Technical Report: Analysis of a Postpartum Hemorrhage Prediction Algorithm
James Weimer
University of Pennsylvania

This technical report describes an analysis of a postpartum hemorrhage prediction algorithm developed at the University of Pennsylvania. This technical report documents the performance of the postpartum hemorrhage prediction algorithm as of August 15, 2022.

An analysis of a postpartum hemorrhage (PPH) algorithm was performed using data from a 500+ prospective study that collected 560 nm (green) light on women at 5 Hz from admission to 24 hours post-delivery using a Samsung Galaxy Watch Active at the Pennsylvania Hospital in Philadelphia. For the purposes of development and analysis, the data was labeled by the clinical team leveraging the electronic health records. Specifically, the following labels were automatically extracted from the electronic health record for each subject: (i) delivery datetime; (ii) PPH interventions and their datetime (if any); and (iii) the estimated blood loss amounts and datetimes. The delivery datetime was recorded as the combined date and time of birth. The PPH interventions were identified as medications given beyond standard of care for birth at the Pennsylvania Hospital. Specifically, a second dose of Pitocin, or any administration of medications used to intervene in a PPH (e.g., methylergonovine, misoprostol, hemabate, Txa). The PPH intervention datetime was recorded as the combined date and time of medication administration. The estimated blood loss amounts and datetimes were measured in mL and the datetime was recorded as the combined date and time of recorded estimated blood loss.

For the purposes of algorithm design, validation and testing we set aside 50% of the subjects as "test subjects" and used the remaining as "training/validation subjects". For the training/validation subjects we initially only considered subjects that experienced a standard vaginal delivery (i.e., we removed deliveries by cesarian section). The reasoning behind only considering standard vaginal deliveries was motivated by the fact that the heart rate varies more significantly during a standard vaginal delivery than a cesarian section due to the different levels of pain experienced and physical effort required by the subject. Within the standard vaginal delivery training/validation subjects, we defined "control" subjects to be subjects who did not receive a PPH intervention (as defined above) and did not have an estimated blood loss greater than 1000 mL. Only including control subjects for training/validation that had a standard vaginal delivery and an estimate blood loss of less than 1000 mL ensures that any outlier subjects who experienced a clinically significant PPH but (for one reason or another) did not receive an intervention were omitted – as they would negatively bias algorithm training and validation. Within the standard vaginal delivery training/validation subjects, we defined "case" subjects to be subjects who received a PPH intervention (as defined above), regardless of estimated blood loss. The reason for omitting estimated blood loss from our definition of case subjects is motivated by the fact that a clinician made the decision to intervene. This logically implies that there was at least increased concern about the potential of PPH that warranted intervention by the clinical team at the Pennsylvania Hospital in Philadelphia. Moreover, the interventions – if effective – can prevent significant blood loss and the corresponding subjects should be considered "case" subjects since an intervention was taken regardless of estimated blood loss.

For the standard vaginal delivery training/validation control subjects (here forth referred to as "control subjects"), all data collected from admission to 24 hours postpartum was utilized for training/validation. Differently, for the standard vaginal delivery training/validation case subjects (here forth referred to as "case subjects"), only the data prior to intervention was utilized for training/validation. The reasoning for limiting the case subject data is to ensure that the developed algorithm is predictive (i.e., predicts an intervention) rather than responsive. By only training and validating on the data prior to intervention (if one exists) the algorithm predictive performance is anticipated to generalize to clinical deployments.

For training/validation, we utilized a standard '5-fold cross validation approach'. For each of the five folds, the remaining four folds were used to learn a proprietary model to discriminate case and control subjects that was evaluated on the original fold. For each fold, the proprietary model architecture was trained on the other 4 data folds using Binary Cross Entropy loss. Once trained, a hold-one-out cross validation was performed on the original fold to choose a threshold to achieve a desired specified commercially viable false alarm rate by leveraging all but one subject in the original fold to select a threshold for prediction and then apply that threshold to the hold out subject. Leveraging a combination of 5-fold cross validation for training and a hold-one-out cross validation for threshold selection improves the likelihood of generalization. Said differently, has we utilized the other 4 data folds to train the algorithm and select the threshold, the resulting model may not generalize well since the prediction threshold would have been chosen based on the same data used for training.

Leveraging the trained models and predictive results, performance metrics were calculated. Specifically, we focused on sensitivity (i.e., the percent of PPH interventions correctly predicted to need a PPH intervention) and specificity (i.e., the percent of control subjects that were not predicted to need a PPH intervention).

Analysis Results

Appling the analysis described above, there were 35 case subjects and 113 control subjects. For the case subjects, the algorithm correctly predicted 26 subjects as needing an intervention. For the control subjects, the algorithm correctly predicted 57 subjects as not needing an intervention. This corresponds to a sensitivity of 74% and a specificity of 50%. Stated differently, the algorithm correctly predicts which half of all subjects will contain nearly 75% of women needing an intervention.