Machine Learning for Prediction of Maternal Hemorrhage and Transfusion

Running Title: Machine learning for maternal hemorrhage

Homa K. AHMADZIA, MD, MPH,[1] Ms. Alexa C. DZIENNY, BS,[2] Mike Bopf, MS,[3] Jerome J FEDERSPIEL, MD, PhD,[4] Richard AMDUR, PhD,[5] Madeline Murguia RICE, PhD,[6] Laritza RODRIGUEZ, MD, PhD[2]

1. Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, George Washington University, Washington, DC

2. The George Washington University School of Medicine and Health Sciences, Washington, DC

3. Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD

4. Department of Obstetrics and Gynecology Division of Maternal-Fetal Medicine Duke University, Durham, NC

5. Medical Faculty Associates, The George Washington University School of Medicine and Health Sciences, Washington, DC

6. George Washington University Biostatistics Center, Washington DC

Disclosure Statement: The authors report no conflict of interest

Funding information: HKA was supported by K23HL141640, and JJF was supported by TL1TR002555

Corresponding author:
Homa K. Ahmadzia, MD, MPH
Division of Maternal Fetal Medicine
Department of Obstetrics and Gynecology
The George Washington University School of Medicine and Health Sciences
2100 Eye Street, NW, 20052
Phone: 202-741-2500
Email: hahmadzia@mfa.gwu.edu

ABSTRACT

Rationale, Aims and Objectives: Postpartum hemorrhage (PPH) is a leading cause of maternal morbidity and mortality worldwide. One approach to reduce PPH burden is to tailor management based on predicted PPH risk. However, current risk stratification algorithms are based on risk factors identified through traditional statistical approaches or expert opinion. Machine learning offers the potential to optimize PPH prediction by allowing for more complex model structures. We sought to improve PPH prediction and to compare machine learning and traditional statistical methods.

Method: We developed predictive models using the Consortium for Safe Labor (2002-2008) dataset, including data from 12 hospitals across the United States. The primary outcome of interest was a composite of transfusion of blood products or PPH (estimated blood loss ≥ 1,000 mL). The secondary outcome was transfusion of any blood products. Fifty maternal antepartum and intrapartum characteristics as well as hospital characteristics were selected using Cramer's phi and Theil index. Logistic regression, support vector machines, multi-layer perceptron, random forest, and gradient boosting were used to generate models to predict transfusion and to predict the transfusion-PPH composite. Receiver operating characteristic area under the curve (ROC-AUC) and precision/recall area under the curve (PR-AUC; commonly used in low incidence outcomes) were used to compare model performance.

Results: Among 228,438 births, 5,760 women (3.1%) had a postpartum hemorrhage, 5,170 women (2.8%) had blood product transfusion, and 10,344 women (5.6%) met criteria for the transfusion-PPH composite. Overall, models predicting our primary transfusion-PPH composite outcome using antepartum and intrapartum maternal

features had the best positive predictive values with the gradient boosting machine learning model performing the best overall (ROC-AUC=0.833, 95% CI [0.828 - 0.838]; PR-AUC=0.210 95% CI [0.201 - 0.220]). However, model calibration was challenging for all methods. The most predictive maternal features in the gradient boosting model for prediction of transfusion-PPH composite were mode of delivery, oxytocin incremental dose for labor (mU/min), intrapartum tocolytic use, presence of anesthesia nurse, and hospital type.

Conclusion: Machine learning and data-driven statistical modeling offer more objective and discriminative prediction of PPH based on individual antepartum and intrapartum patient features. In PPH risk-based analyses, blood transfusion is not an optimal outcome measure to use since subgroup effects predominate, decreased overall model accuracy and limited generalizability of findings.

Keywords: Postpartum Hemorrhage, Machine Learning, Prediction

INTRODUCTION

Maternal morbidity and mortality have been regarded as a reflection of national healthcare quality. Among lower income countries, postpartum hemorrhage (PPH) is typically the most common cause of maternal mortality, and remains among the top causes in higher-income countries. In the United States, hemorrhage accounted for 11.0% of deaths between 2011 to 2016.[3] To address maternal hemorrhage, maternal hemorrhage protocols have been implemented, which incorporate prospective PPH risk assessment to tailor PPH prophylactic and management approaches to the patient's individual risk profile. However, these protocols are based often based on observational studies that approximated strength of associations with hemorrhage via logistic regression models, and combined the results of multiple studies together in a linear fashion.[5,6,7] However, "standard" logistic regression assumes, unless extra efforts are undertaken, a linear relationship between predictors and the log odds of outcomes and independent relationships between predictors. Additionally, logistic regression and related models often perform poorly with large numbers of included variables.[8,9] Consequently, current risk stratification models fail to accurately ascertain pregnant patient's risk of hemorrhage.[10] Studies attempting to validate these methods have instead identified gaps in the efficacy of these models, as a large number to the majority of patients with postpartum hemorrhage and transfusions were stratified to low or moderate risk groups.[11,12]

Machine learning offers an advantage to current risk assessment methods through its ability to create a robust model based on larger numbers of predictors, with nonlinear relationships and interactions between variables included in analyses.[13] Our

objective in this analysis was to create a validated prediction model using machine learning for postpartum hemorrhage and transfusion to optimize risk-based triage and inform policy makers and stakeholders who aimed to further reduce maternal morbidity and mortality associated with hemorrhage.

METHODS

Data for this analysis were extracted from the Consortium for Safe Labor (CSL) dataset created by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD). It includes antepartum, intrapartum, and postpartum medical histories of 224,438 women from 12 hospitals in the United States. The data deidentified and available for research under request from the NICHD. Women with only one recorded pregnancy in the dataset were included for analysis in the dataset; if women had more than one pregnancy then only the first one was used in the analysis. We selected maternal variables as candidates to build the prediction model for transfusion risk.

Missing data. Machine Learning methods are known to create errors in the presence of missing values.[14] To avoid this, we imputed values as follows: categorical variables with missing and unknown values were assigned to an "unknown" category; continuous variables with missing and unknown values were coded to the median value. Continuous variables for maternal age and body mass index (BMI) were coded into ordinal categories (age <20, ≥20 and <40, ≥40 and <45, ≥45 years old; BMI ≤20, >20 and ≤40, >40 and ≤50, >50 kg/m^2). Imputing estimated blood loss (EBL) as the median value (350 ml) meant that missing values were assumed to be < 1000 mL.

Feature selection. We used the Cramér's V index of nominal association for variable selection.[15] Features were classified into antepartum and intrapartum variables. Two different prediction models were constructed: 1) an antenatal-only model intended to be used in the clinic setting to inform appropriate patient referral and 2) an intrapartum model that included both antepartum and intrapartum characteristics. Individual antepartum and intrapartum maternal variables included for model development are included in the Supplementary Table 1.

Outcomes. Separate models were constructed to predict two target outcomes. The first and primary outcome included a composite and included all those that received transfusion of any blood products or had a postpartum hemorrhage defined by documented blood loss of 1,000 mL or more during or after delivery, and the second outcome was all patients that received transfusion of any blood products.

Data analysis. For each of the four combinations of predictors and outcomes (for predictors, antepartum versus antepartum and intrapartum; for outcomes, transfusion and blood loss greater than a liter versus transfusion alone), the data were split so that 70% of the observations were used for training and 30% were used for testing, with both sets the same outcome rate. We applied a number of methods, including logistic regression (LR), support vector machines (SVC), multi-layer perceptron (MLP), random forest (RF) and gradient boosting (GB), as well as deep learning algorithms including tensorflow imbalanced (TFIM) and learned embedding (Emb). Hyperparameters were tuned for each algorithm using a customized grid search technique. The model performance for each combination of outcome and algorithm was measured with Matthews correlation coefficient (MCC), receiver operating characteristic -area under

the curve (ROC-AUC), precision / recall – area under the curve (PR-AUC), and modified F-score skewed towards recall (F2). The algorithmic processing and results analysis were performed using Python (version 3.6), Pandas (version 1.2), scikit-learn (version 0.24), and Tensorflow (version 2.2).

The primary study question was to identify the strongest set of pre- and intra-operative predictors of hemorrhage/transfusion and the strongest modeling technique. Secondary questions included the level of agreement between metrics for model evaluation, and the extent to which any technique produced results that are clinically useful. This analysis was exempt from review by the George Washington University Institutional Review Board.

RESULTS

Of 228,438 births included in the CSL cohort, we included 185,413 patients once excluding patients with more than one delivery (n=43,025). Maternal age ranged from 11 to 58 with a median of 27 years; 32% were publicly insured, 49% were white non-Hispanic, 22% Black and 17% Hispanic. Of the 185,413 women included for analysis, 71% had vaginal delivery (n=131,130), 29% had cesarean delivery (n=54,283). 5,170 (3%) experienced the primary outcome of transfusion of any blood products. 5,760 (3.11%) had a postpartum hemorrhage defined by an estimated blood loss ≥ 1,000 mL, and 10,344 (6%) experienced the secondary composite outcome of transfusion or estimated blood loss of loss ≥ 1,000 mL. Additional demographic data is summarized in Supplemental Table 2.

After building the models in an iterative process, the performance of the models for predicting both the primary and secondary outcomes were compared using variety of

metrics. The metrics (ROC_AUC, PR-AUC, MCC, F2), as well as sensitivity and specificity at a probability cut point of 50% (Table 1). Overall, models developed using intrapartum maternal variables (Supplementary Table 1 for list of variables) to predict the primary outcome performed better with higher positive predictive values compared with solely using antepartum maternal variables (Supplementary Table 3). For the secondary outcome of transfusion alone, there was little difference in model performance when compared on several performance metrics. In contrast, for the primary outcome of transfusion of blood products or postpartum hemorrhage, the machine learning technique gradient boosting using intrapartum maternal variables had the highest positive predictive value (PR-AUC=0.21, 95% CI [0.20 to 0.22]; ROC-AUC=0.83, 95% CI [0.828 to 0.838]; Figure 2). For this reason, we focus the remainder of our results on this outcome. Both random forest and gradient boosting had significantly higher positive predictive values for predicting the composite transfusion or postpartum hemorrhage compared with logistic regression (PR-AUC= 0.18, 95% CI [0.17 to 0.19]; ROC-AUC=0.81, 95% CI [0.808 to 0.818]).

Figure 3 reveals the calibration curves for the models constructed with intrapartum maternal variables and predicting the transfusion-postpartum hemorrhage composite. Calibration curves portray the *predicted* PPH risk versus the *observed* PPH rate across a range of predicted PPH values. There is better agreement between the models with lower fraction of positives, and none of the models were able to reach the standard curve – for all models, the predicted PPH rate overestimated the observed PPH rate across the range of predicted values. Figure 4 displays the top 25 predictive variables included for model development using antepartum and intrapartum features

for prediction of the transfusion-postpartum hemorrhage composite. As the machine learning gradient boosting model was the best performing model overall, the variables in the figure are in order of the variable importance within the gradient boosting model. The top 10 variables from most predictive weight to least predictive rate for intrapartum prediction of the transfusion-postpartum hemorrhage composite using the gradient boost model are mode of delivery, oxytocin incremental dose for labor (mU/min), intrapartum tocolytic use, use of anesthesia nurse, hospital type, trial of labor, insurance, most serious diabetes control, education, and history of prior cesarean sections. The results of the models for antepartum-only models are listed in Supplemental Table 3. The ROC_AUC and PR_AUC did not perform as well for the models using antepartum only variables, though this was less obvious for the models predicting transfusion only. Of note, upon further sensitivity analysis we also determined that some of the top variables in the model were site specific (i.e. oxytocin incremental dose for labor, intrapartum tocolytic use, use of anesthesia nurse and hospital type) for transfusion outcome specifically (data not included).

DISCUSSION

Principal findings

In this study, logistic regression and machine learning techniques were analyzed and compared to develop prediction models for postpartum hemorrhage and transfusions. We found that the machine learning techniques, particularly gradient boosting, performed best to predict postpartum hemorrhage when postpartum hemorrhage was defined as blood transfusion or blood loss greater than one liter.

However, all prediction models had difficulties with calibration when predicting the rare outcome of transfusion alone.

Clinical Implications

Risk assessment for PPH has been shown in a pre- and post- study to reduce rates of blood transfusion and PPH. However, the risk stratification approaches most commonly used for postpartum hemorrhage in the United States were developed and implemented based on expert opinion, and subsequent validation studies have revealed the limitations of these tools.[19,20] Validation studies using the CMQCC risk assessment tool found that while the tool did produce populations with different rates of hemorrhage among those identified low, medium, and high-risk groups, but the rate of PPH among women identified as in the high-risk group for postpartum hemorrhage was found to be only 22%.[21] Others have found that the AUC-ROC for the CMQCC and AWHONN tools for predicting severe postpartum hemorrhage, defined by transfusion of at least 4 units packed red blood cells during postpartum period, were relatively modest at 0.77 and 0.69, respectively.[22] Furthermore, parameters which are included in PPH risk models based on univariate association with PPH risk may not be independent predictors when incorporated into multivariate models.[22] For these reasons, improvements in PPH risk models are a promising target for improving PPH care.

A previously-published risk assessment for PPH using the CSL dataset demonstrated exceptional model performance, but model performance was dramatically lower in an external validation cohort. This study augments this prior work via incorporation of antepartum and intrapartum risk factors. Nonetheless, additional work is needed before such a model can be implemented in clinical practice. In particular, it will

be important to develop prediction models which are implementable either through straightforward bedside data entry or can be automated via real time electronic medical record data capture, which are well validated in a variety of hospital settings, and ideally, which are paired with recommended risk-based interventions to reduce hemorrhage risk and mitigate hemorrhage which occurs.

Research Implications

For all the intrapartum methods that we tested predicting transfusion or hemorrhage, the ROC-AUCs values were greater than 0.80, which is often cited as a threshold indicating adequate discrimination. However, this conclusion is misleading because in a situation where incidence of the outcome is low (here it was ~3% for transfusion or hemorrhage alone), the positive predictive value (PPV), also known as 'precision', is likely to be quite low. Our precision for the best model was ~13%, meaning that of those predicted to be positive for the outcome, 13% were positive and 87% were negative. This may be satisfactory for clinical uses where preventive interventions have very low cost (in terms of both financial cost and added risk to the patient) but would not be acceptable when the intervention is higher risk or more expensive. In this situation, the PR-AUC provided a more realistic measure of model quality. Precision/recall plots positive predictive value (aka precision), as a function of sensitivity (aka recall) values, thus it accounts for true positives among positive predictions, whereas the ROC AUC emphasizes specificity, which is likely to be very high when true positives are rare.[16,17] The metric with the largest difference between the best and worst-performing models is Precision-Recall AUC (0.16 vs 0.21). Perhaps this

metric should be used more frequently in modeling studies when the outcome of interest occurs less than or equal to 6%.

Strengths and Limitations

Strengths of this study include the use of a large, national multicenter dataset to develop a data-driven model that can predict postpartum hemorrhage using antepartum and intrapartum factors using cutting edge machine learning techniques. Furthermore, we looked at both commonly used end points like estimated blood loss greater than one liter and clinically relevant end points like transfusion; concluding that due to a less frequent occurrence and transfusion practice variation made it more challenging to come up with reliable model for transfusion only.

Limitations of the study include the low reported precision of algorithms. Sensitivity is prioritized for prediction, as clinically missing PPH has more consequences than a false positive. Therefore, the algorithms are trained to be biased towards predicting positives resulting in lower false negative rates at the risk of higher false positive rates and decreased precision. As a result, as shown in the calibration plots, the models systematically overstate hemorrhage risk. In this study, the outcomes of interest were either a composite of transfusion or blood loss greater than or equal to one liter or transfusion only. Our PPH definition was based on the American College of Obstetricians and Gynecologists' reVITALize program that defined postpartum hemorrhage as blood loss greater than or equal to one liter or loss of blood with clinical signs of hypovolemia within 24 hours of delivery. This definition deviates from older traditional definitions that defined PPH as greater than or equal to 500 mL for vaginal delivery and 1000 mL for cesarean delivery.[25] Therefore, the recording of EBL and

transfusions could have been guided by older definitions, as the Consortium of Safe Labor data set was collected between 2002 and 2008.[26] Beyond that, measures of EBL have been shown to be imprecise with low volumes overestimated and high volumes of blood loss underestimated[27]. Furthermore, transfusion was used as a proxy for PPH, and transfusion thresholds vary depending on institution and provider. In addition, the machine learning algorithms are limited by the variables measured and accurately recorded in the data set.

Conclusions

In conclusion, machine learning and data-driven statistical modeling may offer more objective and discriminative prediction of postpartum hemorrhage based on individual antepartum and intrapartum patient features, compared to expert opinion and may improve upon traditional regression models. This can increase the opportunity for precision medicine and improved clinical care to reduce the burden of postpartum hemorrhage as a leading cause of maternal morbidity and mortality.

REFERENCES

1. MacDorman MF, Declercq E, Cabral H, Morton C. Recent Increases in the U.S. Maternal Mortality Rate: Disentangling Trends From Measurement Issues. Obstet Gynecol. 2016 Sep;128(3):447-55. doi: 10.1097/AOG.0000000000001556. PMID: 27500333; PMCID: PMC5001799
2. Neggers YH. Trends in maternal mortality in the United States. Reprod Toxicol. 2016 Sep;64:72-6. doi: 10.1016/j.reprotox.2016.04.001. Epub 2016 Apr 7. PMID: 27063184.
3. Pregnancy Mortality Surveillance System | Maternal and Infant Health | CDC. Published November 25, 2020. Accessed January 9, 2021. https://www.cdc.gov/reproductivehealth/maternal-mortality/pregnancy-mortality-surveillance-system.htm
4. Collier AY, Molina RL. Maternal Mortality in the United States: Updates on Trends, Causes, and Solutions. Neoreviews. 2019 Oct;20(10):e561-e574. doi: 10.1542/neo.20-10-e561. PMID: 31575778; PMCID: PMC7377107.
5. Shields LE, Wiesner S, Fulton J, Pelletreau B. Comprehensive maternal hemorrhage protocols reduce the use of blood products and improve patient safety. Am J Obstet Gynecol. 2015 Mar;212(3):272-80. doi: 10.1016/j.ajog.2014.07.012. Epub 2014 Jul 12. PMID: 25025944.
6. Al-Zirqi I, Vangen S, Forsen L, Stray-Pedersen B. Prevalence and risk factors of severe obstetric haemorrhage. BJOG. 2008 Sep;115(10):1265-72. doi: 10.1111/j.1471-0528.2008.01859.x. PMID: 18715412.
7. Bateman BT, Berman MF, Riley LE, Leffert LR. The epidemiology of postpartum hemorrhage in a large, nationwide sample of deliveries. Anesth Analg. 2010 May 1;110(5):1368-73. doi: 10.1213/ANE.0b013e3181d74898. Epub 2010 Mar 17. PMID: 20237047.
8. Lee HC, Yoon SB, Yang SM, Kim WH, Ryu HG, Jung CW, Suh KS, Lee KH. Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model. J Clin Med. 2018 Nov 8;7(11):428. doi: 10.3390/jcm7110428. PMID: 30413107; PMCID: PMC6262324.
9. Deo RC. Machine Learning in Medicine. Circulation. 2015 Nov 17;132(20):1920-30. doi: 10.1161/CIRCULATIONAHA.115.001593. PMID: 26572668; PMCID: PMC5831252.
10. Pressly MA, Parker RS, Waters JH, Beck SL, Jeyabalan A, Clermont G. Improvements and limitations in developing multivariate models of hemorrhage and transfusion risk for the obstetric population. Transfusion. 2020 Dec 11. doi: 10.1111/trf.16216. Epub ahead of print. PMID: 33305364.
11. Ruppel H, Liu VX, Gupta NR, Soltesz L, Escobar GJ. Validation of Postpartum Hemorrhage Admission Risk Factor Stratification in a Large Obstetrics Population. Am J Perinatol. 2020 May 26:10.1055/s-0040-1712166. doi: 10.1055/s-0040-1712166. Epub ahead of print. PMID: 32455467; PMCID: PMC7688483.
12. Hussain SA, Guarini CB, Blosser C, Poole AT. Obstetric Hemorrhage Outcomes by Intrapartum Risk Stratification at a Single Tertiary Care Center. Cureus. 2019;11(12):e6456. Published 2019 Dec 24. doi:10.7759/cureus.6456

13. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016 Sep 29;375(13):1216-9. doi: 10.1056/NEJMp1606181. PMID: 27682033; PMCID: PMC5070532.
14. Rosati G. Dealing with Missing Data [Internet]. 2019. Available from: https://towardsdatascience.com/dealing-with-missing-data-17f8b5827664
15. Cramér H. Mathematical methods of statistics. 19. printing. Princeton: Princeton Univ. Press; 1999. 575 p. (Princeton landmarks in mathematics and physics).
16. Leonard SA, Kennedy CJ, Carmichael SL, Lyell DJ, Main EK. An Expanded Obstetric Comorbidity Scoring System for Predicting Severe Maternal Morbidity. *Obstet Gynecol*. 2020;136(3):440-449. doi:10.1097/AOG.0000000000004022
17. ROCSaito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. Brock G, ed. *PLoS ONE*. 2015;10(3):e0118432. doi:10.1371/journal.pone.0118432
18. Choi SR, Yoon NR, Hwang SO. Is a cause of postpartum hemorrhage of using of tocolytic agents in preterm delivery? *J Matern Fetal Neonatal Med*. 2017;30(22):2711-2715. doi:10.1080/14767058.2016.1261397
19. Bingham D, Melsop K, Main E. CMQCC obstetric hemorrhage hospital level implementation guide. Available at: https://www.cmqcc.org/resource/1489/download.
20. The AWHONN Postpartum Hemorrhage Project. Postpartum hemorrhage (PPH) risk assessment table 1.0. Available at: https://mygnosis.com/Content/Chunks/3504/assets/pdfs/PPH_Risk_Assessment_Table-7-17-15.pdf.
21. Dilla AJ, Waters JH, Yazer MH. Clinical Validation of Risk Stratification Criteria for Peripartum Hemorrhage. *Obstetrics and gynecology (New York 1953)*. 2013;122(1):120-126. doi:10.1097/AOG.0b013e3182941c78
22. Kawakita T, Mokhtari N, Huang JC, Landy HJ. Evaluation of Risk-Assessment Tools for Severe Postpartum Hemorrhage in Women Undergoing Cesarean Delivery. *Obstetrics & Gynecology*. 2019;134(6):1308-1316. doi:10.1097/AOG.0000000000003574
23. Kawakita T, Mokhtari N, Huang JC, Landy HJ. Evaluation of Risk-Assessment Tools for Severe Postpartum Hemorrhage in Women Undergoing Cesarean Delivery. *Obstetrics & Gynecology*. 2019;134(6):1308-1316. doi:10.1097/AOG.0000000000003574
24. Herrin J, Abraham NS, Yao X, et al. Comparative Effectiveness of Machine Learning Approaches for Predicting Gastrointestinal Bleeds in Patients Receiving Antithrombotic Treatment. *JAMA Netw Open*. 2021;4(5):e2110703. doi:10.1001/jamanetworkopen.2021.10703
25. Venkatesh KK, Strauss RA, Grotegut C, et al. Machine Learning and Statistical Models to Predict Postpartum Hemorrhage. *Obstet Gynecol*. 2020;135(4):935-944. doi:10.1097/AOG.0000000000003759

26. Postpartum Hemorrhage. Accessed July 11, 2021. https://www.acog.org/en/clinical/clinical-guidance/practice bulletin/articles/2017/10/postpartum-hemorrhage
27. NICHD DASH - Study Overview. Accessed July 11, 2021. https://dash.nichd.nih.gov/study/2331
28. DILDY GA, PAINE AR, GEORGE NC, VELASCO C. Estimating blood loss: Can teaching significantly improve visual estimation? *Obstetrics and gynecology (New York 1953)*. 2004;104(3):601-606. doi:10.1097/01.AOG.0000137873.07820.34

ACKNOWLEDGEMENTS

Dr. Ahmadzia's effort was supported by the National Heart Lung and Blood Institute of the National Institutes of Health under award K23HL141640 and Dr. Federspiel's effort was supported by the National Center for Advancing Translation Sciences under award TL1TR002555. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to acknowledge Dr. Christian Macedonia and Dr. Chad Grotegut for their insights on initial model design.

CONFLICT OF INTEREST STATEMENTS

None of the authors have any conflicts of interest pertinent to the subject of this manuscript.

TABLES

Table 1. Performance of machine learning and statistical models based on antepartum and intrapartum maternal variables at predicting transfusion and/or postpartum hemorrhage

Algorithm	# True Positives*	# True Negatives*	# False Positives*	# False Negatives*	Precision	Recall	Specificity	AUC-ROC	AUC-PR	MCC	F2
Primary Outcome: Blood Transfusion or Blood Loss >= 1 liter											
GB	50	6	318	626	0.135	0.889	0.663	0.833	0.210	0.260	0.419
RF	50	6	339	605	0.138	0.857	0.641	0.830	0.204	0.261	0.409
Emb	46	10	296	649	0.134	0.821	0.687	0.813	0.181	0.246	0.406
MLP	49	7	335	609	0.127	0.875	0.645	0.808	0.149	0.245	0.402
TFIM	48	8	323	619	0.129	0.861	0.655	0.822	0.194	0.245	0.403
SVM	49	6	349	595	0.124	0.886	0.630	0.804	0.159	0.242	0.397
LR	46	10	314	631	0.129	0.830	0.668	0.813	0.177	0.238	0.393
Secondary Outcome: Blood Transfusion											
GB	24	4	235	737	0.093	0.866	0.758	0.860	0.111	0.234	0.325
RF	25	3	251	721	0.090	0.887	0.742	0.862	0.107	0.232	0.319
Emb	22	6	223	750	0.090	0.789	0.771	0.837	0.096	0.215	0.309
MLP	24	4	237	735	0.091	0.849	0.756	0.845	0.095	0.227	0.318
TFIM	24	4	240	732	0.091	0.859	0.753	0.855	0.111	0.229	0.319
SVM	24	4	244	728	0.091	0.871	0.749	0.852	0.116	0.230	0.320
LR	24	3	250	722	0.089	0.876	0.743	0.853	0.111	0.228	0.317

*Normalized values per 1000, so easier to compare across different models, the actual N=55,624.GB: Gradient boosting, RF: Random forests, Emb: learned embedding, MLP: Multi-layer percepton, TFIM: Tensorflow imbalanced, SVM: support vector machines, LR: logistic regression, AUC-ROC: Area under the receiver operating curve, AUC-PR: Area under precision-recall curve, MCC: Matthews correlation coefficient, F2: modified F-score skewed towards recall

FIGURE LEGENDS

Figure 1: Derivation of study cohort for women with transfusion and/or postpartum hemorrhage

Figure 2: Receiver Operating Curve (Panel A) and Area under the Precision-Recall Curves (Panel B) for different models using intrapartum maternal variables predicting the composite of transfusion or postpartum hemorrhage

Figure 3: Calibration curves for models using intrapartum maternal variables to predict the composite of transfusion or postpartum hemorrhage

Figure 4: Top 25 predictors based on each model using intrapartum maternal factors predicting the composite of transfusion or postpartum hemorrhage

Figure 1

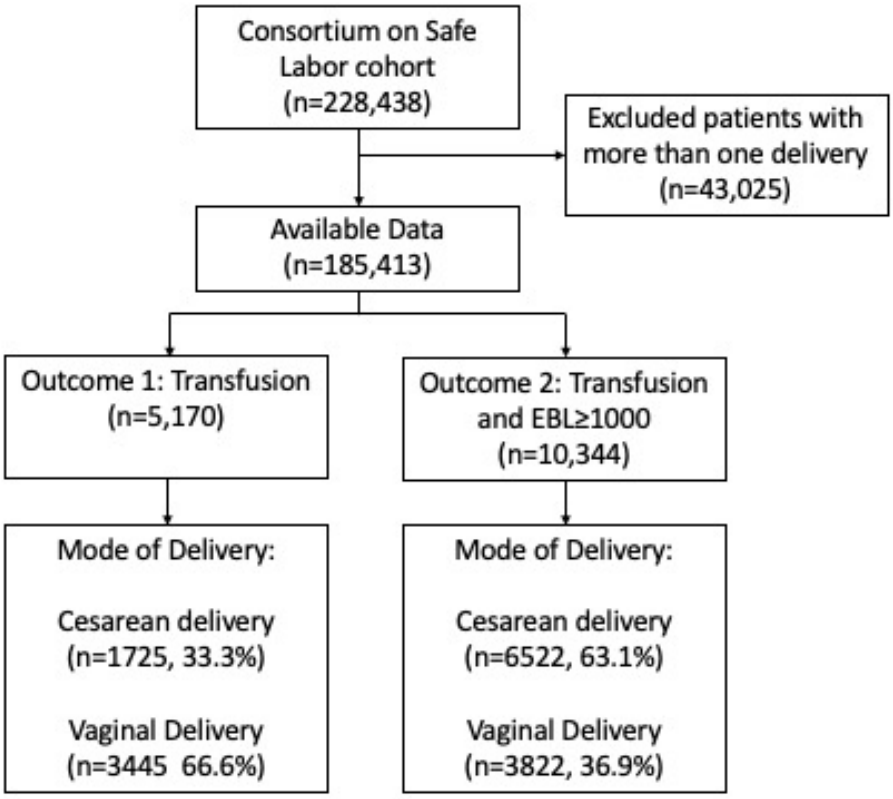

Consortium on Safe Labor cohort (n=228,438)
Excluded patients with more than one delivery (n=43,025)
Available Data (n=185,413)
Outcome 1: Transfusion (n=5,170)
Outcome 2: Transfusion and EBL≥1000 (n=10,344)
Mode of Delivery:
Cesarean delivery (n=1725, 33.3%)
Vaginal Delivery (n=3445 66.6%)
Mode of Delivery:
Cesarean delivery (n=6522, 63.1%)
Vaginal Delivery (n=3822, 36.9%)

Figure 2

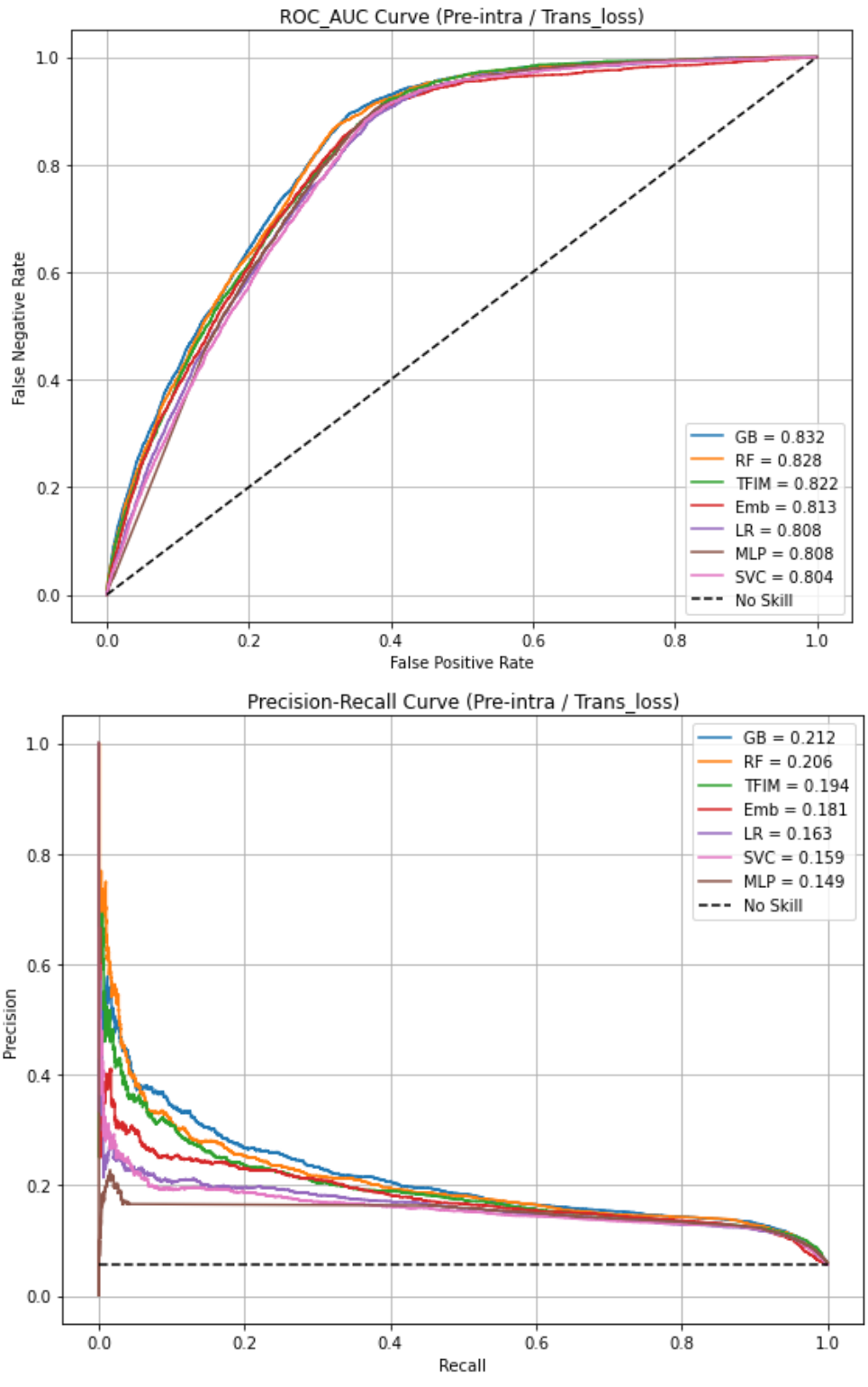

ROC_AUC Curve (Pre-intra / Trans_loss)
False Negative Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
GB = 0.832
RF = 0.828
TFIM = 0.822
Emb = 0.813
LR = 0.808
MLP = 0.808
SVC = 0.804
No Skill
Precision-Recall Curve (Pre-intra / Trans_loss)
Precision
Recall
GB = 0.212
RF = 0.206
TFIM = 0.194
Emb = 0.181
LR = 0.163
SVC = 0.159
MLP = 0.149
No Skill

Figure 3

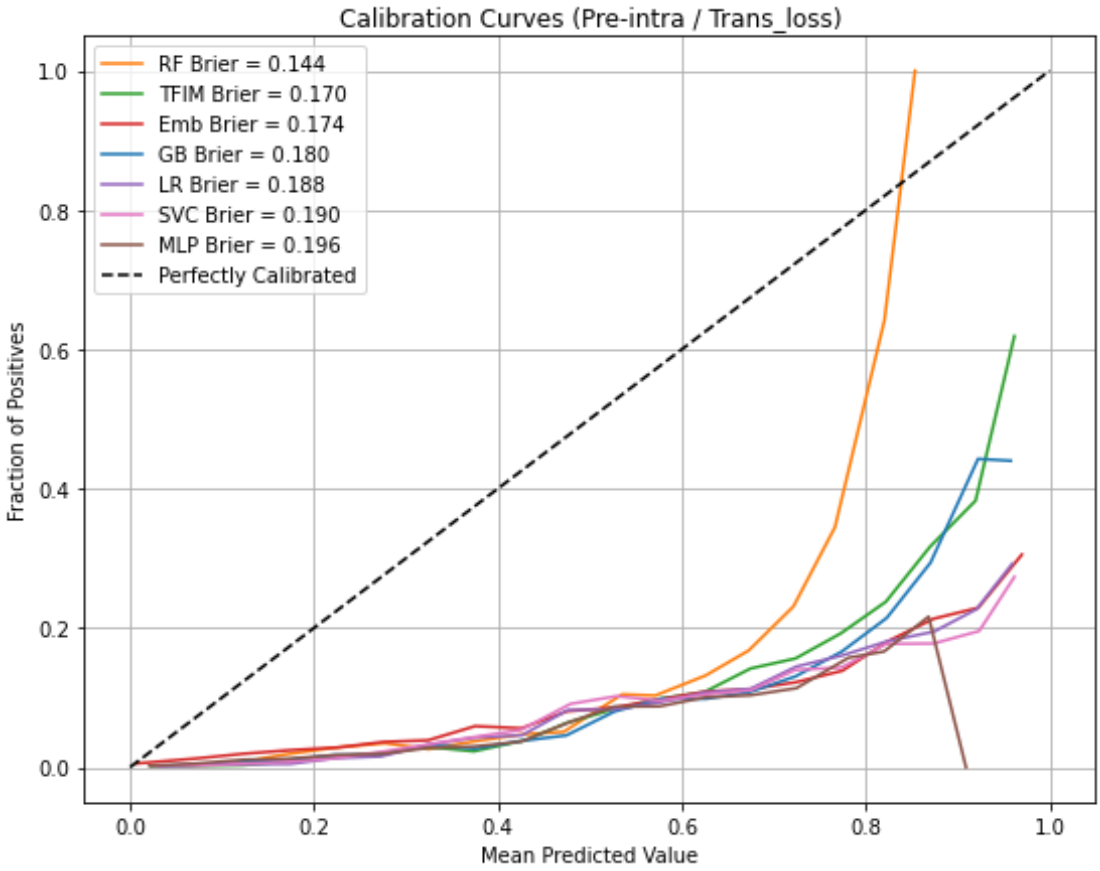
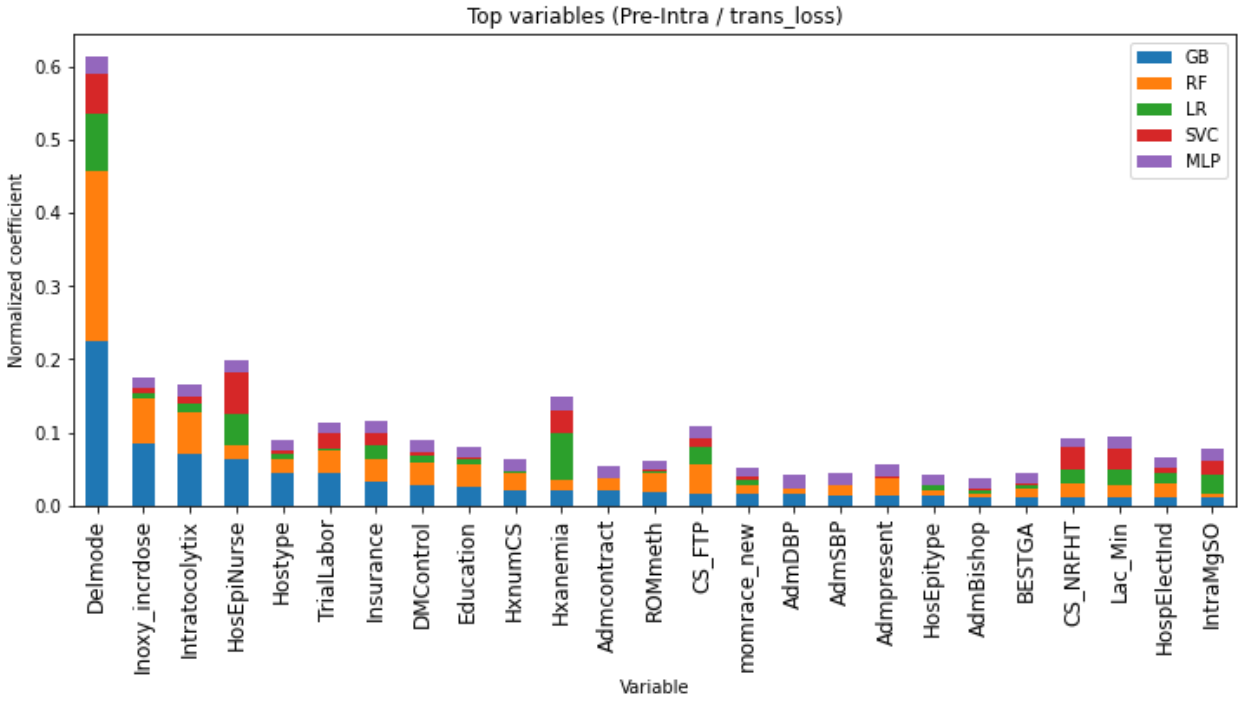

Calibration Curves (Pre-intra / Trans_loss)
RF Brier = 0.144
TFIM Brier = 0.170
Emb Brier = 0.174
GB Brier = 0.180
LR Brier = 0.188
SVC Brier = 0.190
MLP Brier = 0.196
Perfectly Calibrated
Fraction of Positives
Mean Predicted Value
0.0
0.2
0.4
0.6
0.8
1.0

Figure 4


Top variables (Pre-Intra / trans_loss)
GB
RF
LR
SVC
MLP
Normalized coefficient
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Delmode
Inoxy_incrdose
Intratocolytix
HosEpiNurse
Hostype
TrialLabor
Insurance
DMControl
Education
HxnumCS
Hxanemia
Admcontract
ROMmeth
CS_FTP
momrace_new
AdmDBP
AdmSBP
Admpresent
HosEpitype
AdmBishop
BESTGA
CS_NRFHT
Lac_Min
HospElectInd
IntraMgSO
Variable