Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis

Feras Hatib, Ph.D., Zhongping Jian, Ph.D., Sai Buddi, Ph.D., Christine Lee, M.S., Jos Settels, M.S., Karen Sibert, M.D., F.A.S.A., Joseph Rinehart, M.D., Maxime Cannesson, M.D., Ph.D.

ABSTRACT

Background: With appropriate algorithms, computers can learn to detect patterns and associations in large data sets. The authors' goal was to apply machine learning to arterial pressure waveforms and create an algorithm to predict hypotension. The algorithm detects early alteration in waveforms that can herald the weakening of cardiovascular compensatory mechanisms affecting preload, afterload, and contractility.

Methods: The algorithm was developed with two different data sources: (1) a retrospective cohort, used for training, consisting of 1,334 patients' records with 545,959 min of arterial waveform recording and 25,461 episodes of hypotension; and (2) a prospective, local hospital cohort used for external validation, consisting of 204 patients' records with 33,236 min of arterial waveform recording and 1,923 episodes of hypotension. The algorithm relates a large set of features calculated from the high-fidelity arterial pressure waveform to the prediction of an upcoming hypotensive event (mean arterial pressure < 65 mmHg). Receiver-operating characteristic curve analysis evaluated the algorithm's success in predicting hypotension, defined as mean arterial pressure less than 65 mmHg.

Results: Using 3,022 individual features per cardiac cycle, the algorithm predicted arterial hypotension with a sensitivity and specificity of 88% (85 to 90%) and 87% (85 to 90%) 15 min before a hypotensive event (area under the curve, 0.95 [0.94 to 0.95]); 89% (87 to 91%) and 90% (87 to 92%) 10 min before (area under the curve, 0.95 [0.95 to 0.96]); 92% (90 to 94%) and 92% (90 to 94%) 5 min before (area under the curve, 0.97 [0.97 to 0.98]).

Conclusions: The results demonstrate that a machine-learning algorithm can be trained, with large data sets of high-fidelity arterial waveforms, to predict hypotension in surgical patients' records. (ANESTHESIOLOGY 2018; 129:663-74)

PREDICTION of adverse events, from tornadoes to tsunamis, makes life-saving advance preparation possible. Yet in the operating room or the intensive care unit, clinicians often must manage the onset of arterial hypotension with essentially no warning. Hypotension during surgery, defined as mean arterial pressure (MAP) less than 65 mmHg,[1] is associated with increased rates of postoperative myocardial infarction[2] and acute kidney injury,[3] both predictors of poor long-term patient outcome.[4,5] In the intensive care unit setting, hypotension has been linked to an increased incidence of acute kidney injury.[6] The risk of serious complications increases with the duration of hypotension, but it can begin to develop within only a few minutes.[3] Advance warning that hypotension is imminent, even if the warning comes only 10 to 15 min ahead, could facilitate diagnostic and therapeutic measures to lessen the clinical impact.

Machine learning—a discipline within computer science used to analyze large data sets and develop predictive models—has evident applications to health care.[7–10] In the intensive care unit and operating room settings, physiologic waveforms represent a major source of information.[11,12]

Editor's Perspective

What We Already Know about This Topic

- The ability to predict intraoperative hypotension may advance the ability to prevent hypotension-associated complications effectively
- The extent to which advanced waveform analysis of invasive arterial lines may provide meaningful forewarning remains unknown

What This Article Tells Us That Is New

- A machine-learning algorithm based on thousands of arterial waveform features can identify an intraoperative hypotensive event 15 min before its occurrence with a sensitivity of 88% and specificity of 87%
- Further studies must evaluate the real-time value of such algorithms in a broader set of clinical conditions and patients

Typically, clinical monitors analyze physiologic waveforms to extract and display data that clinicians use to make decisions.[13,14] In 2009, an open challenge from PhysioNet and Computers in Cardiology prompted participants to develop tools to forecast acute hypotensive episodes, and

---

This article is featured in "This Month in Anesthesiology," page 1A. Corresponding articles on pages 619, 649, and 675. Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text on the Journal's Web site (www.anesthesiology.org). This article has a visual abstract available in the online version.

10 different approaches were presented.[15] Most of these techniques were based on the analysis of static or absolute measures obtained from arterial pressure waveforms. However, recent studies have suggested that the prodromal stage of hemodynamic instability is characterized by subtle, complex changes in different physiologic variables. These changes reflect altered compensatory mechanisms resulting in unique dynamic signatures in arterial waveforms.[16,17] Although the overt clinical signs of hypotension occur late, dynamic changes in the variability, complexity, and physiologic associations of features in the arterial pressure waveform can herald the occurrence of hypotensive events.[17] Recently, machine learning and complex feature extraction techniques have been proposed to make use of this subtle information contained in arterial waveforms in a way that was previously impossible.[16–21]

In this study, machine learning was used to fine-tune an algorithm, the Hypotension Prediction Index, based on the complex analysis of features in high-fidelity arterial pressure waveform recordings. The algorithm was developed to observe subtle signs that could predict the onset of hypotension in surgical and intensive care unit patients, and subsequent analysis validated the performance of the algorithm in two unique data sets.

Materials and Methods

Databases

This manuscript follows the "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View."[22] Data used in the development and internal validation came from the Multiparameter Intelligent Monitoring in Intensive Care II[11,23] database (specifically the Multiparameter Intelligent Monitoring in Intensive Care II Waveform Database Matched Subset; collaboration between Massachusetts Institute of Technology, Boston, Massachusetts, and Beth Israel Medical Center in Boston, Massachusetts),[24] a freely accessible critical care database, and from an Edwards Lifesciences database (Edwards Lifesciences, USA) of operating room and intensive care unit patients. The data for external validation came from the University of California at Irvine Medical Center (Irvine, California) as part of an ongoing data collection effort (fig. 1). In all cases, the data consisted of Health Insurance Portability and Accountability Act–compliant patient demographic information and arterial pressure waveforms (sampling rates, 100 to 500 Hz). For patients from all data sets whose data were collected during surgery, waveform data between anesthesia induction and tracheal extubation were used. For intensive care unit patients from the Edwards database, waveform data between 24 and 72 h postarterial line placement were used. For intensive care unit patients from the Multiparameter Intelligent Monitoring in Intensive Care II database, waveform data from admission until arterial line withdrawal were used.

The Multiparameter Intelligent Monitoring in Intensive Care II Waveform Database Matched Subset database contains 4,897 waveform records and 5,266 numeric records matched with 2,809 Multiparameter Intelligent Monitoring in Intensive Care II clinical database records.[12,25] From the 2,809 patient records, 2,100 waveform and numeric records were randomly selected and processed by the Multiparameter Intelligent Monitoring in Intensive Care II Waveform Database software package for MATLAB (version R2014a, The Mathworks Inc., USA). Of the 2,100 processed records, 326 included invasive arterial blood pressure recordings, and those
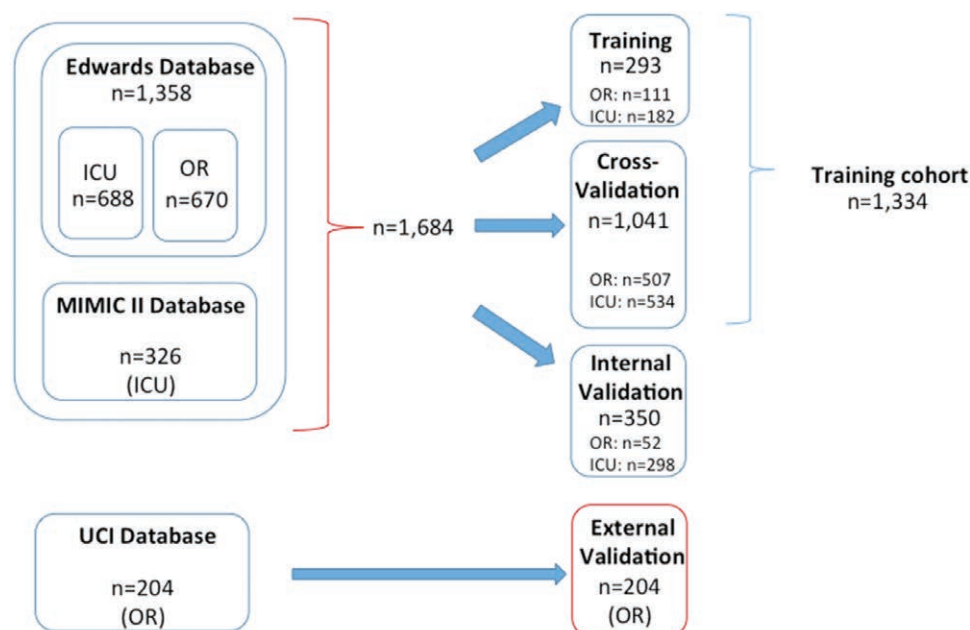
Fig. 1. Patient flow diagram. ICU, intensive care unit; MIMIC II, Multiparameter Intelligent Monitoring in Intensive Care II database; OR, operating room; UCI, University of California Irvine.
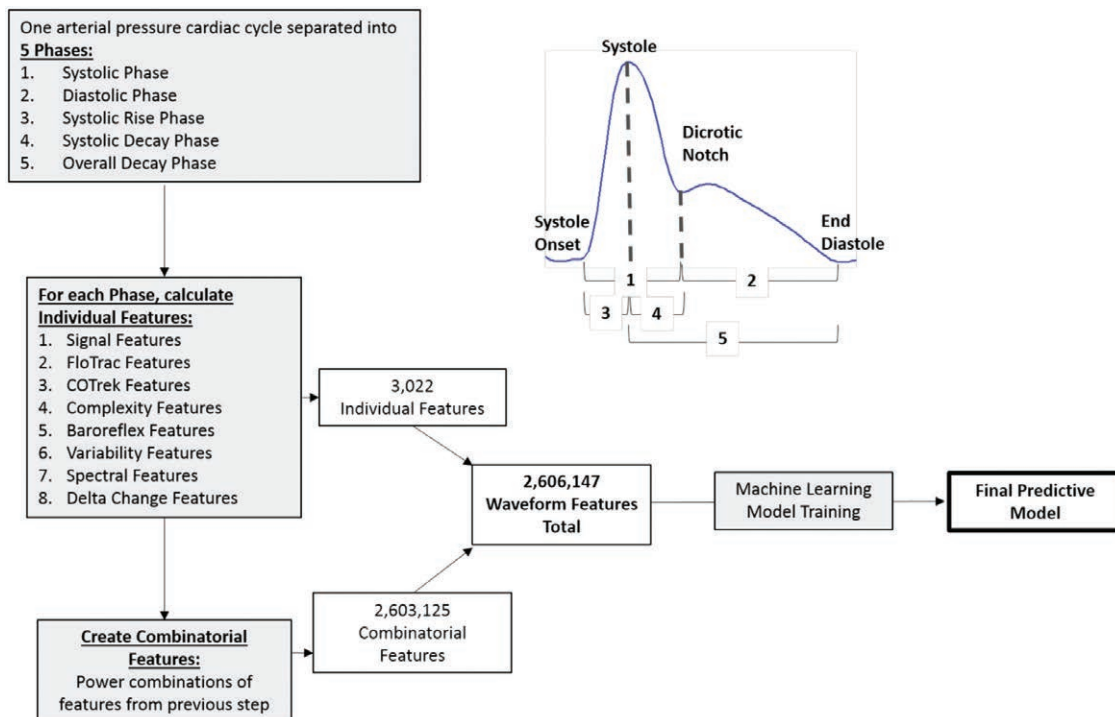
Fig. 2. Summarizing arterial pressure waveform featurization and model training. Higher-level overview of the predictive model development.

patient records were used in our analysis. The Edwards database contains 1,358 records of operating room and intensive care unit patients with invasive arterial pressure waveform recordings. These data sets were collected from 35 sites worldwide between 2005 and 2014, and all data were deidentified according to Health Insurance Portability and Accountability Act protocols. Ethics review and institutional review board exemptions were obtained with Quorum: Seattle Board, North American Board, and Daily Board (Quorum, USA) coordinating 12 Institutional Review Boards (United States), 15 Ethics Committees (European Union), and 1 Ethics Board (Canada). Before model training, 350 patient records were randomly selected from the Edwards and Multiparameter Intelligent Monitoring in Intensive Care II databases and set aside for the model testing and internal validation cohort. The remaining 1,334 records were used for the training cohort. This cohort was randomly divided between 293 patient records used for model training and 1,041 records used for cross-validation to adjust the model.

The University of California at Irvine database, accessed for external validation of the algorithm, contains records of surgical patients more than 18 yr old, collected prospectively from December 2015 to January 2017 at University of California at Irvine Medical Center as part of an Institutional Review Board–approved data collection study (HS#2011-1924). Waveform data were collected with a data integration system (Bernoulli, Cardiopulmonary Corp., USA). The external validation cohort consisted of 204 University of California at Irvine patients (fig. 1) whose arterial waveform records were included after informed, written consent was obtained.

Development of the Model

Our new algorithm is based on a machine-learning model for classification (see also Supplemental Digital Content, http://links.lww.com/ALN/B732). This model relates a large set of features calculated from the arterial pressure waveform to the prediction of an upcoming hypotensive event. In the learning phase, available clinical databases containing arterial pressure waveforms were first prepared for training. Periods of hypotension and nonhypotension were annotated in the databases to serve as the training data set. The arterial pressure waveforms were then processed to extract waveform features. These features were mapped for prediction of hypotensive events with the training data. Figure 2 shows a higher-level overview of the predictive model development. The key steps in development of the algorithm are summarized as follows:

1. Data conditioning, including signal preprocessing, heartbeat detection, and data selection;
2. Featurization of the arterial pressure waveform (extraction of key features or signatures);
3. Annotation of the training data set for periods of hypotension and nonhypotension;
4. Model training.

The development steps are explained briefly below and described in detail in the Supplemental Digital Content (http://links.lww.com/ALN/B732).

Data Conditioning: Signal Preprocessing, Heartbeat Detection, and Data Selection. All arterial pressure waveforms data were downsampled to 100 Hz. The arterial pressure waveforms were then processed through the Edwards (FloTrac,

Edwards Lifesciences) algorithm, as explained in detail in the Supplemental Digital Content (http://links.lww.com/ALN/B732).

Featurization of the Arterial Pressure Waveform (Feature Extraction).

The arterial pressure waveform was first divided into unique beats and then separated into five phases (fig. 2 and Supplemental Digital Content, http://links.lww.com/ALN/B732), which paved the way for calculating the hemodynamic parameters used as model features:

1. Arterial pressure waveform time, amplitude, area, and slope features;
2. FloTrac algorithm features;
3. CO-Trek features;
4. Complexity features;
5. Baroreflex features;
6. Variability features;
7. Spectral features;
8. "Delta-change" features;
9. Combinatorial features.

Briefly, the Edwards FloTrac algorithm computes key hemodynamic parameters, such as cardiac output, stroke volume, vascular tone (the Kai-factor[26]), Windkessel compliance,[27] systemic vascular resistance, stroke volume variation,[28] and several measures of the morphology of the arterial pressure waveform.[26] The Edwards CO-Trek algorithm is the pulse-contour cardiac output algorithm obtained from the ClearSight noninvasive arterial pressure monitoring system (ClearSight, Edwards Lifesciences, formerly Nexfin, Bmeye BV, The Netherlands). Please refer to the Supplemental Digital Content (http://links.lww.com/ALN/B732) for further detail.

After feature extraction of 3,022 individual features, base features were determined by performing receiver-operating characteristic analysis for each of the individual features on the positive and negative data segments of the training data set (see section below on "Model Feature Selection and Training" for the definitions of positive and negative data segments). The 51 individual features with an area under the receiver-operating characteristic curve greater than 0.85 were selected as the base features. All permutations of the 51 base features were computed with either one, two, or a maximum of three features at a time, and power levels in the range of [−2, −1, 0, 1, 2]. The permutation process generated a total of 2.6 million combinatorial features, which we created to ensure features that captured linear and nonlinear interactions. The exact number of combinatorial features obtained this way was

$$\binom{51}{3} \times 5^3 = \frac{51(51-1)(51-2)}{3!} \times 5^3 = 2,603,125$$

Please see figure 2 and Supplemental Digital Content (http://links.lww.com/ALN/B732) for details.

Primary Outcome. MAP was calculated directly from arterial pressure waveform data. Hypotension was defined as any period with MAP < 65 mmHg for at least 1 min, based on studies suggesting that MAP < 65 mmHg is the threshold at which the probability of acute kidney injury and myocardial injury increase.[1] MAP > 75 mmHg was considered nonhypotension. Clearly, the real-world, clinical definition of hypotension cannot be based on a purely binary, all-or-none threshold. We considered that the flex point where the incidence of complications rises is in the MAP range between 65 to 75 mmHg, which may be thought of as a "gray zone" in which ambiguity and some risk coexist.[29] Yet in any binary classification problem, it is important to have two easily separable and mutually exclusive labels. In the interests of precision, we based our model only on definite hypotension (MAP < 65 mmHg) and definite nonhypotension (MAP > 75 mmHg) data.

To eliminate the effect of sudden drops in pressure due to artifact or external event, rather than to the patient's own physiologic responses, hypotensive data segments with a rate of decrease in MAP faster than 0.5 mmHg/s were excluded from the analysis (see also Supplemental Digital Content, http://links.lww.com/ALN/B732). A rate of decrease greater than 0.5 mmHg/s is equivalent to a decline in MAP greater than 30 mmHg in 1 min, which we considered outside the scope of prediction of our algorithm, as it would relate more likely to an acute event (*e.g.,* sudden blood loss or alteration of transducer height) than to progressive onset of hypotension.

To ensure clarity with relation to the algorithm, the early identification period for hypotension was defined as 15 min before an actual event where MAP fell less than 65 mmHg for at least 1 min. For comparison, we also assessed whether hypotensive events could be predicted with percent change in MAP (ΔMAP). Four different ΔMAPs were calculated and evaluated: ΔMAP_{20s}, ΔMAP_{1min}, ΔMAP_{3min}, and ΔMAP_{5min} (the difference between two MAP values that are 20 s, 1, 3, and 5 min apart, respectively).

Model Feature Selection and Training. A hypotensive event was calculated by identifying a section of at least 1-min duration such that all data points in the section showed MAP < 65 mmHg. An event, or positive data point, was chosen as the sample recorded 5, 10, or 15 min before the hypotensive event. A nonhypotensive event was calculated by identifying a 30-min continuous section of data points such that the section was at least 20 min apart from any hypotensive event, and all data points in that section showed MAP > 75 mmHg. A nonevent, or negative data point, was the center point of the nonhypotensive event.

Model Feature Selection. Totals of 3,022 individual features and 2,603,125 combinatorial features were extracted from the arterial pressure waveforms of the training data set (see also Supplemental Digital Content, http://links.lww.com/ALN/B732). For model training, a data matrix of these features for the positive and negative data points was used with a logistic regression model with "binomial" distribution and without regularization. This process

of training was repeated several times, with different patient subsets from the training data set, and also with different definitions for positive and negative data points. Performance of each model after features selection and training was evaluated with the cross-validation data set. The final model was chosen based on the performance, both in terms of prediction error and general behavior in each patient, on the cross-validation data set. To keep only the most useful features, these features were put through a two-step feature selection process: (1) features were retained where the area under the curve was greater than 0.8 for positive and negative data segments of the training data set; and (2) sequential forward features were selected with logistic regression.

Model Type and Output. Machine learning was used to map the arterial pressure waveform features into a prediction of hypotension. We chose logistic regression as our model type because it outputs an interpretable prediction of an event and because it provides smooth transitions between two classes. Logistic regression is a classification method for prediction of a binary response based on one or more model input features. It has the benefit of generating a numerical score to reflect the degree of the severity. This is achieved by using the "logit" transformation of the dependent binary variable and conducting a linear regression. The following equation represents this concept mathematically:

$$\log\left(\frac{h_\omega(x)}{1 - h_\omega(x)}\right) = \omega^T x$$

where "x" is the independent variable (features) vector, "ω" the corresponding vector of coefficients, and $h_\omega(x)$ is the logistic model for the dependent variable. Solving for the logistic function yields:

$$h_\omega(x) = \frac{1}{1 + e^{-\omega^T x}}$$

The logistic function is a continuous function within the range [0, 1]. During model training, the optimum coefficient vector "ω" is calculated from a set of the model input feature vectors "x_i" and the corresponding observed class of the training set "y_i" (0 = no event, 1 = event) for i = 1, 2, …, N, with a log-likelihood cost function shown below:

$$J(\omega) = -\frac{1}{N}\left[\sum_{i=1}^{N} y_i \log\left(h_\omega(x_i)\right) + \left(1 - y_i\right) \log\left(1 - h_\omega(x_i)\right)\right]$$

The solution is obtained by minimizing the above convex cost function with respect to the coefficient vector "ω." Once "ω" is determined with available training data, the logistic model is used on new data (x_t) for calculating the prediction of event "$p(x_t)$."

$$p(x_t) = \frac{1}{1 + e^{-\omega^T x_t}}$$

The prediction produced by the logistic regression model, ranging from 0 to 1, is then multiplied by 100 for scaling. We name the resulting prediction the Hypotension Prediction Index.

Statistical Methods

Data are expressed as mean ± SD (median and [25th–75th] quartiles) and/or number (percentage). (See also Supplemental Digital Content, http://links.lww.com/ALN/B732.) All statistics were performed with MATLAB (version R2014a).

Validation Data and Methods. Receiver-operating characteristic curve analysis was performed to evaluate the performance of the algorithm and ΔMAP (ΔMAP$_{20s}$, ΔMAP$_{1min}$, ΔMAP$_{3min}$). Sensitivity and specificity were calculated from receiver-operating characteristic curves with thresholds that minimized the difference between sensitivity and specificity. To correctly assess the receiver-operating characteristic performance of the algorithm, it is critical to correctly define hypotension (positives) and nonhypotension (negatives) to ensure that the positives are truly positives and the negatives are truly negatives (see also Supplemental Digital Content, http://links.lww.com/ALN/B732).

As outlined above in the section on "Model Feature Selection and Training," a hypotensive event was calculated by identifying a section of at least 1-min duration, with MAP < 65 mmHg for all data points in the section. A positive data point was chosen as the sample recorded at 5-, 10-, and 15-min intervals before the hypotensive event. All positive data points were included in the analysis regardless of their MAP values. A nonhypotensive event was calculated by identifying a 30-min continuous section of data points at least 20 min apart from any hypotensive event and MAP > 75 mmHg for all data points in that section. A nonevent, or negative data point, was the center point of the nonhypotensive event. This was done in order to reduce the impact of intraclass correlation. The algorithm generates a value every 20 s, but we selected only one data point out of a 30-min window because selecting all the data points within the window would have introduced statistical bias. We considered that adjacent 20-s data points are nearly at the same hemodynamic state, and including all data points would have introduced repetitive information. We selected the middata point of a 30-min window, but a random data point or the average could have been selected as well.

From a clinical perspective, the most important objective of the receiver-operating characteristic analysis is to test whether hypotensive events can be predicted by the algorithm x-min in advance, independent of the current MAP level. The next most important feature of the receiver-operating characteristic analysis is to test whether the algorithm correctly predicts period of hemodynamic stability when hypotension will not occur, according to our definition of negative events. A true positive is any event data point with the algorithm value greater than or equal to a chosen

threshold. Sensitivity is the ratio of true positives to all events. A true negative is any nonevent data point with the algorithm value less than a chosen threshold. Specificity is the ratio of true negatives to all nonevents. Positive predictive value was calculated as the ratio of the true positives to all positives (true positives + false positives). Negative predicted value was calculated as the ratio of the true negatives to all negatives (true negatives + false negatives).

Our receiver-operating characteristic analysis captures false negatives, or occurrences when the algorithm is erroneously low before hypotension, including both the unequivocal hypotension zone of MAP < 65 mmHg and the borderline hypotension gray zone between 65 and 75 mmHg. Our receiver-operating characteristic analysis also captures false positives, or occurrences when the algorithm is erroneously high, when hypotension does not occur, and MAP is above 75 mmHg. The only limitation in the receiver-operating characteristic analysis of selecting negatives as data points of MAP > 75 mmHg is that the receiver-operating characteristic analysis may not show the impact of false positives in the borderline 65 to 75 mmHg range. We did not consider this as a fundamental limitation. Clinically, the MAP between 65 and 75 mmHg remains an important intermediate zone, where the risk of complications may still exist, and in which a false positive could be beneficial if it prompts heightened attention to the patient's hemodynamic profile. Of note, in a recent publication on early warning systems by Scully and Daluwatte from the U.S. Food and Drug Administration, negatives are not even considered in analyses of performances of early warning systems.[30]

Analysis of Hypotension Prediction Compared to Actual Occurrence of Hypotensive Events (Algorithm Output and Frequency of Hypotension Analysis).

In this analysis, we plotted the frequency of occurrence of hypotensive events in the data samples at different ranges of the algorithm output. The analysis was performed as follows: (1) hypotensive episodes were defined as MAP < 65 mmHg for at least 1 min; (2) event samples were taken going back exactly "t" min (t = 5, 10, 15 min) before the start of a hypotensive episode; (3) nonhypotensive episodes with MAP > 75 mmHg were at least 20 min apart from any hypotensive episode; (4) nonevent samples were taken as the midpoint of every 30-min nonhypotensive episode; (5) all algorithm output values for the above event and nonevent samples, for a given data set, were accumulated and segmented into algorithm output bins; and (6) for each bin, the percentage of event samples in that bin was the rate of events, as the event samples have an event happening in "t" min (see Supplemental Digital Content, http://links.lww.com/ALN/B732).

Results

Descriptive Statistics

Model Training, Cross-validation, Internal Validation, and External Validation Data.

The model training cohort (n = 293 patients) presented 25,461 positive segments (total duration, 127,921 min) and 56,143 negative segments (total duration, 418,038 min). These included 87 ± 248 (12 [0, 60]) hypotensive events per patient representing $10 \pm 21\%$ (4 [0, 17]) of each patient's monitoring time. Each hypotensive event lasted 7.1 ± 7.3 min (5.2 [2.7, 7.9]). See table 1 for detailed results.

The cross-validation cohort (n = 1,041 patients) contained 33,915 positive segments (total duration, 239,629 min) and 87,902 negative segments (total duration, 708,535 min), including 33 ± 59 (14 [4, 36]) hypotensive events per patient representing $13 \pm 22\%$ (11 [2, 30]) of each patient's monitoring time. Each hypotensive event lasted 7.0 ± 10.1 min (4.8 [2.7, 8.0]).

The internal validation cohort (n = 350 patients) contained 14,969 positive segments (total duration, 125,999 min) and 49,011 negative segments (total duration, 391,537 min). These included 43 ± 60 (23 [5, 56]) hypotensive events per patient, representing $12 \pm 20\%$ (5 [1, 21]) of each patient's monitoring time. Each hypotensive event lasted 7.8 ± 7.2 min (4.6 [2.6, 7.1]).

The external validation database (n = 204 patients) contained 1,923 positive segments (total duration, 5,684 min) and 3,731 negative segments (total duration, 27,552 min). These included 9 ± 11 (6 [2, 14]) hypotensive events per patient, representing $7 \pm 9\%$ (3 [1, 8]) of each patient's monitoring time. Each hypotensive event lasted 3.0 ± 3.6 min (2.2 [1.5, 3.4]).

Validation of the Algorithm

Internal Validation. Figure 3 shows the receiver-operating characteristic curves for the algorithm output and all ΔMAPs as classifiers of hypotensive events with the internal validation data set 5, 10, and 15 min before the occurrence of hypotension. Table 2 shows the performance at different single points of time, up to 15 min before the events.
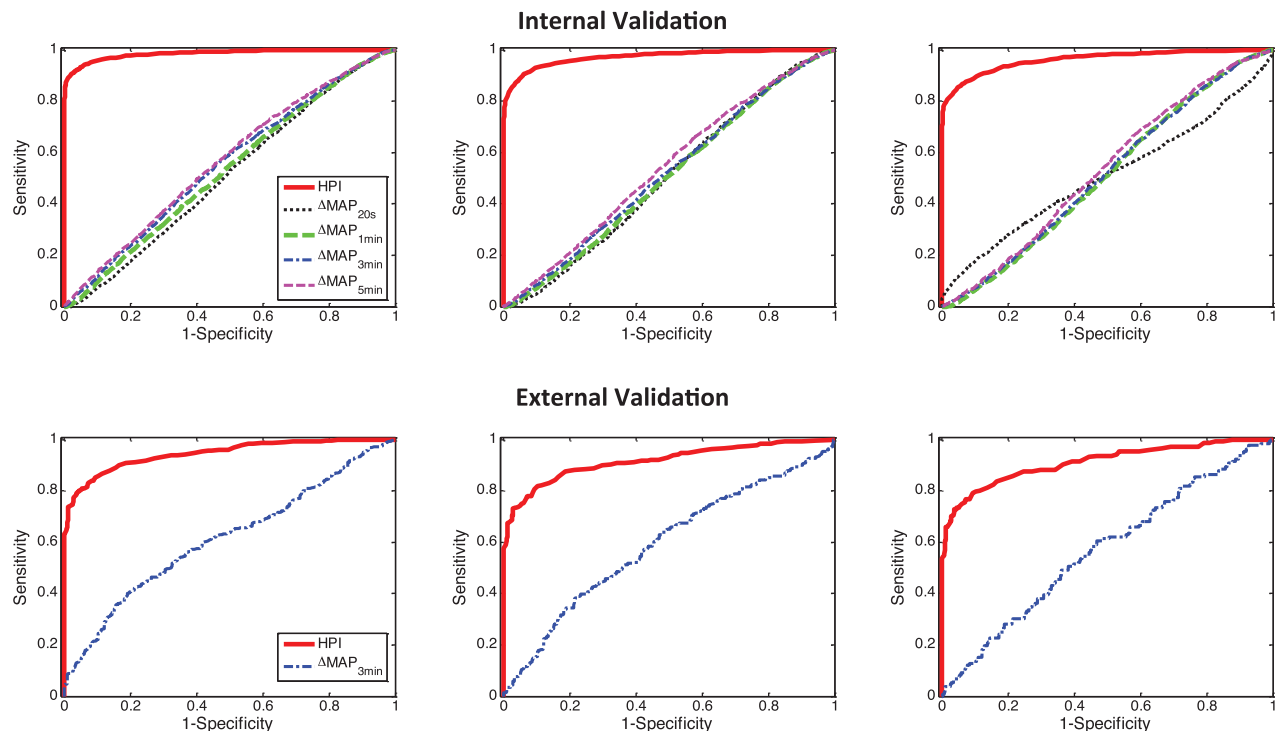
External Validation. Figure 3 shows receiver-operating characteristic curves of the algorithm and ΔMAP_{3min} as classifiers of hypotensive events with the external validation data set 5, 10, and 15 min before the occurrence of hypotension. Table 2 shows the performance at different single points of time, up to 15 min before the events. Figure 4 shows the results of the analysis of hypotension prediction compared to the actual occurrence of hypotensive events. The rate of occurrence of hypotension increases linearly with increase in the algorithm output for intermediate values of the indicator. Figure 5 illustrates a typical example of a developing hypotensive event during surgery. In this example, MAP was stable in the 75 to 80 mmHg range, but 17 min before the hypotensive event the algorithm output increased sharply from about 50 to 95% and remained above 90% until the event occurred. During the hypotensive event, the algorithm output remained at 100%. Simply monitoring MAP in this case would not have indicated an impending event. (See additional results in the Supplemental Digital Content, http://links.lww.com/ALN/B732.)

Table 1. Databases and Cohort Patients' Demographics

	Type	OR + ICU	OR	ICU
MIMIC II and Edwards databases	Number of patients, n	1,684	670	1014
	Sex, male	1097	439	658
	Age, yr	62.5±14.7	61.5±14.4	63.2±14.9
	BSA, m²	1.8±0.3	1.7±0.3	1.9±0.3
Training cohort (n = 1,334)				
Model training	Number of patients, n	293	111	182
	Sex, male	183	68	115
	Age, yr	61.4±16	56.5±16.3	64.5±15.1
	BSA, m²	1.9±0.3	1.8±0.3	1.9±0.3
Cross-validation	Number of patients, n	1,041	507	534
	Sex, male	694	342	352
	Age, yr	63±14.4	62.9±14	63.2±14.8
	BSA, m²	1.8±0.3	1.7±0.2	1.9±0.3
Model testing and internal validation cohort	Number of patients, n	350	52	298
	Sex, male	220	29	191
	Age, yr	61.9±14.7	58.3±11.3	62.6±15.1
	BSA, m²	1.9±0.2	1.8±0.2	1.9±0.3
External validation cohort (UCI data)	Number of patients, n	204	204	
	Sex, male	100	100	
	Age, yr	57±14	57±14	
	BSA, m²	1.9±0.3	1.9±0.3	

Data are presented as mean ± SD.

BSA, body surface area; ICU, intensive care unit; MIMIC II, Multiparameter Intelligent Monitoring in Intensive Care II; OR, operating room; UCI, University of California Irvine.

Fig. 3. Receiver-operating characteristic curves displaying the ability of the algorithm and ΔMAP to predict hypotension before its occurrence in the internal and external validation cohorts. Receiver-operating characteristic curves displaying the ability of the algorithm and ΔMAP to predict hypotension 5, 10, and 15 min (*left*, *center*, and *right*, respectively) before its occurrence in the internal validation cohort (*upper*) and in the external validation cohort (*bottom*). HPI, Hypotension Prediction Index; ΔMAP_{20s}, ΔMAP_{1min}, ΔMAP_{3min}, and ΔMAP_{5min}, the absolute changes in mean arterial pressure more than 20 s, 1 min, 3 min, and 5 min, respectively.

Table 2. Area under the Receiver-operating Characteristic Curve, Sensitivity, and Specificity of the Algorithm and ∆MAP in Internal and External Validation Cohorts

Parameter	Time to Event	AUC	Sensitivity, %	Specificity, %	PPV, %	NPV, %	Threshold
Internal validation cohort							
HPI	5 min	0.97 (0.971–0.977)	91.8 (89.7–94.0)	92.2 (90.3–94.0)	88.8 (87.9–89.7)	94.4 (93.8–94.9)	41
	10 min	0.95 (0.950–0.960)	89.3 (87.2–91.4)	89.5 (87.3–91.7)	76.7 (75.3–78.1)	95.6 (95.1–96.1)	38
	15 min	0.95 (0.940–0.952)	87.5 (85.1–89.9)	87.3 (84.8–89.8)	67.2 (65.6–68.9)	95.9 (95.4–96.4)	36
∆MAP$_{20s}$	5 min	0.51 (0.500–0.522)	51.0 (49.3–52.7)	51 (49.7–52.2)	—	—	0.05 mmHg
	10 min	0.50 (0.490–0.516)	50.5 (48.3–52.7)	50.5 (49.3–51.7)	—	—	0.03 mmHg
	15 min	0.51 (0.494–0.523)	49.4 (46.8–52.1)	49.4 (48.2–50.6)	—	—	–0.01 mmHg
∆MAP$_{1min}$	5 min	0.53 (0.519–0.543)	52.1 (50.5–53.8)	52.1 (50.9–53.4)	—	—	0.15 mmHg
	10 min	0.51 (0.492–0.520)	50.5 (48.2–52.8)	50.5 (49.3–51.7)	—	—	0.08 mmHg
	15 min	0.51 (0.495–0.525)	50.5 (48.1–52.9)	50.5 (49.3–51.7)	—	—	0.08 mmHg
∆MAP$_{3min}$	5 min	0.55 (0.541–0.567)	54.5 (52.7–56.3)	54.5 (53.0–55.9)	—	—	0.29 mmHg
	10 min	0.52 (0.503–0.532)	51.5 (49.1–53.9)	51.5 (50.1–52.8)	—	—	0.14 mmHg
	15 min	0.51 (0.498–0.531)	50.9 (48.0–53.8)	50.9 (49.6–52.2)	—	—	0.11 mmHg
∆MAP$_{5min}$	5 min	0.57 (0.554–0.581)	54.9 (52.6–57.1)	54.9 (53.5–56.3)	—	—	0.31 mmHg
	10 min	0.54 (0.526–0.556)	53.1 (50.3–55.9)	53.1 (51.7–54.5)	—	—	0.19 mmHg
	15 min	0.53 (0.517–0.550)	52.4 (49.5–55.3)	52.4 (51.0–53.8)	—	—	0.14 mmHg
External validation cohort							
HPI	5 min	0.95 (0.933–0.961)	86.8 (83.6–89.9)	88.5 (84.9–92.0)	93.2 (91.0–95.3)	78.6 (74.3–82.9)	39
	10 min	0.92 (0.896–0.943)	84.2 (79.6–88.8)	84.3 (80.2–88.4)	83.6 (79.4–87.8)	84.8 (80.8–88.8)	37
	15 min	0.91 (0.885–0.944)	83.6 (78.2–89.0)	83.3 (78.9–87.8)	74.0 (67.9–80.1)	90.0 (86.5–93.4)	36
∆MAP$_{20s}$	5 min	0.59 (0.554–0.632)	56.8 (52.8–60.8)	56.7 (51.2–62.3)	—	—	0.21 mmHg
	10 min	0.51 (0.466–0.559)	49.7 (44.3–55.0)	50.3 (45.2–55.5)	—	—	0.01 mmHg
	15 min	0.51 (0.462–0.566)	50.8 (44.3–57.4)	51.0 (45.7–56.2)	—	—	0.03 mmHg
∆MAP$_{1min}$	5 min	0.62 (0.586–0.663)	61.0 (56.7–65.3)	60.9 (55.6–66.2)	—	—	0.44 mmHg
	10 min	0.57 (0.525–0.619)	57.4 (51.7–63.1)	57.4 (52.3–62.5)	—	—	0.25 mmHg
	15 min	0.56 (0.505–0.614)	56.3 (49.1–63.5)	56.4 (51.3–61.5)	—	—	0.22 mmHg
∆MAP$_{3min}$	5 min	0.61 (0.571–0.655)	58.4 (53.1–63.7)	58.3 (52.8–63.9)	—	—	0.75 mmHg
	10 min	0.59 (0.542–0.640)	57.1 (50.7–63.5)	57.1 (51.4–62.7)	—	—	0.61 mmHg
	15 min	0.57 (0.510–0.626)	55.3 (46.8–63.8)	55.4 (49.8–61.1)	—	—	0.44 mmHg
∆MAP$_{5min}$	5 min	0.59 (0.550–0.640)	57.2 (51.3–63.2)	57.4 (52.5–62.2)	—	—	0.65 mmHg
	10 min	0.55 (0.499–0.607)	53.7 (45.9–61.5)	53.8 (49.0–58.6)	—	—	0.36 mmHg
	15 min	0.56 (0.501–0.624)	53.9 (44.6–63.2)	53.8 (49.0–58.6)	—	—	0.36 mmHg

Data are presented as mean (95% CI).

AUC, area under the curve; HPI, Hypotension Prediction Index; NPV, negative predictive value; ∆MAP$_{20s, 1min, 3min,}$ and $_{5min}$, percent changes in mean arterial pressure more than 20 s, 1 min, 3 min, and 5 min, respectively; PPV, predictive positive value.

Discussion

These results demonstrate that it is possible to train a machine-learning model, with large data sets of high-fidelity arterial pressure waveforms, to predict arterial hypotension events in the physiologic data sets of surgical patients up to 15 min before they occur. No reliable methods currently exist to predict the likelihood that a patient will become hemodynamically unstable, although several methods are available to monitor hemodynamic parameters, identify cardiovascular volatility, and alert clinicians when it occurs.[16] There is evidence that subtle dynamic linkages or interconnections exist among disparate physiologic variables at the earliest stages of instability that are clinically imperceptible.[16,17] These unique signatures of dynamic interconnections become more defined as pathologic states develop, and their changes over time presage the development of a worsening physiologic state.[16,17,20,31,32]

Several recently published studies have demonstrated that certain patient-related factors may be harbingers of hemodynamic instability. Convertino *et al.* developed a machine-learning model that estimated central blood volume loss with 96.5% accuracy in a hemorrhagic shock model using lower-body negative pressure.[33] Noninvasive hemodynamic features used in the model included standard variables collected from patients during surgery, including blood pressure, end-tidal carbon dioxide, respiratory rate, and pulse character. The same group also applied novel machine-learning methods to plethysmographic waveforms to identify patients who were developing cardiac instability.[34] More specific to the topic of hypotension in the critical care setting, Ghosh *et al.* have used sequential contrast patterns, mining the methodology of blood pressure monitoring to anticipate the onset of hypotension in the intensive care unit.[35,36] While of
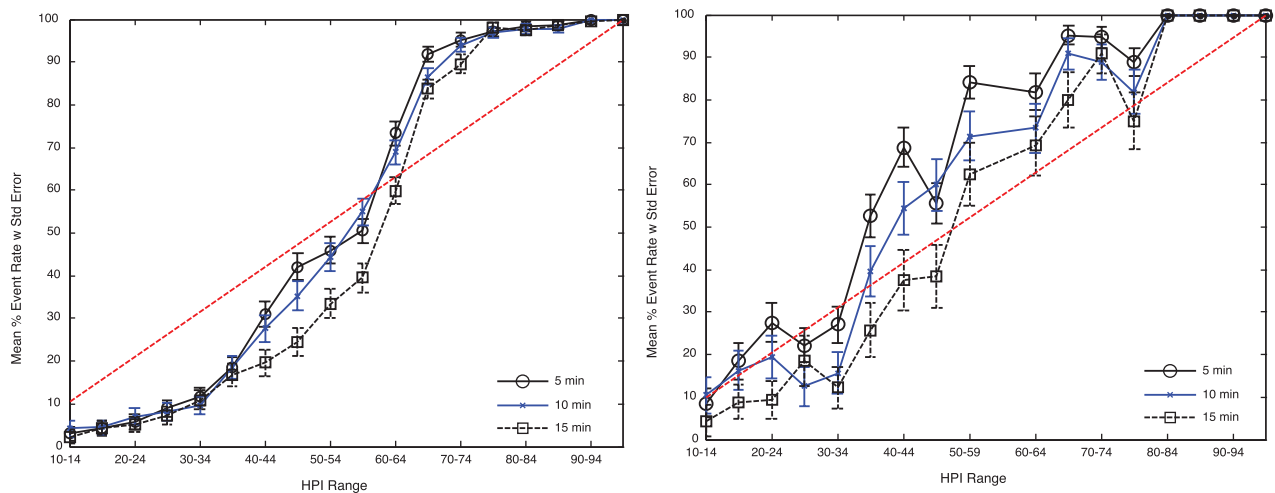
Fig. 4. Analysis of hypotension prediction compared to actual occurrence of hypotensive events for the internal (*left*) and external (*right*) cohorts. The frequency of occurrence of hypotensive events is relatively low when the Hypotension Prediction Index (HPI) is low, and high when HPI is high. The rate of occurrence of hypotension increases linearly with increase in HPI for intermediate values of the indicator.
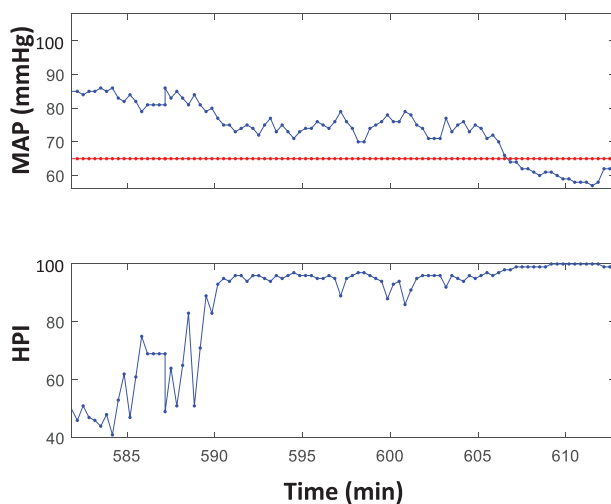
Fig. 5. One illustrative patient record showing the association between the algorithm output (Hypotension Prediction Index [HPI]) and the evolution of mean arterial pressure (MAP) over time.

interest, this approach does not yet allow real-time prediction of hypotension.

Our work centers on the development of a predictive algorithm based on a machine-learning model for potential real-time prediction of hypotension. The algorithm output indicates the likelihood that a patient's condition is trending toward a hypotensive event. When the algorithm output is low, the likelihood of a hypotensive event is also low, and the time-to-event interval tends to be long. Conversely, when the algorithm output is high, the likelihood of a hypotensive event is high, and the time-to-event interval tends to be shorter.

The algorithm is based on detection of physiologic signatures in high-resolution arterial pressure waveforms caused by weakening of the cardiovascular compensatory mechanisms that typically occur before hypotension and that affect cardiac preload, afterload, and contractility. The early stage of instability appears to be characterized by subtle, complex changes in the associations among different physiologic variables.[16,17] Dynamic changes in the variability, complexity, and physiologic associations of features in the arterial pressure waveform occur before the obvious clinical occurrence of hypotensive events.[17] The fundamental emphasis of the algorithm is to detect the earliest appearance of these dynamic changes in the arterial pressure waveform and to use them to predict an upcoming hypotensive event. Specifically, the algorithm detects dynamic changes corresponding to physiologic interactions among left ventricular contractility, preload, and afterload. The main challenge in detecting these complex changes is that they are highly multivariate. They are not only indiscernible to the human eye, but also are not detectable by simple signal processing algorithms. In order to detect the multivariate variability and interactions preceding hypotensive events, the algorithm uses complex machine-learning techniques.

Machine-learning methods are powerful mathematical tools that allow accurate quantification of dynamic multivariate interconnections. It is important to emphasize that in the algorithm, machine-learning techniques quantify the complex processes of cardiac compensatory mechanisms mathematically; they do not capture statistical relationships as do most machine-learning–based algorithms used in the past.[15] The assessment of the physiologic associations is critical to the algorithm, as it represents the effect of the dynamic links among thousands of automatically derived hemodynamic features, all from the arterial waveform. The assessment of the physiologic associations included computation of linear and nonlinear combinations of all 3,022 variability/complexity features in the arterial waveform computed initially. These combinatorial features then provided key information on the nonlinear effects and dynamic physiologic

671

interactions among all 3,022 of the individual linear features. We see a 1:1 linear association between the algorithm output and event rate as arterial blood pressure declines to a MAP level of 65 to 70 mmHg (fig. 4). Then the algorithm output sharply rises when approaching a hypotensive event, apparently due to the increase in the interconnection signatures that the model detects in the arterial pressure waveform when hypotension is impending. Our algorithm used this large, comprehensive analysis of interaction effects to assess compensatory mechanisms and capture the cross-correlational changes among thousands of automatically derived hemodynamic features, all from the arterial waveform, that herald the onset of hypotension.

Besides the potential clinical value of this development, we can envision new fields of investigation for basic physiology research by augmenting the analysis of physiologic waveforms with computer science techniques and reverse engineering. While this is a potentially significant step forward, many unanswered questions remain regarding the use of real-time predictive algorithms in the surgical setting as opposed to an intensive care unit or medical care setting. These questions are not specific to the development of a hypotension prediction algorithm and may apply to any predictive algorithm in the rapidly evolving scenario of an invasive procedure.

Study Limitations

We identified seven major limitations in our study, which will need to be explored further in prospective real-time clinical research:

1. Our system was trained and developed from the records of operating room and intensive care unit patients, while the external validation was performed in a surgical patients' physiologic data set. These facts are important to keep in mind when interpreting our results, and further validation in the intensive care unit setting may be needed. Intensive care unit patients undergoing supportive and therapeutic care constitute a different experimental model from patients undergoing an acute surgical intervention.

2. While the algorithm developed in this study is able to predict hypotension, it remains unclear how clinicians would act upon the alarm. The dynamic relationships among early warning systems, clinicians' responses, and changes in physicians' behaviors are not completely understood, and our study does not address how anesthesiologists and intensivists would respond to an alarm triggered by our algorithm. It is unclear whether they would respond or not, and if they respond it is unclear what they would do. After descriptive and predictive analytics, the next two steps would be prescriptive analytics and cognitive analytics. In association with a hypotension prediction algorithm, a decision-support tool could suggest interventions or treatment alternatives

that could prevent or reduce the severity of hypotension. These considerations are beyond the scope of this study.

3. The benefits of initiating treatment before the onset of hypotension are not yet clear. It seems likely that a predictive algorithm could decrease the duration of hypotension during surgery and in the intensive care unit. However, even though the relationship between hypotension and complications is statistically significant, direct causality has not been established. It is not yet known to what degree decreasing the incidence and duration of hypotension would improve outcome.

4. As described in this manuscript, our algorithm depends on invasive arterial line waveforms. A small minority of patients currently receives arterial line catheters during surgery. Developing a similar algorithm from noninvasive arterial pressure waveform recordings is needed to expand to applicability of this approach.

5. We did not include in our study any hypotensive events caused by clinical interventions (*e.g.,* laparoscopic insufflation, liver manipulation, or vascular clamping or unclamping). The algorithm we present is based on predicting hypotension that occurs outside of such events. This limitation, together with the absence of clinical context in our algorithm development, emphasizes the *in silico* nature of the current work as opposed to *in vivo* testing. Of note, the question of whether the algorithm can predict hypotension in the period immediately after the induction of anesthesia remains unanswered, as it was not formally tested in this study.

6. We used proprietary Edwards Lifesciences algorithms (FloTrac, CO-Trek) that may not be available to others who wish to replicate this study.

7. We present a study that has potential for real-time prediction of hypotension, but we have not analyzed data streams in real time to predict hypotension. We also used an arbitrary definition of hypotension (hypotensive events defined as MAP < 65 mmHg and nonhypotensive events defined as MAP > 75 mmHg), which creates an artificial testing environment—albeit one that is necessary at this level of development. These precise although arbitrary definitions likely improved the objective assessment of the algorithm's successes and shortcomings.

Conclusions

Our results demonstrate that a machine-learning algorithm can be trained, with large data sets of high-fidelity arterial pressure waveforms, to predict arterial hypotension events in a surgical patients' physiologic data set.

Acknowledgments

Research Support

Competing Interests

Correspondence

Address correspondence to Dr. Cannesson: Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, 757 Westwood Plaza, Los Angeles, California 90095. mcannesson@mednet.ucla.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

References

1. Salmasi V, Maheshwari K, Yang D, Mascha EJ, Singh A, Sessler DI, Kurz A: Relationship between intraoperative hypotension, defined by either reduction from baseline or absolute thresholds, and acute kidney and myocardial injury after noncardiac surgery: A retrospective cohort analysis. ANESTHESIOLOGY 2017; 126:47–65

2. van Waes JA, van Klei WA, Wijeysundera DN, van Wolfswinkel L, Lindsay TF, Beattie WS: Association between intraoperative hypotension and myocardial injury after vascular surgery. ANESTHESIOLOGY 2016; 124:35–44

3. Walsh M, Devereaux PJ, Garg AX, Kurz A, Turan A, Rodseth RN, Cywinski J, Thabane L, Sessler DI: Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: Toward an empirical definition of hypotension. ANESTHESIOLOGY 2013; 119:507–15

4. Kazmers A, Jacobs L, Perkins A: The impact of complications after vascular surgery in Veterans Affairs Medical Centers. J Surg Res 1997; 67:62–6

5. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I; Devereaux PJ, Chan MT, Alonso-Coello P, Walsh M, Berwanger O, Villar JC, Wang CY, Garutti RI, Jacka MJ, Sigamani A, Srinathan S, Biccard BM, Chow CK, Abraham V, Tiboni M, Pettit S, Szczeklik W, Lurati Buse G, Botto F, Guyatt G, Heels-Ansdell D, Sessler DI, Thorlund K, Garg AX, Mrkobrada M, Thomas S, Rodseth RN, Pearse RM, Thabane L, McQueen MJ, VanHelder T, Bhandari M, Bosch J, Kurz A,

6. Izawa J, Kitamura T, Iwami T, Uchino S, Takinami M, Kellum JA, Kawamura T: Early-phase cumulative hypotension duration and severe-stage progression in oliguric acute kidney injury with and without sepsis: An observational study. Crit Care 2016; 20:405

7. Medicine. NUa: New opportunities, new challenges: The changing nature of biomedical science. In: U.S. Committee on the Organizational Structure of the National Institutes of Health, eds. Enhancing the Vitality of the National Institutes of Health: Organizational Change to Meet New Challenges. Washington, D.C., National Academies Press, 2003.

8. Cannesson M, Tanabe M, Suffoletto MS, McNamara DM, Madan S, Lacomis JM, Gorcsan J III: A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. J Am Coll Cardiol 2007; 49:217–26

9. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, Bhatt DL, Fonarow GC, Laskey WK: Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: Comparison of machine learning and other statistical approaches. JAMA Cardiol 2017; 2:204–9

10. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316:2402–10

11. Lee J, Scott DJ, Villarroel M, Clifford GD, Saeed M, Mark RG: Open-Access MIMIC-II Database for Intensive Care Research. 2011 Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc) 2011:8315–8

12. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG: Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database. Crit Care Med 2011; 39:952–60

13. Bickler PE, Cannesson M, Shelley KH: Trends and challenges in clinical monitoring: Papers from the 2015 IAMPOV symposium. Anesth Analg 2017; 124:2–3

14. Shelley K, Cannesson M: "Off-label" use of clinical monitors: What happens when new physiologic understanding meets state-of-the-art technology. Anesth Analg 2014; 119:1241–2

15. PhysioNet and Computers in Cardiology Challenge: Predicting acute hypotensive episodes. Available at: https://physionet.org/challenge/2009/. Accessed November 7, 2017

16. Pinsky MR, Dubrawski A: Gleaning knowledge from data in the intensive care unit. Am J Respir Crit Care Med 2014; 190:606–10

17. Pinsky MR: Complexity modeling: Identify instability early. Crit Care Med 2010; 38(10 Suppl):S649–55

18. Chen L, Dubrawski A, Wang D, Fiterau M, Guillame-Bert M, Bose E, Kaynar AM, Wallace DJ, Guttendorf J, Clermont G, Pinsky MR, Hravnak M: Using supervised machine learning to classify real alerts and artifact in online multisignal vital sign monitoring data. Crit Care Med 2016; 44:e456–63

19. Fiterau M, Dubrawski A, Chen L, Hravnak M, Clermont G, Pinsky M: Automatic identification of artifacts in monitoring critically ill patients. Intensive Care Med 2013; 39:S470

20. Guillame-Bert M, Dubrawski A, Wang D, Hravnak M, Clermont G, Pinsky MR: Learning temporal rules to forecast instability in continuously monitored patients. J Am Med Inform Assoc 2017; 24:47–53

21. Hravnak M, Chen L, Bose E, Fiterau M, Guillame-Bert M, Dubrawski A, Clermont G, Pinsky M: Artifact patterns in continuous noninvasive monitoring of patients. Intensive Care Med 2013; 39(Suppl 2):S405

22. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S, Berk M: Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. J Med Internet Res 2016; 18:e323

23. Scott DJ, Lee J, Silva I, Park S, Moody GB, Celi LA, Mark RG: Accessing the public MIMIC-II intensive care relational database for clinical research. BMC Med Inform Decis Mak 2013; 13:9

24. PhysioNet: MIMIC II Waveform Database Matched Subset. Available at: https://www.physionet.org/physiobank/database/mimic2wdb/matched/. (Accessed November 7, 2017

25. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE: PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000; 101:E215–20

26. Pratt B, Roteliuk L, Hatib F, Frazier J, Wallen RD: Calculating arterial pressure-based cardiac output using a novel measurement and analysis method. Biomed Instrum Technol 2007; 41:403–11

27. Langewouters GJ, Wesseling KH, Goedhard WJ: The static elastic properties of 45 human thoracic and 20 abdominal aortas *in vitro* and the parameters of a new model. J Biomech 1984; 17:425–35

28. Cannesson M, Musard H, Desebbe O, Boucau C, Simon R, Hénaine R, Lehot JJ: The ability of stroke volume variations obtained with Vigileo/FloTrac system to monitor fluid responsiveness in mechanically ventilated patients. Anesth Analg 2009; 108:513–7

29. Ray P, Le Manach Y, Riou B, Houle TT: Statistical evaluation of a biomarker. ANESTHESIOLOGY 2010; 112:1023–40

30. Scully CG, Daluwatte C: Evaluating performance of early warning indices to predict physiological instabilities. J Biomed Inform 2017; 75:14–21

31. Guillame-Bert M, Dubrawski A, Chen L, Hravnak M, Pinsky M, Clermont G: Learning temporal rules to forecast instability in intensive care patients. Intensive Care Med 2013; 39(Suppl 2):S470

32. Hravnak M, Chen L, Dubrawski A, Bose E, Pinsky MR: Temporal distribution of instability events in continuously monitored step-down unit patients: Implications for rapid response systems. Resuscitation 2015; 89:99–105

33. Convertino VA, Moulton SL, Grudic GZ, Rickards CA, Hinojosa-Laborde C, Gerhardt RT, Blackbourne LH, Ryan KL: Use of advanced machine-learning techniques for non-invasive monitoring of hemorrhage. J Trauma 2011; 71(1 Suppl):S25–32

34. Convertino VA, Grudic G, Mulligan J, Moulton S: Estimation of individual-specific progression to impending cardiovascular instability using arterial waveforms. J Appl Physiol (1985) 2013; 115:1196–202

35. Ghosh S, Feng M, Nguyen H, Li J: Risk prediction for acute hypotensive patients by using gap constrained sequential contrast patterns. AMIA Annu Symp Proc 2014; 2014:1748–57

36. Ghosh S, Feng M, Nguyen H, Li J: Hypotension risk prediction *via* sequential contrast patterns of ICU blood pressure. IEEE J Biomed Health Inform 2016; 20:1416–26