



Published in final edited form as:

*Obstet Gynecol.* 2020 April ; 135(4): 935–944. doi:10.1097/AOG.0000000000003759.

## Machine Learning and Statistical Models to Predict Postpartum Hemorrhage

Kartik K Venkatesh, MD, PhD<sup>1</sup>, Robert A Strauss, MD<sup>1</sup>, Chad Grotegut, MD<sup>2</sup>, R Philip Heine, MD<sup>3</sup>, Nancy C Chescheir, MD<sup>1</sup>, Jeffrey SA Stringer, MD<sup>1</sup>, David M Stamilio, MD, MSCE<sup>3</sup>, M Katherine Menard, MD, MPH<sup>1</sup>, J. Eric Jelovsek, MD, MMed, MSDS<sup>2</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, University of North Carolina Chapel Hill (Chapel Hill, NC)

<sup>2</sup>Department of Obstetrics and Gynecology, Duke University (Durham, NC)

<sup>3</sup>Department of Obstetrics and Gynecology, Wake Forest University (Winston-Salem, NC)

### Abstract

**Objective:** To predict a woman's risk of postpartum hemorrhage at labor admission using machine learning and statistical models.

**Methods:** Predictive models were constructed and compared using data from 10 of 12 sites in the U.S. Consortium for Safe Labor Study (2002–2008) that consistently reported estimated blood loss at delivery. The outcome was postpartum hemorrhage, defined as an estimated blood loss 1,000 mL. Fifty-five candidate risk factors routinely available on labor admission were considered. We used logistic regression with and without lasso regularization (lasso regression) as the two statistical models and random forest and Extreme Gradient Boosting as the two machine learning models to predict postpartum hemorrhage. Model performance was measured by C statistics (i.e., concordance index), calibration, and decision curves. Models were constructed from the first phase (2002–2006) and externally validated (i.e., temporally) in the second phase (2007–2008). Further validation was performed combining both temporal and site-specific validation.

**Results:** Of the 152,279 assessed births, 7,279 (4.8%, 95% CI 4.7 to 4.9) had a postpartum hemorrhage. All models had good to excellent discrimination. The Extreme Gradient Boosting model had the best discriminative ability to predict postpartum hemorrhage (C statistic: 0.93; 95% CI: 0.92 to 0.93) followed by random forest (C statistic: 0.92; 95% CI: 0.91 to 0.92). The lasso regression model (C statistic: 0.87; 95% CI: 0.86 to 0.88) and logistic regression (C statistic: 0.87; 95% CI: 0.86 to 0.87) had lower but good discriminative ability. The above results held with validation across both time and sites. Decision curve analysis demonstrated that while all models provided superior net benefit when clinical decision thresholds were between 0 to 80% predicted risk, the Extreme Gradient Boosting model provided the greatest net benefit.

---

Corresponding author: Kartik K. Venkatesh, MD, PhD, Department of Obstetrics and Gynecology, University of North Carolina Chapel Hill, Chapel Hill, NC; kartik.k.venkatesh@gmail.com.

Each author has confirmed compliance with the journal's requirements for authorship.

For a list of institutions involved in the Consortium on Safe Labor, see Appendix 1 online at <http://links.lww.com/xxx>.

Financial Disclosure

The other authors did not report any potential conflicts of interest.

**Conclusion:** Postpartum hemorrhage on labor admission can be predicted with excellent discriminative ability using machine learning and statistical models. Further clinical application is needed, which may assist health care providers to be prepared and triage at risk women.

## PRECIS

Postpartum hemorrhage can be predicted on labor admission with excellent discriminative ability using both machine learning and statistical models.

---

## INTRODUCTION

Postpartum hemorrhage is the primary source of maternal morbidity and mortality worldwide, accounting for nearly a third of deaths of pregnant and postpartum women.<sup>1</sup> In the United States (U.S), the rate of postpartum hemorrhage has increased by at least 26% in the past decade.<sup>2,3</sup> Although maternal death is a rare outcome in the U.S., blood transfusion after hemorrhage, which is 50 times more common than death, is the main diagnosis associated with severe maternal morbidity in the U.S.<sup>4,5</sup> While the methods to estimate blood loss at delivery continue to evolve,<sup>6</sup> the American College of Obstetricians and Gynecologists (ACOG) currently defines a postpartum hemorrhage as a cumulative blood loss 1,000 mL or signs or symptoms of hypovolemia within 24 hours after birth.<sup>7</sup>

Predicting a woman's risk of postpartum hemorrhage on labor admission requires the obstetrician to incorporate known risk factors and then to approximate the probability of hemorrhage by using a risk strata scheme.<sup>7,8</sup> With an increasing focus on standardized guidelines to prevent and manage postpartum hemorrhage,<sup>9,10</sup> limited tools exist to accurately predict which women are at the highest risk for hemorrhage.<sup>7</sup> Current risk-based stratification guidelines adopted by ACOG and California Maternal Quality Care Collaborative (CMQCC) include decision tree algorithms based on clinical consensus,<sup>11</sup> expert opinion, and prior observational data.<sup>2,12,13</sup> An accurate and validated clinical prediction model that could be deployed on the labor and delivery unit (L&D) for postpartum hemorrhage is lacking.<sup>14-17</sup>

Current methods for predicting postpartum hemorrhage are based on risk stratification methods. Improved predictive ability could be achieved by applying traditional statistical and machine learning methods.<sup>18</sup> Recent advances in machine learning, which employs advanced computer-driven algorithms aimed at detecting patterns in data, have increasingly attracted attention because of their superior predictive ability primarily in determining intensive care unit admission and hospital readmission compared to statistical models.<sup>19,20</sup> However, these innovative approaches have yet to be widely tested in obstetrics.<sup>21</sup> Advantages of machine learning include the ability to process non-additive relationships and incorporate complex interactions between factors that do not need to be pre-specified.<sup>20</sup> For these reasons, it is possible that machine learning approaches could accurately identify women at highest risk of postpartum hemorrhage and improve obstetric decision making,<sup>21,22</sup> and possibly improved clinical outcomes.

Our objective was to develop and validate prediction models for postpartum hemorrhage on labor admission. We quantified how closely the diagnosis of postpartum hemorrhage

classified by machine learning and statistical models matched the known diagnosis of an estimated blood loss  $\geq 1,000$  mL.

## METHODS

Data from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Consortium on Safe Labor (CSL) were used for the development and validation of the prediction models. Briefly, the CSL was a retrospective cohort study of women delivering at  $\geq 23$  weeks gestation between 2002 and 2008 at 12 clinical sites with 19 hospitals across 9 ACOG districts in the U.S.<sup>23</sup> This cohort included data abstracted from electronic medical records, including demographics, prenatal complications, labor and delivery information, and maternal and neonatal outcomes. The CSL included a total of 228,438 deliveries at  $\geq 23$  weeks gestation, with 9.5% of women (N=5,053) contributing  $>1$  birth over the study period. The present analysis was performed using a de-identified dataset under a waiver of informed consent and was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill. Methods and reporting guidelines were followed as proposed in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement.<sup>24</sup>

The candidate predictors of postpartum hemorrhage for all models were chosen from routinely available data on labor admission and were compiled from expert opinion, consensus statements, prior observational cohorts, and collected from the CSL database.<sup>25–27</sup> We first included risk factors for postpartum hemorrhage *a priori* identified by the CMQCC and ACOG<sup>7,28</sup> (see Table 1 in Practice Bulletin No. 183, available at <https://www.acog.org/Clinical-Guidance-and-Publications/Practice-Bulletins-List>). Additional predictors included available patient socio-demographics (e.g., age, race), obstetric diagnoses (e.g., placenta previa, fetal macrosomia, preeclampsia), comorbid conditions (e.g., chronic hypertension, diabetes), and vital signs on labor admission (Appendix 2, available online at <http://links.lww.com/xxx>). Risk factors in the CSL were captured by searching for specific *International Classification of Diseases, 9<sup>th</sup> Revision, Clinical Modification* (ICD-9-CM) codes as well as from documentation in the L&D clinical record. Some previously identified risk factors as outlined by CMQCC and ACOG that would be available to obstetric providers on labor admission were not available in this dataset, including a history of prior postpartum hemorrhage and thrombocytopenia on admission.

The primary outcome of postpartum hemorrhage was defined as a cumulative estimated blood loss (EBL) of  $\geq 1,000$  mL, regardless of mode of delivery. This outcome was selected because it was consistent with the most recent definition of the ACOG reVITALize program,<sup>28</sup> as well as the consensus case definition of the Brighton Collaboration Primary Postpartum Hemorrhage Working Group,<sup>25</sup> the Royal College of Obstetricians and Gynaecologists (UK),<sup>26</sup> and the World Health Organization (WHO).<sup>27</sup>

We developed two statistical models using logistic regression and logistic regression with lasso regularization (lasso regression)<sup>29</sup> and two machine learning models using random forest<sup>30</sup> and Extreme Gradient Boosting algorithms.<sup>31</sup> Lasso regression is also referred to as penalized regression because a penalty is imposed on variables with high variance in

order to eliminate the number of variables and improve model predictions. Recent review articles provide a framework for interpreting clinical studies that use machine learning methods for clinical readers.<sup>29,32,33</sup>

Variable reduction was performed during logistic regression using backwards stepwise elimination. The variable selection process started with the full model and a bootstrap bias-corrected concordance index was used as the stopping criteria. Variables with individual P values that were  $>.05$  were left in the model if they offered information to improve the overall model. The removal of each variable was evaluated by determining which variable had the smallest effect on the adjusted  $R^2$  and was stopped when the bootstrap concordance index had a change of 0.001. Variable reduction was performed during lasso regression by 10-fold cross-validation of the lambda value and the final model incorporated the variables that were most predictive within one standard error of the best value.

Model performance was assessed using three recommended measures<sup>24</sup>: 1) the C statistic, or area under the receiver operating characteristic (ROC) curve, 2) calibration curves, and 3) decision curves. The C statistic measures the model's overall ability to discriminate between high- and low-risk patients, but does not allow one to understand how the model performs across the entire range of possible predictions, which is measured by the model's calibration curve. Calibration curves were plotted to show the relationship between the model's predicted outcomes against the cohort's observed outcome, where a perfectly calibrated model follows a 45° line.<sup>34</sup> In this study, Decision curve analysis (DCA) was used to quantify the net benefit of using each model and to visually compare the models.<sup>35</sup> DCA evaluates the benefits of a diagnostic test, or a prediction model in this case, across a range of patient preferences for accepting risk of undertreatment and overtreatment to facilitate decisions about test selection and use.<sup>35</sup> DCA assesses the value of information provided by the model by considering the likely range of a patient's risk and benefit preferences, without the need for actually measuring these preferences for a particular patient.<sup>35</sup> The net benefit is determined by calculating the difference between the expected benefit and the expected harm associated with each proposed testing and possible treatment strategies. The expected benefit is represented by the number of patients who have the outcome and who will receive treatment (true positives) using the proposed strategy.<sup>36</sup> The net benefit is calculated as: Net benefit = true positive rate - (false positive rate x weighting factor) in which the weighting factor = Threshold probability/1-threshold probability and the threshold probability is a level of certainty above which the patient or physician would choose to intervene. Variable importance, a scaled measure with a maximum value of 100,<sup>31</sup> was plotted to understand the contribution of each predictor in the machine learning models. All models were internally validated using bootstrapping or cross-validation to measure optimism-corrected reliability. Missing predictor values were imputed using Multiple Imputed Chained Equations (MICE).<sup>37</sup> The frequency of missing data for each variable generally did not vary over time between the two time periods used in temporal validation (2002–2006 and 2007–2008) (Appendix 3, available online at <http://links.lww.com/xxx>).

Measures, such as sensitivity, specificity, and false positive and negative probabilities are not recommended when reporting performance of clinical prediction models because they are performance measures after introducing one or more artificial probability thresholds or

categories.<sup>24</sup> Although these are useful for estimating accuracy or classification measures often reported in a single diagnostic test or prognostic factor studies, such dichotomization and related classification measures lead to loss of information when providing a prediction for the future and introducing such a threshold implies that it is relevant to clinical practice, which often is not the case.<sup>24</sup>

The dataset was split by time (i.e., temporal validation) in which we used the first phase (2002–2006) for model derivation and the second phase (2007–2008) for model validation. When the sample size is very large, this approach has been shown to be methodologically more rigorous than a simple random split of the dataset.<sup>38,39</sup> Next, we combined both site-specific and temporal validation by using each site once as a validation sample, with the remaining sites used for model derivation during the first phase. Site-specific estimates of discrimination (C statistic) and calibration were pooled and tested in the second phase. All analyses were performed with R statistical software version 3.4.1 (R Foundation for Statistical Computing).

## RESULTS

Of 228,438 births in the CSL cohort across 12 sites, data were available from 10 of 12 sites for estimated blood loss (152,279, or 66.6%) (Figure 1). The mean estimated blood loss was 445 mL (SD: 2,327), and was two times greater for cesarean than vaginal deliveries (769.9 mL vs. 315.0 mL) (Appendix 4, available online at <http://links.lww.com/xxx>). Of the 152,279 assessed births, 7,279 (4.7%, 95% CI 4.6 to 4.9) had a postpartum hemorrhage as defined by an estimated blood loss > 1,000 mL, which was higher for cesarean (15.2%, 95% CI: 14.9 to 15.6) than vaginal delivery (0.6%, 95% CI: 0.5 to 0.6). Of 145,943/152,279 (95.8%) births with available transfusion data, 11.3% (n=789) with a postpartum hemorrhage had a transfusion versus 1.9% (n=2,724) without a postpartum hemorrhage.

A total of 55 candidate predictors of postpartum hemorrhage available on labor admission were assessed for possible model inclusion, including socio-demographic, obstetric, clinical, and physiologic variables (Appendix 2, <http://links.lww.com/xxx>). Variables associated with postpartum hemorrhage included maternal age, pre-pregnancy body mass index, black race, residing in the South, delivery at a community hospital, gestational age at birth, maternal comorbid conditions (e.g., pregestational and gestational diabetes, chronic and gestational hypertension), prior preterm birth, antepartum admission, threatened preterm birth, antenatal steroids, breech presentation, prelabor rupture of membranes, whether trial of labor was attempted, and whether labor was initiated spontaneously. The variables included in each model are presented in Appendix 5, available online at <http://links.lww.com/xxx>.

After temporal and site validation, the best performing model was the Extreme Gradient Boosting model with the highest discriminative ability (C statistic 0.93, 95% CI: 0.92 to 0.93) (Table 1). The random forest model also had a high discriminative ability (C statistic 0.92; 95% CI: 0.91 to 0.92). The lasso regression (C statistic: 0.87; 95% CI: 0.86 to 0.88) and the logistic regression model had lower discriminative ability (C statistic: 0.87; 95% CI: 0.86 to 0.87). Appendix 6, available online at <http://links.lww.com/xxx>, displays the ROC curves for the 2 machine learning and 2 statistical models.

Models were further validated both over time (i.e., temporally) and by site for predicting postpartum hemorrhage. As above, we similarly noted good to excellent discrimination for prediction of postpartum hemorrhage with Extreme Gradient Boosting (C statistic: 0.93; 95% CI: 0.92 to 0.94), followed by random forest (C statistic: 0.92; 95% CI: 0.91 to 0.92). Logistic regression (C statistic 0.87; 95% CI: 0.86 to 0.87) and lasso regression (C statistic: 0.87; 95% CI: 0.86 to 0.88) had lower discriminative ability.

Figure 2 demonstrates the overall calibration curves with 95% confidence intervals of the four models. The calibration curve shows the variation in performance of each model in comparison to perfect agreement between the predicted probability of the model and the actual probability. We also present the best performing model, Extreme Gradient Boosting, by each of the ten assessed sites in Appendix 7, available online at <http://links.lww.com/xxx>. This model assigned an accurate probability of postpartum hemorrhage when prediction ranged from 0 to 70–80%.

In the decision curve analysis (Figure 3), all models provided superior net benefit when clinical decision thresholds were between 0 to 80%. The net benefit for the Extreme Gradient Boosting model was greatest across the range of threshold probabilities compared with the other models (i.e., lasso, random forest, and logistic regression).

Figure 4 displays the variable importance in the Extreme Gradient Boosting model for postpartum hemorrhage, which was the best performing model. The top 10 variables ranked from most to least important included pre-pregnancy maternal weight, admission maternal weight, prenatal diagnosis of fetal macrosomia, admission temperature, attempted trial of labor on admission, pre-pregnancy maternal body mass index, admission systolic blood pressure, multiple gestation, anemia diagnosis during pregnancy, and spontaneous labor on admission.

We were unable to compare our models to current risk stratification strategies as outlined by CMQCC (i.e., low, medium, and high risk for postpartum hemorrhage) as we only had two-thirds of these variables available in the dataset and many of these variables (such as uterine fibroids, known bleeding disorder or coagulopathy) were not routinely assessed on labor admission and therefore were not considered as candidate variables for our models.

## DISCUSSION

We found that machine learning and statistical models can accurately predict postpartum hemorrhage using data available at the time of admission for labor. Machine learning models performed the best but at the cost of possibly increased complexity and minimal clinical significance. Extreme Gradient Boosting and random forest models provided excellent discriminative ability to predict postpartum hemorrhage. Importantly, these models achieved high predictive performance using clinical and physiologic data readily available to the obstetric provider at the time of labor admission.

Clinical application could allow obstetric providers to be prepared and in some instances (i.e., when not in spontaneous labor) to triage women at high-risk of postpartum hemorrhage to the appropriate level of maternity care.<sup>40</sup> It would be reasonable to integrate the models

into an online calculator or automated input in the electronic medical record for immediate use on labor admission. At the current time, statistical models are easier to integrate in this manner compared to many machine learning algorithms. Given that the statistical models performed well this should also be considered. Regardless, any model will need to be prospectively assessed in local contemporary cohorts of pregnant women across regional settings. Our model focused on prediction at the time of labor admission and hence employed variables available at that time, and it may also be reasonable to build models that include intrapartum variables that affect hemorrhage risk, such as length of labor and mode of delivery.

To date, accurate prediction models using statistical models for postpartum hemorrhage have been lacking. Dilla et al. found that CMQCC risk strata of low-, medium-, and high-risk generally predicted postpartum hemorrhage at a tertiary care center, but this analysis did not involve development of a prediction model that could be prospectively tested.<sup>41</sup> Betts et al. aimed to identify women at risk for common postpartum complications using Australian administrative data and found good discrimination for postpartum hypertension and surgical site infection (both with AUC > 0.80), but not postpartum hemorrhage.<sup>42</sup> Similarly, an earlier model did not accurately predict postpartum hemorrhage using data from the HYPITAT trial of women with gestational hypertension or preeclampsia in the Netherlands (AUC 0.59).<sup>43</sup> Albright et al. built a logistic regression model that adequately predicted transfusion after cesarean delivery using data from the MFMU Cesarean Registry (AUC 0.82).<sup>44</sup> Recently, models using traditional statistical methods been developed with fair to good predictive ability for intensive care unit admission (AUC 0.81) and failed induction among obese women (AUC 0.79) using population-based administrative data.<sup>45,46</sup> In comparison to these prior models, our data suggest that both machine learning and statistical models can provide superior discriminative ability in the case of postpartum hemorrhage and a final model should be chosen based on a combination of discrimination, calibration (especially accurate calibration in the range of predictions where clinical decisions will be affected), ease of use in the clinical setting, and acceptability by clinicians and patients.

Prediction of postpartum hemorrhage on labor admission could allow for optimizing L&D unit healthcare resources, risk mitigation, and timely care.<sup>47</sup> Currently, health care providers rely largely on their clinical judgment and recognition of non-specific risk factors. To further systematize this process, the Alliance for Innovation on Maternal Health (AIM) has established an obstetric hemorrhage patient safety bundle, including policies, guidelines, and algorithms, to aid in the prompt recognition and management of postpartum hemorrhage.<sup>8,28</sup> The CMQCC categorizes pregnant women into low-, medium-, or high-risk based on clinical or laboratory risk factors,<sup>48</sup> and state-wide implementation of their hemorrhage protocol has been associated with a reduction in severe maternal morbidity.<sup>49</sup> Despite increasing impetus for wider adoption across the U.S.,<sup>11</sup> these risk-based stratification guidelines or decision tree algorithms are based on expert opinion and clinical consensus and these methods do not provide an individualized risk prediction. In the current study, we did not formally compare our models to the CMQCC risk strata for both methodological (i.e., availability of only about two-thirds of these variables) as well as pragmatic (i.e., many of these variables are not routinely assessed on labor admission) considerations. Our definition of hemorrhage did not incorporate transfusion of blood products and further

models to accurately predict transfusion beyond current guidelines are needed as recently highlighted.<sup>50</sup> We did note a 2% rate of transfusion among those women without postpartum hemorrhage, which may reflect outcome misclassification or cases of transfusion that occurred after delayed or late postpartum hemorrhage.

There are several study limitations. Missing data is an important limitation of this analysis, including restricting this analysis to the subset of the original cohort with available blood loss data and the substantial proportion of covariates with incomplete data, albeit we used up-to-date imputation techniques. The proportion of missing data is a limitation to the general application of these models affecting generalizability of these results. However, it is likely that missing data or incomplete ascertainment will continue to be a limitation when applying these models in real time with EHR data.

Similar to clinical applications of precision medicine in other medical specialties, whether improved prediction will impact clinical outcomes and patient disposition compared with conventional clinical practice remains to be studied.<sup>51</sup> Alternatively stated, accurate prediction of postpartum hemorrhage may not necessarily change care or produce better patient outcomes. It is possible that inclusion of more predictors in the model, such as serial measurement of vital signs, physical examination findings, and laboratory data, may improve prediction, but may not be feasible and could potentially delay intervention. The impact of a model on clinical decision making depends on multiple provider and care environment characteristics, including capacity to formulate a timely clinical response, weighing of risks and benefits of intervention, ability to execute that action, and patient (or health care provider) adherence to the recommended intervention.<sup>19</sup> Additional environmental constraints may include personnel, space, and equipment, which are not integrated into current prediction models. Furthermore, estimated blood loss is known to be imprecise, inaccurate, and often underestimated,<sup>52</sup> and there can be substantial variability in the relationship between blood loss and clinical signs and symptoms.<sup>53</sup>

Our definition of postpartum hemorrhage is consistent with current clinical guidelines; however, we did not assess other relevant clinical measures of acute blood loss, including a higher blood loss threshold, hemorrhage resulting in transfusion, drop of hemoglobin pre- to post- delivery, or change in vital signs to suggest hemodynamic instability. Predictive models for these outcomes will need to be tested while using newer, more precise methods of assessing blood loss at delivery, including weight-based or photographic, colorimetric quantitative blood loss approaches. Finally, machine learning approaches are data driven and depend on accurate data. Some important clinical variables were not measured in the CSL dataset (e.g., postpartum hemorrhage in prior pregnancy, thrombocytopenia, thromboprophylactic drug treatment, placental characteristics, and uterine fibroids). Some assessed variables, such as a prenatal diagnosis of macrosomia, were likely not universally captured and it was unclear how they were defined. Additionally, the dataset is now a decade old. These models will need to be replicated in more contemporary prospective datasets, ideally in real time as part of an integrated electronic medical record.

Strengths of this study include creation of a generalizable model drawn from a large dataset from multiple hospitals across the U.S. over nearly a decade, as well as utilization of an



analytical approach that has yet to widely studied and implemented in obstetrics. Our findings reinforce that machine learning approaches can be used to improve clinical prediction in obstetrics, and this “proof of concept” will need to be prospectively tested.

In conclusion, these findings are an opportunity to apply novel prediction approaches to support decision making on labor admission in an era of rising U.S. maternal morbidity and mortality. As predictive tools become more widely used in obstetric care,<sup>45</sup> they can be included in guidelines and care pathways for clinical use and further testing. Identification of women at high risk of postpartum hemorrhage on labor admission using both machine learning and statistical models could allow for more prompt diagnosis and possibly intervention, which may result in more accurate clinical care, improved patient outcomes, and better resource allocation.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Funding:** Supported by the Intramural Research Program of the National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD). The Consortium on Safe Labor was funded by the Intramural Research Program of the NICHD, through Contract No. HHSN267200603425C. The study was funded by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. Intramural investigators designed the study and data was collected by clinical site investigators. The corresponding author has full access to the data and final responsibility for preparation and submission of the paper for publication.

David M. Stamilio reports receiving funds for being a symposium lecturer at the VI Congreso Javeriano de Ginecología y Obstetricia Sep 12–14, 2019, and he was compensated for travel expenses by Pontificia Universidad Javeriana, Bogota, Colombia. He was also a co-investigator on a Bill & Melinda Gates Foundation grant (OPPI191684; PI Jeff Stringer) from July 12, 2018–April 30, 2019.

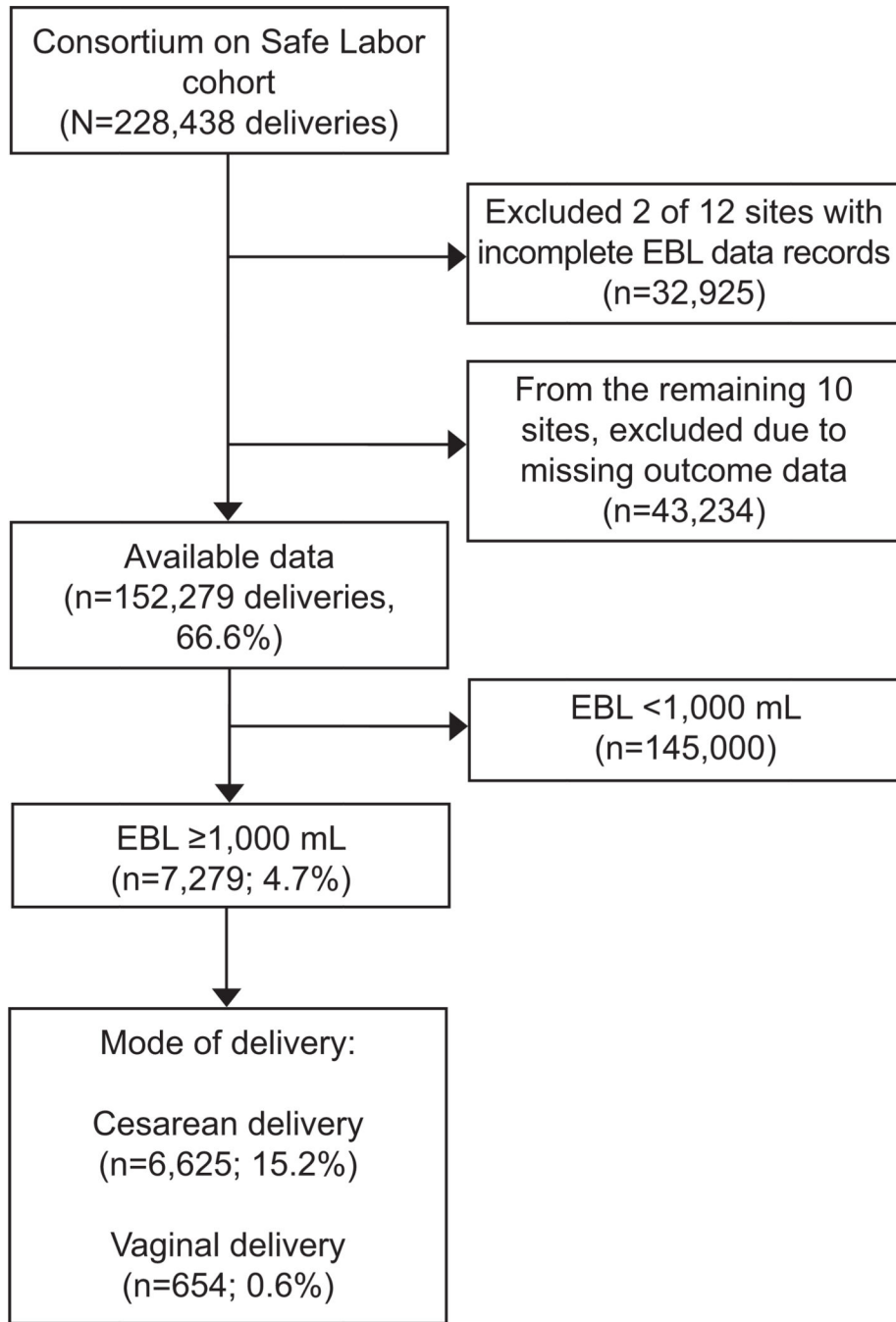
## REFERENCES

1. Say L, Chou D, Gemmill A, Tunçalp ÖZ, Moller AB, Daniels J, Gülmezoglu AM, Temmerman M, Alkema L. Global causes of maternal death: a WHO systematic analysis. *Lancet Global Health* 2014;2:e323–33. [PubMed: 25103301]
2. Bateman B, Berman MF, Riley LE, Leffert LR. The epidemiology of postpartum hemorrhage in a large, nationwide sample of deliveries. *Anaesthesia and analgesia* 2010;110:1368–73.
3. Callaghan W, Kuklina EV, Berg CJ. Trends in postpartum hemorrhage: United States, 1994–2006. *American Journal of Obstetrics and Gynecology* 2010;202:353. [PubMed: 20350642]
4. Creanga A, Berg CJ, Ko JY, Farr SL, Tong VT, Bruce FC, Callaghan WM. Maternal mortality and morbidity in the United States: where are we now? *Journal of Womens Health* 2014;23:3–9.
5. Callaghan W, Mackay AP, Berg CJ. Identification of severe maternal morbidity during delivery hospitalizations, United States, 1991–2003. *American Journal of Obstetrics and Gynecology* 2008;199:133. [PubMed: 18279820]
6. Hamm R, Wang E, Romanos A, O'Rourke K, Srinivas SK. Implementation of Quantification of Blood Loss Does Not Improve Prediction of Hemoglobin Drop in Deliveries with Average Blood Loss. *American journal of perinatology* 201;35:134–9. [PubMed: 28838005]
7. Committee on Practice Bulletins-Obstetrics. Practice Bulletin No. 183: Postpartum Hemorrhage. *Obstetrics and gynecology* 2017;130:e168–e86. [PubMed: 28937571]
8. Obstetric hemorrhage. 2019 (Accessed March 25, 2019, at <https://safehealthcareforeverywoman.org/patient-safety-bundles/obstetric-hemorrhage/>.)

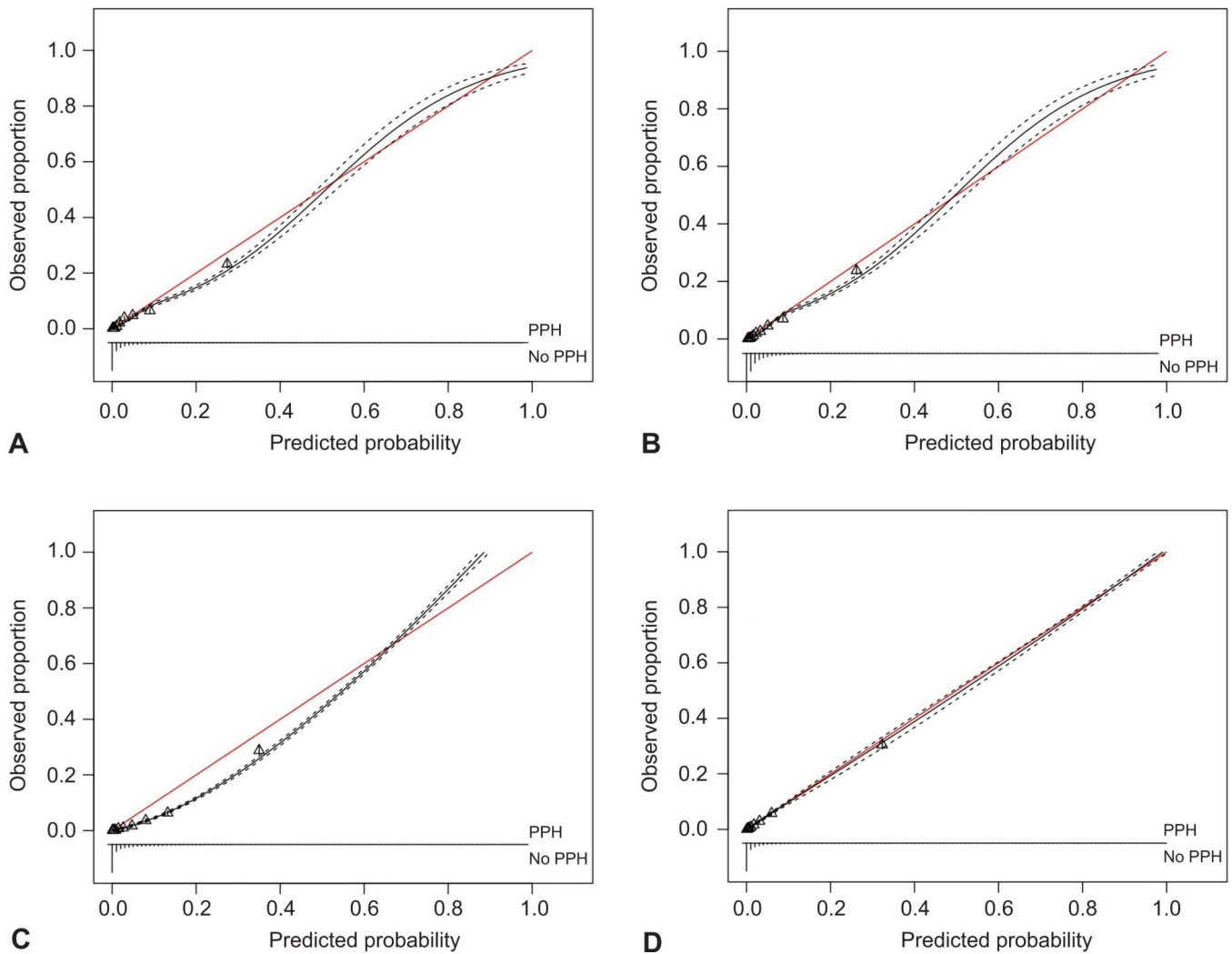
9. Dahlke J, Mendez-Figueroa H, Maggio L, Hauspurg AK, Sperling JD, Chauhan SP, Rouse DJ. Prevention and management of postpartum hemorrhage: a comparison of 4 national guidelines. *American Journal of Obstetrics and Gynecology* 2015;213:76. [PubMed: 25731692]
10. Shields L, Wiesner S, Fulton J, Pelletreau B. Comprehensive maternal hemorrhage protocols reduce the use of blood products and improve patient safety. *American Journal of Obstetrics and Gynecology* 2015;212:272–80. [PubMed: 25025944]
11. Main E, Goffman D, Scavone BM, Low LK, Bingham D, Fontaine PL, Gorlin JB, Lagrew DC, Levy BS; National Partnership for Maternal Safety; Council on Patient Safety in Women's Health Care. National Partnership for Maternal Safety: Consensus Bundle on Obstetric Hemorrhage. *Obstetrics and gynecology* 2015;126:155–62. [PubMed: 26241269]
12. Wetta L, Szychowski JM, Seals S, Mancuso MS, Biggio JR, Tita AT. Risk factors for uterine atony/postpartum hemorrhage requiring treatment after vaginal delivery. *American Journal of Obstetrics and Gynecology* 2013;209:51. [PubMed: 23507549]
13. Kramer M, Berg C, Abenhaim H, Dahhou M, Rouleau J, Mehrabadi A, Joseph KS. Incidence, risk factors, and temporal trends in severe postpartum hemorrhage. *American Journal of Obstetrics and Gynecology* 2013;209:449. [PubMed: 23871950]
14. Prata N, Hamza S, Bell S, Karasek D, Vahidnia F, Holston M. Inability to predict postpartum hemorrhage: insights from Egyptian intervention data. *BMC Pregnancy and Childbirth* 2011;11. [PubMed: 21276263]
15. Mousa H, Cording V, Alfirevic Z. Risk factors and interventions associated with major primary postpartum hemorrhage unresponsive to first-line conventional therapy. *Acta obstetrica et gynecologica Scandinavica* 2008;87:652–61. [PubMed: 18568465]
16. Biguzzi E, Franchi F, Ambrogi F, Ibrahim B, Bucciarelli P, Acaia B, Radaelli T, Biganzoli E, Mannucci PM. Risk factors for postpartum hemorrhage in a cohort of 6011 Italian women. *Thrombosis Research* 2012;129:e1–7. [PubMed: 22018996]
17. Helman S, Drukker L, Fruchtman H, Ioscovich A, Farkash R, Avitan T, Samueloff A, Grisar-Granovsky S. Revisit of risk factors for major obstetric hemorrhage: insights from a large medical center. *Archives of Obstetrics and Gynecology* 2015;292:819–28.
18. Cuocolo R, Perillo T, De Rosa E, Ugga L, Petretta M. Current applications of big data and machine learning in cardiology. *Journal of Geriatric Cardiology* 2019;16:601–7. [PubMed: 31555327]
19. Shah N, Milstein A, Bagley SC. Making Machine Learning Models Clinically Useful. *Journal of the American Medical Association* 2019;322:1351.
20. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine Learning-Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. *JAMA Open Network* 2019;2:e186937.
21. Escobar G, Gupta NR, Walsh EM, Soltesz L, Terry SM, Kipnis P. Automated early detection of obstetric complications: theoretic and methodologic considerations. *American Journal of Obstetrics and Gynecology* 2019;220:297–307. [PubMed: 30682365]
22. Fohner A, Greene JD, Lawson BL, Chen JH, Kipnis P, Escobar GJ, Liu VX. Assessing clinical heterogeneity in sepsis through treatment patterns and machine learning. *Journal of the American Medical Informatics Association* 2019;In press.
23. Zhang J, Troendle J, Reddy UM, Laughon SK, Branch DW, Burkman R, Landy HJ, Hibbard JU, Haberman S, Ramirez MM, Bailit JL, Hoffman MK, Gregory KD, Gonzalez-Quintero VH, Kominiarek M, Learman LA, Hatjis CG, van Veldhuisen P; Consortium on Safe Labor. Contemporary cesarean delivery practice in the United States. *American Journal of Obstetrics and Gynecology* 2010;203:326. [PubMed: 20708166]
24. Moons K, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine* 2015;162:W1–73. [PubMed: 25560730]
25. Kerr R, Eckert LO, Winikoff B, Durocher J, Meher S, Fawcus S, Mundle S, Mol B, Arulkumaran S, Khan K, Wandwabwa J, Kochhar S, Weeks A; Brighton Collaboration Primary Postpartum Haemorrhage Working Group. Postpartum haemorrhage: Case definition and guidelines for data

- collection, analysis, and presentation of immunization safety data. *Vaccine* 2016;34:6102–9. [PubMed: 27431424]
26. Postpartum Haemorrhage, Prevention and Management (Green-top Guideline No. 52). 2016 (Accessed March 16, 2019, at <https://www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg52/>.)
  27. World Health Organization. The prevention and management of postpartum haemorrhage: report of a technical working group, Geneva 3–6, July 1989. 1989.
  28. Menard M, Main EK, Currihan SM. Executive summary of the reVITALize initiative: standardizing obstetric data definitions. *Obstetrics and gynecology* 2014;124:150–3. [PubMed: 24901267]
  29. Tibshirani R Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society* 1996;Series B:267–88.
  30. Breiman L Random Forests. *Machine Learning* 2001;45:5–32.
  31. xgboost: extreme gradient boosting. 2018 (Accessed October 12, 2019, at <https://cran.r-project.org/web/packages/xgboost/index.html>.)
  32. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology* 2019;In press.
  33. Liu Y, Chen PHC, Krause J, Peng L. How to Read Articles That Use Machine Learning Users' Guides to the Medical Literature. *Journal of the American Medical Association* 2019;322:1806–16. [PubMed: 31714992]
  34. Jelovsek J, Hill AJ, Chagin KM, Kattan MW, Barber MD. Predicting Risk of Urinary Incontinence and Adverse Events After Midurethral Sling Surgery in Women. *Obstetrics and gynecology* 2016;127:330–40. [PubMed: 26942362]
  35. Vickers A, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* 2006;26:565–74. [PubMed: 17099194]
  36. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *Journal of the American Medical Association* 2015;313:409–10. [PubMed: 25626037]
  37. White I, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* 2011;30:377–99. [PubMed: 21225900]
  38. Roberts D, Bahn V, Ciuti S, Boyce M, Elith J, Guillera-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 2017;40:913–29.
  39. Steyerberg E Validation of Prediction Models. *Clinical Prediction Models*. New York: Springer; 2019.
  40. Society for Maternal-Fetal Medicine. Obstetric Care Consensus, Number 9: Levels of Maternal Care. *American Journal of Obstetrics and Gynecology* 2019;In press.
  41. Dilla A, Waters JH, Yazer MH. Clinical validation of risk stratification criteria for peripartum hemorrhage. *Obstetrics and gynecology* 2013;122:120–6. [PubMed: 23743452]
  42. Betts K, Kisely S, Alati R. Predicting common maternal postpartum complications: Leveraging health administrative data and machine learning. *BJOG* 2019;In press
  43. Koopmans C, van der Tuuk K, Groen H, Doornbos JP, de Graaf IM, van der Salm PC, Porath MM, Kuppens SM, Wijnen EJ, Aardenburg R, van Loon AJ, Akerboom BM, van der Lans PJ, Mol BW, van Pampus MG; HYPITAT study group. Prediction of postpartum hemorrhage in women with gestational hypertension or mild preeclampsia at term. *Acta obstetrica et gynecologica Scandinavica* 2014;93:399–407. [PubMed: 24575790]
  44. Albright C, Spillane TE, Hughes BL, Rouse DJ. A Regression Model for Prediction of Cesarean-Associated Blood Transfusion. *American journal of perinatology* 2019;36:879–85. [PubMed: 30743270]
  45. Rossi R, Hall E, Dufendach K, DeFranco EA. Predictive Model of Factors Associated With Maternal Intensive Care Unit Admission. *Obstetrics and gynecology* 2019;134:216–24. [PubMed: 31306325]
  46. Rossi R, Requarth EW, Warshak CR, Dufendach K, Hall ES, DeFranco EA. Predictive Model for Failed Induction of Labor Among Obese Women. *Obstetrics and gynecology* 2019;134:485–93. [PubMed: 31403588]

47. Merriam A, Wright JD, Siddiq Z, D'Alton ME, Friedman AM, Ananth CV, Bateman BT. Risk for postpartum hemorrhage, transfusion, and hemorrhage-related morbidity at low, moderate, and high volume hospitals. *Journal of Maternal Fetal and Neonatal Medicine* 2018;31:1025–34. [PubMed: 28367647]
48. Improving health care response to obstetric hemorrhage version 2.0. A California quality improvement toolkit. . 2015 (Accessed February 1,9, at <https://www.cmqcc.org/resources-tool-kits/toolkits/ob-hemorrhage-toolkit>.)
49. Main E, Cape V, Abreo A, Vasher J, Woods A, Carpenter A, Gould JB. Reduction of severe maternal morbidity from hemorrhage using a state perinatal quality collaborative. *American Journal of Clinical Nutrition* 2017;216:298.
50. Kawakita T, Mokhtari N, Huang JC, Landy HJ. Evaluation of Risk-Assessment Tools for Severe Postpartum Hemorrhage in Women Undergoing Cesarean Delivery. *Obstetrics and gynecology* 2019;134:1308–16. [PubMed: 31764744]
51. Emanuel E, Wachter RM. Artificial Intelligence in Health Care: Will the Value Match the Hype? *Journal of the American Medical Association* 2019;321:2281. [PubMed: 31107500]
52. Gr Dildy, Paine AR George NC, Velasco C. Estimating blood loss: can teaching significantly improve visual estimation? *Obstetrics and gynecology* 2004;104:601–6. [PubMed: 15339775]
53. Pacagnella R, Souza JP, Durocher J, Perel P, Blum J, Winikoff B, Gülmezoglu AM. A systematic review of the relationship between blood loss and clinical signs. *PLoS ONE* 2013;8:e57594.

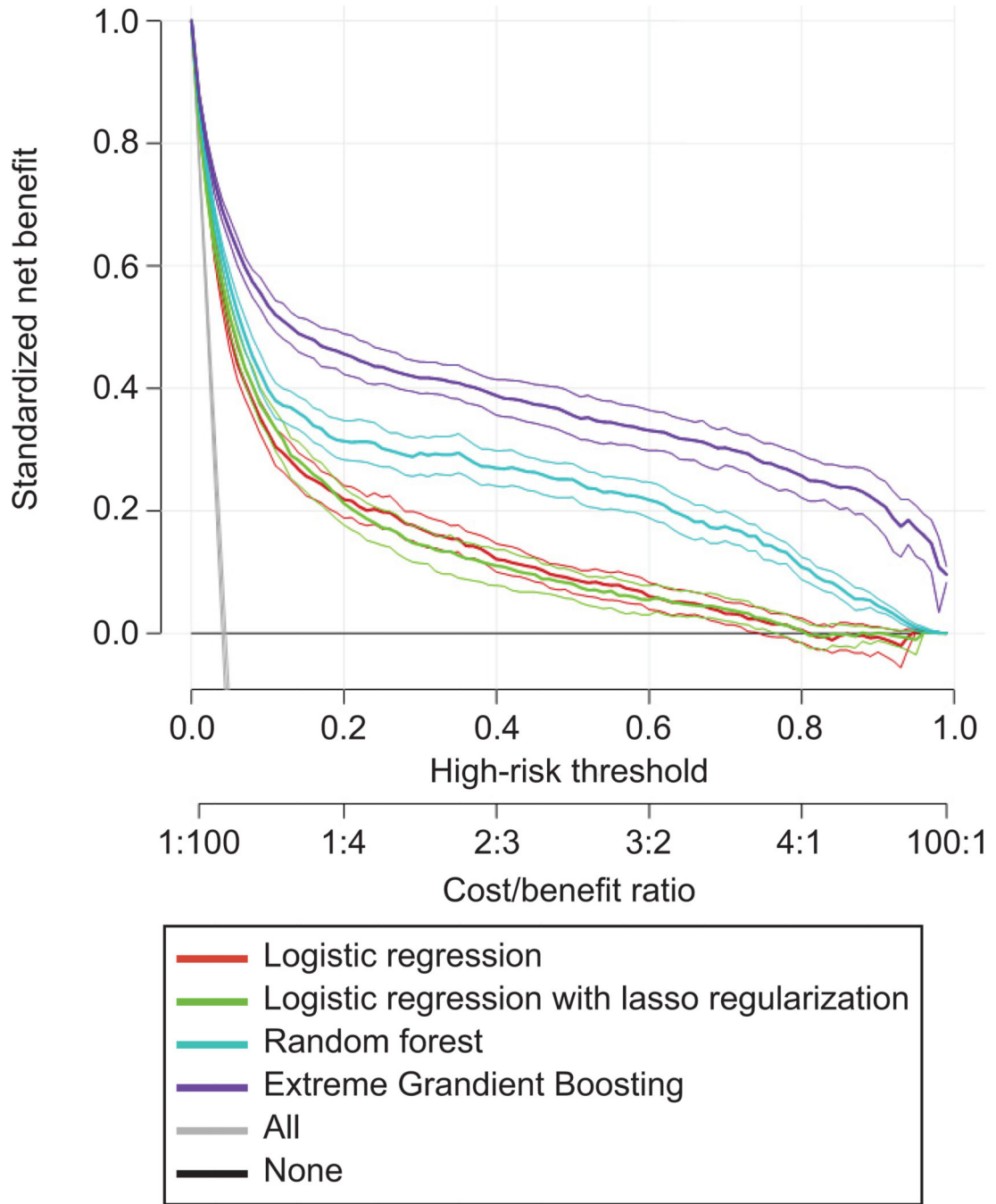


**Figure 1.** Flow chart of women with postpartum hemorrhage (estimated blood loss [EBL]  $\geq$  1,000 mL).



**Figure 2.**

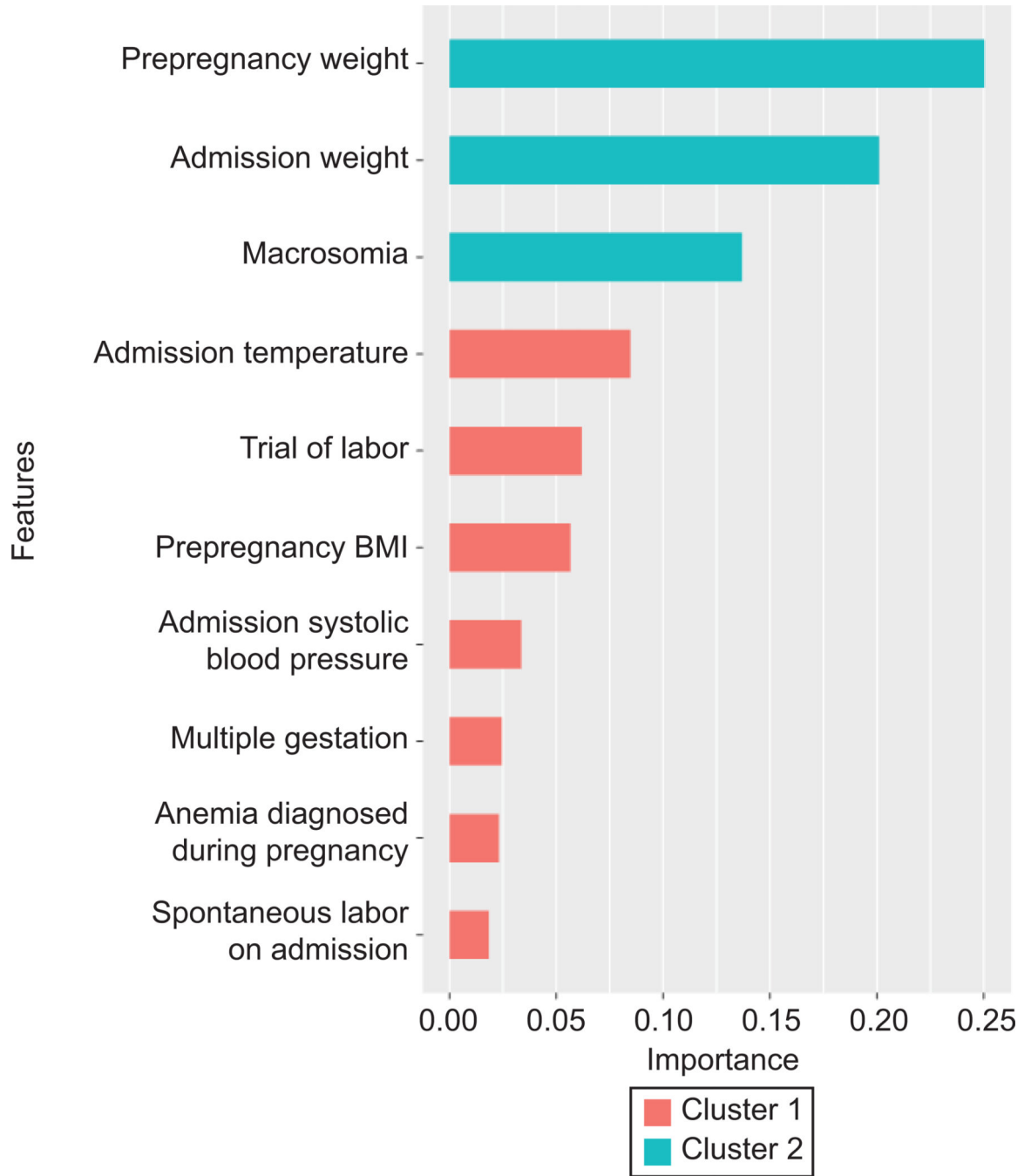
Calibration curves demonstrating the performance of predicting postpartum hemorrhage (PPH) for all four models: Logistic regression (**A**), logistic regression with lasso regularization (**B**), random forest (**C**), and extreme gradient boosting (**D**). The figure demonstrates the variation in each model's performance. The *red line* indicates perfect agreement between the predicted probability of the model and the actual probability. The *black line bounded by two dotted lines* indicates the overall calibration with 95% CIs of each model. Each *triangle* represents a group of individuals risk. There are 10 triangles and each triangle represents a decile of risk. **A.** Calibration (intercept:  $-0.17$  [ $-0.22$  to  $-0.11$ ]; slope:  $0.96$  [ $0.92$  to  $1.00$ ]); discrimination (C-statistic:  $0.87$  [ $0.86$  to  $0.87$ ]). **B.** Calibration (intercept:  $-0.20$  [ $-0.26$  to  $-0.15$ ]; slope:  $1.08$  [ $1.04$  to  $1.12$ ]); discrimination (C-statistic:  $0.87$  [ $0.86$  to  $0.88$ ]). **C.** Calibration (intercept:  $-0.60$  [ $-0.66$  to  $-0.54$ ]; slope:  $1.28$  [ $1.23$  to  $1.33$ ]); discrimination (C-statistic:  $0.92$  [ $0.91$  to  $0.92$ ]). **D.** Calibration (intercept:  $-0.13$  [ $-0.19$  to  $-0.06$ ]; slope:  $1.02$  [ $0.98$  to  $1.06$ ]); discrimination (C-statistic:  $0.93$  [ $0.92$  to  $0.94$ ]).



**Figure 3.** Decision curve analysis of predicting postpartum hemorrhage by all models. The x-axis indicates the threshold probability for postpartum hemorrhage outcome. The threshold probability is a level of certainty above which the patient or physician would choose to intervene. The probability threshold captures the relative value the patient or physician places on receiving an intervention for the outcome, if present, to the value of avoiding an intervention if the outcome is not present. The y-axis indicates the net benefit. The net benefit is calculated as true positive rate – (false positive rate x weighting factor). Weighting

factor is calculated as  $\text{threshold probability} / (1 - \text{threshold probability})$ . For example, when threshold probability is 0.1, weighting factor is  $0.1 / (1 - 0.1) = 0.1 / 0.9$ . The decision curves indicate the net benefit of each model as well as two clinical alternatives (classifying no women as having the outcome vs. classifying all women as having the outcome) over a specified range of threshold probabilities of outcome. Compared with the clinical alternatives, the net benefit for the Extreme Gradient Boosting model was greatest across the range of threshold probabilities.





**Figure 4.** Importance of each predictor in the Extreme Gradient Boosted model to predict risk of postpartum hemorrhage. The variable importance is a measure scaled to have a maximum value of one. Cluster indicates features that are similar to one another in importance value. BMI, body mass index.

**Table 1.**

Discrimination between women with and without postpartum hemorrhage using two machine learning and two statistical models

Model.	Temporal Validation <sup>1</sup>		Temporal and site validation <sup>2</sup>	
	<2007	>= 2007	<2007	>= 2007
<b>Extreme Gradient Boosting</b>	0.95 (0.95 to 0.95)	0.93 (0.92 to 0.93)	0.93 (0.92 to 0.94)	0.93 (0.92 to 0.94)
<b>Random forest</b>	0.99 (0.99 to 1)	0.92 (0.91 to 0.92)	0.92 (0.91 to 0.92)	0.92 (0.91 to 0.92)
<b>Logistic regression with lasso regularization</b>	0.88 (0.87 to 0.88)	0.87 (0.86 to 0.88)	0.87 (0.86 to 0.88)	0.87 (0.86 to 0.88)
<b>Logistic regression model</b>	0.87 (0.87 to 0.88)	0.87 (0.86 to 0.87)	0.87 (0.86 to 0.87)	0.87 (0.86 to 0.87)

<sup>1</sup>In temporal validation, models were constructed from the first phase (2002–2006) and externally validated (i.e., temporally) in the second phase (2007–2008).

<sup>2</sup>In temporal and site validation, both clinical site-specific (total of 10 sites) and temporal validation were combined by using each site once as a validation sample, with the remaining sites used for model derivation during the first phase.

Data are presented as concordance index with 95% confidence intervals.