



Research and Applications

Improving postpartum hemorrhage risk prediction using longitudinal electronic medical records

Amanda B. Zheutlin ^{1,*}, Luciana Vieira^{2,*}, Ryan A. Shewcraft^{1,*}, Shilong Li¹, Zichen Wang¹, Emilio Schadt¹, Susan Gross^{1,3}, Siobhan M. Dolan^{2,3}, Joanne Stone², Eric Schadt^{1,3,*}, and Li Li ^{1,3,*}

¹Sema4, Stamford, Connecticut, USA, ²Division of Maternal Fetal Medicine, Department of Obstetrics, Gynecology, and Reproductive Science, Icahn School of Medicine at Mount Sinai, New York, New York, USA and ³Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA

*Equal contribution.

Corresponding Author: Li Li, MD, Sema4, 333 Ludlow St., North Tower, 7th Floor, Stamford, CT 06902, USA; li.li@sema4.com

Received 16 March 2021; Revised 30 June 2021; Editorial Decision 14 July 2021; Accepted 26 July 2021

ABSTRACT

Objective: Postpartum hemorrhage (PPH) remains a leading cause of preventable maternal mortality in the United States. We sought to develop a novel risk assessment tool and compare its accuracy to tools used in current practice.

Materials and Methods: We used a PPH digital phenotype that we developed and validated previously to identify 6639 PPH deliveries from our delivery cohort (N = 70 948). Using a vast array of known and potential risk factors extracted from electronic medical records available prior to delivery, we trained a gradient boosting model in a subset of our cohort. In a held-out test sample, we compared performance of our model with 3 clinical risk-assessment tools and 1 previously published model.

Results: Our 24-feature model achieved an area under the receiver-operating characteristic curve (AUROC) of 0.71 (95% confidence interval [CI], 0.69–0.72), higher than all other tools (research-based AUROC, 0.67 [95% CI, 0.66–0.69]; clinical AUROCs, 0.55 [95% CI, 0.54–0.56] to 0.61 [95% CI, 0.59–0.62]). Five features were novel, including red blood cell indices and infection markers measured upon admission. Additionally, we identified inflection points for vital signs and labs where risk rose substantially. Most notably, patients with median intrapartum systolic blood pressure above 132 mm Hg had an 11% (95% CI, 8%–13%) median increase in relative risk for PPH.

Conclusions: We developed a novel approach for predicting PPH and identified clinical feature thresholds that can guide intrapartum monitoring for PPH risk. These results suggest that our model is an excellent candidate for prospective evaluation and could ultimately reduce PPH morbidity and mortality through early detection and prevention.

Key words: postpartum hemorrhage, phenotype, electronic medical records, risk assessment, clinical decision support

INTRODUCTION

Postpartum hemorrhage (PPH) remains a leading preventable cause of maternal morbidity and mortality in the United States and worldwide.^{1–4} Recent trends in the United States suggest rates of PPH are

rising, with reported increase in prevalence from 2.9% in 2010 to 3.2% in 2014, which constitutes a 13.0% increase.^{5,6} Though the risk of maternal mortality has remained stable,^{4,7,8} PPH still repre-

sents 11.2% of maternal deaths in the United States,⁹ highlighting the need for accurate risk prediction and prevention.

PPH is most commonly defined using blood loss and clinical signs of hemodynamic compromise, though specific criteria vary.¹⁰ The American College of Obstetricians and Gynecologists reVITALize program defines obstetric hemorrhage as a cumulative estimated blood loss ≥ 1000 mL or blood loss associated with signs or symptoms of hypovolemia within 24 hours of delivery.¹¹ The World Health Organization defines postpartum hemorrhage as a cumulative blood loss of ≥ 500 mL within 24 hours after birth.¹² Differences and inconsistencies in definitions complicate identification and subsequent prediction of PPH.¹³

Nonetheless, stratification tools based on known risk factors are used to identify women at high risk of obstetric hemorrhage, promoting clinical awareness and prompting measures to mitigate risk.^{11,14} The California Maternal Quality Care Collaborative (CMQCC)¹⁵; Association of Women's Health, Obstetric and Neonatal Nurses (AWHONN)¹⁶; and New York Safety Bundle for Obstetric Hemorrhage (NYSBOH)¹⁷ are widely used in the United States at admission and during labor. These guidelines stratify women into low-, medium-, and high-risk groups based on characteristics such as low platelet count, multiple gestation, prior cesarean delivery or uterine surgery, and prior history of PPH to determine the need for pretransfusion testing.^{18,19}

Evaluation of these risk-stratification tools revealed that they have limited clinical utility.^{18,20} Assessments of the CMQCC have shown a statistically significant difference in risk for PPH between women in low-, medium-, and high-risk groups. However, this same tool has consistently classified more than 40% of PPH cases as low risk because they had no risk factors upon admission.^{18,21,22} The tool underestimates risk for various definitions of severe PPH, as well as PPH defined as >1000 mL cumulative blood loss. An assessment of the CMQCC, AWHONN, and NYSBOH toolkits for predicting severe PPH (blood transfusion of ≥ 4 units of blood) in women undergoing cesarean delivery reported better performance with only 4% to 17% of cases misclassified as low risk.²⁰ While these cases are among the most critical to detect, they account for a relatively small proportion of total PPH cases. Despite these limitations, modified versions of the CMQCC provide guidance nationally for pretransfusion testing.¹¹

To improve the accuracy of PPH risk prediction, novel approaches to predict PPH have utilized large datasets to identify more specific risk factors for PPH. Venkatesh et al²³ found that machine learning and statistical models utilizing data available at time of labor admission accurately predict PPH. However, these types of models have not been systemically compared with risk stratification tools currently being used as part of standard of care.

Previously, we developed and validated an accurate digital phenotyping algorithm to ascertain PPH from comprehensive electronic medical record (EMR) data that incorporates not only cumulative blood loss, but also other important diagnostic and treatment-related features indicating PPH, such as use of uterotonics and hemorrhage-related procedures.²⁴ The combination of machine learning methods and EMR data allows for the construction of predictive models based on population-scale analyses that involve more precisely defined outcomes and exposures. We hypothesized that EMR data from our large, diverse health system may provide for a richer feature set to construct more predictive models for prospectively identifying those at risk for PPH. Our aim was to create a comprehensive predictive tool to determine risk of PPH prior to delivery using integrated clinical features from large-scale, high-dimensional clinical data derived from the Mount Sinai Health System EMR

database. Furthermore, we compare the performance of our model with existing risk assessment tools.

MATERIALS AND METHODS

We aimed to build a novel informatics-based tool to assess risk of PPH (Figure 1). We derived thousands of potentially informative features from clinical information recorded prior to delivery for patients in our delivery cohort. Using PPH status determined by our physician-validated digital phenotyping algorithm,²⁴ which considers medication dosing and timing, fluctuations in lab values during labor and delivery, and medical observations across labor and delivery admission, in addition to the standard estimated blood loss metric, for retrospectively identifying deliveries with PPH, we built a model to estimate PPH risk prior to delivery and compared results with a previously published model²³ and 3 widely used clinical risk toolkits.^{15–17} Our approach differed from others in that we combined results from multiple machine learning methods (tree-based and regression-based) to robustly select features derived from real world data in a large healthcare system, and thus refer to this model as an “integrated machine learning” (IML) model. We received approval from the Icahn School of Medicine at Mount Sinai Institutional Review Board (IRB-17-01245) to conduct this study. Further details on all methods can be found in the [Supplementary Methods](#).

Experimental design

For each pregnancy journey, we used clinical information available from 8 months prior to pregnancy (estimated based on gestational age at delivery) up to and including delivery time. This time frame captures events occurring prior to and during pregnancy while limiting the variation in data availability across patients, which can often be substantial. To allow for independent testing of our model and to facilitate comparison across other risk assessment tools, we divided our pregnancy-delivery cohort into training (80%) and testing (20%) sets. Patients were randomly assigned to training or testing; all deliveries for a given patient were assigned to the same set.

Feature engineering and selection

Current risk stratification efforts for PPH have relied primarily on known risk factors (eg, prior cesarean delivery).²⁰ While these factors are significant predictors of PPH, in many cases, PPH occurs in women with no known risk factors.^{11,18,21} Considering this, we included both known risk factors and thousands of potential predictors extracted from EMR to maximize our ability to detect any patterns of clinical information that increase risk for PPH. These predictors spanned the total set of disease categories, generic medications, procedures, lab values, and vital signs ever detected in our cohort and included multiple ways of measuring many factors (eg, minimum and maximum pulse). Of note, we generated functional principal components to summarize multiple vital sign and lab measurements for each patient over time. We excluded any feature with fewer than 5 individuals with nonmissing values because these features are unlikely to be informative. Among the remaining features, we used a combination of gradient boosting, adaptive lasso regression, and logistic regression to select a small subset of features for input into our risk model.

Learning algorithm

Selected features were used to train gradient boosted decision tree models with 100-fold cross-validation employed within the training

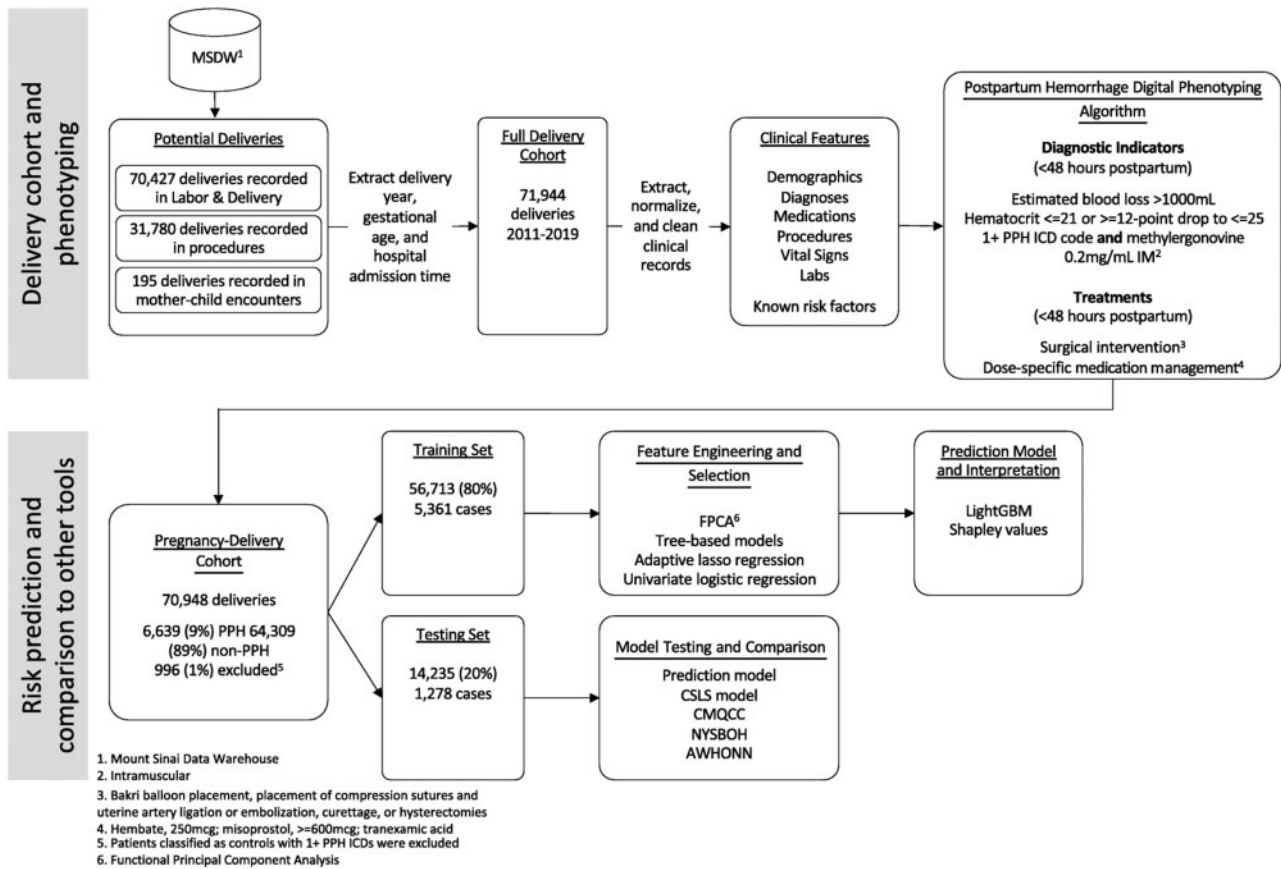


Figure 1. Overview of study design and model development. AWHONN: Association of Women's Health, Obstetric and Neonatal Nurses; CMQCC: California Maternal Quality Care Collaborative; CSLS: U.S. Consortium for Safe Labor Study; ICD: International Classification of Diseases; NYSBOH: New York Safety Bundle for Obstetric Hemorrhage; PPH: postpartum hemorrhage.

dataset with the python package LightGBM. We used stratified k-fold splits to retain PPH rates across folds and balanced sample weights to boost the relative importance of PPH deliveries in classification. We set iterations to continue only until there was no improvement in area under the receiver-operating characteristic curve (AUROC) in validation. Within folds, the best model was selected based on the F1 score, which reflects the weighted average of precision and recall. We reported average sensitivity, specificity, positive predictive value, negative predictive value, AUROC, and F1 score across folds. The final model was estimated using the full training sample with the average number of best iterations from cross-validation and then applied without modification to the test dataset. We estimated 95% confidence intervals (CIs) for test performance metrics via bootstrapping (1000 samples with replacement).

Model interpretation and simplification

Knowing what information drives prediction is key for adoption by medical professionals. Here, we used Shapley values to estimate relative importance of each feature using the python package SHAP.^{25,26} Overall feature importance can be calculated by taking the mean across individual Shapley values for each feature. Shapley values can also be transformed to relative risk scores and plotted against patients' values to show inflection points above or below which risk is substantially increased or decreased. Changes in absolute risk were calculated by multiplying Shapley-derived relative risk

score by the prevalence of PPH across the entire cohort. Finally, we used Shapley values to simplify our model to the minimum necessary features for maximum performance to improve its clinical utility and ease of interpretation.

Comparison with other risk assessment tools

We extracted risk factors for 3 commonly used PPH risk assessments: the CMQCC, AWHONN, and NYSBOH.¹⁵⁻¹⁷ We assigned women to low-, medium-, or high-risk groups based on 16 to 17 largely overlapping binary criteria that evaluate medical history, obstetric complications, and current vital signs and labs.²⁰ We used intrapartum versions when available as this timing better aligns with the timing of our model. We used both high and medium risk as cut-offs for computing sensitivity, specificity, positive predictive value, negative predictive value, and AUROC for each of the 3 toolkits.

We also implemented a previously published EMR-based risk prediction model that was based on the U.S. Consortium for Safe Labor Study (CSLS), 2002 to 2008, as an additional comparison.²³ The authors used several statistical methods to analyze 55 risk factors. Here, we trained a gradient boosting machine (their highest performing method) using all available risk factors (54 of 55; 1 variable, parity, is not available in our dataset) in our training sample and applied this model to our test sample for a direct comparison. Significant differences between our model and all other risk assess-

ment tools were assessed using a 2-sided DeLong test for comparing AUROCs within the same sample.²⁷

Comparison with an alternative phenotype

Because the PPH phenotype we used here is newly developed, we evaluated our model and all comparison risk assessment tools against a more commonly used phenotype, estimated blood loss (EBL) ≥ 1000 mL, as well. For these analyses, we followed the same procedures as described previously, but using only patients with nonmissing EBL values ($n = 63\,348$) and using EBL alone to define our outcome (deliveries with EBL ≥ 1000 mL were labeled PPH, mean 1204 ± 496 mL; deliveries with EBL < 1000 mL were labeled controls, mean 414 ± 229 mL).

RESULTS

Patient cohort demographics and clinical characteristics

Our final pregnancy-delivery cohort included 70 948 deliveries (79.6%, $n = 56\,509$ of 70 948 independent patients), and PPH prevalence was 9% (Figure 1). PPH and non-PPH deliveries differed significantly in several demographic and clinical characteristics (Table 1). Additional summary statistics were reported previously.²⁴

Selected features and model performance

We generated 5327 features in our training sample, 3982 of which had at least 5 nonmissing values. Of these features, we selected 219 that showed significant association with PPH across 2 selection procedures (gradient boosting and either adaptive lasso or logistic

regression). These 219 derived features represented 98 unique raw features (eg, minimum and maximum pulse would be counted as 2 features, but only 1 unique feature). These derived features excluded any features that were used for the PPH digital phenotype.²⁴

All selected features were included in cross-validation training, but only 80 unique raw features (178 total derived features) received nonzero importance scores (Supplementary Table S2). Using these 80 features, our IML model achieved an AUROC of 0.73 (95% CI, 0.72-0.74) in the training set and an AUROC of 0.72 (95% CI, 0.70-0.73) in the test set. Performance across subsets of features based on their importance scores suggested that there was minimal additional value added beyond the top 29 derived features, representing 24 unique raw features (Supplementary Figure S1). Using only these top 24 features, AUROC was 0.71 (95% CI, 0.69-0.72) in the test set. Additional performance metrics were listed in Supplementary Table S1 (training) and Table 2 (testing).

Top 24 features highlight novel risk markers measured at admission and intrapartum

Among the top 24 unique features, 7 were lab results, 6 were diagnoses, 4 were vital signs, 4 were demographic variables, and 3 were medications. Considering the full list of 29 features, 19 (66%) features were ones used in current clinical practice or reported previously, including anemia, preeclampsia, and Cesarean delivery (Supplementary Table S4). Five (17%) features were novel measures of risk factors previously reported. For example, we reported antepartum pulse as a top feature; pulse is a known risk factor, however,

Table 1. Demographics and clinical characteristics for the Mount Sinai Health System delivery cohort

	Pregnancy-Delivery Cohort	PPH	Non-PPH
Demographics			
Number of deliveries	70 948 (100)	6639 (9)	64 309 (89)
Age at delivery, y ^a	32 \pm 6	33 \pm 6	32 \pm 6
Race^a			
White	39 977 (56)	3176 (48)	36 801 (57)
African American	7318 (10)	911 (14)	6407 (10)
Asian	5728 (8)	622 (9)	5106 (8)
Native American	278 (<1)	25 (<1)	253 (<1)
Other	13 256 (19)	1495 (22)	11 761 (18)
Unknown	4391 (6)	410 (6)	3981 (6)
Ethnicity^a			
Non-Hispanic	40 058 (57)	3444 (55)	36 629 (57)
Hispanic	11 313 (16)	1269 (19)	10 044 (16)
Unknown	19 577 (28)	15 686 (25)	17 891 (28)
Insurance^a			
Private	41 443 (59)	3633 (55)	37 810 (59)
Medicaid or Medicare	23 301 (33)	2474 (37)	20 827 (32)
Uninsured	464 (1)	46 (1)	418 (1)
Other or missing	5740 (8)	486 (7)	5254 (8)
Clinical characteristics at hospital admission			
Body mass index, kg/m ^{2a}	29 \pm 5	30 \pm 6	29 \pm 5
SBP, mm Hg ^a	121 \pm 14	125 \pm 16	121 \pm 14
DBP, mm Hg ^a	73 \pm 11	75 \pm 12	72 \pm 11
Temperature, °F ^a	98.2 \pm 0.4	98.3 \pm 0.5	98.2 \pm 0.4
Hematocrit, % ^a	36 \pm 3	35 \pm 4	36 \pm 3
Platelets, 10 ⁻⁹ /L ^a	207 \pm 57	202 \pm 61	207 \pm 57
Gestational weeks at delivery ^a	39 \pm 2	38 \pm 3	39 \pm 2

Values are n (%) or mean \pm SD.

DBP: diastolic blood pressure; PPH: postpartum hemorrhage; SBP: systolic blood pressure.

^aSignificant difference between cases and controls, $P < .001$.

Table 2. Vital signs and lab values show discrete increases in relative risk for PPH

Feature During Hospital Admission	SHAP-Based Cut Point	Increase in Relative Risk (Absolute Risk) for PPH		
		Quartile 1	Median	Quartile 3
Minimum SBP	132 mm Hg	9% (0.9%)	11% (1.0%)	13% (1.2%)
Minimum DBP	85 mm Hg	2% (0.2%)	2% (0.2%)	3% (0.3%)
Median pulse	90 beats/min	3% (0.3%)	4% (0.4%)	5% (0.5%)
Minimum hematocrit	30%	2% (0.1%)	5% (0.4%)	7% (0.7%)
Minimum hemoglobin	10.0 g/dL	2% (0.2%)	3% (0.3%)	6% (0.6%)
Minimum platelets	$150 \times 10^{-9}/L$	1% (0.1%)	1% (0.1%)	2% (0.2%)

DBP: diastolic blood pressure; SBP: systolic blood pressure.

Table 3. Performance matrix across risk assessment tools in our independent test set

Risk Assessment Tool	Sensitivity	Specificity	PPV	NPV	AUROC
Integrated Machine Learning (all 80 variables)	0.57 (0.54–0.60)	0.73 (0.72–0.74)	0.17 (0.16–0.18)	0.95 (0.94–0.95)	0.72 (0.70–0.73)
Integrated Machine Learning (top 24 variables)	0.58 (0.55–0.61)	0.71 (0.70–0.72)	0.17 (0.15–0.18)	0.95 (0.94–0.95)	0.71 (0.69–0.72)
Consortium for Safe Labor Study	0.56 (0.53–0.59)	0.69 (0.68–0.70)	0.15 (0.14–0.16)	0.94 (0.94–0.95)	0.67 (0.66–0.69)
Intrapartum CMQCC—high risk	0.27 (0.24–0.30)	0.88 (0.88–0.89)	0.19 (0.17–0.20)	0.92 (0.92–0.93)	0.58 (0.56–0.59)
Intrapartum CMQCC—medium risk	0.63 (0.59–0.68)	0.58 (0.57–0.59)	0.13 (0.12–0.14)	0.94 (0.94–0.95)	0.61 (0.59–0.62)
Intrapartum NYSBOH—high risk	0.22 (0.20–0.25)	0.87 (0.87–0.88)	0.15 (0.13–0.16)	0.92 (0.91–0.92)	0.55 (0.54–0.56)
Intrapartum NYSBOH—medium risk	0.60 (0.56–0.65)	0.61 (0.60–0.62)	0.13 (0.12–0.14)	0.94 (0.93–0.94)	0.60 (0.59–0.62)
Admission AWHONN—high risk	0.43 (0.40–0.47)	0.75 (0.74–0.76)	0.15 (0.13–0.16)	0.93 (0.93–0.94)	0.59 (0.58–0.61)
Admission AWHONN—medium risk	0.89 (0.84–0.94)	0.20 (0.20–0.21)	0.10 (0.09–0.10)	0.95 (0.94–0.96)	0.55 (0.54–0.56)

AUROC: area under the receiver-operating characteristic curve; AWHONN: Association of Women's Health, Obstetric and Neonatal Nurses; CMQCC: California Maternal Quality Care Collaborative; NPV: negative predictive value; NYSBOH: New York Safety Bundle for Obstetric Hemorrhage; PPH: postpartum hemorrhage; PPV: positive predictive value.

antepartum measures have not been previously considered. As such, we find that patient data in the 24 feature model is more complete than the full, 80-feature model (Supplementary Figure S2). Finally, 5 (17%) features represent newly identified risk factors that may warrant further investigation including red blood cell count, mean corpuscular hemoglobin, red blood cell distribution width, absolute neutrophil count, and white blood cell count, many of which are measured in complete blood count panels. These features were available in 62.5% of PPH deliveries and 55.0% of non-PPH deliveries. To illustrate how each unique feature contributed to individual predicted risk, we plotted patients' Shapley values for the most important version of each of the top 24 unique raw features, excluding functional principal components, colored by the feature value itself (Figure 2). Shapley values reflect the relative contribution of the feature to the patient's predicted risk score; often, high feature values had a larger impact on an individual's risk score.

Blood pressure and pulse are measured frequently across the duration of hospital admission to monitor patient well-being. We calculated 3-hour moving averages and SDs for cases and controls during the 12 hours preceding delivery (Figure 3). Fixed-effects models confirmed patients who developed PPH had consistently higher blood pressure and pulse in the hours prior to delivery (and

hemorrhage) than patients who delivered without PPH ($P_s < 2 \times 10^{-16}$) (Supplementary Table S4). Additionally, we found that relative risk for PPH increased, sometimes dramatically, when values for these key vital signs, as well as for the primary labs used for PPH risk assessment upon admission, passed certain inflection points (Table 2, Figure 4). As noted with the shaded areas in each plot, these points did not always align with the reference range for normal pregnancy values.

Predictive risk model outperforms previously published and existing risk assessment tools

Our 24-feature IML model achieved a significantly higher AUROC than all other risk assessment tools we applied ($P_s < 1.94 \times 10^{-7}$ from 2-sided DeLong test) (Table 3, Supplementary Figure S3). The CSLS model achieved an AUROC of 0.67 (95% CI, 0.66–0.69) and the clinical risk assessment toolkits yielded AUROC values between 0.55 (95% CI, 0.54–0.56) and 0.61 (95% CI, 0.59–0.62) across toolkits and case classification thresholds (Table 3). Because the CSLS and our models had no clear risk category thresholds, we assigned risk category labels using deciles of predicted risk (top 10% = high risk; 60%–90% = medium risk, <60% = low risk) to compare precision across risk tools (see

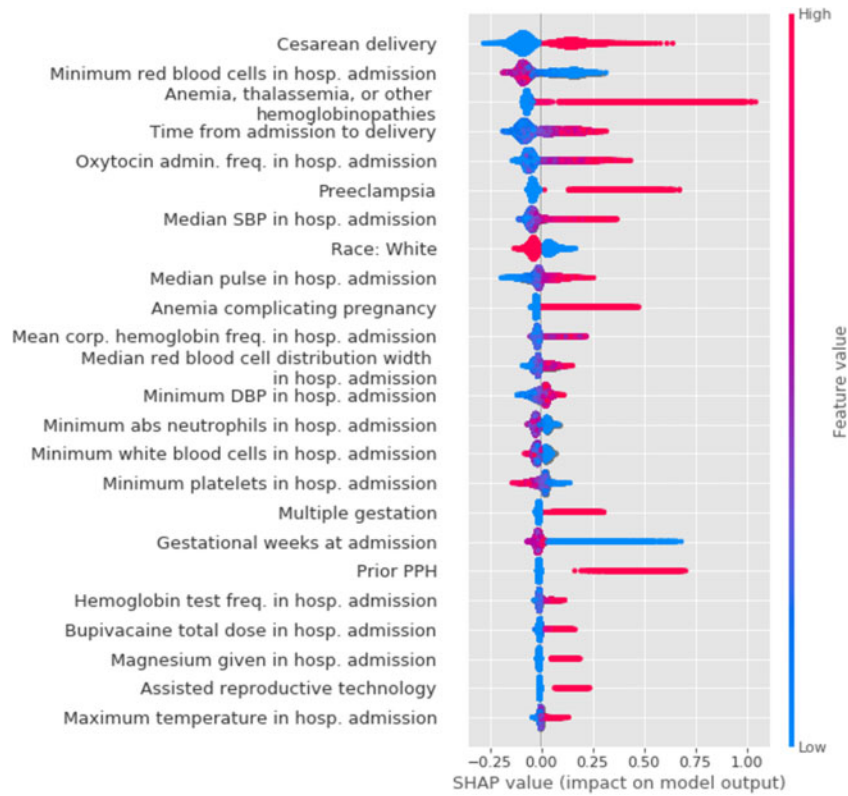


Figure 2. SHAP summary plot. SHAP summary plot for top 24 clinical features for postpartum hemorrhage (PPH) prediction shows the SHAP values for the most important features from Gradient boosting model in the training data. Features in the summary plot (y-axis) are ordered by the mean absolute SHAP values (x-axis), which represents the importance of the feature in driving the PPH prediction. Values of the feature for each patient are colored by their relative value, with red color indicating high value and blue color indicating low value. Positive SHAP values indicate increased risks for PPH and negative values indicate protective effects to PPH. DBP: diastolic blood pressure; freq.: frequency; hosp.: hospital; SBP: systolic blood pressure.

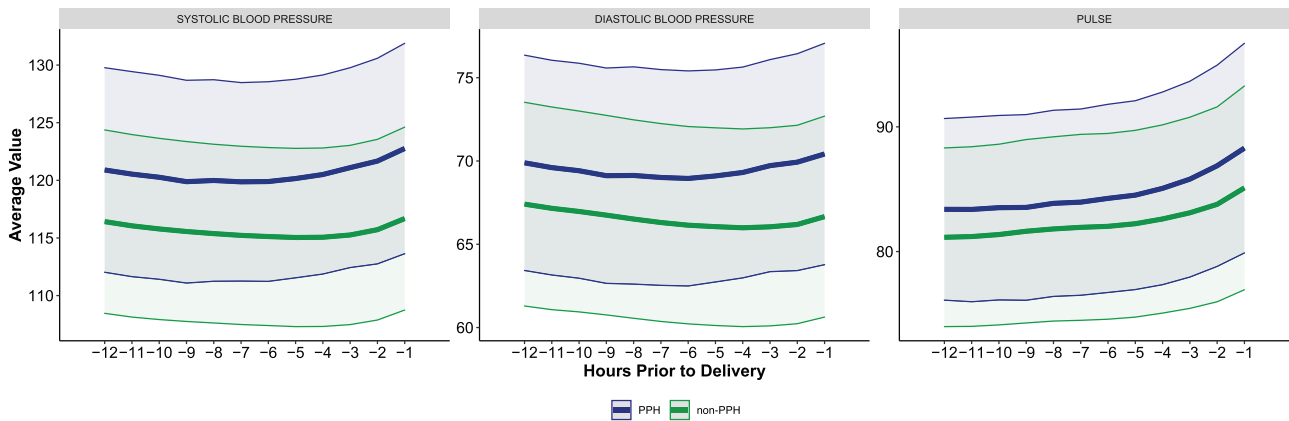


Figure 3. Dynamic changes of 3 vital signs consistently measured prior to delivery. Moving averages and standard deviations with 3-hour windows across the 12 hours prior to delivery were computed for cases and controls. PPH: postpartum hemorrhage.

Supplementary Table S4). Among the high-risk category, PPV was 28% for our model, 24% for the CSLS model, and 15% to 19% for the clinical toolkits (Figure 5).

Finally, we found a similar pattern of results using a more commonly implemented phenotype, EBL ≥ 1000 mL, rather than our digital phenotyping algorithm (Supplementary Table S5). Case prevalence using EBL alone was 7% overall (n = 4182). Our IML model achieved an AUROC of 0.85 (95% CI, 0.84-0.87) using all 80

unique features, with no loss in AUROC when restricting the top 24 unique features (0.85; 95% CI, 0.84-0.86); this was significantly higher than all other risk tools ($P_s < 2.2 \times 10^{-16}$ from 2-sided DeLong test). The CSLS model had an AUROC of 0.77 (95% CI, 0.75-0.78), and the clinical toolkits had AUROCs ranging from 0.55 (95% CI, 0.54-0.56) to 0.66 (95% CI, 0.64-0.57) depending on the toolkit and case classification threshold (Supplementary Table S5).

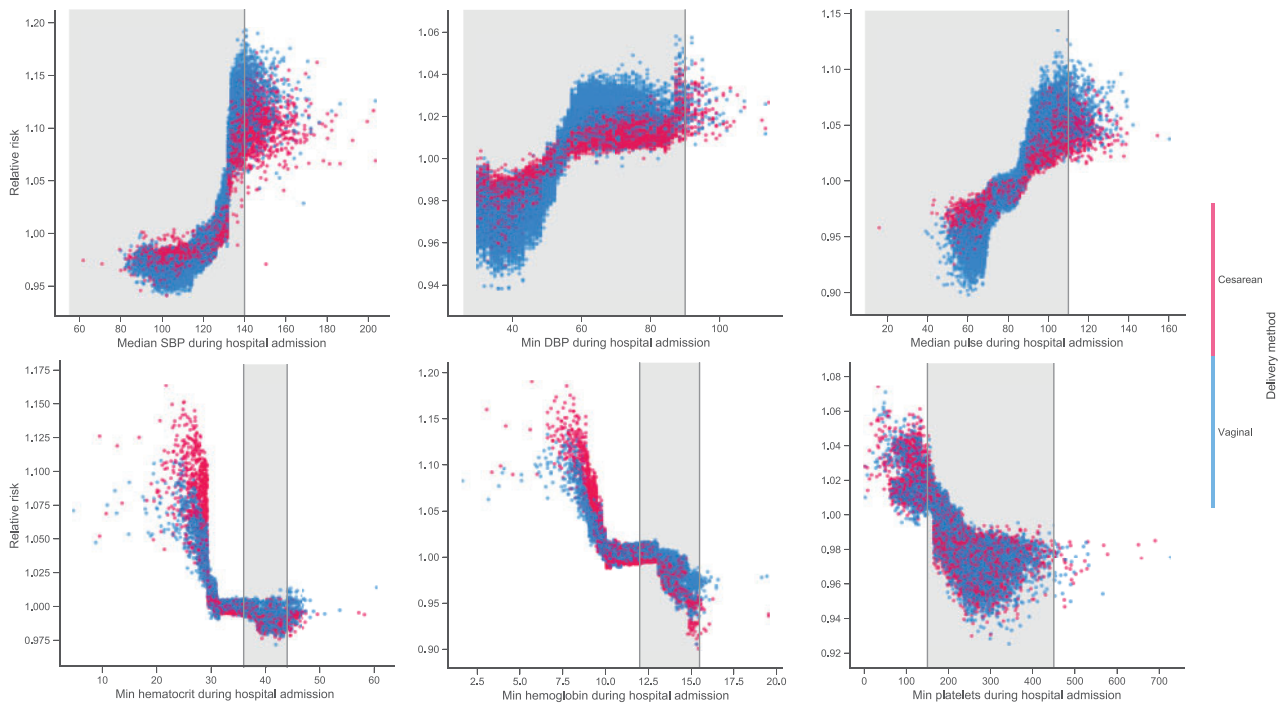


Figure 4. SHAP dependency plot. SHAP scores (relative risks, y-axis) for postpartum hemorrhage (PPH) prediction was plotted against feature values (x-axis) for patients in the training data. The plot shows how different values of the features can affect relative risks and ultimately impact classifier decision for 6 vital signs and lab measurements stratified by type of delivery. The shaded gray area reflects the reference ranges for the corresponding vital signs or lab measures. Data points are colored by the delivery method (Cesarean or vaginal). DBP: diastolic blood pressure; SBP: systolic blood pressure.

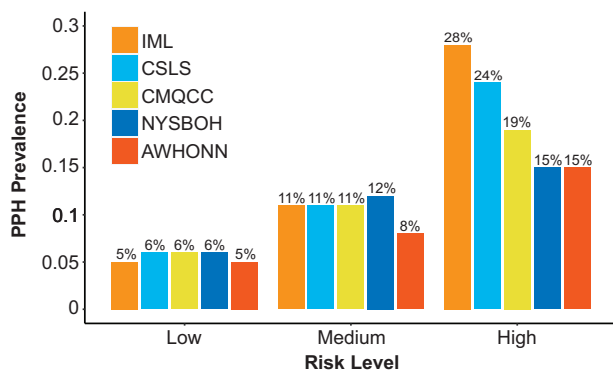


Figure 5. Postpartum hemorrhage (PPH) prevalence among patients of different risk groups varies by risk assessment tools. Case prevalence within each risk category for each risk tool was calculated. Risk categories were assigned using deciles for U.S. Consortium for Safe Labor Study (CSLS) and Sema4 models (high risk = top 10%, medium risk = 60%-90%, low risk = <60%). AWHONN: Association of Women's Health, Obstetric and Neonatal Nurses; CMQCC: California Maternal Quality Care Collaborative; IML: integrated machine learning; NYSBOH: New York Safety Bundle for Obstetric Hemorrhage.

DISCUSSION

This study highlights a novel approach to predicting postpartum hemorrhage utilizing a rich, diverse EMR dataset spanning 9 years of deliveries. Our IML model used only 24 clinical variables—all passively collected through routine clinical care—to evaluate risk for PPH more accurately than 3 currently used risk assessment tools and a previously developed EMR-based model. When we used deciles of predicted risk to assign patients to risk categories, we found that PPH prevalence was 28% in our high-risk category, nearly twice as

high as when risk was determined using clinical risk tools (15%-19%) (Figure 5). Additionally, we identified inflection points for vital signs and labs where risk for PPH increased. These can help to guide risk management for PPH, while several novel risk factors may be additionally useful for monitoring risk during hospital admission. Finally, we showed that phenotype sensitivity can have a high impact on risk prediction research by comparing models using a highly accurate digital phenotype vs a less sensitive, blood loss-based one.

One significant advantage of our model is that it substantially outperformed currently used clinical tools with only 24 features, all of which are generally assessed prior to delivery. Our feature selection process was key to this success. All other risk assessment tools were based on expert opinion and clinical consensus, which does not always result in a set of features that maximize predictive accuracy. By assaying thousands of potential risk factors and using a suite of data-driven approaches to find the optimal set, we selected 24 known and novel risk factors that together delivered the highest performance of the tools we tested, including a model with more than twice as many features (CSLS, 55 features).

Our study also offered high-resolution analysis of the PPH status-dependent temporal trends for vital signs and labs that are assessed and used for monitoring the peripartum period. In the obstetric population, there is a wide range of hemodynamic changes associated with blood loss.^{28,29} Although hemodynamic changes encompass the reVITALize definition for PPH, there are no discrete definitions for these changes. Our model shows significant increasing trends for systolic blood pressure (SBP), diastolic blood pressure (DBP), and pulse in cases prior to delivery (Supplementary Table S4). As depicted in Figure 3, SBP, DBP, and pulse begin rising at approximately 5 hours prior to delivery. There is significant variability

in the relationship between blood loss and clinical signs, but these trends may provide useful insights for clinical care.²⁹

Additionally, we highlighted some notable transition points for important features such as platelets, hemoglobin, hematocrit, SBP, DBP, and pulse (Table 2; Figure 4). Notably, platelet count >150 appears to be protective for PPH. Recent data has shown that mild thrombocytopenia is an independent risk factor for PPH.^{30,31} Nonetheless, our cohort provided robust data clearly showing an increasing risk of hemorrhage with decreasing platelet for both vaginal and cesarean deliveries. Furthermore, we found that hematocrit and hemoglobin values reflected differential risks for PPH for cesarean relative to vaginal deliveries. Risk for all patients increased with values lower than 30% or 10.0 g/dL, respectively (Table 2), but the increase was steeper for cesarean deliveries than for vaginal deliveries (Figure 4). Cesarean deliveries were performed in 35.3% (n = 25 074 of 70 948) of patients. Finally, we found that risk increased at different thresholds for SBP, DBP, and pulse than what would be expected based on reference ranges for normal values (Table 2). For SBP, we observed an 11% median increase in risk for PPH when values were above 132 mm Hg, despite a typical cutoff of <140 mm Hg in pregnancy. Similarly, high pulse and DBP increased in risk at values within the normal range: 90 beats/min and 85 mm Hg, respectively, while >110 beats/min and >90 mm Hg are considered high in pregnancy. We also note differences in vital signs between vaginal and cesarean deliveries which reflect physiologic differences in mode of delivery (active pushing in vaginal delivery vs closely monitored blood pressures intraoperatively with neuraxial or general anesthesia). Nonetheless, our data suggest that intrapartum vital signs can capture risk for PPH and using PPH-specific risk guidelines may enhance monitoring relative to using reference ranges derived from the general population.

We also highlight several novel risk factors that are not currently monitored but are available from routine lab measurements and provide additional information for clinicians. We found that red blood cells indices (red blood cell distribution width, red blood cell count, and mean corpuscular hemoglobin) were key risk factors for PPH in our model. Red blood cell distribution width can be a marker for disease severity for underlying conditions such as diabetes, cardiovascular disease, and chronic kidney disease, as well.^{32,33} These indices are likely associated with anemia, which impairs the physiologic response to blood loss and leads to worse prognosis. Further, infectious markers (white blood cell count and absolute neutrophil count) also had a high importance. These markers may indicate underlying inflammatory process, which also may alter or blunt the maternal response to blood loss leading to PPH. While these factors are already routinely assessed upon admission to Labor & Delivery, they are not currently used for monitoring risk. This may be a promising future direction for detecting individuals at risk, especially individuals presenting without other known risk factors.

Our final contribution was evaluating all risk prediction tools against a highly accurate PPH phenotype. Any assessment of the performance of a risk tool depends on how accurately the outcome it aims to predict is measured. Here, we compared performance of all risk tools for 2 definitions of PPH: (1) a digital phenotype that we developed²⁴ and (2) a phenotype based exclusively on a blood loss threshold for hemorrhage (≥ 1000 mL) recommended by American College of Obstetricians and Gynecologists¹¹ and assessed by the CCLS model.²³ We have previously compared these phenotypes to chart review labels and found that the digital phenotype was significantly more accurate (AUROC of 0.85 vs 0.67).²⁴ A key driver of this discrepancy is that many patients with PPH do not pass this

blood loss threshold (but can be identified using other measures of blood loss or receipts of treatment) and thus are misclassified as controls.¹³ Here, using the same features and statistical methods for each tool, all risk models achieved higher performance for the EBL-based phenotype. Our model improved 14 points in AUROC (0.85 [95% CI, 0.84-0.86] vs 0.71 [95% CI, 0.69-0.72]) by switching phenotypes. However, this high performance is not particularly meaningful given the low accuracy of the phenotype and underlines the importance of both developing high-quality phenotypes and building risk assessment tools based on well-measured outcomes. Most prior work aimed at improving risk prediction for PPH has not benefited from using robust phenotypes, and this remains a major barrier to achieving that goal.

Our study comes with several limitations: it is retrospective, relies on data from a single healthcare system, and does not utilize free-text notes. Our study may not reflect performance in a real-world clinical setting, and prospective validation of this work with model calibration³⁴ in diverse healthcare systems is critical to evaluate any potential clinical utility, although currently no such prognostic model exists for the general obstetric population.³⁵ Nonetheless, EMR in Mount Sinai Health System is the one of the largest and most comprehensive EMR systems, representing racial and ethnic diversity in New York City, as well as EMR implementation from various data sources. In the current study, we did not assess for differences in prediction by race or ethnicity. Although we note differences in rates of postpartum hemorrhage by race, ethnicity, and insurance type, we have not accounted for potential differences in baseline characteristics or other risk factors that may contribute to these differences. Given this, it may be premature to incorporate these differences into our model and conclusions.³⁶ This would be an important area for future study. Finally, while our use of deidentified data can facilitate the application of our tools in research settings, systematic incorporation of data from clinical notes could bring valuable information particularly for patients missing key delivery details or without additional confirmation of PPH.

In summary, we utilized a large, robust, and diverse dataset to develop a novel risk prediction model for PPH, one of the leading causes of maternal mortality in the United States. In comparison with existing risk assessment tools, including those currently used in clinical practice, we are able to achieve a higher AUROC with relatively few features using a highly accurate digital phenotype for PPH. We further identified inflection points for vital signs and lab values in which risk for PPH begins to increase, which can be used as guidelines for monitoring risk intrapartum. The improvements in risk assessment afforded by our approach will require real-time, prospective evaluation in a hospital setting.³⁷ This tool affords integration of multiple important clinical factors to best predict risk of clinically significant PPH. Our results suggest that using this model could facilitate early identification of PPH and allocation of appropriate resources,³⁸ such as increased personnel, quick access to uterotonics, and identification of patients who need close observation. Ultimately, this may lead to reduced incidence, lower severity, and lower rates of maternal mortality.

FUNDING

This project was performed in collaboration with Sema4. Sema4 is a company that integrates genetic testing and data analytics to improve diagnosis, treatment, and prevention of disease. The Icahn School of Medicine at Mount Sinai holds equity in this for-profit company.

AUTHOR CONTRIBUTIONS

LL and ErS were involved in concept and design. ABZ, RAS, SL, ZW, and EmS were involved in data acquisition, cleaning, and interpretation of data. ABZ and RAS were involved in data analysis. ABZ, LV, RAS, and LL were involved in drafting of the manuscript. All authors were involved in critical revision of the manuscript for important intellectual content. LL and ErS were involved in supervision. All authors take responsibility for the final, published version and are accountable for all aspects of the work.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

We thank Mount Sinai Data Warehouse physician team for validating data accuracy and facilitating the chart review process. We also thank the Sema4 IT team for infrastructural and computational support.

CONFLICT OF INTEREST STATEMENT

Noe declared.

DATA AVAILABILITY STATEMENT

We used several open-source libraries to build our machine learning model, namely “fdapace” (<https://github.com/functionaldata/tPACE>) in R for FPCA and LightGBM (<https://lightgbm.readthedocs.io/en/latest/>) and scikit-learn (<https://scikit-learn.org/stable/>) in Python and will release the code under the CC BY-NC-SA 3.0 license (<https://creativecommons.org/licenses/by-nc-sa/3.0/>). However, our data collection, cleaning, and quality control framework makes use of proprietary data structures and libraries, so we are not releasing or licensing this code. We provide implementation details in the methods section and in the [supplementary information](#) to allow for independent replication.

REFERENCES

- Say L, Chou D, Gemmill A, *et al*. Global causes of maternal death: a WHO systematic analysis. *Lancet Glob Heal* 2014; 2 (6): e323–33.
- Khan KS, Wojdyla D, Say L, *et al*. WHO analysis of causes of maternal death: a systematic review. *Lancet* 2006; 367 (9516): 1066–74.
- Building U.S. Capacity to Review and Prevent Maternal Deaths. 2018. http://reviewtoaction.org/2018_Report_from_MMRCs. Accessed December 12, 2020.
- Jayakumaran JS, Schuster M, Ananth CV. 260: Postpartum hemorrhage and its risk of maternal deaths in the US. *Am J Obstet Gynecol* 2020; 222 (1): S178–9.
- Reale SC, Easter SR, Xu X, *et al*. Trends in postpartum hemorrhage in the United States from 2010 to 2014. *Anesth Analg* 2020; 130 (5): e119–22.
- Data on Selected Pregnancy Complications in the United States. Atlanta, GA: Centers for Disease Control and Prevention. <https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pregnancy-complications-data.htm>. Accessed December 10, 2020.
- Marshall AL, Durani U, Bartley A, *et al*. The impact of postpartum hemorrhage on hospital length of stay and inpatient mortality: a National Inpatient Sample-based analysis. *Am J Obstet Gynecol* 2017; 217 (3): 344.e1–6.
- Creanga AA, Syverson C, Seed K, *et al*. Pregnancy-related mortality in the United States, 2011–2013. *Obstet Gynecol* 2017; 130 (2): 366–73.
- Petersen EE, Davis NL, Goodman D, *et al*. Vital signs: pregnancy-related deaths, united states, 2011–2015, and strategies for prevention, 13 states, 2013–2017. *MMWR Morb Mortal Wkly Rep* 2019; 68: 423–9.
- Quantitative blood loss in obstetric hemorrhage. *Obstet Gynecol* 2019; 134: 1368–9.
- Shields LE, Goffman D, Caughey AB. Practice Bulletin No. 183: Postpartum hemorrhage. *Obstet Gynecol* 2017; 130: e168–86.
- Vogel JP, Williams M, Gallos I, *et al*. WHO recommendations on uterotonics for postpartum haemorrhage prevention: what works, and which one? *BMJ Glob Health* 2019; 4 (2): e001466.
- Clapp MA, McCoy TH, James KE, *et al*. The utility of electronic health record data for identifying postpartum hemorrhage. *Am J Obstet Gynecol MFM* 2021; 3 (2): 100305.
- Main EK, Goffman D, Scavone BM, *et al*; Council on Patient Safety in Women’s Health Care. National partnership for maternal safety. *Obstet Gynecol* 2015; 126 (1): 155–62.
- Bingham D, Melsop K, Main EK. *CMQCC Obstetric Hemorrhage Hospital Level Implementation Guide*. Stanford University, Palo Alto, CA: California Maternal Quality Care Collaborative; 2010.
- Association of Women’s Health Obstetric and Neonatal Nurses. Postpartum Hemorrhage Project: A Multi-Hospital Quality Improvement Program. <https://www.awhonn.org/postpartum-hemorrhage-pph/>. Accessed May 5, 2021.
- American Congress of Obstetricians and Gynecologists. Maternal Safety Bundle for Obstetric Hemorrhage. 2019. <https://www.acog.org/-/media/project/acog/acogorg/files/forms/districts/smi-ob-hemorrhage-bundle-risk-assessment-ld-admin-intrapartum.pdf>. Accessed May 5, 2021.
- Dilla AJ, Waters JH, Yazer MH. Clinical validation of risk stratification criteria for peripartum hemorrhage. *Obstet Gynecol* 2013; 122 (1): 120–6.
- Bingham D, Scheich B, Bateman BT. Structure, process, and outcome data of AWHONN’s postpartum hemorrhage quality improvement project. *J Obstet Gynecol Neonatal Nurs* 2018; 47 (5): 707–18.
- Kawakita T, Mokhtari N, Huang JC, *et al*. Evaluation of risk-assessment tools for severe postpartum hemorrhage in women undergoing cesarean delivery. *Obstet Gynecol* 2019; 134 (6): 1308–16.
- Ruppel H, Liu VX, Gupta NR, *et al*. Validation of postpartum hemorrhage admission risk factor stratification in a large obstetrics population. *Am J Perinatol* 2020 May 26 [E-pub ahead of print]. doi: 10.1055/s-0040-1712166.
- Kramer MS, Berg C, Abenheim H, *et al*. Incidence, risk factors, and temporal trends in severe postpartum hemorrhage. *Am J Obstet Gynecol* 2013; 209 (5): 449.e1–7.
- Venkatesh KK, Strauss RA, Grotegut CA, *et al*. Machine learning and statistical models to predict postpartum hemorrhage. *Obstet Gynecol* 2020; 135 (4): 935–44.
- Zheutlin AB, Viera L, Shewcraft RA, *et al*. A comprehensive digital phenotype for postpartum hemorrhage. *medRxiv*, doi: <https://medrxiv.org/content/10.1101/2021.03.01.21252691>, 4 Mar 2021, preprint: not peer reviewed.
- Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, *et al*, eds. *Advances in Neural Information Processing Systems 30*. Red Hook, NY: Curran Associates, Inc.; 2017.
- Lundberg SM, Nair B, Vavilala MS, *et al*. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018; 2 (10): 749–60.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics* 1988; 44 (3): 837–45.
- Borovac-Pinheiro A, Pacagnella RC, Cecatti JG, *et al*. Postpartum hemorrhage: new insights for definition and diagnosis. *Am J Obstet Gynecol* 2018; 219 (2): 162–8.
- Pacagnella RC, Souza JP, Durocher J, *et al*. A systematic review of the relationship between blood loss and clinical signs. *PLoS One* 2013; 8 (3): e57594.
- Attali E, Epstein D, Reicher L, *et al*. Mild thrombocytopenia prior to elective cesarean section is an independent risk factor for blood transfusion.

- Arch Gynecol Obstet* 2021; 304 (3): 627–32. doi: 10.1007/s00404-021-05988-x.
31. Govindappagari S, Moyle K, Burwick RM. Mild thrombocytopenia and postpartum hemorrhage in nulliparous women with term, singleton, vertex deliveries. *Obstet Gynecol* 2020; 135 (6): 1338–44.
 32. Al-Kindi SG, Refaat M, Jayyousi A, *et al.* Red cell distribution width is associated with all-cause and cardiovascular mortality in patients with diabetes. *Biomed Res Int* 2017; 2017: 5843702.
 33. Ferreira JP, Lamiral Z, Bakris G, *et al.* Red cell distribution width in patients with diabetes and myocardial infarction: An analysis from the EXAMINE trial. *Diabetes, Obes Metab* 2021; 23 (7):1580–7.
 34. Van Calster B, McLernon DJ, van Smeden M, *et al.*; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; 17 (1): 230.
 35. Neary C, Naheed S, McLernon D, *et al.* Predicting risk of postpartum haemorrhage: a systematic review. *BJOG* 2021; 128 (1): 46–53.
 36. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med* 2020; 383 (9): 874–82.
 37. Vogenberg FR. Predictive and prognostic models: implications for health-care decision-making in a modern recession. *Am Heal Drug Benefits* 2009; 2: 218–22.
 38. Kleinrouweler CE, Cheong-See FM, Collins GS, *et al.* Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol* 2016; 214 (1): 79–90.e36.