30. Obstetricians ACo, Gynecologists. Low-dose aspirin use during pregnancy. ACOG Committee Opinion No. 743. Obstet Gynecol. 2018;132:e44–52.
31. Li Y-x, Shen X-p, Yang C, et al. Novel electronic health records applied for prediction of pre-eclampsia: machine-learning algorithms. Pregnancy Hypertension 2021;26:102–9.
32. Bertini A, Salas R, Chabert S, Sobrevia L, Pardo F. Using machine learning to predict complications in pregnancy: a systematic review. Front Bioeng Biotechnol 2021;9:780389. https://doi.org/10.3389/fbioe.2021.780389.
33. Jhee JH, Lee S, Park Y, et al. Prediction model development of late-onset preeclampsia using machine learning-based methods. PLoS One 2019;14(8):e0221202.
34. Marić I, Tsur A, Aghaeepour N, et al. Early prediction of preeclampsia via machine learning. Am J Obstetr Gynecol MFM 2020;2 (2):100100.
35. Sufriyana H, Wu Y-W, Su EC-Y. Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia. EBioMedicine 2020;54:102710.
36. Sandström A, Snowden JM, Höijer J, Bottai M, Wikström A-K. Clinical risk assessment in early pregnancy for preeclampsia in nulliparous women: A population based cohort study. PloS one 2019;14(11):e0225716.
37. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of pre-eclampsia. Am J Obstetr Gynecol 2020;223 (1):12–23.e7.
38. Wright D, Wright A, Tan MY, Nicolaides KH. When to give aspirin to prevent pre-eclampsia: application of Bayesian decision theory. Am J Obstetr Gynecol 2022;226(2): S1120–5.
39. Ying W, Catov JM, Ouyang P. Hypertensive disorders of pregnancy and future maternal cardiovascular risk. J Am Heart Assoc 2018;7 (17):e009382.
40. Grobman WA, Rice MM, Reddy UM, et al. Labor induction versus expectant management in low-risk nulliparous women. New Engl J Med 2018;379(6):513–23.
41. Sutton MY, Anachebe NF, Lee R, Skanes H. Racial and Ethnic Disparities in Reproductive Health Services and Outcomes, 2020. Obstet Gynecol 2021;137(2):225–33. https://doi.org/10.1097/aog.0000000000004224.
42. Hernández-Díaz S, Toh S, Cnattingius S. Risk of pre-eclampsia in first and subsequent pregnancies: prospective cohort study. Bmj 2009;338.