Archival Report

Biological Psychiatry: GOS

Predicting Prenatal Depression and Assessing Model Bias Using Machine Learning Models

Yongchao Huang, Suzanne Alvernaz, Sage J. Kim, Pauline Maki, Yang Dai, and Beatriz Peñalver Bernabé

ABSTRACT

BACKGROUND: Perinatal depression is one of the most common medical complications during pregnancy and postpartum period, affecting 10% to 20% of pregnant individuals, with higher rates among Black and Latina women who are also less likely to be diagnosed and treated. Machine learning (ML) models based on electronic medical records (EMRs) have effectively predicted postpartum depression in middle-class White women but have rarely included sufficient proportions of racial/ethnic minorities, which has contributed to biases in ML models. Our goal is to determine whether ML models could predict depression in early pregnancy in racial/ethnic minority women by leveraging EMR data.

METHODS: We extracted EMRs from a large U.S. urban hospital serving mostly low-income Black and Hispanic women (n = 5875). Depressive symptom severity was assessed using the Patient Health Questionnaire-9 self-report questionnaire. We investigated multiple ML classifiers using Shapley additive explanations for model interpretation and determined prediction bias with 4 metrics: disparate impact, equal opportunity difference, and equalized odds (standard deviations of true positives and false positives).

RESULTS: Although the best-performing ML model's (elastic net) performance was low (area under the receiver operating characteristic curve = 0.61), we identified known perinatal depression risk factors such as unplanned pregnancy and being single and underexplored factors such as self-reported pain, lower prenatal vitamin intake, asthma, carrying a male fetus, and lower platelet levels. Despite the sample comprising mostly low-income minority women (54% Black, 27% Latina), the model performed worse for these communities (area under the receiver operating characteristic curve: 57% Black, 59% Latina women vs. 64% White women).

CONCLUSIONS: EMR-based ML models could moderately predict early pregnancy depression but exhibited biased performance against low-income minority women.

https://doi.org/10.1016/j.bpsgos.2024.100376

Perinatal depression (PND), depression during pregnancy and up to 1 year postpartum, is one of the most common complications during the perinatal period (1). In the US, the rate of PND is 10% to 20% (1) and has increased more than 3-fold from 2000 to 2015 (2,3). The rates of PND among Black and Latina women are 2- to 5-fold higher than non-Hispanic White (NHW) women (4–8). In our own longitudinal studies in low-income women of color, the rate of PND is 23% (9), whereas the US average rate is 12% (10). During the COVID-19 pandemic, the prevalence of PND rose to 27% to 32% (11–13), highlighting the importance of environmental stressors in these mood disorders (9). PND confers significant obstetric risks of low birth weight (14), preterm labor (14,15), higher maternal morbidity and mortality, longer hospital stays post-delivery and higher delivery costs (16), lower initiation and duration of breastfeeding (17), and poor maternal-fetal attachment (18). Infants born from women with PND have increased risks of stunted growth, inadequate cognitive development, altered stress response, underdeveloped socio-emotional behavior, and future mental disorders (16,19–24). In extreme cases, PND can lead to suicide, a leading cause of maternal mortality in the first year after delivery (25).

Multiple individual-level factors have been linked to increased PND risk, such as lack of a partner and social support (26), unplanned pregnancy (27), young age (28), history of depression or trauma (28), and adverse childhood experiences (29,30). Despite a higher prevalence of depression during pregnancy in Black and Latina women, they are less likely to be screened (31–33) and to share their PND symptoms (34–36). Providers commonly screen for depression at least once during pregnancy and postpartum (37). Common PND screening methods include self-reported questionnaires, such as the Edinburgh Postpartum Depression Score (38) and the Patient Health Questionnaire-9 (PHQ-9) (39). However, the utility of these self-reported questionnaires depends on the accurate disclosure of symptoms (34–36). For instance, Black women might share their PND symptoms with family but are more reluctant to seek help from medical professionals than NHW women due to social stigma (34,40) and concern of being viewed as lacking strength and resilience (41–43). Fear of the

ISSN: 2667-1743

consequences of symptom disclosure, such as losing the infant's custody (34), and medical mistrust are also important factors for which pregnant individuals might decide to underreport or completely deny their symptoms (44). Furthermore, communication of depressive symptoms may be influenced by cultural factors that are not fully captured by Edinburgh Postpartum Depression Score or PHQ-9 (44–48). For example, studies have shown that low-income urban Black women endorse lower levels of depressive symptom severity in self-reported questionnaires than other populations (46–48), despite having the same clinical diagnosis, hence leading to more false negative results in Black women than other groups.

Computational approaches, such as machine learning (ML) using electronic medical records (EMRs), have been able to predict pregnancy outcomes, including gestational diabetes (49–51), preterm birth (52), and suicidal thoughts (53). EMR-based ML models for PND have generally focused on predicting postpartum depression (54–64) and rarely include racial and ethnic minorities (55,60). In middle-class White women, EMR-based ML models to predict postpartum depression can perform relatively well, with area under the receiver operating characteristic curve (AUROC) $>$0.75, using various approaches such as random forest (55,59,60,64–66), artificial neural networks (56,57,67), and logistic regression (55,63,68). However, the performance of EMR-based ML models for predicting depression in pregnancy for low-income minority women remains unexplored.

Biases in prediction performance exist in ML models, which refer to the disparate levels of model prediction of the outcome of interest for certain sociodemographic variables, called protected variables, such as sex, gender, age, race, ethnicity, or socioeconomic status (69). Given that EMRs are not collected from well-designed balanced studies, the model performance can better predict or be biased toward groups more represented in EMRs. In addition, some groups might have more available data (e.g., more clinical encounters and diagnostic tests), or their EMR information is of higher quality (70). White women are commonly disproportionally overrepresented in EMRs. Consequently, ML models based on EMRs have lower performance at predicting PND risk in Black and Latina women than White individuals (60). Other reasons for poor performance include the lower quality and quantity of EMRs for racial/ethnic minorities than those of White women because of lower access to care (60), e.g., no documented medical history in the EMRs, underutilization of health care due to work conditions, lack of childcare or transportation, reluctance to reveal certain sensitive information (34), and/or implicit bias by health care providers (34,44). Thus, understanding the sources of bias in ML models is essential for the equitable prediction of PND risk in diverse populations.

Despite the negative consequences of depression during pregnancy for both the mother and the infant that disproportionally affect low-income women of color, no study has thus far used EMRs enriched by women of color from underresourced communities to predict depression early in pregnancy and examined the model bias. Here, we constructed ML models to predict depression symptom severity early in pregnancy for urban low-income women of color who received care in the same outpatient obstetric clinic, and we subsequently assessed model performance and racial and ethnic biases.

METHODS AND MATERIALS

Study Population

EMRs were extracted from patients who received obstetric care at the University of Illinois Chicago Hospital Health Sciences System, an urban academic hospital in Chicago, where the population served is 52% non-Hispanic Black (NHB), 29% Hispanic (H), 9% NHW, and 10% Asian and Native American, and we further selected EMRs from patients who were 15 years or older and received their obstetric care and delivered at University of Illinois Chicago between 2014 and 2020. We obtained information on the patient's sociodemographic characteristics, medical history, mental health assessments, health behaviors (e.g., substance abuse, vital records, laboratory tests, medications, prior and current obstetric complications, and delivery and fetus information) (Tables S1 and S2). Depressive symptom severity was assessed using the self-reported PHQ-9 (39). The EMR extraction was approved by the University of Illinois Institutional Review Board (IRB # 2020-0553).

Preprocessing of EMR Data

Patients were included if 1) their PHQ-9 scores were between 1 and 4 or 9 and above and 2) their first obstetric visit was before 24 gestational weeks given that depressive symptoms later in pregnancy might be distinct from those in early pregnancy. We excluded patients with mild levels of depression (PHQ-9 scores between 5 and 8) or those who reported no depression symptoms (PHQ-9 score 0) to avoid incorporating possible false negatives due to underreported or denial of symptoms, respectively. Patients with a PHQ-9 score $>$ 9 were categorized as cases, and those with a PHQ-9 score between 1 and 4 as controls (Figures S1 and S2).

EMR features recorded after the first perinatal appointment were excluded from downstream analysis (e.g., gestational diabetes or preeclampsia diagnosis), as well as features that were missing in more than 60% of patients (Figure S3 and Table S3). Prescribed medications were grouped into 29 broad classes based on their most common use, mechanism of action, and targeted organ system (Table S2). Qualitative features (e.g., race, insurance, and medications) were hard coded, and continuous features were transformed using min-max normalization. Missing data were imputed using MICE (version 3.15.0) (71) (Figure S4). To partially mitigate bias, ML models were trained by excluding self-reported race and ethnicity results from genetic tests to assess ancestry and preferred language to communicate with their providers. See Supplemental Methods for more details.

ML Model Selection and Training

We explored different ML models, including random forest, elastic net, and XGBoost, to identify the most suitable ones for predicting PND. We evaluated and selected models based on their accuracy, AUROC, positive predictive value, negative predictive value, specificity, and model bias. We also computed the harmonic mean of precision and recall (F1) and the area under the precision-recall curve. Model sensitivity was fixed at 85% for all models to maximize the identification of true depression cases (Figure S5).

We employed a nested *k*-fold cross-validation approach for model evaluation and hyperparameter tuning. This approach

consisted of 2 layers: a 10-fold outer loop and a 5-fold inner loop. In the outer loop, the data were divided into a training set and a test set. The training set from the outer loop was then used in the inner loop to be further split into subtraining and validation sets. These subsets were used to train models and optimize their hyperparameters. The model with the best hyperparameters was applied to the test set from the outer loop to assess performance, ensuring that every patient in the dataset was evaluated at once (Figure S6). To maintain the representativeness of the original dataset in each fold, we stratified the data according to race/ethnicity and the classification of individuals as either a case or a control (Table S4). For hyperparameter optimization, we leveraged the GridSearchCV function from the scikit-learn Python package (version 1.2.0), which conducted a 5-fold cross-validation within the inner loop. See Supplemental Methods for more details.

Identification of the Most Important Features

To identify the importance and directionality with respect to depressive symptom severity of the EMR features, we calculated the Shapley values using Shapley additive explanations (SHAP) (72). The Shapley value, Φ_i, is defined as the estimated contribution of feature i in all samples to the depression outcome. For feature dependence analysis, we followed the procedure developed by Artzi *et al.* (49) to convert the Shapley values into relative risk (RR). Briefly, in the SHAP analysis, the log-odds of the predicted probability are calculated as Φ_i; then the predicted probability of a single feature i is

$$P_i = S(\Phi_0 + \Phi_i) \tag{1}$$

where

$$S(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

and Φ_0 is the base Shapley value, i.e., the logit of the population prevalence (denoted as P_0). Therefore,

$$\text{logit}\,(P_0) = \ln\,(P_0 / (1 - P_0)) = \Phi_0 \tag{3}$$

$$P_0 = S(\Phi_0) \tag{4}$$

Here P_0 was set as 0.14, which was the PND prevalence estimated from the EMRs of the current study.

RR of a single feature i was calculated as follows:

$$RR_i = \frac{P_i}{P_0} = \frac{S(\Phi_0 + \Phi_i)}{S(\Phi_0)} \tag{5}$$

When calculating the RR of PND in relation to a set of features A, equation 5 can be extended as follows:

$$RR_A = \frac{P_A}{P_0} = \frac{S\left(\Phi_0 + \sum_{n \in A} \Phi_n\right)}{S(\Phi_0)} \tag{6}$$

We employed the function *dependence_plot()* within the *shap* package in Python (version 0.39.0) to determine the correlation between the most important features and the rest of the EMR variables based on their Shapley values.

Statistical Analysis

The Mann-Whitney test was used for continuous variables to identify differences among independent groups. Distinctions between models were assessed using the Wilcoxon signed-rank test. A two proportion z test was used to evaluate the proportion differences for categorical or dichotomous variables. Pairwise Tukey's honestly significant difference was applied to compare Shapley values across race/ethnicity for each feature and the Bonferroni correction to adjust for multiple comparisons. To establish the relationships between each feature and the outcome, we used Spearman correlation to estimate the direction of the effect.

Bias Assessments

We employed 4 common metrics to assess model bias: the disparate impact (DI) (73), the equal opportunity difference (EOD) (74), and the standard deviation (std) of true positives and false positives (equalized odds [EO]) across race/ethnicity (75):

$$\text{DI} = \frac{\Pr(y = 1|\text{unprivileged})}{\Pr(y = 1|\text{privileged})} \tag{7}$$

$$\begin{aligned}\text{EOD} = {} & \Pr(\text{y} = 1|\text{Y} = 1, \text{privileged}) \\ & - \Pr(\text{y} = 1|\text{Y} = 1, \text{unprivileged})\end{aligned} \tag{8}$$

Here, Y and y denote the true and predicted outcome of depression (case = 1, control = 0), respectively, and Pr is probability.

$$EO(FPR) = std\{\text{FPR}_{\text{NHB}}, \text{FPR}_{\text{NHW}}, \text{FPR}_{\text{H}}\} \tag{9}$$

$$EO(TPR) = std\{\text{TPR}_{\text{NHB}}, \text{TPR}_{\text{NHW}}, \text{TPR}_{\text{H}}\} \tag{10}$$

where FPR is the false positive rate and TPR is the true positive rate.

The privileged group was selected as the racial/ethnic group with the highest AUROC. In our study, the privileged group was NHW women, and the unprivileged groups were NHB and Latina women. The calibration curve (76) was calculated using the *calibration_curve()* function with 100 quantile bins.

RESULTS

EMRs Were Extracted From an Understudied Population of Urban Low-Income Women of Color

The extracted data included 5875 pregnant women and 694 EMR features (Figure 1; Tables S3 and S5). After data preprocessing, a total sample of 2414 women and 74 EMR features was employed for downstream analysis (Table 1). Most of the EMRs belonged to low-income women of color, 54% NHB and 27% NHW women, and more than 72% of women were in federal aid health care plans (Medicaid and Medicare). We observed statistically significant differences between racial/ethnic groups (Tables S6 and S7), with NHB women being

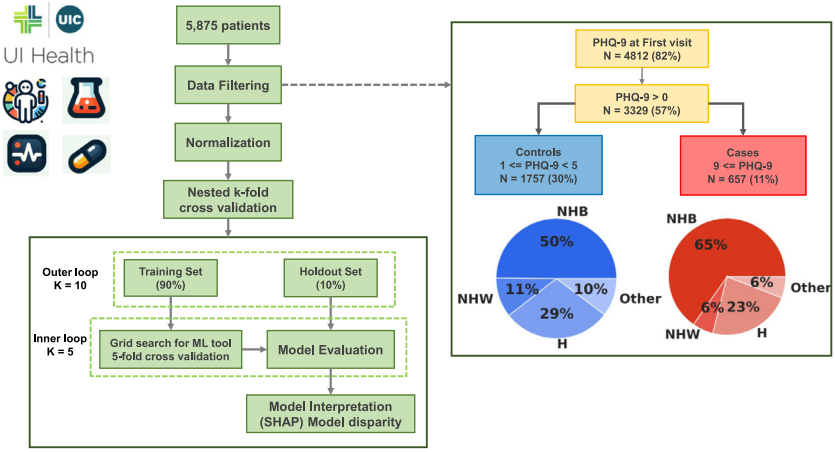

Figure 1. Overall workflow. Electronic medical records were extracted from 5875 patients ages 15 years or older who received obstetric care and delivered at University of Illinois (UI) Health between 2014 and 2020. Patients were included based on their Patient Health Questionnaire-9 (PHQ-9) scores and timing of their first obstetric visit, before 24 weeks. The preprocessed electronic medical record data underwent normalization and nested *k*-fold cross-validation to develop machine learning (ML) models for predicting perinatal depression. This approach allowed for the identification of important features contributing to perinatal depression prediction and the assessment of potential model disparities across different racial/ethnic groups. H, Hispanic or Latina; NHB, non-Hispanic Black; NHW, non-Hispanic White; SHAP, Shapley additive explanations.

more likely to be single or unemployed or having an unplanned pregnancy (all adjusted $p < .01$). We also detected statistically significant differences between cases and controls overall as well as when segregating the data by race/ethnicity, with women reporting high levels of depressive symptoms being more likely to have an unplanned pregnancy and to smoke independently of their race/ethnicity (adjusted $p < .01$).

Elastic Net Was the Best Model to Predict Depression in Early Pregnancy

First, we determined the most adequate ML model to predict depressive symptom severity early in pregnancy in our sample, including random forest (77), XGBoost (78), and elastic net (79). While both elastic net and random forest had similar performance when models were agnostic to patient's race/ethnicity, the calibration analysis suggests that the elastic net model's predictive accuracy is moderately consistent across different risk levels of PND with slope close to 1 and intercept close to 0 (Figure S7; Table S8). Elastic net also had significantly lower bias, EO (TPR), and EO (FPR) predicting PND. Thus, it was chosen for downstream analysis (Table S8).

Identification of Well-Known and Novel Features Associated With Prenatal Depressive Symptom Severity

To identify the features that were most predictive of depressive symptom severity early in pregnancy and the directionality of their associations, we determined their Shapley values. Based on game theory, Shapley values provide an estimate of the contribution made by each feature toward the overall prediction of the model. The top 20 most predictive features based on their mean absolute Shapley values included well-known sociodemographic factors associated with PND, such as having an unplanned pregnancy (80); being single (81), of young age (28), or unemployed (82); and use of federal aid insurance (proxy for poverty) (83) and of tobacco (84) (Figure 2A), yet our model also identified features that have either not been previously associated with depressive symptom severity in pregnancy or only reported in a few studies. For instance, we discovered that elevated depressive symptoms were positively associated with self-reported levels of pain; an asthma diagnosis; carrying a male fetus (85); use of antihistamines, analgesics, antibiotics, or mood and anxiety medication; and higher mean platelet volumes in blood (86,87). Features related to preventive care, such as prenatal vitamin intake (88) and immunization against influenza and tetanus, were negatively correlated with depressive symptom severity.

Inspection of the contribution of top significant features to PND across patient's race/ethnicity (Figure 2B–D; Figures S8 and S9) and sociodemographic factors, such as being single, having an unplanned pregnancy, unemployment, and low socioeconomic status, were weaker predictors of depressive symptoms in NHB women than NHW or Hispanic women (adjusted $p < .01$). However, tobacco use, infections, asthma diagnosis, gastrointestinal disturbance, and elevated self-reported levels of pain had significantly higher contributions to PND in NHB women than the other 2 groups (adjusted $p < .01$). Noticeably, mood and anxiety medication, prenatal vitamin intake, and immunization had higher significance in NHW women (adjusted $p < .01$).

Predicted RR of Prenatal Depression

Next, we determined the RR associated with the most predictive features of depressive symptom severity based on their Shapley value (Figure 3 and Figure S10). High self-reported pain levels were positively and linearly correlated with an increased risk of PND, independently of whether the patient was receiving any pain medication (Figure 3A). NHB individuals more often reported higher levels of pain (85%, score ≥ 6) than any other group (Figure 3A). A history of medication for mood disorders was associated with an increased risk of PND (RR = 1.22 $\pm$ 0.02) (Figure 3B). As previously reported (80,81,84), tobacco use (RR = 1.1 $\pm$ 0.01) (Figure 3C), unplanned pregnancy (RR = 1.03 $\pm$ 0.01) (Figure 3D), and being single (RR = 1.02 $\pm$ 0.01) (Figure 3E) were associated with increased risk of PND.

Despite the Study Sample Being Enriched With Low-Income Women of Color, the Model Was Biased Against Black Women

Finally, we examined whether ML models had the same predictive capability to predict PND without including the individual's race/ethnicity in the feature set for model training

Table 1. Sociodemographic Characteristics of the Sample Used for Model Optimization

	Cases, n = 657				Controls, n = 1757			
	All cases	NHB, 429 (65%)	NHW, 38 (6%)	H, 151 (23%)	Other, 39 (6%)	All controls	NHB, 876 (50%)	NHW, 188 (11%)	H, 512 (29%)	Other, 181 (10%)
Single	565 (86%)	398 (93%)	23 (61%)	121 (80%)	23 (59%)	1258 (72%)	785 (90%)	71 (38%)	327 (64%)	75 (41%)
Federal Aid Insurance	534 (81%)	365 (85%)	22 (58%)	125 (83%)	22 (56%)	1209 (69%)	701 (80%)	64 (34%)	353 (69%)	91 (50%)
Unplanned Pregnancy	516 (79%)	372 (87%)	18 (47%)	105 (70%)	21 (54%)	1040 (59%)	688 (79%)	45 (24%)	241 (47%)	66 (36%)
Unemployed	564 (86%)	371 (89%)	29 (79%)	129 (87%)	35 (90%)	1352 (77%)	699 (83%)	122 (70%)	396 (83%)	135 (75%)
Age, Years	30 (16–49)	29 (17–47)	35 (23–49)	30 (18–48)	31 (18–46)	31 (15–55)	30 (15–48)	36 (15–50)	31 (15–53)	33 (15–55)
BMI	29 (15–74)	29 (16–73)	26 (19–51)	29 (15–74)	23 (18–40)	29 (16–75)	30 (17–73)	27 (18–68)	29 (16–74)	25 (16–54)
PHQ-9	11 (9–27)	11 (9–27)	11 (9–25)	11 (9–27)	13 (9–24)	2 (1–4)	2 (1–4)	2 (1–4)	2 (1–4)	2 (1–4)

Values are presented as n (%) or median (range). Sociodemographic feature distributions (n = 2414) after preprocessing.
BMI, body mass index; H, Hispanic or Latina; NHB, non-Hispanic Black; NHW, non-Hispanic White; PHQ, Patient Health Questionnaire.

(Figure 4A, B). The model performance for any patient, independent of their racial/ethnic background, was moderate (area under the precision-recall curve = 50%; AUROC = 66%) (Model 1). However, when stratifying by racial/ethnic groups, elastic net predictions of PND status for NHW women were significantly higher than those for NHB women (Model 1 AUROC: NHW [69%], NHB [55%], H [63%], respectively; adjusted $p < .001$). Furthermore, the performance of the ML model for NHW women was better than for NHB and Latina women when compared by specificity, accuracy, and negative predictive value (Figure S11). Notably, our model predicted NHB women with PND with a sensitivity of 90% although the positive predictive value and specificity were low (Model 1 positive predictive value: NHB [36%], NHW [31%], H [31%]; specificity: NHB [20%], NHW [67%], H [48%]) (Figure S11B, D, G).

To further estimate the inequality of model prediction or model bias, we calculated 4 common bias metrics: EOD—the true positive difference between the privileged and unprivileged group; DI—the ratio of positive predictions between the privileged and unprivileged groups; and EO/standard deviation of TPR and FPR across race/ethnicity. Removal of race/ethnicity training features from the models—specifically, when comparing Model 1 with Model 2—moderately reduced model performance disparities as measured by DI, EOD, EO (TPR), and EO (FPR), with more significant improvements for NHB women (adjusted $p < .05$) than NHW women (Figure 4C, D and Figure S12).

DISCUSSION

In this study, we examined the capability of interpretable ML models to predict the risk of depression during early pregnancy in low-income women of color using EMRs, demonstrating that EMR-based ML models were moderately predictive of depressive symptom severity in early pregnancy in this high-risk population. In addition, using the SHAP analysis, we identified factors associated with the risk of early PND including the directionality of their associations. Our models not only revealed well-known factors, such as unplanned pregnancy, history of medication use to treat mood disorders, or young age, but also captured novel or underexplored markers, such as self-reported pain levels, asthma diagnosis, carrying a male fetus, or systemic platelet volume. Importantly, our results highlighted the significant performance disparity in model prediction among racial/ethnic groups, with women of color at greater disadvantages in model predictions.

ML models to predict PND in racial and ethnic minorities are scarce. EMR-based ML models to predict postpartum depression have been developed in a wide variety of populations [e.g., Chinese (59), European (56,57,61,64), NHW women in the US (65)]. These EMR-based ML models can perform relatively well. However, previous studies have not demonstrated the predictive capability of EMR-based ML models to estimate PND status in low-income minority women. Our results filled this current gap and showed that EMR-based models trained with a sample primarily composed of low-income minority women moderately predict PND status, yet the elastic net model had a better prediction performance for the less prevalent group in the sample, NHW women

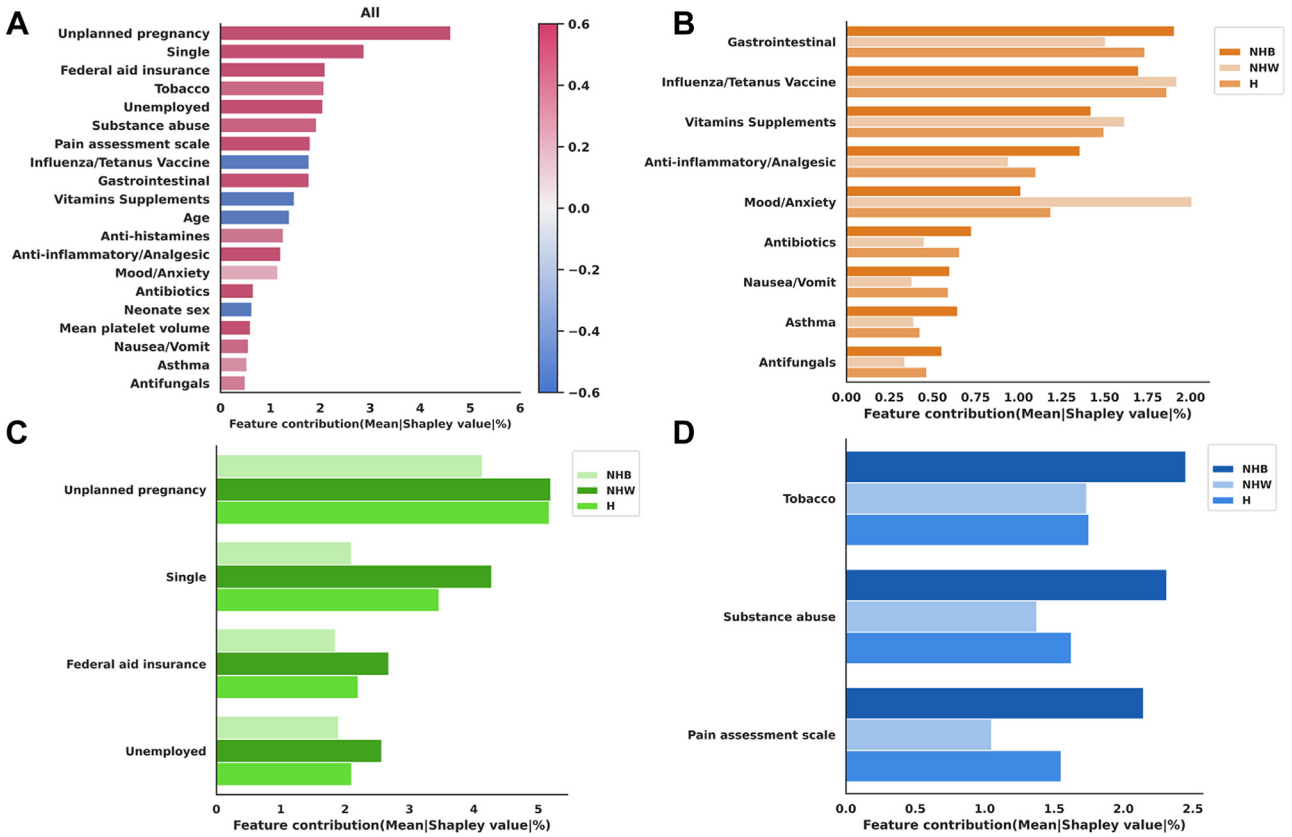

Figure 2. Multiple known and unexplored electronic medical record features were associated with perinatal depression (PND) status. **(A)** Top 20 most important features based on their Shapley value to predict PND status. The color of the bar represents the correlation between each feature and PND status; red indicates a positive association with outcome, and blue indicates a negative association. **(B–D)** Features whose importance is significantly different by race/ethnicity in terms of medication use, sociodemographic factors, substance use, and pain assessment scale. Pairwise Tukey's honestly significant difference was used to determine features included in **(B–D)**. H, Hispanic or Latina; NHB, non-Hispanic Black; NHW, non-Hispanic White.

(AUROC = 69%), than NHB women (AUROC = 55%) and Latina women (AUROC = 63%).

Our results are interpretable, a very important capability to identify markers that providers can intervene upon. Using SHAP analysis, we revealed several established markers that increased the risk of depression in early pregnancy, such as unplanned pregnancy (80) or being single (81). Our results based on minority populations who are at high risk of PND also suggest that individuals who do not take vitamin supplements during early pregnancy may be at increased risk of PND. This observation aligns with results from a systematic review performed by Sparling *et al.* (88). The authors found that lower levels of folate, vitamin D, iron, selenium, zinc, fats, and fatty acids were associated with increased risks of PND; those studies were limited in their population diversity, which either failed to include minority populations or had limited sample sizes of minority groups.

Importantly, our analysis uncovered several new or underexplored markers that increased the predicted risk of depression during early pregnancy, such as self-reported pain levels, mean platelet volume in blood (86,87), carrying a male fetus (85), and having a diagnosis of asthma. Higher levels of pain have been linked to postpartum depression (89), and our findings suggest that this relationship also holds during the early stages of pregnancy. Similarly, patients with major depression exhibit a higher mean platelet volume (86,87), but this phenomenon has not yet been investigated in the context of PND. Recently, Myers *et al.* (85) reported that carrying a male fetus increases the odds of postnatal depression. While the mechanisms that link the fetus's sex and depression status are unclear, women who carry male fetuses have lower estradiol levels in the blood than those carrying female fetuses (90) with lower levels of estrogens outside pregnancy associated with higher levels of depressive symptoms (91). In addition, women carrying male fetuses have higher levels of inflammatory markers, such as interleukin 1β, in early pregnancy, and higher levels of inflammation have been linked to PND (92,93).

Despite our original expectation, employing samples enriched in individuals from underserved populations did not mitigate EMR-based ML model's performance bias. Our results revealed a significant disparity in model performance among race/ethnicity groups to predict PND. NHW women, the least represented group in the sample, were predicted significantly better than NHB women, who represented more than 54% of the sample. This contradicts prior studies and

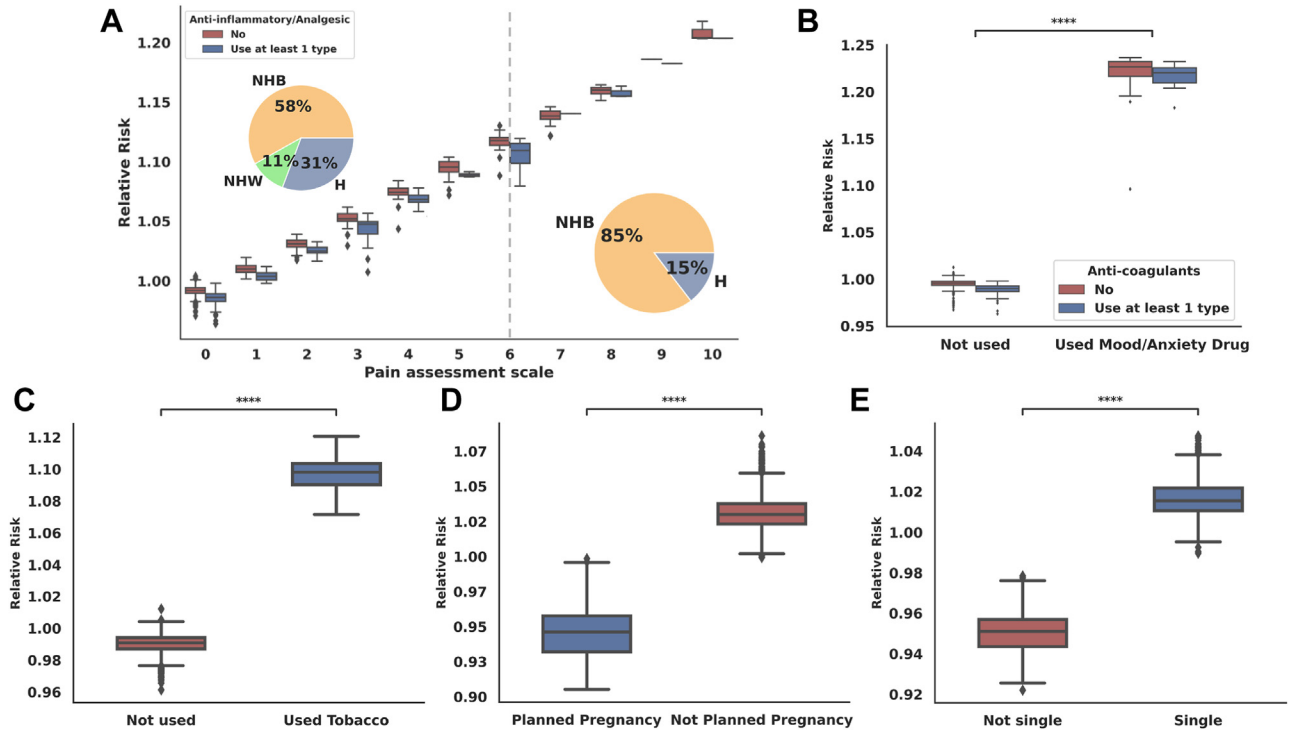

Figure 3. The predicted relative risk of top selected features. **(A)** Predicted relative risk of self-reported pain level and its interaction with the use of anti-inflammatory and analgesic drugs (e.g., ibuprofen). The pie chart represents the racial/ethnic distribution of pain assessment levels below and equal to or above 6. **(B–E)** Predicted relative risk of mood/anxiety medication use and their interactions with anticoagulant medication, of tobacco use, of unplanned pregnancy, and of being single, respectively. Two proportion *z* test was used in **(A)**. Mann-Whitney *U* test was used in **(B–E)**. ****, adjusted $p < .05$. H, Hispanic or Latina; NHB, non-Hispanic Black; NHW, non-Hispanic White.

suggests that even with a majority of data from underprivileged groups with similar EMR quality, ML models can indeed exhibit performance bias.

Multiple reasons might explain our observed results. For instance, due to cultural differences, social stigma, or medical mistrust, NHB women might underreport their symptoms more often than NHW women. Studies indicate that urban low-income women of color report lower levels of depressive symptoms than NHW women using self-reported questionnaires despite having the same diagnosis when assessed by a

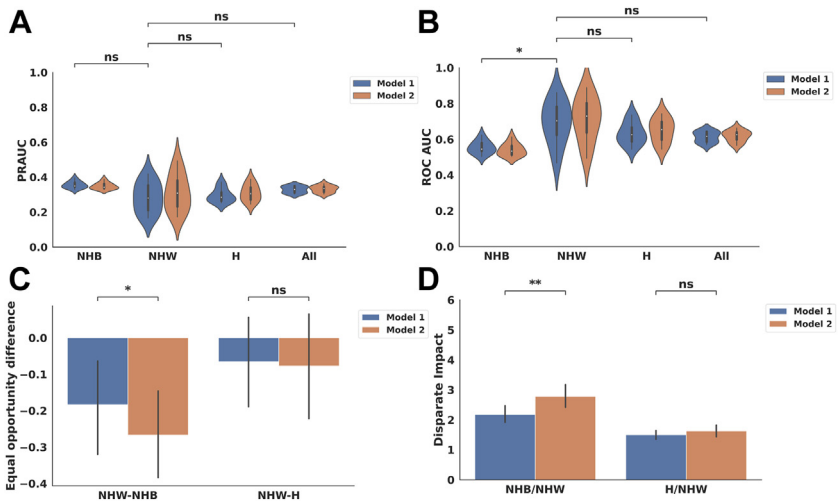

Figure 4. Electronic medical record–based machine learning models are biased against low-income minority women. **(A)** Area under the precision-recall curve (PRAUC) for different racial/ethnic groups. **(B)** Area under the receiver operating characteristic curve (ROC AUC) for different race/ethnicity groups. **(C)** Equal opportunity difference between unprivileged and privileged groups. **(D)** Disparate impact values between unprivileged and privileged groups. Privileged group: non-Hispanic White (NHW) women; unprivileged groups, non-Hispanic Black (NHB) and Hispanic (H) women. Model 1: trained without including race or ethnicity as variables in the model (blue). Model 2: trained including race and ethnicity as variables in the model (orange). Mann-Whitney *U* test used for **(A, B)**. Wilcoxon signed-rank test used for **(C, D)**. *$p < .05$, **$p < .01$. ns, nonsignificant.

clinical provider (46). Thus, a lesser number of NHB women with PND might be referred to mental health providers. Our results agree with this. Although our model was able to detect elevated depressive symptoms in NHB women with a high sensitivity, it exhibited a low positive predictive value and specificity. In other words, the model predicted a large proportion of NHB women as cases instead of as controls, suggesting that a larger proportion of NHB women may have experienced higher levels of depressive symptoms. However, our results are based on self-report screening tools and should be confirmed using clinical diagnosis. Another plausible reason for the lower predictive performance of our EMR-based ML model for women of color might be the different use of medical services. Although all the EMRs used in this study were from the same clinic, it was observed that NHB women, predominantly covered by federal aid insurance, tended to have their initial obstetric appointment later in their pregnancy, in contrast to NHW women.

Another possibility is that EMRs might not contain all the necessary information to predict PND. For example, certain neighborhood characteristics, such as violence rate, living in poverty areas, or levels of air pollution, are known to increase chronic stress and, consequently, chronic inflammation (94), one of the hallmarks of PND (95). Thus, chronic stress and chronic inflammation might mediate the negative effects of structural inequalities in depression during pregnancy. Therefore, future research should include neighborhood factors that could boost model prediction performance in women who are exposed to higher contextual risks, e.g., racial minority and low-income women living in highly segregated urban poverty areas (96). Bias can also be addressed using computational approaches from simple algebraic transformations, such as the geometric projection (97), to more sophisticated deep learning approaches, such as the adversarial framework (75).

Our study has several strengths, including 1) the exploration of the predictive capabilities of ML models using EMR to identify depression early in pregnancy in low-income women of color, given that most of the current studies aimed to predict postpartum depression in middle-class NHW women; 2) the use of SHAP analysis to make model results interpretable and thus actionable; and 3) assessing model biases. Given that the current study has not been externally validated through other datasets, our models should be confirmed in other populations with larger sample sizes, including other groups that are at high risk of PND (e.g., housing instability) using clinical diagnoses of PND, such as ICD-9/10 codes, instead of self-report depression questionnaires.

Conclusions

Interpretable ML models based on available EMRs can aid in identifying women at high risk of PND in their early pregnancy. This will enable providers to intervene proactively early enough to prevent the negative consequences of PND for both the mother and the child, such as prescribing appropriate medications, recommending therapy, or providing guidance on preventive measures. However, new tools and approaches are necessary to increase the prediction performance of EMR-based ML models to reduce model biases so that the risk of PND can be equitably predicted for all pregnant individuals.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was funded through the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant No. R21HD110779 [to BPB and YD]). BPB has been supported by the National Institute on Minority Health and Health Disparities (Grant No. U54MD012523) and K12 BIRCWH Award (Grant No. K12HD101373). In addition, this work has been also supported by the National Center for Advancing Translational Sciences, National Institutes of Health (Grant No. UL1TR002003). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

BPB and YD conceptualized the idea; BPB, YD, and YH designed the methodology for model creation and validation; BPB, YH, and SA curated and preprocessed the data; YH performed the analysis and subsequent validation; BPB, YD, and YH interpreted the results and wrote the original draft; and all the authors critically reviewed and edited the manuscript.

We thank Dr. Subhash Kumar Kolar Rajanna for his assistance in extracting the electronic medical records used in this manuscript.

Data are available to the research community upon approval of the University of Illinois Chicago Institutional Review Board (IRB 2020-0553). Code is available at https://github.com/LabBea/EMR_project.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Biomedical Engineering, Colleges of Engineering and Medicine, University of Illinois, Chicago, Illinois (YH, SA, YD, BPB); Division of Health Policy and Administration, School of Public Health, University of Illinois, Chicago, Illinois (SJK); Department of Psychiatry, College of Medicine, University of Illinois, Chicago, Illinois (PM); Department of Psychology, College of Medicine, University of Illinois, Chicago, Illinois (PM); Department of Obstetrics and Gynecology, College of Medicine, University of Illinois, Chicago, Illinois (PM); and Center of Bioinformatics and Quantitative Biology, University of Illinois, Chicago, Illinois (YD, BPB).

YD and BPB are joint senior authors.

Address correspondence to Beatriz Peñalver Bernabé, M.S., Ph.D., at penalver@uic.edu, or Yang Dai, Ph.D., at yangdai@uic.edu.

Received Nov 16, 2023; revised Jul 16, 2024; accepted Jul 21, 2024.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsgos.2024.100376.

REFERENCES

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T (2005): Perinatal depression: A systematic review of prevalence and incidence. Obstet Gynecol 106:1071–1083.
2. Pino EC, Zuo Y, Schor SH, Zatwarnicki S, Henderson DC, Borba CP, *et al.* (2019): Temporal trends of co-diagnosis of depression and/or anxiety among female maternal and non-maternal hospitalizations: Results from Nationwide Inpatient Sample 2004–2013. Psychiatry Res 272:42–50.
3. MacDorman MF, Declercq E, Cabral H, Morton C (2016): Recent increases in the U.S. maternal mortality rate: Disentangling trends from measurement issues. Obstet Gynecol 128:447–455.
4. Mukherjee S, Trepka MJ, Pierre-Victor D, Bahelah R, Avent T (2016): Racial/ethnic disparities in antenatal depression in the United States: A systematic review. Matern Child Health J 20:1780–1797.
5. Gavin AR, Melville JL, Rue T, Guo Y, Dina KT, Katon WJ (2011): Racial differences in the prevalence of antenatal depression. Gen Hosp Psychiatry 33:87–93.
6. Kuo WH, Wilson TE, Holman S, Fuentes-Afflick E, O'Sullivan MJ, Minkoff H (2004): Depressive symptoms in the immediate postpartum period among Hispanic women in three U.S. cities. J Immigr Health 6:145–153.
7. Grobman WA, Parker CB, Willinger M, Wing DA, Silver RM, Wapner RJ, *et al.* (2018): Racial disparities in adverse pregnancy outcomes and psychosocial stress. Obstet Gynecol 131:328–335.
8. Toledo-Corral CM, Gao L, Chavez T, Grubbs B, Habre R, Dunton GF, *et al.* (2022): Role of race, ethnicity, and immigration in perceived

stress and depressive symptomatology trends during pregnancy. J Immigr Minor Health 24:561–569.
9. Wenzel E, Penalver Bernabe B, Dowty S, Nagelli U, Pezley L, Gibbons R, Maki P (2022): Using computerised adaptive tests to screen for perinatal depression in underserved women of colour. Evid Based Ment Health 25:23–28.
10. Shorey S, Chee CYI, Ng ED, Chan YH, Tam WWS, Chong YS (2018): Prevalence and incidence of postpartum depression among healthy mothers: A systematic review and meta-analysis. J Psychiatr Res 104:235–248.
11. Adrianto N, Caesarlia J, Pajala FB (2022): Depression in pregnant and postpartum women during COVID-19 pandemic: Systematic review and meta-analysis. Obstet Gynecol Sci 65:287–302.
12. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G (2020): Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. J Affect Disord 277:5–13.
13. Mateus V, Cruz S, Costa R, Mesquita A, Christoforou A, Wilson CA, *et al.* (2022): Rates of depressive and anxiety symptoms in the perinatal period during the COVID-19 pandemic: Comparisons between countries and with pre-pandemic data. J Affect Disord 316:245–253.
14. Jarde A, Morais M, Kingston D, Giallo R, MacQueen GM, Giglia L, *et al.* (2016): Neonatal outcomes in women with untreated antenatal depression compared with women without depression: A systematic review and meta-analysis. JAMA Psychiatry 73:826–837.
15. Madigan S, Oatley H, Racine N, Fearon RMP, Schumacher L, Akbari E, *et al.* (2018): A meta-analysis of maternal prenatal depression and anxiety on child socioemotional development. J Am Acad Child Adolesc Psychiatry 57:645–657.e8.
16. McKee K, Admon LK, Winkelman TNA, Muzik M, Hall S, Dalton VK, Zivin K (2020): Perinatal mood and anxiety disorders, serious mental illness, and delivery-related health outcomes, United States, 2006–2015. BMC Womens Health 20:150.
17. Grigoriadis S, VonderPorten EH, Mamisashvili L, Tomlinson G, Dennis CL, Koren G, *et al.* (2013): The impact of maternal depression during pregnancy on perinatal outcomes: A systematic review and meta-analysis. J Clin Psychiatry 74:e321–e341.
18. Lindgren K (2001): Relationships among maternal-fetal attachment, prenatal depression, and health practices in pregnancy. Res Nurs Health 24:203–217.
19. Yonkers KA, Wisner KL, Stewart DE, Oberlander TF, Dell DL, Stotland N, *et al.* (2009): The management of depression during pregnancy: A report from the American Psychiatric Association and the American College of Obstetricians and Gynecologists. Obstet Gynecol 114:703–713.
20. Grote NK, Bridge JA, Gavin AR, Melville JL, Iyengar S, Katon WJ (2010): A meta-analysis of depression during pregnancy and the risk of preterm birth, low birth weight, and intrauterine growth restriction. Arch Gen Psychiatry 67:1012–1024.
21. Henrichs J, Schenk JJ, Roza SJ, van den Berg MP, Schmidt HG, Steegers EA, *et al.* (2010): Maternal psychological distress and fetal growth trajectories: The Generation R Study. Psychol Med 40: 633–643.
22. Li D, Liu L, Odouli R (2009): Presence of depressive symptoms during early pregnancy and the risk of preterm delivery: A prospective cohort study. Hum Reprod 24:146–153.
23. Murray L, Cooper PJ, Wilson A, Romaniuk H (2003): Controlled trial of the short- and long-term effect of psychological treatment of postpartum depression: 2. Impact on the mother–child relationship and child outcome. Br J Psychiatry 182:420–427.
24. Gentile S (2017): Untreated depression during pregnancy: Short- and long-term effects in offspring. A systematic review. Neuroscience 342:154–166.
25. Palladino CL, Singh V, Campbell J, Flynn H, Gold KJ (2011): Homicide and suicide during the perinatal period: Findings from the National Violent Death Reporting System. Obstet Gynecol 118:1056–1063.
26. Glazier RH, Elgar FJ, Goel V, Holzapfel S (2004): Stress, social support, and emotional distress in a community sample of pregnant women. J Psychosom Obstet Gynaecol 25:247–255.
27. Biaggi A, Conroy S, Pawlby S, Pariante CM (2016): Identifying the women at risk of antenatal anxiety and depression: A systematic review. J Affect Disord 191:62–77.
28. Melville JL, Gavin A, Guo Y, Fan MY, Katon WJ (2010): Depressive disorders during pregnancy: Prevalence and risk factors in a large urban sample. Obstet Gynecol 116:1064–1070.
29. Racine N, Zumwalt K, McDonald S, Tough S, Madigan S (2020): Perinatal depression: The role of maternal adverse childhood experiences and social support. J Affect Disord 263:576–581.
30. Atzl VM, Narayan AJ, Rivera LM, Lieberman AF (2019): Adverse childhood experiences and prenatal mental health: Type of ACEs and age of maltreatment onset. J Fam Psychol 33:304–314.
31. Skaer TL, Sclar DA, Robison LM, Galin RS (2000): Trends in the rate of depressive illness and use of antidepressant pharmacotherapy by ethnicity/race: An assessment of office-based visits in the United States, 1992–1997. Clin Ther 22:1575–1589.
32. Borowsky SJ, Rubenstein LV, Meredith LS, Camp P, Jackson-Triche M, Wells KB (2000): Who is at risk of nondetection of mental health problems in primary care? J Gen Intern Med 15:381–388.
33. Sidebottom A, Vacquier M, LaRusso E, Erickson D, Hardeman R (2021): Perinatal depression screening practices in a large health system: Identifying current state and assessing opportunities to provide more equitable care. Arch Womens Ment Health 24:133–144.
34. Russell S, Lang B, Clinton J, Adams C, Lamb J (2013): Perinatal Mental Health experiences of women and health professionals. Boots Family Trust Alliance. Available at: https://basw.co.uk/sites/default/files/resources/basw_40423-4_0.pdf. Accessed •••.
35. Henshaw C (2014): Screening and risk assessment for perinatal mood disorders. In: Barnes DL, editor. Women's Reproductive Mental Health Across the Lifespan. Cham: Springer International Publishing, 91–108.
36. Chew-Graham CA, Sharp D, Chamberlain E, Folkes L, Turner KM (2009): Disclosure of symptoms of postnatal depression, the perspectives of health professionals and women: A qualitative study. BMC Fam Pract 10:7.
37. Anonymous (2018): ACOG Committee Opinion No. 757: Screening for perinatal depression. Obstet Gynecol 132:e208–e212.
38. Cox JL, Holden JM, Sagovsky R (1987): Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 150:782–786.
39. Kroenke K, Spitzer RL, Williams JB (2001): The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 16:606–613.
40. Liu CH, Giallo R, Doan SN, Seidman LJ, Tronick E (2016): Racial and ethnic differences in prenatal life stress and postpartum depression symptoms. Arch Psychiatr Nurs 30:7–12.
41. Liao KYH, Wei M, Yin M (2020): The misunderstood schema of the strong Black woman: Exploring its mental health consequences and coping responses among African American women. Psychol Women Q 44:84–104.
42. Woods-Giscombé CL (2010): Superwoman schema: African American women's views on stress, strength, and health. Qual Health Res 20:668–683.
43. Amankwaa LC (2003): Postpartum depression, culture and African-American women. J Cult Divers 10:23–29.
44. Conteh N, Gagliardi J, McGahee S, Molina R, Clark CT, Clare CA (2022): Medical mistrust in perinatal mental health. Harv Rev Psychiatry 30:238–247.
45. Peñalver Bernabé B, Maki PM, Cunningham JL, Eisenlohr-Moul T, Tussing-Humphreys L, Carroll IM, *et al.* (2023): Interactions between perceived stress and microbial-host immune components: Two demographically and geographically distinct pregnancy cohorts. Transl Psychiatry 13:3.
46. Chaudron LH, Szilagyi PG, Tang W, Anson E, Talbot NL, Wadkins HIM, *et al.* (2010): Accuracy of depression screening tools for identifying postpartum depression among urban mothers. Pediatrics 125:e609–e617.
47. Yang J, Martinez M, Schwartz TA, Beeber L (2017): What is being measured? A comparison of two depressive symptom severity instruments with a depression diagnosis in low-income high-risk mothers. J Womens Health (Larchmt) 26:683–691.

48. Tandon SD, Cluxton-Keller F, Leis J, Le HN, Perry DF (2012): A comparison of three screening tools to identify perinatal depression among low-income African American women. J Affect Disord 136:155–162.
49. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, *et al.* (2020): Prediction of gestational diabetes based on nationwide electronic health records. Nat Med 26:71–76.
50. Ye Y, Xiong Y, Zhou Q, Wu J, Li X, Xiao X (2020): Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: A retrospective cohort study. J Diabetes Res 2020:4168340.
51. Qiu H, Yu HY, Wang LY, Yao Q, Wu SN, Yin C, *et al.* (2017): Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. Sci Rep 7:16417.
52. Gao C, Osmundson S, Velez Edwards DR, Jackson GP, Malin BA, Chen Y (2019): Deep learning predicts extreme preterm birth from electronic health records. J Biomed Inform 100:103334.
53. Zhong QY, Mittal LP, Nathan MD, Brown KM, Knudson González D, Cai T, *et al.* (2019): Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: Towards a solution to the complex classification problem. Eur J Epidemiol 34:153–162.
54. Zhang Y, Tayarani M, Wang S, Liu Y, Sharma M, Joly R, *et al.* (2021): Identifying urban built environment factors in pregnancy care and maternal mental health outcomes. BMC Pregnancy Childbirth 21:599.
55. Park Y, Hu J, Singh M, Sylla I, Dankwa-Mullan I, Koski E, Das AK (2021): Comparison of methods to reduce bias from clinical prediction models of postpartum depression. JAMA Netw Open 4:e213909.
56. Tortajada S, García-Gomez JM, Vicente J, Sanjuán J, de Frutos R, Martín-Santos R, *et al.* (2009): Prediction of postpartum depression using multilayer perceptrons and pruning. Methods Inf Med 48:291–298.
57. Jiménez-Serrano S, Tortajada S, García-Gómez JM (2015): A mobile health application to predict postpartum depression based on machine learning. Telemed J E Health 21:567–574.
58. Hochman E, Feldman B, Weizman A, Krivoy A, Gur S, Barzilay E, *et al.* (2021): Development and validation of a machine learning-based postpartum depression prediction model: A nationwide cohort study. Depress Anxiety 38:400–411.
59. Zhang W, Liu H, Silenzio VMB, Qiu P, Gong W (2020): Machine learning models for the prediction of postpartum depression: Application and comparison based on a cohort study. JMIR Med Inform 8: e15516.
60. Zhang Y, Wang S, Hermann A, Joly R, Pathak J (2021): Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. J Affect Disord 279:1–8.
61. Amit G, Girshovitz I, Marcus K, Zhang Y, Pathak J, Bar V, Akiva P (2021): Estimation of postpartum depression risk from electronic health records using machine learning. BMC Pregnancy Childbirth 21:630.
62. Wang S, Pathak J, Zhang Y (2019): Using electronic health records and machine learning to predict postpartum depression. Stud Health Technol Inform 264:888–892.
63. Munk-Olsen T, Liu X, Madsen KB, Kjeldsen MZ, Petersen LV, Bergink V, *et al.* (2022): Postpartum depression: A developed and validated model predicting individual risk in new mothers. Transl Psychiatry 12:419.
64. Andersson S, Bathula DR, Iliadis SI, Walter M, Skalkidou A (2021): Predicting women with depressive symptoms postpartum with machine learning methods. Sci Rep 11:7877.
65. Shin D, Lee KJ, Adeluwa T, Hur J (2020): Machine learning-based predictive modeling of postpartum depression. J Clin Med 9.
66. Qasrawi R, Amro M, VicunaPolo S, Abu Al-Halawa D, Agha H, Abu Seir R, *et al.* (2022): Machine learning techniques for predicting depression and anxiety in pregnant and postpartum women during the COVID-19 pandemic: A cross-sectional regional study. F1000Res 11:390.
67. Javed F, Gilani SO, Latif S, Waris A, Jamil M, Waqas A (2021): Predicting risk of antenatal depression and anxiety using multi-layer perceptrons and support vector machines. J Pers Med 11.
68. Weitkamp K, Klein E, Midgley N (2016): The experience of depression: A qualitative study of adolescents with depression entering psychotherapy. Glob Qual Nurs Res 3:2333393616649548.
69. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH (2018): Ensuring fairness in machine learning to advance health equity. Ann Intern Med 169:866–872.
70. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G (2018): Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med 178:1544–1547.
71. Zhang Z (2016): Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med 4:30.
72. Lundberg S, Lee S (2017): A unified approach to interpreting model predictions. arXiv https://doi.org/10.48550/arXiv.1705.07874.
73. Friedler S, Scheidegger C, Venkatasubramanian S (2014): Certifying and removing disparate impact. arXiv https://doi.org/10.48550/arXiv.1412.3756.
74. Hardt M, Price E, Srebro N (2016): Equality of opportunity in supervised learning. arXiv https://doi.org/10.48550/arXiv.1610.02413.
75. Yang J, Soltan AAS, Eyre DW, Yang Y, Clifton DA (2023): An adversarial training framework for mitigating algorithmic biases in clinical machine learning. NPJ Digit Med 6:55.
76. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative (2019): Calibration: The Achilles heel of predictive analytics. BMC Med 17:230.
77. Breiman L (2001): Random forests. Mach Learn 45:5–32.
78. Chen T, Guestrin C (2016): XGBoost: A scalable tree boosting system. arXiv https://doi.org/10.48550/arXiv.1603.02754.
79. Zou H, Hastie T (2005): Regularization and variable selection via the elastic Net. J R Stat Soc B 67:301–320.
80. Muskens L, Boekhorst MGBM, Kop WJ, van den Heuvel MI, Pop VJM, Beerthuizen A (2022): The association of unplanned pregnancy with perinatal depression: A longitudinal cohort study. Arch Womens Ment Health 25:611–620.
81. Bedaso A, Adams J, Peng W, Xu F, Sibbritt D (2022): An examination of the association between marital status and prenatal mental disorders using linked health administrative data. BMC Pregnancy Childbirth 22:735.
82. Aochi Y, Honjo K, Kimura T, Ikehara S, Iso H, Japan Environment and Children's Study Group (2021): Association between maternal employment status during pregnancy and risk of depressive symptomatology 1 month after childbirth: The Japan Environment and Children's Study. J Epidemiol Community Health 75:531–539.
83. Steenland MW, Trivedi AN (2023): Association of Medicaid expansion with postpartum depression treatment in Arkansas. JAMA Health Forum 4:e225603.
84. Anastasopoulou SV, Bonotis KS, Hatzoglou C, Dafopoulos KC, Gourgoulianis KI (2022): Smoking patterns and anxiety factors among women expressing perinatal depression. Womens Health Rep (New Rochelle) 3:198–206.
85. Myers S, Johns SE (2019): Male infants and birth complications are associated with increased incidence of postnatal depression. Soc Sci Med 220:56–64.
86. Ataoglu A, Canan F (2009): Mean platelet volume in patients with major depression: Effect of escitalopram treatment. J Clin Psychopharmacol 29:368–371.
87. Canan F, Dikici S, Kutlucan A, Celbek G, Coskun H, Gungor A, *et al.* (2012): Association of mean platelet volume with DSM-IV major depression in a large community-based population: The MELEN study. J Psychiatr Res 46:298–302.
88. Sparling TM, Nesbitt RC, Henschke N, Gabrysch S (2017): Nutrients and perinatal depression: A systematic review. J Nutr Sci 6:e61.
89. Bair MJ, Robinson RL, Katon W, Kroenke K (2003): Depression and pain comorbidity: A literature review. Arch Intern Med 163:2433–2445.
90. Toriola AT, Vääräsmäki M, Lehtinen M, Zeleniuch-Jacquotte A, Lundin E, Rodgers KG, *et al.* (2011): Determinants of maternal sex steroids during the first half of pregnancy. Obstet Gynecol 118:1029–1036.
91. Dwyer JB, Aftab A, Radhakrishnan R, Widge A, Rodriguez CI, Carpenter LL, *et al.* (2020): Hormonal treatments for major depressive disorder: State of the art. Am J Psychiatry 177:686–705.

92. Osborne LM, Monk C (2013): Perinatal depression – The fourth inflammatory morbidity of pregnancy? theory and literature review. Psychoneuroendocrinology 38:1929–1952.
93. Ramiro-Cortijo D, de la Calle M, Böger R, Hannemann J, Lüneburg N, López-Gimenez MR, *et al.* (2020): Male fetal sex is associated with low maternal plasma anti-inflammatory cytokine profile in the first trimester of healthy pregnancies. Cytokine 136:155290.
94. Giurgescu C, Misra DP, Sealy-Jefferson S, Caldwell CH, Templin TN, Slaughter-Acey JC, Osypuk TL (2015): The impact of neighborhood quality, perceived stress, and social support on depressive symptoms during pregnancy in African American women. Soc Sci Med 130:172–180.
95. Leff-Gelman P, Mancilla-Herrera I, Flores-Ramos M, Cruz-Fuentes C, Reyes-Grajeda JP, García-Cuétara MDP, *et al.* (2016): The immune system and the role of inflammation in perinatal depression. Neurosci Bull 32:398–420.
96. Koritala BSC, Porter KI, Arshad OA, Gajula RP, Mitchell HD, Arman T, *et al.* (2021): Night shift schedule causes circadian dysregulation of DNA repair genes and elevated DNA damage in humans. J Pineal Res 70:e12726.
97. He Y, Burghardt K, Lerman K (2020): A geometric solution to fair representations. In: AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. New York, NY: Association for Computing Machinery, 279–285.