OPEN

Increasing Cardiovascular Data Sampling Frequency and Referencing It to Baseline Improve Hemorrhage Detection

Anthony Wertz, MS[1]; Andre L. Holder, MD[2]; Mathieu Guillame-Bert, PhD[1]; Gilles Clermont, MD, CM, FCCM[2]; Artur Dubrawski, PhD[1]; Michael R. Pinsky, MD, CM, MCCM[2]

Objectives: We hypothesize that knowledge of a stable personalized baseline state and increased data sampling frequency would markedly improve the ability to detect progressive hypovolemia during hemorrhage earlier and with a lower false positive rate than when using less granular data.

Design: Prospective temporal challenge.

Setting: Large animal research laboratory, University Medical Center.

Subjects: Fifty-one anesthetized Yorkshire pigs.

Interventions: Pigs were instrumented with arterial, pulmonary arterial, and central venous catheters and allowed to stabilize for 30 minutes then bled at a constant rate of either 5 mL·min^{-1} ($n = 13$) or 20 ($n = 38$) until mean arterial pressure decreased to 40 or 30 mm Hg in the 5 and 20 mL·min^{-1} pigs, respectively.

Measurements and Main Results: Data during the stabilization period served as baseline. Hemodynamic variables collected at 250 Hz were used to create predictive models of "bleeding" using featurized beat-to-beat and waveform data and compared with models using mean unfeaturized hemodynamic variables averaged over 1-minute as simple hemodynamic metrics using random forest classifiers to identify bleeding with or without baseline data. The robustness of the prediction was evaluated in a leave-one-pig-out cross-validation. Predictive performance of models was compared by their activity monitoring operating characteristic and receiver operating characteristic profiles. Primary hemodynamic threshold data poorly identified bleed onset unless very stable initial baseline reference data were available. When referenced to baseline, bleed detection at a false positive rates of 10^{-2} with time to detect 80% of pigs bleeding was similar for simple hemodynamic metrics, beat-to-beat, and waveform at about 3–4 minutes. Whereas when universally baselined, increasing sampling frequency reduced latency of bleed detection from 10 to 8 to 6 minutes, for simple hemodynamic metrics, beat-to-beat, and waveform, respectively. Some informative features differed between simple hemodynamic metrics, beat-to-beat, and waveform models.

Conclusions: Knowledge of personal stable baseline data allows for early detection of new-onset bleeding, whereas if no personal baseline exists increasing sampling frequency of hemodynamic monitoring data improves bleeding detection earlier and with lower false positive rate.

Key Words: animal model; hemodynamic monitoring; hemorrhage; machine learning; predictive models

Hemodynamic monitoring is done to identify cardiovascular insufficiency and its progression. One of the most common and potentially fatal causes of cardiovascular decompensation is hypovolemia, which accounts for 50% of deaths within the first 24 hours of hospital admission, mainly in trauma and other surgical patients (1). Overt clinical signs of progressive occult hypovolemia can be obscure. Yet delayed diagnosis of hemorrhage is associated with poor outcomes (2, 3).

Several factors in monitored hospital settings contribute to the difficulty of identifying hypovolemia early. First, normal autonomic reflexes tend to mitigate initial volume loss through a variety of inter-related mechanisms, including increased interstitial and cellular fluid translocation into the intravascular space

[1]Auton Lab, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

[2]Cardiopulmonary Research Laboratory, Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA.

Supplemental digital content is available for this article. Direct URL citations appear in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

Supported, in part, by National Institutes of Health grants from HL07820 and GM126811.

The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: pinsky@pitt.edu

Crit Care Expl 2019;1:e0058

DOI: 10.1097/CCE.0000000000000058

and increasing sympathetic tone which sustain effective circulating blood volume even though total circulating blood volume decreases. Dynamic metrics of autonomic tone and venous return like heart rate variability (HRV), pulse pressure variation (PPV), or stroke volume variation (SVV) are more sensitive at identifying deterioration earlier during this compensated phase (4, 5). With further volume loss, cardiac output decreases but arterial pressure remains constant by increasing arterial tone, although measures of tissue perfusion like urine output and mixed venous oxygenation decrease. Second, not all patients are monitored to the same intensity to detect early cardiovascular changes when appropriate treatment should be more effective. Third, patients respond differently to progressive hypovolemia based on their cardiovascular reserve, comorbidities, and coexisting therapies. Finally, in many patients, an initial stable reference period is unavailable to identify instability in personalized relative terms. For example, no baseline data are available for patients presenting in a decompensated state, whereas routine surgical patients and other hospitalized patients will likely have baseline stable data.

Clinically, hemodynamic monitoring usually is linked to a threshold-based alarm system, wherein alerts are made once measured or derived physiologic values exceed some predefined thresholds. These systems tend to create high numbers of false alerts leading to alarm fatigue and cognitive distress (6). We recently demonstrated that a personalized smart alarm system can screen out most artifacts in real time, minimizing artifact-based false alerts (7, 8). Furthermore, applying similar analysis to data cleaned of artifact, we demonstrated that analysis of patient-specific physiologic data using machine learning approaches accurately identified patients' probability for future deterioration (9). Scoring systems like the Modified Early Warning Score (10), and manual centile-based early warning systems (11) perform well, but they are not patient-specific and are labor-intensive in application. Using similar trajectories from other subjects to predict group response of an individual has also be used to varying degrees of success (8). Although we have shown that fused variables derived from physiologic time series data can identify cardiorespiratory insufficiency in step-down unit patients in real time (12) such approaches have not been applied in more acute settings.

We hypothesized that increasing hemodynamic time-series data sampling frequency to allow richer discrimination of derived features, referred to as featurization, and also referencing current subject's data to their stable baseline would decrease bleeding detection latency for the same false positive rate (FPR), compared to less featurized analyses and those without baselines available. To examine this hypothesis, we used our established porcine model of severe hemorrhagic shock where bleeding rate is fixed such that time from onset directly reflects blood lost, although animals exposed to this challenge display a highly variable course from baseline to hypotension (13).

We therefore compared individual single hemodynamic variable thresholds as presently used clinically do create alerts (raw vital sign data) with minimal featurization to progressively increasing data sampling and featurization capacity as low frequency mean values measured every 2 minutes, reflecting the best possible modeling from existing clinical databases of most healthcare systems, beat-to-beat (B2B) data as incorporating the dynamic changes of functional hemodynamic monitoring principles, and continuous high-frequency waveform as the most advanced form of data featurization possible to quantify timeliness of bleed onset detection alerts and FPR with and without baseline reference.

METHODS

Experimental Model and Data Collection

The study protocol was approved by the University of Pittsburgh Institutional Animal Care and Use Committee and previously described in detail (13). Briefly, 51 female Yorkshire pigs (wt. 21.3 ± 0.65 kg) were anesthetized (ketamine, xylazine, and telazol intramuscular) for induction, intubated and ventilated (8 mL/kg tidal volume, F_{IO_2} 0.4, 3 cm H_2O positive end-expiratory pressure) on maintenance anesthesia (2.0–3.0% isoflurane). 0.9% NaCl solution was infused at 1 mL/kg/hr IV prior to the study. A pulmonary artery catheter (PAC) (Vigilance catheter; Edwards LifeSciences, Irvine, CA) was inserted via internal jugular vein, a triple lumen 18-gauge catheter into a femoral artery and a large-bore introducer (8F) in the femoral vein. The arterial pressure signal was also simultaneously recorded on a LiDCOplus monitor (LiDCO, London, United Kingdom). Triplicate bolus thermodilution cardiac output was used to calibrate the PAC continuous cardiac output and LiDCO monitors.

Following surgery, the swine were rested 30 minutes without further manipulation to establish a baseline. Then, 38 pigs were bled using a roller pump (Masterflex L/S easy-load II pump; Cole-Parmer, Vernon Hills, IL) at 20 $mL \cdot min^{-1}$. Additional 13 pigs were bled at 5 mL min^{-1}. The initial bleeding was continued until the mean arterial pressure (MAP) decreased to 30 mm Hg for the 20 $mL \cdot min^{-1}$ bleed and 40 mm Hg for the 5 $mL \cdot min^{-1}$ bleed animals. Data were collected during stabilization periods to obtain baselines and until the end of the bleed.

Data Acquisition

Electrocardiogram and continuous pressure waveforms were collected at 250 Hz, including arterial, central venous, pulmonary arterial pressures, arterial oxygen saturation (measured by pulse oximetry) and its plethysmographic waveform, and electrocardiography (ECG) tracings to define R-R intervals, continuous cardiac output (LiDCO), and mixed venous oxygen saturation. We used an automatic artifact filter that excluded step changes in pressures associated with closing and opening the vascular pressure transducer stopcock to perform blood draws but otherwise used all data. Derived data from the waveforms included systolic and diastolic arterial and pulmonary arterial pressures, MAP, and mean pulmonary artery pressures. Twelve additional variables were obtained on a B2B basis from LiDCOplus monitor including stroke volume, PPV, and SVV monitors. HRV variables were calculated from the R-R interval data.

Feature Groups by Data Density

To assess the impact of data sampling frequency on bleeding onset detection, we trained models on three featurized datasets: simple hemodynamic metrics (SM) with minimal featurization sampled every 1-minute, B2B, and waveform, conveying increasing

richness of data features, with each higher sampling frequency set including features obtained from lower sampling frequency data. Unfeaturized raw vital signals (Raw) were used as univariate predictors for the 20 mL·min^{-1} cohort.

Personalized Baseline Normalization or Universal Normalization

Each feature set was standardized using its mean and variance computed either from a 10-minute period in each subject's stable baseline period (personalized) or from the collective baseline periods of all subjects in the training set (universal). For the personalized baseline, these normalized features were used in addition to the nonnormalized features in model training.

Evaluation Methods

Each data sampling frequency set contains several time series (SM = 13, B2B = 24, waveform = 177), reflecting increasingly detailed featurization. We used machine learning methodology to automatically define optimal bleeding alert criteria separately for each group. Features were computed on rolling time windows and updated at 2 Hz. Classifiers were trained separately for each feature set configuration to estimate instantaneous bleed probability. Although we evaluated several feasible machine learning algorithms, to focus our discussion on the impact of data sampling frequency, we report only results from random forest classifiers because they systematically matched or outperformed the alternatives, consistent with our prior data (14). The model predictions were evaluated using a leave-one-pig-out cross-validation procedure. We also computed two metrics to identify the 10 most important physiologic features in each model: data point count (frequency of usage of a feature in evaluating the test set observations) and maximum information gain (a feature's best observed performance in separating bleeding from non-bleeding).

Model performance was evaluated using two linked approaches: receiver operating characteristic (ROC) curves plotting either true positive rate (TPR) as a function of FPR, or true negative rate (TNR) as a function of false negative rate (FNR); and activity monitoring operating characteristic (AMOC) curves plotting FPR as a function of average time to detection (TTD) measured from the onset of bleeding given a specific minimum detection rate (TPR).

Area under the ROC curve variable reflects the overall capability of a model to separate the two classes, but it does not provide insight on the operationally optimal settings of its sensitivity. In most clinical scenarios, it is of interest to either maximize detectability of positive cases (i.e., bleeding) at the lowest possible false alert rates, or to maximize identifiability of true negatives (non-bleeding, TNR) at the lowest possible miss rates (FNR), or both. To focus on these operationally relevant ranges, we present the ROC diagrams in two views: FPR plotted as function of TPR, and TNR as a function of FNR, both with logarithmically scaled horizontal axes to emphasize the low error ranges. Each pair of such diagrams plots the same data, but they focus on the opposite ends of model settings.

Just as ROC diagrams show trade-offs between event detectability and error rates, AMOC curves depict trade-offs between latency of detection and the frequency of alerts generated by the system. Similar to ROC, AMOC is produced by varying the sensitivity threshold of detection. Typically, the more sensitive the models, the earlier we can detect bleeding, but this comes at the cost of triggering more detection alerts, some of which may be false. Conversely, a less sensitive model will produce fewer alerts, but will likely take more time to report bleeding. The operationally optimal setting for a specific clinical environment can be established from an AMOC plot if the relative costs of handling false alerts and benefits from faster detection of bleeding can be estimated. We report AMOC curves for the average TTD at 80% detection thresholds.

Impact of Noise on Model Performance

To study the impact of the noise associated with unstable baselines prior to bleed, transient intervals of instability and missing data, we reviewed all data records and hand-selected stable baseline intervals, excluding animals with no clear stable interval during baseline or more than two episodes of transient (< 2 min each) instability during bleed, to create a "clean" cohort of pigs for model building to compare it to the full experimental data group. This resulted in an exclusion of eleven 20 mL·min^{-1} bleed pigs for this clean group analysis. These cleaned data were then used to assess the ability of primary hemodynamic data thresholds to identify bleeding onset.

RESULTS

Impact of Data Sampling Frequency and Personalized Normalization

By AMOC curve analysis for 5 and 20 mL·min^{-1} bleed group (Figs. 1 and 2, respectively) with personalized and universal baseline, greater personalized feature richness from higher sampling frequency data yielded better performance (i.e., same or lower TTD) for the same FPR as compared with models using features at lower sampling frequencies. This is generally the case for operationally relevant low error rates (e.g., FPR $\leq 10^{-3}$) underscoring how such model-based alerts would greatly minimize potential alarm fatigue. This difference is not seen in the universal normalized 5 mL·min^{-1} models. When minimal discriminative differences exist for group normalized data, they tend to be with B2B and waveform compared with SM models. An example of four animals AMOC is shown in **Figure E3** (Supplemental Digital Content 1, http://links.lww.com/CCX/A108; **legend**, Supplemental Digital Content 2, http://links.lww.com/CCX/A109).

The logarithmically scaled ROC curves (Figs. 3 and 4) reveal that waveform features provide a significant boost in the detection of negatives at very low FNRs for the normalized models. There appears to be some advantage to waveform models in detection of negatives without personalized normalization, but this benefit is not large. An example of model performance in three representative animals is displayed in **Figure E1** (Supplemental Digital Content 1, http://links.lww.com/CCX/A108; legend, Supplemental Digital Content 2, http://links.lww.com/CCX/A109).

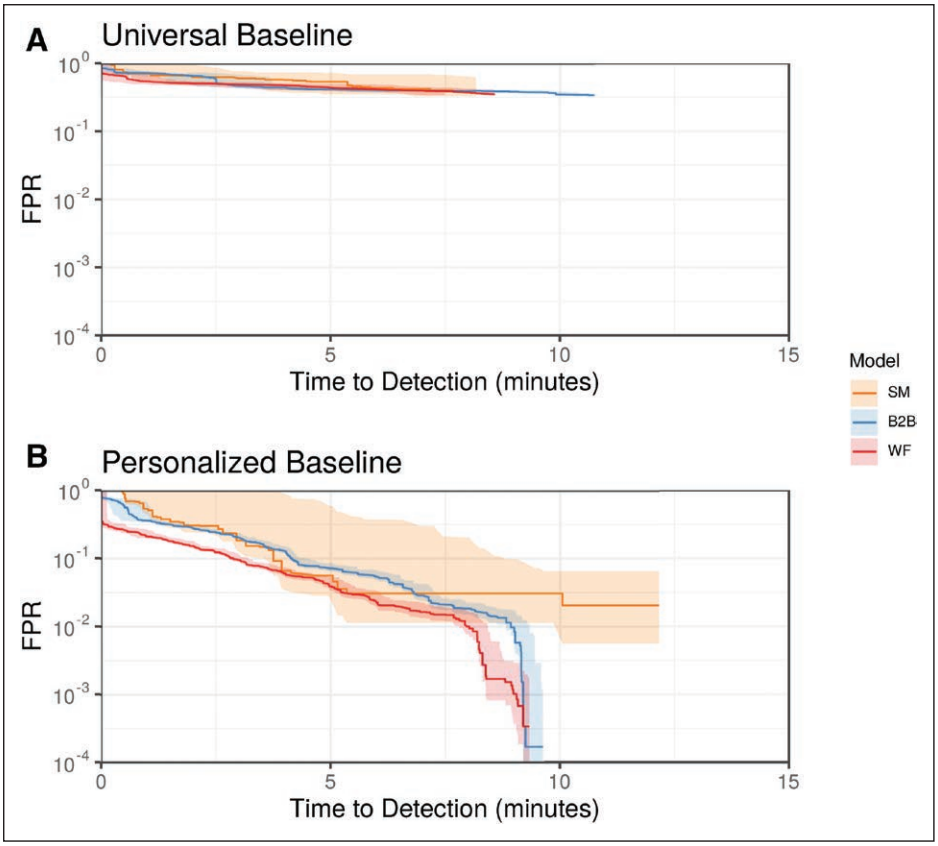

Figure 1. Activity monitoring operating characteristic curves showing the average time to detection needed to detect the onset of 80% of animals for the 5 mL·min^{-1} cohort for simple metrics (SM), beat-to-beat (B2B), and waveform (WF) using a universal baseline (**A**), or personalized baseline (**B**). FPR = false positive rate.

Comparison Between 5 and 20 mL·min^{-1} Bleed Groups

The nonnormalized models trained on the 20 mL·min^{-1} bleed group have an earlier TTD for the same FPR than models trained on 5 mL·min^{-1} (**Fig. E2**, Supplemental Digital Content 1, http://links.lww.com/CCX/A108; legend, Supplemental Digital Content 2, http://links.lww.com/CCX/A109). Interestingly, the detection latencies for the two bleed rates are comparable for personalized normalized models. When detection latency is evaluated on the volume of blood lost (Fig. 5), bleeding can be detected at lower loss for 5 than 20 mL·min^{-1} pigs, presumably because adaptive changes in response to bleeding have more time to manifest and be noticed at the slower bleed rate.

Identifying Relevant Features of Each Model

Tables E1–E4 (Supplemental Digital Content 1, http://links.lww.com/CCX/A108) display the most important measured and derived variables for each model using 5 and 20 mL·min^{-1} bleed data, respectively, quantified as those providing the highest gain of information to separate bleed from non-bleed. We used two metrics to quantify specific features input into the final model's performance. "Maximum information gained" and "Weighted data-point count." In training a decision tree, information gain is used to pick the best feature to split on to minimize the total entropy in the resulting data slices generated by the split. Here we define it as the difference between the entropy before to after the split. Whereas weighted data-point count indicates the number of times, the feature was encounter in making a prediction weighted by the depth at which the data point was encountered. Many of the important features remained common across models and included metrics associated with pulmonary arterial, central venous and systemic arterial pressures, as well as mixed venous oxygen saturation and cardiac output, consistent with known physiology. However, their relevance in each model varied widely suggesting that different aspects of each variable were informative in different contexts.

Effect of Noise on Model Performance

AMOC curve analysis primary hemodynamic data of 38 pig dataset revealed poor bleed detection (Fig. E3, Supplemental Digital Content 1, http://links.lww.com/CCX/A108; legend, Supplemental Digital Content 2, http://links.lww.com/CCX/A109). However, when the pig data set was cleaned of those animals without a good baseline, personalized normalization of pulse pressure, systolic arterial pressure and SVV showed good performance (**Fig. E4**, Supplemental Digital Content 1, http://links.lww.com/CCX/A108; legend, Supplemental Digital Content 2, http://links.lww.com/CCX/A109).

DISCUSSION

Our study demonstrates that for the detection of hemorrhagic shock from hemodynamic monitoring data: 1) increasing data sampling frequency, by allowing for richer feature extraction, improves model performance; 2) referencing models to a prior personalized baseline greatly improves model performance compared with using universal baseline reference values; and 3) the best specific features used to predict bleeding are often different with the level of data sampling frequency.

Impact of Data Density

Detection latencies of bleeding onset in multivariate models of hemodynamic monitoring data streams are superior to those of univariate models and at very low error rates. Since the sparsely collected data of SM models and the more complex B2B and waveform models all draw from the same data streams, there is little foreseeable disadvantage and potentially substantial benefits in the data hierarchy approach to bleed detection. Thus, in patients who have the potential of becoming unstable quickly, more intensive sampling frequencies should be considered so clinically relevant instability events can be recognized more quickly. Furthermore, detectable physiologic markers may appear very soon after the onset of bleeding to provide fast, reliable detection, but higher frequency features are needed to make these detections

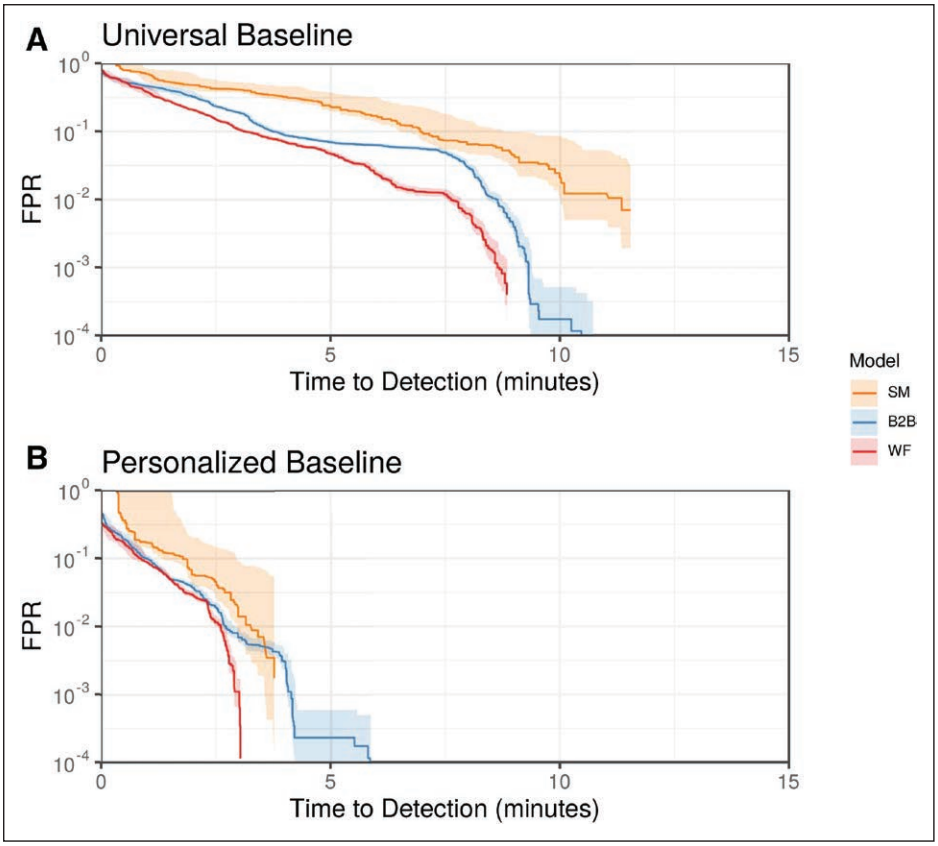

Figure 2. Activity monitoring operating characteristic curves showing the average (time to detection needed to detect the onset of 80% of animals for the 20 mL·min^{-1} cohort for simple metrics (SM), beat-to-beat (B2B), and waveform (WF) using a universal baseline (**A**), or personalized baseline (**B**). FPR = false positive rate.

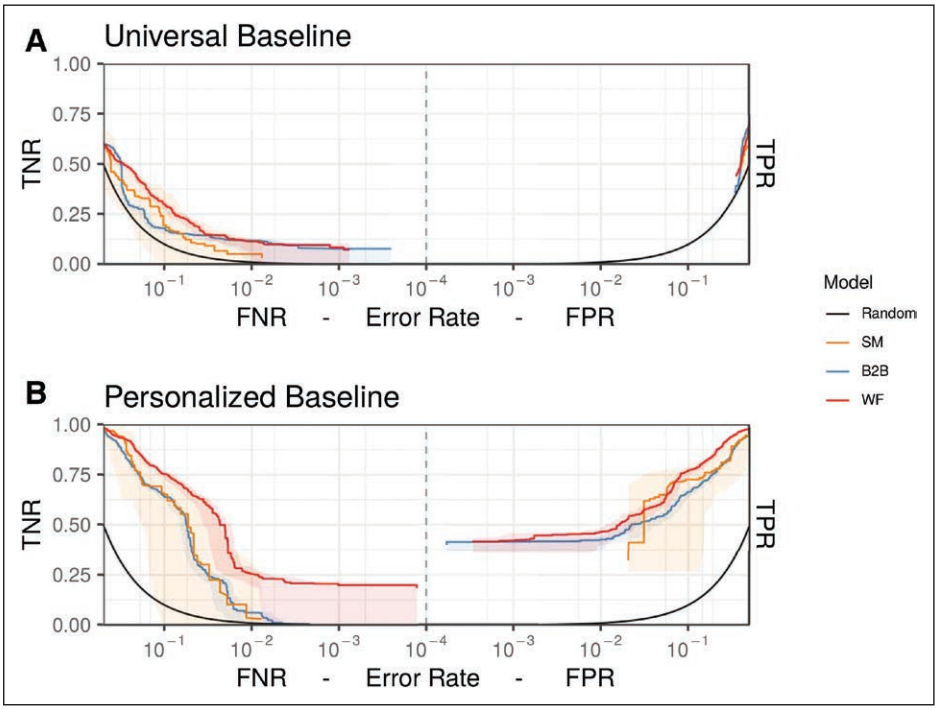

Figure 3. Receiver operating characteristic curves for the 5 mL·min^{-1} cohort for simple metrics (SM), beat-to-beat (B2B), and waveform (WF) models using a universal baseline (**A**), or personalized baseline (**B**). Plots show the tradeoff between detection performance (true positive rate [TPR] and true negative rate [TNR]) and error rate (false positive rate [FPR] and false negative rate [FNR]).

without excessive false alarms. We chose FPR markedly lower than considered clinically acceptable for a detection system to underscore how this approach minimizes potential alarm fatigue. Relaxing these conservative settings would yield faster detections at higher positive predictive values.

Data density affects the ability to create even simple physiologic relevant alert thresholds. Importantly, even the SM featurization models outperform primary hemodynamic data suggesting that using SM data to create models of instability is superior to existing simple threshold-based alerts. Interestingly, primary hemodynamic thresholds performed well if only cleaned data and clear baseline states were available, a set of conditions often not present in clinical scenarios. Our B2B model also included the functional hemodynamic monitoring HRV, PPV, and SVV features. HRV represents a collection of metrics that are a surrogate for autonomic function: a decrease in HRV signifies a loss of complex neurocardiac interactions seen in healthy individuals (15). Decreased HRV is associated with mortality in trauma patients and can predict outcomes up to 12 hours in advance and is superior to traditional vital sign threshold values in detecting hemodynamic decompensation (5). Liu et al (16) concluded that HRV-derived features were able to predict the need for life-saving interventions. But raw HR also proved useful in their models, presumably because significant ECG noise artifact may limit utility of HRV. PPV and SVV under mechanical ventilation identify volume responsiveness (4, 17–19). PPV and SVV-triggered resuscitation protocols reduce perioperative complications (20, 21). The PPV/SVV is a surrogate for dynamic arterial elastance and can predict MAP response to changes in vasopressor dose (22, 23).

We used machine learning methodology to plumb the physiologic signatures and tuned our models for optimal performance for both TNR and TPR for the lowest FNR and FPR. The resulting models depend on the quality of the physiologic data available. Clinicians may accept different tradeoffs of precision and accuracy. The levels of performance and utility of any predictive model will be patient, procedure, location and care specific, as there is no universal level of stability or instability that would be optimal

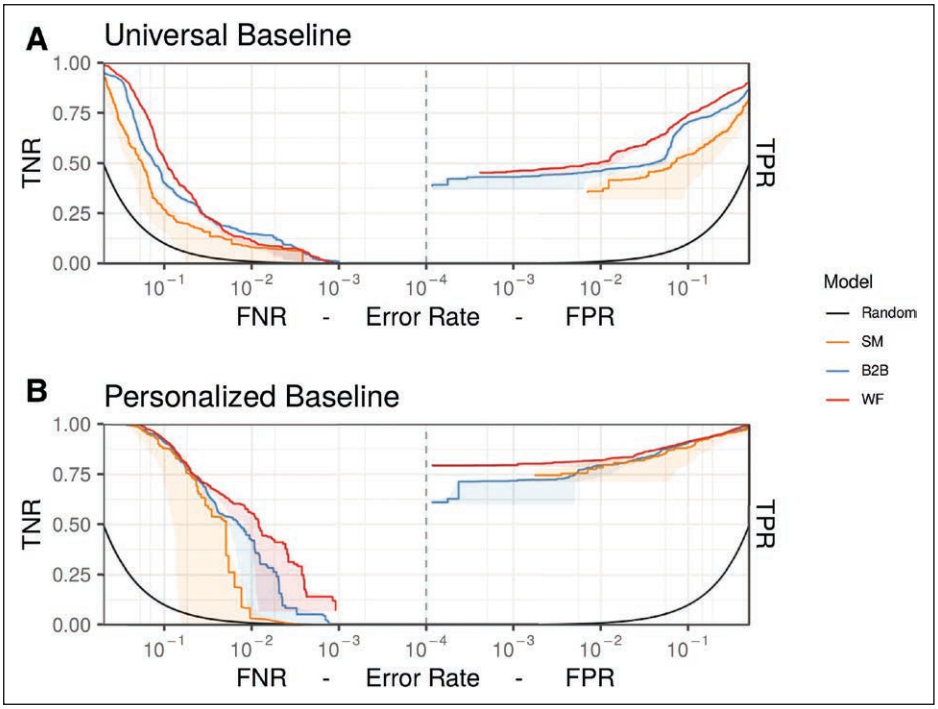

Figure 4. Receiver operating characteristics for the 20 mL·min^{-1} cohort for simple metrics (SM), beat-to-beat (B2B), and waveform (WF) models using a universal baseline (**A**), or personalized baseline (**B**) using the same format as Figure 5. FNR = false negative rate, FPR = false positive rate, TNR = true negative rate, TPR = true positive rate.

for all patients, care environments, and conditions. The consequences of a false positive alert may be minimal in the operating room when attention is already being paid to the patient, whereas it could be disruptive in settings of more remote care and limited resources as in the post-anesthesia or step-down unit. Similarly, false negatives can be disastrous in poorly monitored areas or with less skilled clinicians who might otherwise delay resuscitation until overt circulatory shock was present. We specifically chose models with markedly low FPR and FNR which needed to be associated with slightly prolonged time to bleeding detection. To the extent that excessive false alerts do not promote alarm fatigue, and early detection of hypovolemia is required, these models would still be operational: the operator would just set the detection threshold of a specific algorithm at a higher FPR level. However, our study provides a roadmap to create different models specific to the patient's individual monitoring needs and selecting the appropriate data sampling frequency, models, and their sensitivity to optimize desired performance variables (detection rates, error rates, timeliness of detection) for each specific scenario. In practice, such optimizations can be done on-the-fly and without re-training the models, unless the underlying distributions of data or patient populations vary substantially, as may be the case when transferring the approach to a distinct data environment, for example, in a different medical center.

In our experiments, we used a well-established machine learning methodology. Our selection of random forest classifiers was motivated by our prior success with these types of models, but we do not expect to see qualitatively different outcomes when using similarly established alternatives, such as k-nearest neighbors (24–26) or neural networks (16, 27), which had been used for bleed detection in the past.

The Value of Personalized Baselines

Personal baselined high-density data models predict the onset of bleeding with markedly lower latency and fewer false alarms than models based on lower density data and universal normalization. Although derived B2B variables like PPV and SVV can predict volume responsiveness, volume responsiveness is also present in healthy subjects (28), whereas increasing PPV and SVV suggest progressive hypovolemia. Baseline data may not be available in acute triage settings, in which case we can fall back on a universal baselined reference. However, elective surgical patients usually have preoperative baseline, as do hospitalized patients at increased risk of subsequent cardiovascular insufficiency. Importantly, featurizations used in this analysis did not provide much separation in the case of universally normalized data. Without personalized normalization, high frequency models are better at identification of negatives (prebleed state) with lower FNR (Figs. 3 and 4). This is expected because most of the variability seen in the predictions occurs when comparing individuals before bleeding. After bleeding starts, inter-subject variability decreases. Different featurizations or more sophisticated detection methodologies on universal normalized data might be able to reestablish the good performance and better separation seen with normalization but would require a more extensive patient data library of similar patient cohorts to mitigate detrimental effects of patient diversity. This hypothesis has profound clinical relevance and needs to be studied.

Interaction Between Data Sampling Frequency and Feature Performance

Analyzing data density requires consideration of at least two attributes: richness of featurization and evaluation interval (29). The density of the collected data impacts what kinds of features can be realistically derived. Computations of sample entropy, frequency decomposition, or even basic windowed statistics (e.g., rolling mean) are unrealistic at low frequency collection rates because data are coming in too sparsely to reflect rapid changes. The relative importance of features driving prediction models changes as one increases data sampling frequency (Tables E1–E4, Supplemental Digital Content 1, http://links.lww.com/CCX/A108), underscoring the concept that richer featurization of physiologic data reported on hemodynamic monitors improves instability detection. On the other hand, more complex feature spaces allow more complex models with elevated risk of overfitting, making thorough cross-validation a must. To mitigate overfitting, we used a leave-one-subject-out cross-validation, so that no animal's data was used to create the model with which they were tested. Beneficial net effects of more sophisticated featurization

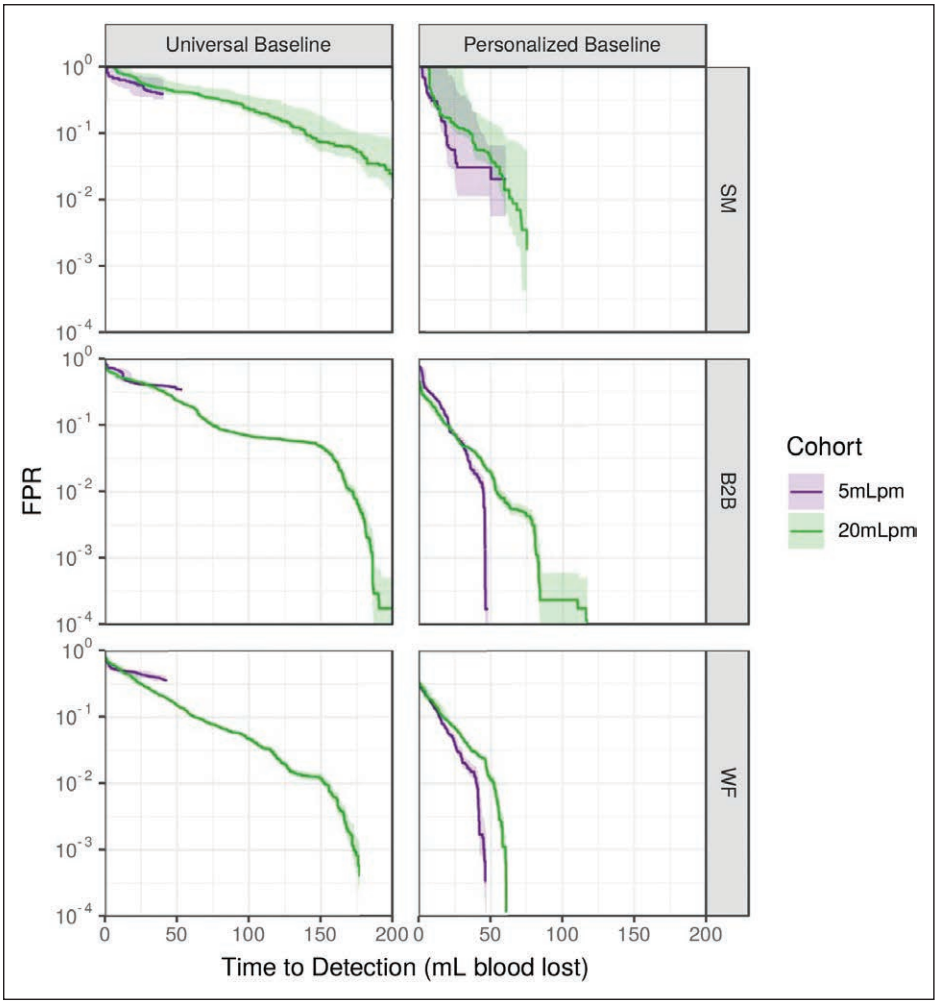

Figure 5. Activity monitoring operating characteristic curves comparison of the detection performance in terms of volume of blood lost before detection between the 5 and 20 mL·min^{-1} cohorts for the simple metrics (SM) (*top*), beat-to-beat (B2B) (*middle*), and waveform (WF) (*bottom*) models using either a universal baseline (*left*) or personalized baseline (*right*). FPR = false positive rate.

account for the observed differences between SM, B2B, and waveform models. The example of central venous pressure (CVP) is especially illustrative. Primary hemodynamic data, SM and B2B models found CVP values nondiscriminative, whereas with normalized waveform models CVP featurization became important, suggesting that although CVP is a poor marker of volume status even when featurized on a B2B basis, its waveform featurization allows finer discrimination of the interplay between CVP and the pressure gradient for venous return, the primary process altered by hemorrhage. The relative contribution of waveform pulmonary artery pressure featurization also underscores this concept. Although nonspecific use of PAC data as presently used does not provide earlier diagnostic benefit. This hypothesis too has profound clinical relevance and needs to be studied further.

The Downside of Waveform Data Collection

Less time between evaluations in waveform data means less time between potential alarms. The false alarm rate may not change, just the time between consecutive alarms. This trade-off must be weighed against the benefit of reduced latency in the TTD. Higher data density also carries a greater cost of storage, transmission, and processing, while collecting it may expose patients to more invasive procedures. These limitations may in practice favor the use of low frequency data despite the availability of devices that can collect information at higher levels of detail. Still, there is no reason high frequency data cannot be temporarily buffered and featurized at less frequent intervals to reduce data storage burden. Choosing the interval at which to evaluate a model (and store patient data) becomes an intriguing variable to understand.

Limitations

The study has several limitations. First, we performed it in healthy pigs under light general anesthesia. Thus, the predictive value of similar analyses in conscious patients with comorbidities remains undefined. Still, the concepts that data density and personalized baseline reference values matter remain valid. Second, we used a fixed bleeding rate. Bleeding is rarely constant in realistic scenarios. Still, the bleed rate analyzed were slow enough to mimic clinically relevant progressive hypovolemic states. Third, we included all hemodynamic monitoring data in our analysis to build predictive models. Patients are often monitored with a minimal number of concurrent devices. We did not analyze the relative benefits of each device or combinations on model performance. Clinically, monitored physiologic signals are often intermittent and subject to artifact. Although, a research animal laboratory reflects the optimal data collection environment, like a cardiac catheterization laboratory, these models must degrade as the validity of the data degrades in more operationally realistic settings. As mentioned in the introduction, we recently demonstrated that we could filter over 80% of all physiologic vital sign artifact using machine learning algorithms (7), but thorough sensitivity analysis needs to be conducted in each application. Finally, we did not parse out our analysis by specific monitoring devices but included potentially all routine monitoring currently available. It will be interesting to see if limited datasets from one or lesser number of monitoring devices would perform similarly. That analysis remains to be done.

SUMMARY

Bleed detection in a porcine model of either 5 or 20 mL·min^{-1} blood loss demonstrated that the greater the data density and its associated richer featurization the better the ability to identify bleeding earlier with low FPR. Also, if personalized baseline normalization is available, detection models can perform markedly

better. Without such a personalized baseline higher data density is needed to improve model performance. Finally, the features useful in models of varying density may be different even though the same sources of physiologic data are used in all.

REFERENCES

1. Kauvar DS, Lefering R, Wade CE: Impact of hemorrhage on trauma outcome: An overview of epidemiology, clinical presentations, and therapeutic considerations. *J Trauma* 2006; 60:S3–S11
2. Calzavacca P, Licari E, Tee A, et al: A prospective study of factors influencing the outcome of patients after a medical emergency team review. *Intensive Care Med* 2008; 34:2112–2116
3. McNicholl BP. The golden hour and prehospital trauma care. *Injury* 1994; 25:251–254
4. Yi L, Liu Z, Qiao L, et al: Does stroke volume variation predict fluid responsiveness in children: A systematic review and meta-analysis. *Plos One* 2017; 12:e0177590
5. Norris PR, Morris JA Jr, Ozdas A, et al: Heart rate variability predicts trauma patient outcome as early as 12 h: Implications for military and civilian triage. *J Surg Res* 2005; 129:122–128
6. Ruskin KJ, Hueske-Kraus D: Alarm fatigue: Impacts on patient safety. *Curr Opin Anaesthesiol* 2015; 28:685–690
7. Chen L, Dubrawski A, Wang D, et al: Using supervised machine learning to classify real alerts and artifact in online multisignal vital sign monitoring data. *Crit Care Med* 2016; 44:e456–e463
8. Ebadollahi S, Sun J, Gotz D, et al: Predicting patient's trajectory of physiological data using temporal trends in similar patients: A system for near-term prognostics. *AMIA Annu Symp Proc* 2010; 2010:192–196
9. Chen L, Ogundele O, Clermont G, et al: Dynamic and personalized risk forecast in step-down units. Implications for monitoring paradigms. *Ann Am Thorac Soc* 2017; 14:384–391
10. Loftus TJ, Mira JC, Ozrazgat-Baslanti T, et al: Sepsis and critical illness research center investigators: Protocols and standard operating procedures for a prospective cohort study of sepsis in critically ill surgical patients. *BMJ Open* 2017; 7:e015136
11. Watkinson PJ, Pimentel MAF, Clifton DA, et al: Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. *Resuscitation* 2018; 129:55–60
12. Hravnak M, Devita MA, Clontz A, et al: Cardiorespiratory instability before and after implementing an integrated monitoring system. *Crit Care Med* 2011; 39:65–72
13. Gómez H, Mesquida J, Hermus L, et al: Physiologic responses to severe hemorrhagic shock and the genesis of cardiovascular collapse: Can irreversibility be anticipated? *J Surg Trauma* 2012; 178:358–369
14. Hravnak M, Chen L, Dubrawski A, et al: Real alerts and artifact classification in archived multi-signal vital sign monitoring data: Implications for mining big data. *J Clin Monit Comput* 2016; 30:875–888
15. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J* 1996; 17:354–381
16. Liu NT, Holcomb JB, Wade CE, et al: Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock* 2014; 42:108–114
17. Michard F, Boussat S, Chemla D, et al: Relation between respiratory changes in arterial pulse pressure and fluid responsiveness in septic patients with acute circulatory failure. *Am J Respir Crit Care Med* 2000; 162:134–138
18. Slagt C, Malagon I, Groeneveld AB: Systematic review of uncalibrated arterial pressure waveform analysis to determine cardiac output and stroke volume variation. *Br J Anaesth* 2014; 112:626–637
19. Marik PE, Cavallazzi R, Vasu T, et al: Dynamic changes in arterial waveform derived variables and fluid responsiveness in mechanically ventilated patients: A systematic review of the literature. *Crit Care Med* 2009; 37:2642–2647
20. Benes J, Chytra I, Altmann P, et al: Intraoperative fluid optimization using stroke volume variation in high risk surgical patients: Results of prospective randomized study. *Crit Care* 2010; 14:R118
21. Salzwedel C, Puig J, Carstens A, et al: Perioperative goal-directed hemodynamic therapy based on radial arterial pulse pressure variation and continuous cardiac index trending reduces postoperative complications after major abdominal surgery: A multi-center, prospective, randomized study. *Crit Care* 2013; 17:R191
22. Monge MI, Cano AG, Gracia M: Dynamic arterial elastance to predict arterial pressure response to volume loading in preload-dependent patients. *Crit Care* 2011; 15:R15
23. Guinot PG, Bernard E, Levrard M, et al: Dynamic arterial elastance predicts mean arterial pressure decrease associated with decreasing norepinephrine dosage in septic shock. *Crit Care* 2015; 19:14
24. Chen L, Gribok A, Reisner AT, et al: Exploiting the existence of temporal heart-rate patterns for the detection of trauma-induced hemorrhage. *Conf Proc IEEE Eng Med Biol Soc* 2008; 2008:2865–2868
25. Glass TF, Knapp J, Amburn P, et al: Use of artificial intelligence to identify cardiovascular compromise in a model of hemorrhagic shock. *Crit Care Med* 2004; 32:450–456
26. Lu K, Shoemaker WC, Wo CC, et al: A mathematical program to predict survival and to support initial therapeutic decisions for trauma patients with long-bone and pelvic fractures. *Injury* 2007; 38:318–328
27. Roberts DA, Holcomb JB, Parker BE Jr, et al: The use of polynomial neural networks for mortality prediction in uncontrolled venous and arterial hemorrhage. *J Trauma* 2002; 52:130–135
28. Godfrey GE, Dubrey SW, Handy JM: A prospective observational study of stroke volume responsiveness to a passive leg raise manoeuvre in healthy non-starved volunteers as assessed by transthoracic echocardiography. *Anaesthesia* 2014; 69:306–313
29. Pinsky MR, Dubrawski A: Gleaning knowledge from data in the intensive care unit. *Am J Respir Crit Care Med* 2014; 190:606–610