Supervised Machine Learning Predictive Analytics for Prediction of Postinduction Hypotension

Samir Kendale, M.D., Prathamesh Kulkarni, Ph.D., Andrew D. Rosenberg, M.D., Jing Wang, M.D., Ph.D.

ABSTRACT

Background: Hypotension is a risk factor for adverse perioperative outcomes. Machine learning methods allow large amounts of data for development of robust predictive analytics. The authors hypothesized that machine learning methods can provide prediction for the risk of postinduction hypotension

Methods: Data was extracted from the electronic health record of a single quaternary care center from November 2015 to May 2016 for patients over age 12 that underwent general anesthesia, without procedure exclusions. Multiple supervised machine learning classification techniques were attempted, with postinduction hypotension (mean arterial pressure less than 55 mmHg within 10 min of induction by any measurement) as primary outcome, and preoperative medications, medical comorbidities, induction medications, and intraoperative vital signs as features. Discrimination was assessed using cross-validated area under the receiver operating characteristic curve. The best performing model was tuned and final performance assessed using split-set validation.

Results: Out of 13,323 cases, 1,185 (8.9%) experienced postinduction hypotension. Area under the receiver operating characteristic curve using logistic regression was 0.71 (95% CI, 0.70 to 0.72), support vector machines was 0.63 (95% CI, 0.58 to 0.60), naive Bayes was 0.69 (95% CI, 0.67 to 0.69), k-nearest neighbor was 0.64 (95% CI, 0.63 to 0.65), linear discriminant analysis was 0.72 (95% CI, 0.71 to 0.73), random forest was 0.74 (95% CI, 0.73 to 0.75), neural nets 0.71 (95% CI, 0.69 to 0.71), and gradient boosting machine 0.76 (95% CI, 0.75 to 0.77). Test set area for the gradient boosting machine was 0.74 (95% CI, 0.72 to 0.77).

Conclusions: The success of this technique in predicting postinduction hypotension demonstrates feasibility of machine learning models for predictive analytics in the field of anesthesiology, with performance dependent on model selection and appropriate tuning. **(Anesthesiology 2018; XXX:00-00)**

HYPOTENSION has been demonstrated to be an independent risk factor for adverse perioperative outcomes.[1–5] Despite careful administration of medications by anesthesiologists during induction of general anesthesia, hypotension is often an unintended consequence. Thus, early recognition of intraoperative hypotension may lead to preventive measures to improve anesthetic and surgical outcome.

There are currently few available methods used in clinical practice for prediction of postinduction hypotension.[6,7] It is likely that a variety of factors are involved in the precipitation of postinduction hypotension, including patient comorbidities, home medications taken on the day of surgery, and medications used for induction of anesthesia. Due to this complexity, simple modeling techniques may not be sufficient for risk prediction of postinduction hypotension.

Machine learning methods provide an opportunity for large amounts of data to be incorporated into development of robust predictive analytics, often without many of the pitfalls and restrictions of standard modeling techniques.[8,9]

What We Already Know about This Topic

- The ability to predict postinduction hypotension remains limited and challenging due to the multitude of data elements that may be considered
- Novel machine learning algorithms may offer a systematic approach to predict postinduction hypotension, but are understudied

What This Article Tells Us That Is New

- Among 13,323 patients undergoing a variety of surgical procedures, 8.9% experienced a mean arterial pressure less than 55 mmHg within 10 min of induction start
- While some machine learning algorithms perform worse than logistic regression, several techniques may be superior
- Gradient boosting machine, with tuning, demonstrates a receiver operating characteristic area under the curve of 0.76, a negative predictive value of 19%, and positive predictive value of 96%

These techniques are increasingly being used in various fields of medicine, including in the diagnosis of primary

<zdoi;10.1097/ALN.0000000000002374>

hyperparathyroidism, prognosis in stage III colorectal cancer, discharge disposition following craniotomy, mortality prediction after cardiac surgery, and identifying graft failure after liver transplantation.[10–16] Modern electronic health records with integrated anesthesiology intraoperative records provide the opportunity for complex clinical decision support tools to aid clinicians in making decisions based on objective evidence and data, in addition to training and experience. There is now access to large amounts of perioperative data *via* the electronic health record. Before implementation of these tools, however, there must be proof-of-concept explorations followed by rigorous validation ("bench work") to confirm reliability in clinical practice ("bedside").

In the current study, we hypothesized that we could develop a highly discriminative machine learning model for prediction of postinduction hypotension using information readily available from the electronic health record to demonstrate the viability of machine learning methods for intraoperative predictive analytics.

Materials and Methods

The data did not contain any direct patient identifiers, and no direct interaction with human subjects was involved. Our Institutional Review Board (New York University Langone Health, New York, New York) does not directly review such analyses, and requires certification that no direct patient identifier is accessible *via* the data kept in record. This study was considered exempt from review and written informed consent was consequently waived.

Patient Population

Patients who underwent any procedure requiring general anesthesia from November 2015 to May 2016 were included. Only patients over the age of 12 were included. There were no procedure-based exclusions.

Primary Outcome

The target output (or primary outcome) was postinduction hypotension, defined as any single mean arterial pressure (MAP) less than 55 mmHg on any noninvasive or continuous blood pressure measurement within 10 min of the recorded induction time of general anesthesia. Designation of hypotension was limited only to the available recorded data. Any recorded blood pressure measurement was included, whether it was from an arterial line or from a non-invasive blood pressure cuff. Any blood pressure taken by non-invasive blood pressure cuff was recorded, so the frequency would depend on the cycle time designated by the anesthesiologist. Continuous blood pressure measurement was recorded at a resolution of 1 min. Multiple thresholds for the adverse potential for hypotension have also been explored, and relationships have been drawn between MAP less than 49 and postoperative mortality. In addition, systolic blood pressure and relative blood pressure changes have also been explored in association with adverse postoperative outcome. This particular threshold was

chosen based on the association of MAP less than 55 mmHg with adverse perioperative outcomes.[1,3,4,17] Specifically, MAP less than 55 mmHg has been associated with both acute kidney injury and myocardial infarction in multiple previous explorations of large data sets in association with outcomes. Because this was a binary categorical designation (hypotension either occurred or it did not), only classification methods of machine learning were considered.

Data Source

Data were aggregated from the electronic health record (Epic; Epic Systems, USA) from a single large academic institution, including a six-month range from November 2015 to May 2016. The data were made available by unified reports initially created by the hospital information technology department for reporting and research purposes. There were reports generated for: demographic data, intraoperative medications administered, intraoperative event times, preoperative medical comorbidities in the form of the problem list as of the time of the encounter, intraoperative vital signs and ventilator data, and preoperative medications as of the time of the encounter. Each set of values was assessed individually for validity, including by visualization of summary data and case sampling, in which 100 cases were assessed manually.

Data Elements

From demographic data, the following were extracted: age, sex, body mass index, time of surgery, and American Society of Anesthesiologists (ASA) Physical Status score. From intraoperative medications, by using the time of medication administration, only medications given between entry into the operating area and 10 min after induction start were included. Intraoperative vital signs were restricted to the same time period (from start of anesthesia to 10 min after induction start). At our institution, preoperative IV medications are not given before entry into the operating room. Induction start is a single manually entered event that indicates the beginning of administration of induction medication, and records that did not include an induction time were not included. We do not indicate induction end within the electronic health record at our institution. The data frame was truncated to include only the relevant time period. From the preoperative medical comorbidities, the following were extracted by text search from the problem list, which is included in every patient electronic health record and can be populated as discrete fields by any documenting provider: coronary artery disease, hypertension, congestive heart failure, atrial fibrillation, chronic kidney disease, asthma, chronic obstructive pulmonary disease, gastroesophageal reflux disease, obstructive sleep apnea, diabetes mellitus, and aortic stenosis. These exact terms were used for searching the problem list, as only these discrete terms can be entered to define the given comorbidity within the problem list (*i.e.*, coronary artery disease is listed as "coronary artery disease" or "coronary artery disease [CAD]" and not only "CAD"). The entire list of preoperative medications, as separated into pharmacologic classes,

was retained for more specific feature selection. Pharmacologic classes were determined by the underlying Epic categorization system, more specifically the Clarity (Epic Systems) designations intended for querying and reporting of medications. Clarity is the relational database that houses data from the Epic electronic health record. Clarity categorizes any given medication by therapeutic class, pharmacologic class, and pharmacologic subclass. For the purposes of this study, we used the most specific category: the pharmacologic subclass. Only electronic health record data were used, and no waveform data from the monitor were analyzed. Data was obtained exactly as it was recorded in the electronic health record.

Statistical Analysis

Data Cleaning and Feature Selection. Although machine learning is expected to help in unbiased feature selection, our initial feature selection for inclusion was decided based on clinical judgment of factors that are potential contributors to postinduction hypotension as a basis for data extraction, as well as what was available for extraction from the electronic health record. The previously mentioned medical comorbidities, age, sex, body mass index, time of surgery, and ASA Physical Status score were included. Time of surgery was represented by the hour during which the patient entered the procedure area, as a continuous variable (*i.e.,* entering the room at 7:26 AM was encoded as 7; entering the room at 10:15 PM was encoded as 22). ASA Physical Status score was represented as a categorical variable. Age and body mass index were included as continuous variables, while the remaining were considered categorical. Medications used during induction of general anesthesia were investigated, and only the most common medications were included: midazolam, propofol, etomidate, fentanyl, rocuronium, and succinylcholine. All medications were included as continuous variables indicating dose administered. In circumstances where there was no data for an intraoperative medication, such as no value for administration of propofol, that value was converted to zero. The data were inspected and continuous data did not indicate a need for normalization. Blood pressures that were obviously out of physiologic range were excluded (MAP less than 20 mmHg, MAP greater than 200 mmHg, or pulse pressure less than 20 mmHg), but no other attempt was made at artifact detection. No other preprocessing was performed. Consequently, all data were used without modification, and no additional measures were taken regarding potentially "missing" or misclassified data. This was done to preserve application of real-world data. Features were examined in relation to the primary outcome for data leakage or perfect separation potential. Data leakage refers to information within the training set that leads to excessively optimistic predictions. This data may be information that is not available in the real world setting, or data that contains the information that is to be predicted. For example, in this setting, data leakage could have occurred if lowest MAP had been included as a feature, as this value is directly related to the binary definition of hypotension used in the models. Perfect separation refers to data that clearly forces the outcomes of the algorithm into one classification or another (*i.e.,* if variable $x = 1$, then outcome $y = 1$ always). No features were considered to be a risk for data leakage or perfect separation. From intraoperative vital signs and ventilator data, the first MAP, maximum end tidal volatile anesthetic concentration, and mean peak inspiratory pressure were included, all of which were included as continuous variables. Recursive feature elimination was used as a wrapper method on top of random forest for feature selection, using the "rfe" function from the "caret" package, resulting in a subset of the available features for inclusion within the machine learning models.

Model Selection. Because not all machine learning methods have robust internal validation, data was randomly separated into 70/30 training and test sets for validation. Specifically, 70% of the data was used for training the machine learned models, and 30% was held out for the test set (fig. 1). In no predetermined order, machine learning algorithms were trained on the training set, using tenfold cross-validation repeated three times to minimize initial overfitting. Because of concerns of how various machine learning models treat class imbalance, area under the receiver operating characteristic curve was used as the primary performance metric due to its threshold-independence instead of a simple accuracy metric, which may not be reflective of performance in the setting of class imbalance. In addition, the threshold-dependent measures of sensitivity and specificity at the "best" thresholds were computed for each model. "Best" threshold refers to the threshold at which sensitivity and specificity are both maximized, not necessarily the optimal threshold for clinical integration. The following machine learning algorithms were trained: logistic regression, support vector machines, naive Bayes, k-nearest neighbor, linear discriminant analysis, random forest, neural nets, and stochastic gradient boosting machine. Although the goal was developing a predictive model, an additional aim was to explore how various machine learning algorithms compare with respect to handling of pre- and intraoperative data.

All available features after recursive feature elimination were used for all training algorithms. Some machine learning algorithms perform built-in feature selection (*i.e.,* random forest), and this behavior was not restricted. The "caret" package in R was used for initial training and tenfold cross-validation, using receiver operating characteristic as the performance metric, and basic tuning on parameters specific to each method of machine learning (https://CRAN.R-project.org/package=caret, accessed May 4, 2018). Tuning was performed by a limited grid-search as dictated by the defaults in the package. Functions, packages, and tuning parameters used for each machine learning method are shown in appendix 1.[18–21] (https://CRAN.R-project.org/package=gbm, accessed May 4, 2018). Receiver operating characteristic curves were generated using the "pROC" package.[22] Bootstrap 95% CI were computed with 2,000 stratified bootstrap replicates by the "ci" function in the "pROC" package.
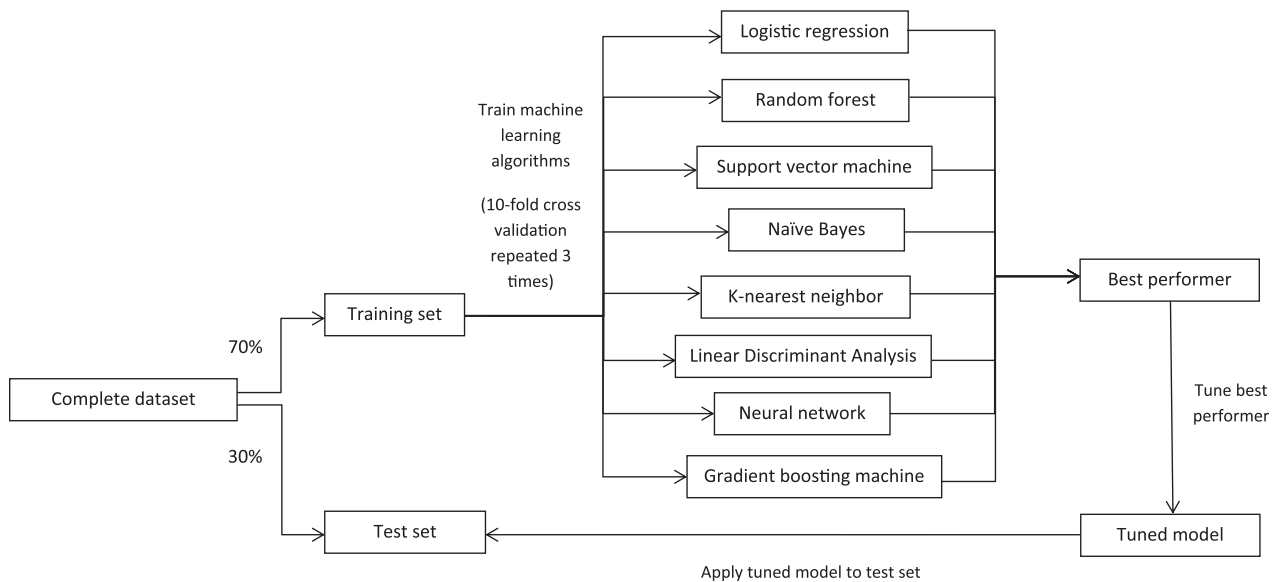
Fig. 1. Diagram of methods. The complete data set was split into training and test sets. The machine learning methods were trained on the training set and the best performer selected for additional parameter tuning before being applied to the test set for validation.

Model Tuning and Testing. After model selection, which involved a course-level tuning of the models, the best performing algorithm as determined by highest area under the receiver operating characteristic curve was fine-tuned further for parameters specific to the method. Tuning refers to optimization of the algorithm by modification of parameters in order to achieve the best performance. Modifiable parameters are specific to each machine learning algorithm (*i.e.,* number of trees for random forest, distance and kernel for k-nearest neighbor, *etc.*). Tuning parameters were assessed by expanded manual grid search, in which large but realistic ranges of values are given for each tuning parameter, and performance of the resulting models is compared. Variable importance was determined for the final model if this is a possibility for the given machine learning algorithm, as not all machine learning algorithms are amenable to computing variable importance. Variable importance is computed based on how important any given feature is to aid in the classification process when the classifier is built, determined by its effect on the performance measure. Generally, variable importance helps to assess the impact of any given variable on the performance of the algorithm. If a variable with high importance is permuted or removed from the model, the performance decreases. The greater the importance, the more essential the variable is to the performance of the model. Nonetheless, assumptions about effect size cannot be drawn directly about the relationship of variable importance to the primary outcome.[23] The "varImp" function from the "caret" package was used for variable importance (https://CRAN.R-project.org/package=caret, accessed May 4, 2018). The final model was then simulated on the test set to determine generalizability of the algorithm, and assess whether the model was overfitted.[24] The process is depicted in figure 1. All statistical operations were performed using the R statistical software (R Foundation for Statistical Computing, version 3.3.2, Austria). A sample data set (Supplemental Digital Content 1, http://links.lww.com/ALN/B773) and sample code of the primary analysis (Supplemental Digital Content 2, http://links.lww.com/ALN/B774) are provided.

Sensitivity Analyses
Different Definition of Hypotension. We used MAP less than 55 mmHg as the definition of hypotension due to its association with certain postoperative outcomes. Other studies suggest a more conservative definition of hypotension, namely MAP less than 65 mmHg, as being associated with harm, namely myocardial and kidney injury.[17] Because of this, we undertook a sensitivity analysis in which MAP less than 65 mmHg was considered as the definition for hypotension, and trained the best performing algorithm on this new definition to generate a new model.

Treated Hypotension. We initially only aimed to identify patients that experienced hypotension, as in those cases there was an implication that the hypotension was unanticipated. In cases of anticipated or suspected hypotension, the anesthesiologist is likely to have treated the hypotension. Sometimes the treatment is successful and other times not. This creates a few potential outcomes: untreated hypotension, unsuccessful treatment of hypotension, and successful treatment of hypotension. Our initial model would only capture the first two options. We undertook a sensitivity analysis in which the third outcome (successful treatment) is explored by including administration of phenylephrine or ephedrine as part of the target output definition in addition to MAP less than 55 mmHg, and trained the best performing algorithm on this new definition to generate a new model.

Table 1. Data Set Population Characteristics and Characteristics of Patients Who Experienced and Did Not Experience Postinduction Hypotension

	All Cases (n = 13,323)	No Hypotension (n = 12,138)	Hypotension (n = 1,185)	*P* Value
Age, yr, mean (SD)	50 (18)	49 (18)	55 (16)	< 0.001
Sex (male)	5,882 (44)	5,469 (45)	424 (36)	< 0.001
Body mass index, kg/m^2, mean (SD)	28 (6.4)	28 (6.3)	27 (6.8)	0.006
ASA score				< 0.001
Unknown	336 (2.5)	312 (2.6)	24 (2.0)	
I	2,616 (20)	2,482 (20)	134 (11)	
II	6,669 (50)	6,143 (51)	526 (44)	
III	3,103 (23)	2,719 (22)	384 (32)	
IV	590 (4.4)	481 (4.0)	109 (9.2)	
V	9 (0.07)	1 (0.008)	8 (0.68)	
Surgical service				< 0.001
Unknown	50 (0.38)	41 (0.34)	9 (0.76)	
Cardiovascular	468 (3.5)	387 (3.2)	81 (6.8)	
Gastroenterology	172 (1.3)	160 (1.3)	12 (1.0)	
Bariatrics	625 (4.7)	558 (4.6)	67 (5.7)	
General endocrine	221 (1.7)	201 (1.7)	20 (1.7)	
Oral surgery	66 (0.50)	63 (0.52)	3 (0.25)	
General pediatric	69 (0.52)	59 (0.49)	10 (0.84)	
General robotic	79 (0.59)	76 (0.63)	3 (0.25)	
General surgical oncology	486 (3.6)	445 (3.7)	41 (3.4)	
Transplant	46 (0.35)	38 (0.31)	8 (0.68)	
Wound care	71 (0.53)	59 (0.49)	12 (1.0)	
General	1,305 (9.8)	1,216 (10)	89 (7.5)	
Gynecology	1,550 (12)	1,463 (12)	87 (7.3)	
Movement disorder	110 (0.83)	99 (0.82)	11 (0.93)	
Neurosurgical craniotomy	338 (2.5)	311 (2.6)	27 (2.3)	
Pediatric neurosurgery	93 (0.70)	80 (0.66)	13 (1.1)	
Neurosurgical spine	342 (2.6)	282 (2.3)	60 (5.1)	
Obstetrics	9 (0.07)	5 (0.04)	4 (0.34)	
Ophthalmology	130 (0.98)	124 (1.0)	6 (0.51)	
Orthopedic foot/hand	175 (1.3)	164 (0.84)	11 (0.08)	
Pediatric orthopedic	61 (0.46)	55 (0.51)	6 (0.51)	
Orthopedic shoulder/elbow	22 (0.17)	22 (0.18)	0 (0)	
Orthopedic spine	898 (6.7)	785 (6.5)	113 (9.5)	
Sports	1,839 (14)	1,753 (14)	86 (7.3)	
Orthopedic tumor	51 (0.38)	48 (0.40)	3 (0.25)	
Orthopedic total joint	787 (5.9)	723 (6.0)	64 (5.4)	
Orthopedic trauma/fracture	178 (1.3)	150 (1.2)	28 (2.4)	
Otolaryngology head and neck	97 (0.73)	87 (0.72)	10 (0.84)	
Otolaryngology	853 (6.4)	781 (6.4)	72 (6.1)	
Pediatric oncology	26 (0.20)	18 (0.15)	8 (0.68)	
Plastic cosmetic	163 (1.2)	153 (1.3)	10 (0.84)	
Plastic reconstructive	396 (3.0)	328 (3.0)	32 (2.8)	
Pulmonary	170 (1.3)	155 (1.3)	15 (1.3)	
Thoracic	338 (2.5)	290 (2.4)	48 (4.1)	
Urology	923 (6.9)	832 (6.9)	91 (7.7)	
Vascular	114 (0.86)	91 (0.75)	23 (1.9)	

(Continued)

Table 1. (Continued)

	All Cases (n = 13,323)	No Hypotension (n = 12,138)	Hypotension (n = 1,185)	P Value
Medical comorbidities				
Coronary artery disease	439 (3.3)	373 (3.1)	66 (5.6)	< 0.001
Hypertension	2,065 (15)	1,794 (15)	271 (23)	< 0.001
Congestive heart failure	339 (2.5)	281 (2.3)	58 (4.9)	< 0.001
Atrial fibrillation	424 (3.2)	365 (3.0)	59 (5.0)	< 0.001
Chronic kidney disease	101 (0.74)	89 (0.73)	12 (1.0)	0.37
Asthma	440 (3.3)	397 (3.3)	43 (3.6)	0.57
Chronic obstructive pulmonary Disease	220 (1.7)	187 (1.5)	33 (2.8)	0.002
Gastroesophageal reflux disease	638 (4.8)	563 (4.6)	75 (6.3)	0.01
Obstructive sleep apnea	482 (3.6)	419 (3.5)	63 (5.3)	0.001
Diabetes mellitus	1,102 (8.3)	962 (7.9)	140 (12)	< 0.001
Aortic stenosis	94 (0.69)	82 (0.68)	12 (1.0)	0.25
Preoperative medications				
ACE inhibitors	1,313 (9.8)	1,159 (9.5)	154 (13)	< 0.001
Anxiolytics	2,156 (16)	1,928 (16)	228 (19)	0.003
Anticonvulsants	1,509 (11)	1,340 (11)	169 (14)	0.001
Antidepressants	1,989 (15)	1,796 (15)	193 (16)	0.18
Antiplatelet	512 (3.8)	419 (3.5)	93 (7.8)	< 0.001
Angiotensin receptor blockers	1,505 (11)	1,276 (11)	229 (19)	< 0.001
Short acting inhaled beta agonists	1,132 (8.5)	1,024 (8.4)	108 (9.1)	0.46
Inhaled glucocorticoids	790 (5.9)	713 (5.9)	77 (6.5)	0.42
Beta blockers	2,337 (18)	2,011 (17)	326 (28)	< 0.001
Biguanides	1,057 (7.9)	926 (7.6)	131 (11)	< 0.001
Systemic glucocorticoids	1,144 (8.9)	1,013 (8.3)	131 (11)	0.002
Leukotrienes	430 (3.2)	391 (3.2)	39 (3.3)	0.97
Antiprostatic hypertrophy	872 (6.5)	773 (6.4)	99 (8.4)	0.01
Thiazide diuretics	1,208 (9.1)	1,046 (0.40)	162 (0.42)	< 0.001
Levothyroxine	1,495 (11)	1,308 (11)	187 (16)	< 0.001
Loop diuretics	747 (5.6)	630 (5.2)	117 (9.9)	< 0.001
Proton pump inhibitors/H2 blockers	3,298 (25)	2,917 (24)	381 (32)	< 0.001
Statins	2,966 (22)	2,605 (21)	361 (30)	< 0.001
Aspirin	3,750 (28)	3,361 (28)	389 (33)	< 0.001
Calcium channel blockers	1,604 (12)	1,421 (12)	183 (15)	< 0.001
Intraoperative medications				
Midazolam (mg)	2 (0–2)	2 (0–2)	2 (0–2)	0.33
Propofol (mg)	200 (150–200)	200 (150–200)	150 (100–200)	< 0.001
Etomidate (mg)	0 (0-0)	0 (0-0)	0 (0-0)	0.009
Fentanyl (mcg)	100 (50–150)	100 (50–150)	100 (50–150)	< 0.001
Rocuronium (mg)	50 (35–50)	10 (0–50)	30 (0–50)	< 0.001
Succinylcholine (mg)	100 (100-100)	0 (0-0)	0 (0-0)	0.009
Maximum sevoflurane concentration (%)	1.4 (1.0–1.9)	1.2 (0.5–1.7)	1.0 (0.2–1.6)	< 0.001
Maximum desflurane concentration (%)	0 (0-0)	0 (0-0)	0 (0-0)	0.22
Peak inspiratory pressure (mm Hg)	16 (13–19)	16 (13–19)	17 (14–20)	< 0.001
First mean arterial pressure (mm Hg)	83 (74–93)	96 (87–105)	87 (75–97)	< 0.001

Only 20 of the most frequent preoperative medications are presented. Complete data set characteristics of only modeled features are shown in appendix 2. "Intraoperative medications" refers to medications given between entry into the procedure area and 10 min from induction start.

ACE inhibitors = angiotensin converting enzyme inhibitors; ASA score = American Society of Anesthesiologists Physical Status score; H2 blockers = histamine 2 receptor blockers.

Adjusting for Class Imbalance. Although threshold invariant metrics such as area under the receiver operating characteristic curve tend to be more resistant to class imbalance, there are additional methods to reduce the impact of class imbalance on measuring model performance. We undertook a sensitivity analysis in which a down-sampling of the majority class was performed on the training set to reduce initial class imbalance. The test set was not modified at all. The best performing algorithm was trained on the down-sampled training set to generate a new model.

Results

After exclusion of cases without an induction time and patients younger than age 12, there were 13,323 cases remaining, 1,185 (8.9%) of which experienced postinduction hypotension. There were 412 cases (3.0%) with missing

induction time. There were 2,051 (15%) cases that utilized continuous arterial blood pressure monitoring. Ultimately, the training set contained 9,326 cases and the test set contained 3,997 cases. There were 816 (8.7%) cases of postinduction hypotension within the training set, and there were 369 (9.2%) cases of postinduction hypotension within the test set. Data characteristics of the complete data set are detailed in table 1, while data characteristics of the data set only including those features for modeling are detailed in appendix 2. Final feature selection after recursive feature elimination is depicted in figure 2.

After training, area under the receiver operating characteristic curve using logistic regression was 0.71 (95% CI, 0.70 to 0.72); support vector machines was 0.63 (95% CI, 0.58 to 0.60); naive Bayes was 0.69 (95% CI, 0.67 to 0.69); k-nearest neighbor was 0.64 (95% CI, 0.63 to 0.65); linear discriminant analysis was 0.72 (95% CI, 0.71 to 0.73); random forest 0.74 (95% CI, 0.73 to 0.75); neural nets was 0.71 (95% CI, 0.69 to 0.71); and gradient boosting machines was 0.76 (95% CI, 0.75 to 0.77). Receiver operating characteristic curves, as well as sensitivity and specificity at "best" thresholds for each machine learning method, are depicted in figure 3.

Based on the model selection process, it appeared that gradient boosting machine was the strongest initial performer to be a candidate for continuing tuning and further testing. Other parameters that were tuned specific to the gradient boosting machine method were the number of trees (range 50 to 400), interaction depth (range 1 to 8), shrinkage (range 0.01 to 0.3), and the minimum number of variables at terminal node (range 5 to 30). Final tuning resulted in a gradient boosting machine algorithm with 200 trees, interaction depth of 6, shrinkage of 0.05, and 30 minimum variables at terminal node. The final model had an area under the receiver operating characteristic curve of 0.77 (95% CI, 0.75 to 0.78). Final variable importance can be seen in figure 4. The model run on the test set had an area under the receiver operating characteristic curve of 0.74 (95% CI, 0.72 to 0.76), a negative predictive value of 19% (95% CI, 16 to 21%) and a positive predictive value of 96% (95% CI, 95 to 97%). Areas under the receiver operating characteristic curve for all machine learning classifiers run on the test set are presented in table 2 solely for comparison in this setting.

Sensitivity Analyses

The model with a different definition of hypotension than the primary analysis (MAP less than 65 mmHg) had an area under the receiver operating characteristic curve of 0.72 (95% CI, 0.71 to 0.72), with specificity of 65% and sensitivity of 67% at the "best" threshold.
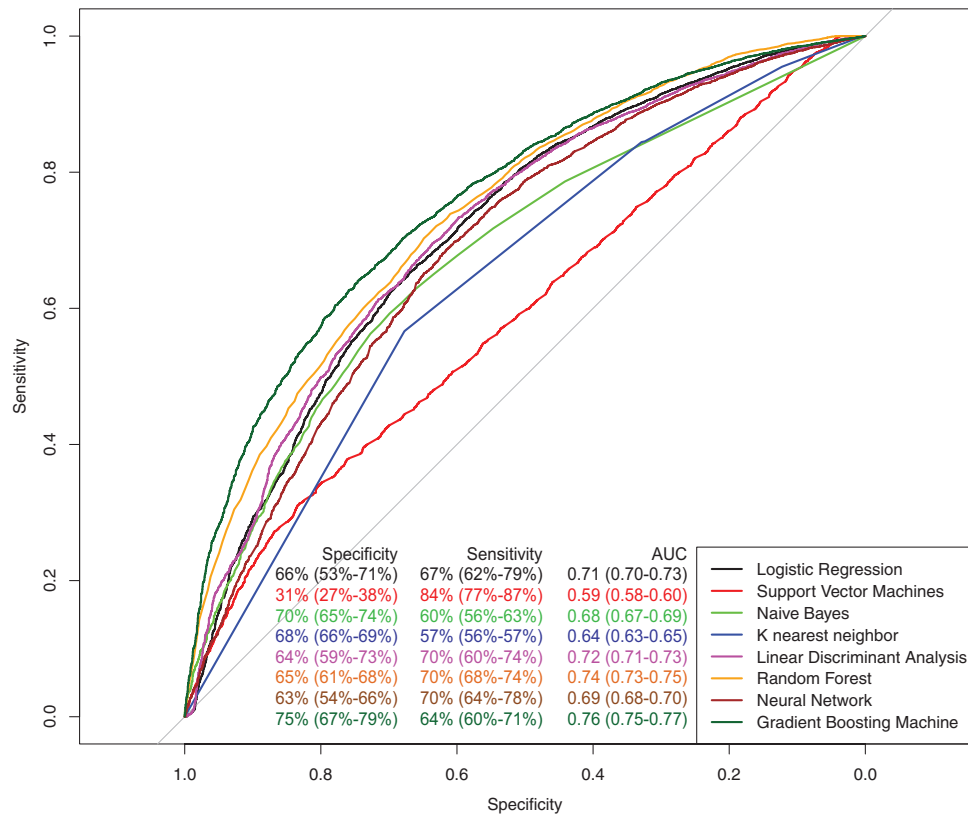
The model that incorporated administration of phenylephrine or ephedrine within the hypotension outcome definition had an area under the receiver operating characteristic curve of 0.75 (95% CI, 0.74 to 0.75), with specificity of 63% and sensitivity of 73% at the best threshold.

Fig. 2. Reduction of dimensionality by recursive feature elimination on the training data set. The number of features used for training was reduced from the list of features on the left to the list of features on the right. Italics indicate features that were eliminated. ACE inhibitors = angiotensin converting enzyme inhibitors; ASA score = American Society of Anesthesiologists Physical Status score.

Specificity	Sensitivity	AUC	
66% (53%–71%)	67% (62%–79%)	0.71 (0.70–0.73)	Logistic Regression
31% (27%–38%)	84% (77%–87%)	0.59 (0.58–0.60)	Support Vector Machines
70% (65%–74%)	60% (56%–63%)	0.68 (0.67–0.69)	Naive Bayes
68% (66%–69%)	57% (56%–57%)	0.64 (0.63–0.65)	K nearest neighbor
64% (59%–73%)	70% (60%–74%)	0.72 (0.71–0.73)	Linear Discriminant Analysis
65% (61%–68%)	70% (68%–74%)	0.74 (0.73–0.75)	Random Forest
63% (54%–66%)	70% (64%–78%)	0.69 (0.68–0.70)	Neural Network
75% (67%–79%)	64% (60%–71%)	0.76 (0.75–0.77)	Gradient Boosting Machine

Fig. 3. Receiver operating characteristic curves of machine learning methods for prediction of postinduction hypotension in the training data set. A greater area under the receiver operating characteristic curve (AUC) represents higher discriminative ability of the model. Area under the receiver operative characteristics curves, as well as specificity and sensitivity of each machine learning model for prediction of postinduction hypotension at "best" threshold are presented with 95% CIs. "Best" threshold refers to the threshold at which specificity and sensitivity are both maximized.

The model that utilized the down-sampled training set had an area under the receiver operating characteristic curve of 0.76 (95% CI, 0.75 to 0.77), with specificity of 69% and sensitivity of 69% at the best threshold.

Discussion

In this study, we examined the use of machine learning methods based on existing information in the electronic health record for intraoperative predictive analytics, specifically prediction of postinduction hypotension. The final model used a gradient boosting machine that demonstrated strong discrimination in both the training (area under the receiver operating characteristic curve 0.76, 95% CI, 0.75 to 0.77) and testing (area under the receiver operating characteristic curve 0.74, 95% CI, 0.72 to 0.77) sets. Gradient boosting machines function as an ensemble method by sequentially adding weak classifiers, in this case decision trees, to reach a final model based on improvement of each classifier. The results of this exploration are not surprising considering the nature of the data, namely that the machine learning methods that handle class imbalance better (gradient boosting machines, random forest, logistic regression) performed better than other methods. Boosting algorithms tend to suffer in cases of highly misclassified data, so the strong performance of this algorithm offers some indication of the veracity of the modeled data. Most of the variables of high importance are also not unexpected as far as clinically credibility; it is realistic to expect that features such as age, induction agents, volatile anesthetic concentration, and mean peak inspiratory pressure would be relevant for prediction of postinduction hypotension. Some were surprising, however, such as the relatively high importance of levothyroxine and bisphosphonates.

Some machine learning algorithms have been reported within the anesthesiology, perioperative care, and pain medicine fields, such as for predicting mortality after cardiac surgery,[13] predicting postoperative sepsis and acute kidney injury,[25] predicting postoperative pain or the need for pain consults,[26,27] or predicting patient controlled analgesia consumption.[28] These methods, however, do not extend into the intraoperative period. There are two major differences between our approach and that of other previous acute hypotension analysis approaches (as from the PhysioNet Challenge[29]). The settings of these explorations are primarily the intensive care unit, wherein data acquisition for processing occurs over a longer period of time and there is not necessarily a discrete inciting event related to hypotension. Additionally, this exploration used electronic health record
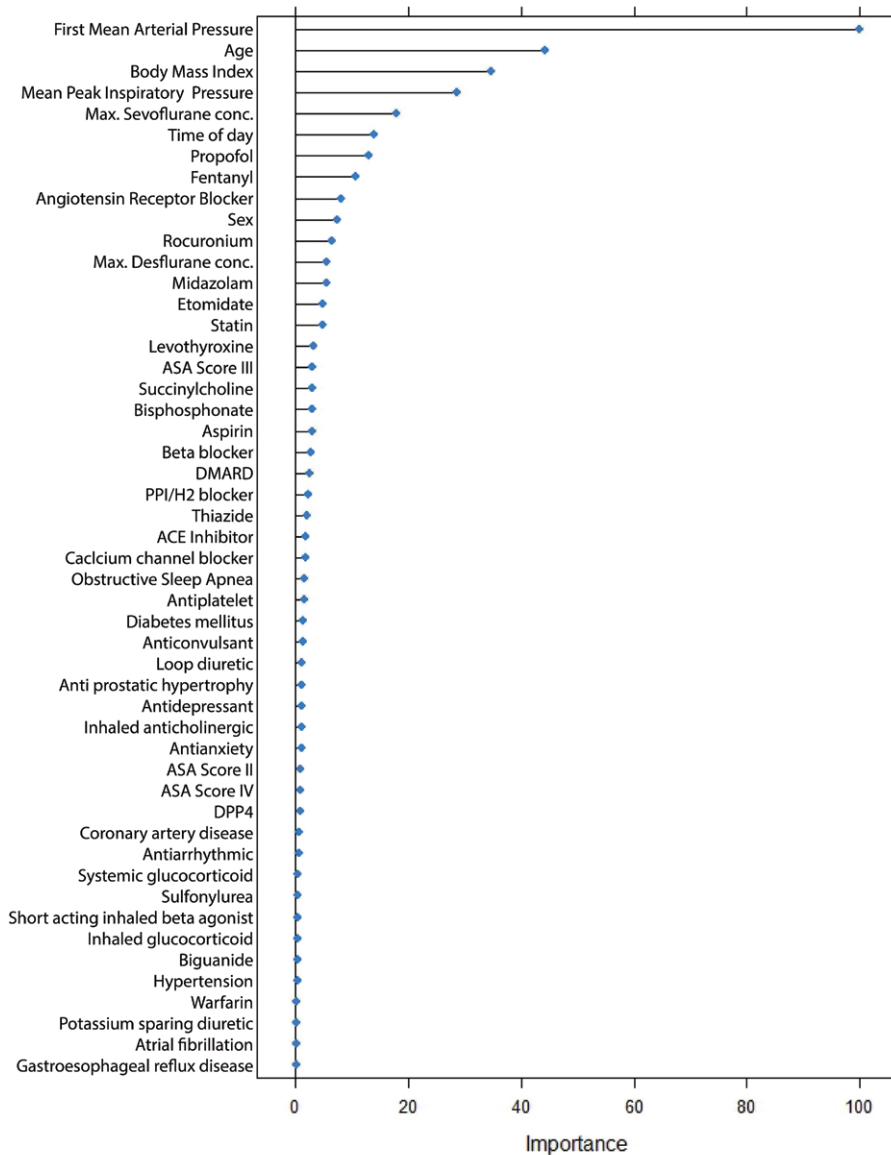
Fig. 4. Variable importance of features included in stochastic gradient boosting machine learning algorithm for prediction of postinduction hypotension. Variable importance is computed based on how important any given feature is to aid in the classification process when the classifier is built, determined by its effect on the performance measure. The greater the importance, the more essential the variable is to the performance of the model. Assumptions about effect size cannot be drawn directly about the relationship of variable importance to the primary outcome. ACE inhibitor = angiotensin converting enzyme inhibitor; ASA score = American Society of Anesthesiologists Physical Status score; DMARD = disease modifying antirheumatic drug; DPP4 = dipeptidyl peptidase-4 inhibitor; max. sevoflurane conc. = maximum sevoflurane concentration; max. desflurane conc. = maximum desflurane concentration; PPI/H2 blocker = proton pump inhibitor/H2 blocker.

data as opposed to waveform data. Waveform data is reliant on the quality of the waveform, as well as presence of invasive blood pressure monitoring, which is not available for all surgical cases.[29,30] While there are some predictive tools for intraoperative hypotension, none currently seem to utilize the electronic health record for clinical decision support integration, or utilize machine learning.[31,32] The benefit of being able to predict postinduction hypotension may ultimately allow clinicians to tailor their induction agents by prepopulating a model and observing the risk

of hypotension, or for triggering an intraoperative alert to notify the clinician of impending hypotension for treatment potential. Our machine learning approach has a strong precedent in both medical and nonmedical fields.

There are a number of benefits to using machine learning for problems such as this. The most obvious of these is the ability to incorporate large amounts of disparate data into a unified algorithm. Most machine learning methods are highly scalable, and thus can handle a variety of problems with differing feature types. The sensitivity analyses

Table 2. Area under the Receiver Operating Characteristic Curves (AUROC) for Each Machine Learning Classifier Run on the Test Data Set

	AUROC (95% CI)
Logistic regression	0.70 (0.67–0.73)
Support vector machines	0.64 (0.61–0.67)
Naïve Bayes	0.67 (0.65–0.70)
K-nearest neighbor	0.65 (0.62–0.68)
Linear discriminant analysis	0.71 (0.67–0.73)
Random forest	0.73 (0.70–0.76)
Neural nets	0.70 (0.67–0.73)
Stochastic gradient boosting	0.74 (0.71–0.76)

demonstrate the flexibility of machine learning approaches to variations in target definition. Machine learning is particularly useful when the limits of human understanding have been superseded. For example, despite a thorough understanding of pharmacology, normal physiology, pathophysiology, and surgical factors, postinduction hypotension still occurs at a surprisingly high rate, likely because the number of variables involved is so vast and complex. Because of this, such a problem is a prime target for machine learning. Although each individual machine learning method may have its own restrictions, most are not bound by the restrictions of classical prediction methods, such as linearity assumptions and the importance of identifying interactions between terms. Circumstances in which regression methods are utilized for machine learning can be preprocessed and tuned using techniques to minimize the impact of those assumptions, though not always entirely eliminated.

There are some disadvantages to machine learning methods, however. Training times can vary widely depending on the methods and tuning parameters, number of complexity of features, computing power, and sheer volume of data. This can make the iterative process required for tuning models relatively time consuming as compared to simple rule-based if–then approaches. As with any other predictive modeling technique, any given machine learning technique may not be an ideal approach for all tasks. For example, there were a number of methods in this study that performed more poorly than logistic regression, which, although considered a machine learning algorithm, is more accessible and familiar to the medical community due to its roots in statistical learning. While some machine learning methods offer information as to the relevance of various features, such as the variable importance shown for the gradient boosting machine algorithm, machine learning tools tend to be "black boxes," effectively. While utility can be measured using performance metrics, the lack of transparency in the algorithms may be inadequate to those who want to have a complete understanding of the clinical implications in order for more specific practice modification to be a possibility. However, the unbiased nature of machine learning algorithms may allow insight into previously unexplored or unexpected factors that may contribute to a given outcome. For example,

exploring why time of day was a variable of high importance may lead to further insight into a potential for modification. As with many other classifiers, threshold-dependent measures such as sensitivity and specificity may not be useful independent of choosing an appropriate threshold to balance desired sensitivity and specificity, which should be done based on clinical guidance and weighing the implications of misclassification as a result of over- or underdiagnosis. Finally, care must be taken when developing models to avoid overfitting, which can happen as a result of data leakage or perfect separation problems, among other causes.[8]

Limitations exist within this project. Although more than 10,000 cases were incorporated into the machine learning, they were extracted only over a six-month period, and within a single institution. A larger data set over a longer time period may have resulted in slightly different results, as practice may have changed over the course of the date range, and a greater amount of data may have led to utilization of more complex relationships among the data for algorithm training. Nonetheless, the expectation is that if incorporated into practice, algorithms would be routinely updated based on the most recent data so as to reflect current practice, as has been suggested in other studies of clinical data source relevance.[33] While we used the area under the receiver operating characteristic curve as a performance metric, there are a number of other acceptable methods for comparison among machine learning methods. The performance metric or metrics should be determined based on the examination of the data and the desired outcome, and no single performance metric is likely to fully encompass the viability of a machine learning application in a given setting. This was a single institution study, which, without external validation with an external database, limits the use of this precise model in another setting. While an area under the receiver operating characteristic curve of 0.76 is demonstrative of reasonably strong discrimination, there still exists substantial potential for improvement in performance before clinical use may be acceptable. There are a number of features that were not easily available as discrete values within the electronic health record for incorporation into the model. For example, numerical indices of heart function such as ventricular ejection fraction, details of airway management, or preoperative lab values may have improved the discrimination of the model. While this may seem to be a limitation, the methods demonstrated herein are reflections of how actual practical model training would occur, with only those features readily available from most electronic health record. With additional availability of data features, the scalability of most machine learning methods allows enhancement (and likely predictive power) of the existing model. As electronic health records evolve, there has also been a transition to including more discrete data fields, which will aid in development of predictive analytics, as well as in other forms of data mining. For data that may not be easily transformed into discrete fields, data aggregation may ultimately have to rely on natural language processing. Finally, our approach

only finely tuned one model. It is possible that if other algorithms were tuned, they may have performed equal to, or better than, the single model we tuned.

Current intraoperative clinical decision support in most electronic health records is primarily rule based, and has demonstrated success in improving adherence to practice guidelines in cases of blood pressure and glucose management, and prophylaxis for postoperative nausea and vomiting and infection.[34–39] Like rule-based clinical decision support systems derived from existing guidelines, machine learning–based systems will need to be grounded in validated methodologies before assimilation into an intraoperative workflow. For this reason, clinical "bedside" application of these tools should be preceded by "bench" validation as described and demonstrated herein, followed by proof of clinical improvement.

Correspondence

Address correspondence to Dr. Kendale: Department of Anesthesiology, Perioperative Care and Pain Medicine, NYU Langone Medical Center, 550 First Avenue, New York, New York, 10016. Samir.Kendale@nyumc.org. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

Appendix 1. Functions, Packages, and Tuning Parameters in the R Statistical Software Used for Each Machine Learning Algorithm

Algorithm	Function	Package	Tuning Parameters
Logistic regression	glm	base	None
Support vector machines	svmRadial	kernlab[18]	Cost, *sigma*
Naïve Bayes	nb	klaR[19]	Distribution type (kernel), *Laplace correction, bandwidth adjustment*
K-nearest neighbor	knn	base	k
Linear discriminant analysis	lda	MASS[20]	none
Random forest	rf	randomForest[21]	Number of randomly selected predictors
Neural nets	nnet	nnet[20]	Number of hidden units, weight decay
Gradient boosting machine	gbm	gbm*	Number of boosting iterations, maximum tree depth, *shrinkage, minimum terminal node size*

Italics indicate tuning parameters that were held constant during initial training.
*https://CRAN.R-project.org/package=gbm, accessed May 4, 2018.

Appendix 2. Data Set Variable Characteristics for Features Included in Modeling, and Characteristics of Patients Who Experienced and Did Not Experience Postinduction Hypotension

	All Cases (n = 13,323)	No Hypotension (n = 12,138)	Hypotension (n = 1,185)	P Value
Age, yr, mean (SD)	50 (18)	49 (18)	55 (16)	< 0.001
Sex (male)	5,882 (44)	5,469 (45)	424 (36)	< 0.001
Body mass index, kg/m², mean (SD)	28 (6.4)	28 (6.3)	27 (6.8)	0.006
ASA score				< 0.001
Unknown	336 (2.5)	312 (2.6)	24 (2.0)	
I	2,616 (20)	2,482 (20)	134 (11)	
II	6,669 (50)	6,143 (51)	526 (44)	
III	3,103 (23)	2,719 (22)	384 (32)	
IV	590 (4.4)	481 (4.0)	109 (9.2)	
V	9 (0.07)	1 (0.008)	8 (0.68)	
Medical comorbidities				
Coronary artery disease	439 (3.3)	373 (3.1)	66 (5.6)	< 0.001
Hypertension	2,065 (15)	1,794 (15)	271 (23)	< 0.001
Congestive heart failure	339 (2.5)	281 (2.3)	58 (4.9)	< 0.001
Atrial fibrillation	424 (3.2)	365 (3.0)	59 (5.0)	< 0.001
Chronic kidney disease	101 (0.74)	89 (0.73)	12 (1.0)	0.37

(Continued)

Appendix 2. (Continued)

	All Cases (n = 13,323)	No Hypotension (n = 12,138)	Hypotension (n = 1,185)	P Value
Asthma	440 (3.3)	397 (3.3)	43 (3.6)	0.57
Chronic obstructive pulmonary disease	220 (1.7)	187 (1.5)	33 (2.8)	0.002
Gastroesophageal reflux disease	638 (4.8)	563 (4.6)	75 (6.3)	0.01
Obstructive sleep apnea	482 (3.6)	419 (3.5)	63 (5.3)	0.001
Diabetes mellitus	1,102 (8.3)	962 (7.9)	140 (12)	< 0.001
Aortic stenosis	94 (0.69)	82 (0.68)	12 (1.0)	0.25
Preoperative medications				
ACE inhibitors	1,313 (9.8)	1,159 (9.5)	154 (13)	< 0.001
Alpha blockers	73 (0.55)	69 (0.57)	4 (0.34)	0.41
Antianginal	166 (1.2)	144 (1.2)	22 (1.9)	0.07
Anxiolytics	2,156 (16)	1,928 (16)	228 (19)	0.003
Antiarrythmics	228 (1.7)	188 (1.5)	40 (3.4)	< 0.001
Anticonvulsants	1,509 (11)	1,340 (11)	169 (14)	0.001
Antidepressants	1,989 (15)	1,796 (15)	193 (16)	0.18
Antineoplastics	386 (2.9)	336 (2.8)	50 (4.2)	0.006
Antiparkinsons	171 (1.3)	153 (1.3)	18 (1.5)	0.54
Antiplatelet	512 (3.8)	419 (3.5)	93 (7.8)	< 0.001
Antiretrovirals	132 (1.0)	116 (1.0)	16 (1.4)	0.25
Angiotensin receptor blockers	1,505 (11)	1,276 (11)	229 (19)	< 0.001
Inhaled anticholinergics	326 (2.4)	279 (2.3)	47 (4.0)	0.001
Long acting inhaled beta agonists	14 (0.11)	9 (0.07)	5 (0.42)	0.002
Short acting inhaled beta agonists	1,132 (8.5)	1,024 (8.4)	108 (9.1)	0.46
Inhaled glucocorticoids	790 (5.9)	713 (5.9)	77 (6.5)	0.42
Beta blockers	2,337 (18)	2,011 (17)	326 (28)	< 0.001
Biguanides	1,057 (7.9)	926 (7.6)	131 (11)	< 0.001
Bisphosphonates	156 (1.2)	121 (1.0)	35 (3.0)	< 0.001
Calcimimetics	29 (0.22)	26 (0.21)	3 (0.25)	1.000
Digoxin	88 (0.66)	74 (0.61)	14 (1.2)	0.03
Disease modifying antirheumatic drugs	96 (0.72)	77 (0.63)	19 (1.6)	< 0.001
Dipeptidyl peptidase-4 inhibitors	334 (2.5)	294 (2.4)	40 (3.4)	0.06
Erythropoetin	55 (0.41)	47 (0.39)	8 (0.68)	0.22
Systemic glucocorticoids	1,144 (8.9)	1,013 (8.3)	131 (11)	0.002
Glucosidases	9 (0.07)	6 (0.05)	3 (0.25)	0.05
Hydralazine	61 (0.46)	56 (0.46)	5 (0.42)	1.000
Immunosuppressants	129 (1.0)	117 (0.96)	12 (1.0)	0.99
Incretins	91 (0.68)	83 (0.68)	8 (0.68)	1.000
Long acting insulin	346 (2.6)	304 (2.5)	42 (3.5)	0.04
Short acting insulin	264 (2.0)	236 (1.9)	28 (2.4)	0.38
Potassium sparing diuretics	255 (1.9)	221 (1.8)	34 (2.9)	0.02
Leukotrienes	430 (3.2)	391 (3.2)	39 (3.3)	0.97
Lithium	32 (0.24)	28 (0.23)	4 (0.34)	0.68
Meglitinides	44 (0.33)	38 (0.31)	6 (0.51)	0.40
Phosphate binders	91 (0.68)	77 (0.63)	14 (1.2)	0.05
Antiprostatic hypertrophy	872 (6.5)	773 (6.4)	99 (8.4)	0.01
Rifamycin	32 (0.24)	27 (0.22)	5 (0.42)	0.30
Sulfonylureas	378 (2.8)	323 (2.7)	55 (4.6)	< 0.001
Thiazide diuretics	1,208 (9.1)	1,046 (0.40)	162 (0.42)	< 0.001
Thrombin inhibitors	40 (0.30)	35 (0.29)	5 (0.42)	0.60
Levothyroxine	1,495 (11)	1,308 (11)	187 (16)	< 0.001
Thiazolidinediones	54 (0.41)	49 (0.40)	5 (0.42)	1.000
Warfarin	323 (2.4)	280 (2.3)	43 (3.6)	0.006
Loop diuretics	747 (5.6)	630 (5.2)	117 (9.9)	< 0.001
Proton pump inhibitors/H2 blockers	3,298 (25)	2,917 (24)	381 (32)	< 0.001
Statins	2,966 (22)	2,605 (21)	361 (30)	< 0.001
Aspirin	3,750 (28)	3,361 (28)	389 (33)	< 0.001

(*Continued*)

Appendix 2. (Continued)

	All Cases (n = 13,323)	No Hypotension (n = 12,138)	Hypotension (n = 1,185)	P Value
Calcium channel blockers	1,604 (12)	1,421 (12)	183 (15)	< 0.001
Factor Xa inhibitors	261 (2.0)	224 (1.8)	37 (3.1)	0.004
Intraoperative medications				
Midazolam (mg)	2 (0–2)	2 (0–2)	2 (0–2)	0.33
Propofol (mg)	200 (150–200)	200 (150–200)	150 (100–200)	< 0.001
Etomidate (mg)	0 (0-0)	0 (0-0)	0 (0-0)	0.009
Fentanyl (mcg)	100 (50–150)	100 (50–150)	100 (50–150)	< 0.001
Rocuronium (mg)	50 (35–50)	10 (0–50)	30 (0–50)	< 0.001
Succinylcholine (mg)	100 (100-100)	0 (0-0)	0 (0-0)	0.009
Maximum sevoflurane concentration (%)	1.4 (1.0–1.9)	1.2 (0.5–1.7)	1.0 (0.2–1.6)	< 0.001
Maximum desflurane concentration (%)	0 (0-0)	0 (0-0)	0 (0-0)	0.22
Peak inspiratory pressure (mmHg)	16 (13–19)	16 (13–19)	17 (14–20)	< 0.001
First mean arterial pressure (mmHg)	83 (74–93)	96 (87–105)	87 (75–97)	< 0.001

"Intraoperative medications" refers to medications given between entry into the procedure area and 10 min from induction start.

ACE inhibitors = angiotensin converting enzyme inhibitors; ASA score = American Society of Anesthesiologists Physical Status score ; H2 blockers = histamine 2 receptor blockers.

References

1. 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

2. Bijker JB, van Klei WA, Vergouwe Y, Eleveld DJ, van Wolfswinkel L, Moons KG, Kalkman CJ: Intraoperative hypotension and 1-year mortality after noncardiac surgery. ANESTHESIOLOGY 2009; 111:1217–26

3. Sun LY, Wijeysundera DN, Tait GA, Beattie WS: Association of intraoperative hypotension with acute kidney injury after elective noncardiac surgery. ANESTHESIOLOGY 2015; 123:515–23

4. Monk TG, Bronsert MR, Henderson WG, Mangione MP, Sum-Ping ST, Bentt DR, Nguyen JD, Richman JS, Meguid RA, Hammermeister KE: Association between intraoperative hypotension and hypertension and 30-day postoperative mortality in noncardiac surgery. ANESTHESIOLOGY 2015; 123:307–19

5. 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

6. Alecu C, Cuignet-Royer E, Mertes PM, Salvi P, Vespignani H, Lambert M, Bouaziz H, Benetos A: Pre-existing arterial stiffness can predict hypotension during induction of anaesthesia in the elderly. Br J Anaesth 2010; 105:583–8

7. Hanss R, Renner J, Ilies C, Moikow L, Buell O, Steinfath M, Scholz J, Bein B: Does heart rate variability predict hypotension and bradycardia after induction of general anaesthesia in high risk cardiovascular patients? Anaesthesia 2008; 63:129–35

8. Kruppa J, Liu Y, Diener HC, Holste T, Weimar C, König IR, Ziegler A: Probability estimation with machine learning methods for dichotomous and multicategory outcome: Applications. Biom J 2014; 56:564–83

9. Steyerberg EW, van der Ploeg T, Van Calster B: Risk prediction with machine learning and regression methods. Biom J 2014; 56:601–6

10. Goldstein BA, Navar AM, Carter RE: Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. Eur Heart J 2017; 38:1805–14

11. Somnay YR, Craven M, McCoy KL, Carty SE, Wang TS, Greenberg CC, Schneider DF: Improving diagnostic recognition of primary hyperparathyroidism with machine learning. Surgery 2017; 161:1113–21

12. Salvucci M, Würstle ML, Morgan C, Curry S, Cremona M, Lindner AU, Bacon O, Resler AJ, Murphy ÁC, O'Byrne R, Flanagan L, Dasgupta S, Rice N, Pilati C, Zink E, Schöller LM, Toomey S, Lawler M, Johnston PG, Wilson R, Camilleri-Broët S, Salto-Tellez M, McNamara DA, Kay EW, Laurent-Puig P, Van Schaeybroeck S, Hennessy BT, Longley DB, Rehm M, Prehn JH: A stepwise integrated approach to personalized risk predictions in stage III colorectal cancer. Clin Cancer Res 2017; 23:1200–12

13. Allyn J, Allou N, Augustin P, Philip I, Martinet O, Belghiti M, Provenchere S, Montravers P, Ferdynus C: A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac aurgery: A decision curve analysis. PLoS One 2017; 12:e0169772

14. Lau L, Kankanige Y, Rubinstein B, Jones R, Christophi C, Muralidharan V, Bailey J: Machine-learning algorithms predict graft failure after liver transplantation. Transplantation 2017; 101:e125–32

15. Muhlestein WE, Akagi DS, Chotai S, Chambless LB: The impact of race on discharge disposition and length of hospitalization after craniotomy for brain tumor. World Neurosurg 2017; 104:24–38

16. Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG, Escobar GJ, Liu VX: Identifying distinct subgroups of ICU patients: A machine learning approach. Crit Care Med 2017; 45:1607–15

17. 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

18. Karatzoglou A, Smola A, Hornik K, Zeileis A: kernlab - an S4 package for kernel methods in R. Journal of Statistical Software 2004; 11(9):1–20

19. Weihs C, Ligges U, Luebke K, Raabe N: klaR: Analyzing German business cycles, Data Analysis and Decision Support. Edited by Daier D, Decker, R, Schmidt-Thieme, L. Springer-Verlag, Berlin, 2005, pp 335–43

20. Venables WN, Ripley BD: Classification, Modern Applied Statistics with S., 4th Edition. Edited by Venables WN, Ripley BD. New York, Springer, 2002, pp 331–42

21. Wiener ALaM: Classification and regression by randomForest. R News 2002; 2:18–22

22. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M: pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12:77

23. Hastie T, Tibshirani T, Friedman J: The elements of statistical learning, 2nd edition, Springer-Verlag, 2009

24. 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

25. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, Rashidi P, Pardalos P, Momcilovic P, Bihorac A: Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. PLoS One 2016; 11:e0155705

26. Tighe PJ, Lucas SD, Edwards DA, Boezaart AP, Aytug H, Bihorac A: Use of machine-learning classifiers to predict requests for preoperative acute pain service consultation. Pain Med 2012; 13:1347–57

27. Tighe PJ, Harle CA, Hurley RW, Aytug H, Boezaart AP, Fillingim RB: Teaching a machine to feel postoperative pain: Combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain. Pain Med 2015; 16:1386–401

28. Hu YJ, Ku TH, Jan RH, Wang K, Tseng YC, Yang SF: Decision tree-based learning to predict patient controlled analgesia consumption and readjustment. BMC Med Inform Decis Mak 2012; 12:131

29. Moody G, Lehman L: Predicting acute hypotensive episodes: The 10th Annual PhysioNet/Computers in Cardiology Challenge. Comput Cardiol 2009; 36:541–4

30. Rocha T, Paredes S, de Carvalho P, Henriques J: Prediction of acute hypotensive episodes by means of neural network multi-models. Comput Biol Med 2011; 41:881–90

31. Cheung CC, Martyn A, Campbell N, Frost S, Gilbert K, Michota F, Seal D, Ghali W, Khan NA: Predictors of intraoperative hypotension and bradycardia. Am J Med 2015; 128:532–8

32. Stapelfeldt WH, Yuan H, Dryden JK, Strehl KE, Cywinski JB, Ehrenfeld JM, Bromley P: The SLUScore: A novel method for detecting hazardous hypotension in adult patients undergoing noncardiac surgical procedures. Anesth Analg 2017; 124:1135–52

33. Chen JH, Alagappan M, Goldstein MK, Asch SM, Altman RB: Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets. Int J Med Inform 2017; 102:71–9

34. Wax DB, Beilin Y, Levin M, Chadha N, Krol M, Reich DL: The effect of an interactive visual reminder in an anesthesia information management system on timeliness of prophylactic antibiotic administration. Anesth Analg 2007; 104:1462–6

35. O'Reilly M, Talsma A, VanRiper S, Kheterpal S, Burney R: An anesthesia information system designed to provide physician-specific feedback improves timely administration of prophylactic antibiotics. Anesth Analg 2006; 103:908–12

36. Sathishkumar S, Lai M, Picton P, Kheterpal S, Morris M, Shanks A, Ramachandran SK: Behavioral modification of intraoperative hyperglycemia management with a novel real-time audiovisual monitor. ANESTHESIOLOGY 2015; 123:29–37

37. Nair BG, Horibe M, Newman SF, Wu WY, Peterson GN, Schwid HA: Anesthesia information management system-based near real-time decision support to manage intraoperative hypotension and hypertension. Anesth Analg 2014; 118:206–14

38. Nair BG, Grunzweig K, Peterson GN, Horibe M, Neradilek MB, Newman SF, Van Norman G, Schwid HA, Hao W, Hirsch IB, Patchen Dellinger E: Intraoperative blood glucose management: Impact of a real-time decision support system on adherence to institutional protocol. J Clin Monit Comput 2016; 30:301–12

39. Ehrenfeld JM, Epstein RH, Bader S, Kheterpal S, Sandberg WS: Automatic notifications mediated by anesthesia information management systems reduce the frequency of prolonged gaps in blood pressure documentation. Anesth Analg 2011; 113:356–63