

# Machine learning using institution-specific multi-modal electronic health records improves mortality risk prediction for cardiac surgery patients



Aaron J. Weiss, MD, PhD,<sup>a,b,c</sup> Arjun S. Yadaw, PhD,<sup>c</sup> David L. Meretzky, MA,<sup>c</sup> Matthew A. Levin, MD,<sup>d</sup> David H. Adams, MD,<sup>a</sup> Ken McCardle, MS,<sup>e</sup> Gaurav Pandey, PhD,<sup>f</sup> and Ravi Iyengar, PhD<sup>c</sup>

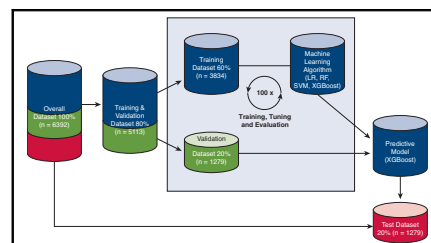
## ABSTRACT

**Background:** The Society of Thoracic Surgeons risk scores are widely used to assess risk of morbidity and mortality in specific cardiac surgeries but may not perform optimally in all patients. In a cohort of patients undergoing cardiac surgery, we developed a data-driven, institution-specific machine learning-based model inferred from multi-modal electronic health records and compared the performance with the Society of Thoracic Surgeons models.

**Methods:** All adult patients undergoing cardiac surgery between 2011 and 2016 were included. Routine electronic health record administrative, demographic, clinical, hemodynamic, laboratory, pharmacological, and procedural data features were extracted. The outcome was postoperative mortality. The database was randomly split into training (development) and test (evaluation) cohorts. Models developed using 4 classification algorithms were compared using 6 evaluation metrics. The performance of the final model was compared with the Society of Thoracic Surgeons models for 7 index surgical procedures.

**Results:** A total of 6392 patients were included and described by 4016 features. Overall mortality was 3.0% (n = 193). The XGBoost algorithm using only features with no missing data (336 features) yielded the best-performing predictor. When applied to the test set, the predictor performed well (F-measure = 0.775; precision = 0.756; recall = 0.795; accuracy = 0.986; area under the receiver operating characteristic curve = 0.978; area under the precision-recall curve = 0.804). eXtreme Gradient Boosting consistently demonstrated improved performance over the Society of Thoracic Surgeons models when evaluated on index procedures within the test set.

**Conclusions:** Machine learning models using institution-specific multi-modal electronic health records may improve performance in predicting mortality for individual patients undergoing cardiac surgery compared with the standard-of-care, population-derived Society of Thoracic Surgeons models. Institution-specific models may provide insights complementary to population-derived risk predictions to aid patient-level decision making. (JTCVS Open 2023;14:214-51)



Machine learning-based risk model development for mortality in patients undergoing cardiac surgery.

## CENTRAL MESSAGE

Institution-specific risk prediction models built on multi-modal EHR data may provide insights complementary to population-derived risk scores to aid patient-level decision making in patients undergoing cardiac surgery.

## PERSPECTIVE

Machine learning models using institution-specific multi-modal EHR data may provide improved performance in predicting mortality for patients undergoing cardiac surgery compared with the standard-of-care STS risk scores derived from population-level data. Institution-specific models may provide insights complementary to population-derived risk predictions to aid patient-level decision making.

From the Departments of <sup>a</sup>Cardiovascular Surgery, <sup>b</sup>Pharmacological Sciences and Institute for Systems Biomedicine, <sup>c</sup>Clinical Operations, and <sup>d</sup>Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; <sup>e</sup>Department of Thoracic and Cardiovascular Surgery, Heart, Vascular & Thoracic Institute, Cleveland Clinic, Cleveland, Ohio; and <sup>f</sup>Division of Cardiothoracic Anesthesia, Department of Anesthesiology and Critical Care, Icahn School of Medicine at Mount Sinai, New York, NY.

Supported by National Institutes of Health Grant P50GM071558 NIGMS Systems Biology Center grant to R.I. that supported this research as part of A.J.W.'s clinical PhD thesis project and R01HG011407 to G.P.

Drs Weiss and Yadaw contributed equally to this article as co-first authors.

Drs Pandey and Iyengar contributed equally as co-senior authors.

Institutional Review Board Number and date of approval: HS-15-00673, 10/29/2015. Received for publication Oct 27, 2022; revisions received Feb 4, 2023; accepted for publication March 16, 2023; available ahead of print March 31, 2023.

Address for reprints: Ravi Iyengar, PhD, One Gustave L. Levy Place, Box 1215, New York, NY 10029 (E-mail: ravi.iyengar@mssm.edu); or Aaron J. Weiss, MD, PhD, 9500 Euclid Avenue, Desk J4-1, Cleveland, OH 44115 (E-mail: weissa2@ccf.org). 2666-2736

Copyright © 2023 The Author(s). Published by Elsevier Inc. on behalf of The American Association for Thoracic Surgery. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). <https://doi.org/10.1016/j.jtc.2023.03.010>

**Abbreviations and Acronyms**

AUROC	= area under the receiver operating characteristic curve
AUPRC	= area under the precision-recall curve
AVR	= aortic valve replacement
CABG	= coronary artery bypass grafting
CPT	= Current Procedural Terminology
EHR	= electronic health record
EuroSCORE	= European System for Cardiac Operative Risk Evaluation
HCUP	= Healthcare Cost and Utilization Project
LR	= logistic regression
MSDW	= Mount Sinai Data Warehouse
MVR	= mitral valve replacement
MVRepa	= mitral valve repair
RFE	= recursive feature elimination
STS	= Society of Thoracic Surgeons
XGBoost	= eXtreme Gradient Boosting

Established risk models, such as the Society of Thoracic Surgeons (STS) risk score<sup>1-3</sup> and the European System for Cardiac Operative Risk Evaluation (EuroSCORE) I<sup>4</sup> and II,<sup>5</sup> are routinely used to assess a cardiac surgery patient's procedural risk. These models aid treatment decisions, profile individual surgeons and institutions, and provide benchmarks for quality improvement initiatives.

The STS models continue to contribute to the field of cardiac surgery; however, they are subject to limitations. Most notably, the models are only applicable to specific case types ("index procedures"), leaving out significant numbers of patients undergoing nonindex procedures or procedural combinations for which no model exists. Furthermore, a single patient's or institution's risk is not readily identified using population-derived, regression-based models.<sup>6,7</sup> Additional challenges include limitations on the number of multi-modal features evaluated, suboptimal handling of missing data, and insufficient incorporation of nonlinear and indirect relationships. As a consequence, the STS models, may fail to accurately predict specific patients with complicated pathologies who require unique tailored preoperative evaluations and complex surgeries.<sup>8-13</sup> Despite these limitations, the STS models continue to provide important benchmarks for hospitals to evaluate and improve their performance.

As cardiac surgery evolves, accurate patient-level risk prediction models applicable across the spectrum of patients and procedures will prove even more important. Machine learning-based models using big electronic health record (EHR) data offer one possible solution for improving

risk prediction. We hypothesized that a rigorous machine learning framework applied to routinely collected, multi-modal EHR data from a large, all-comers cardiac surgery patient cohort could be used to develop a personalized, institution-specific risk prediction model for mortality. We evaluated our model performance in a held-out patient cohort, and compared performance with the population-based STS risk models.

**MATERIALS AND METHODS**

Figure 1 shows the workflow of our study. This retrospective study was approved by the Icahn School of Medicine at Mount Sinai Institutional Review Board and included a waiver of informed consent (HS-15-00673, 10/29/2015).

**Study Cohort**

Consecutive cardiac surgeries performed at the Mount Sinai Hospital from June 1, 2011, to June 1, 2016, were identified using Healthcare Cost and Utilization Project (HCUP) clinical classification software codes<sup>14</sup> including heart valve procedures (HCUP = 43), coronary artery bypass graft (HCUP = 44), other operating room heart procedures (HCUP = 49), and aortic resection, replacement, or anastomosis (HCUP = 52). Duplicates, pediatric surgeries, vascular surgeries, thoracic surgeries, pericardial windows, reexplorations, and noncardiac surgeries were excluded. Patients with more than 1 operation during the study period were treated as separate observations if the procedures occurred during separate hospitalizations. For patients with multiple surgeries during a single hospitalization, only the first surgery was included.

The outcome, postoperative mortality, used the STS definition, namely, death during the same hospitalization as surgery, regardless of timing, or within 30 days of surgery regardless of venue.<sup>1-3</sup>

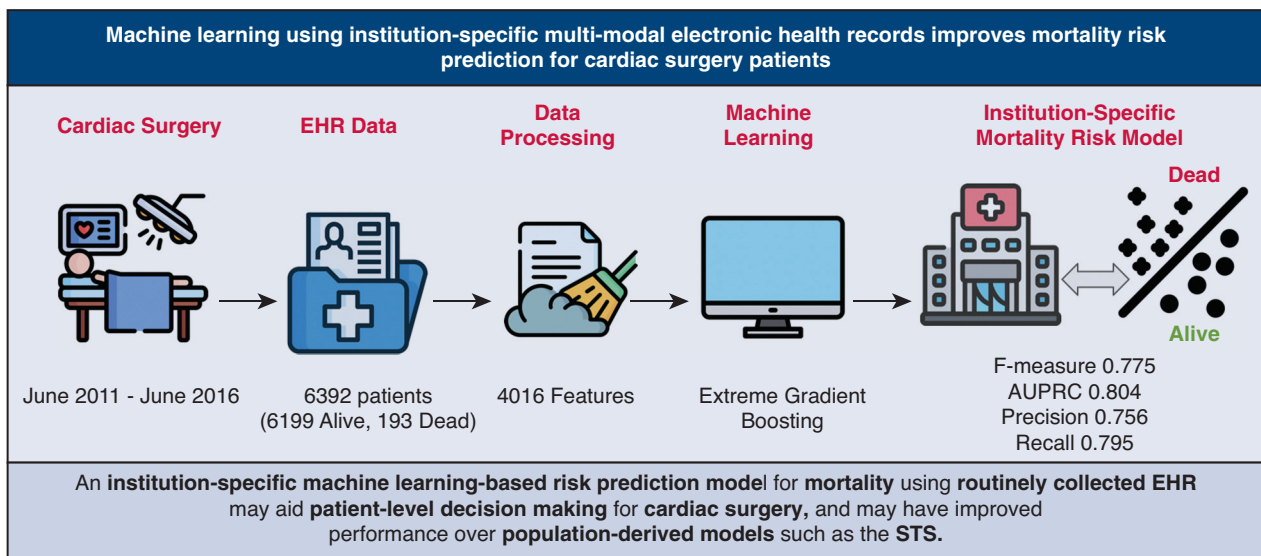
**Electronic Health Record Data Identification, Abstraction, and Cleaning**

A collection of multi-modal clinical features was identified and abstracted through feeds from multiple EHR source systems. After vetting by expert physicians, a final list of multi-modal features comprised of administrative, demographic, clinical, hemodynamic, imaging, laboratory, pharmacological, and procedural features was created (Appendix 1). All observations for all features were uploaded into open-source R software (R Foundation)<sup>15</sup> for systematic cleaning and assessment.

**Data Preparation and Preprocessing**

The generated multi-modal set of features ( $x = 4016$ ) consisted of 3883 preoperative and 133 intraoperative features, along with 1 postoperative outcome (mortality). We used the following multi-step data processing and machine learning strategy to develop a mortality risk prediction model using only the preoperative features and the cardiac surgery patient cohort described (Figure 2). In essence, the model was built without intraoperative surgical procedure or event data providing a true preoperative risk prediction model.

First, label encoding converted into numerical values each of the nonbinary categorical features. Then, the overall cohort ( $n = 6392$ ) was randomly split into an 80% training set ( $n = 5113$  patients) for model development and a 20% holdout test set ( $n = 1279$  patients) for final model validation. The upfront separation of data into the 2 sets ensured the predictive model was rigorously developed and independently validated. The same 3883 preoperative features were available for both sets. The mortality outcome was also available for both sets but used for model development within the training set and for validation within the test set.



\*AUPRC: Area Under Precision-Recall Curve; EHR: Electronic Health Records; STS: Society of Thoracic Surgeons

**FIGURE 1.** Machine learning using institution-specific multi-modal EHRs improves mortality risk prediction for cardiac surgery patients. For a cohort of consecutive cardiac surgery patients from 2011 to 2016, routine EHR data were identified and processed, creating a database of 6392 patients with 4016 features. After applying the machine learning algorithm XGBoost to the training data, a model for predicting mortality in cardiac surgery patients was generated, and this model was then evaluated on a completely independent test set. Various evaluation metrics were used to assess performance of the model, and the model's performance was then compared with that of the STS models for index case types. *EHR*, Electronic health record; *AUPRC*, Area under the precision-recall curve; *STS*, Society of Thoracic Surgeons; *XGBoost*, eXtreme Gradient Boosting.

In the training set, features missing more than 99% of their values were eliminated. Nonmissing categorical features whose values were the same for more than 99% of the patients were also eliminated, because these were likely to be uninformative for outcome prediction. Continuous features were normalized by converting their respective values into z-scores [(value-mean)/(standard deviation)]. To eliminate redundant features, Pearson's correlation coefficients were calculated between each feature pair using only their nonmissing values. Among any pair with correlation greater than 0.9, the feature with the lower correlation with mortality was eliminated. Following these steps, the preprocessed training set included 5113 patient observations with 804 features.

### Missing Value Imputation

The training set was further split into a 75% development set ( $n = 3834$ ) and a 25% validation set ( $n = 1279$ ) that were used to simulate the training and test sets. Next, a stepwise process was implemented to account for missing data patterns and to determine the best missing data cutoff to be included in the model. Using the development set, features with different maximum missing value percentages were identified, and the mean-mode imputation method was applied. Next, 4 machine learning classification algorithms—random Forest (RF), logistic regression (LR), Support Vector Machine, and eXtreme Gradient Boosting (XGBoost)—were trained on the imputed data. The resultant models were evaluated on the validation set using 6 evaluation metrics, namely, F-measure, precision, recall, accuracy, area under the precision-recall curve (AUPRC), and area under the receiver operating characteristic curve (AUROC).<sup>16</sup> Detailed explanations of the machine learning algorithms and the evaluation metrics are provided in [Appendix 2](#).

The imputation and evaluation process was repeated for 0% to 60% missing value levels in 5% increments. To eliminate the dependence on 1 random split, the entire process was repeated 100 times, and the average of each of the metrics for each level was determined. The complete results

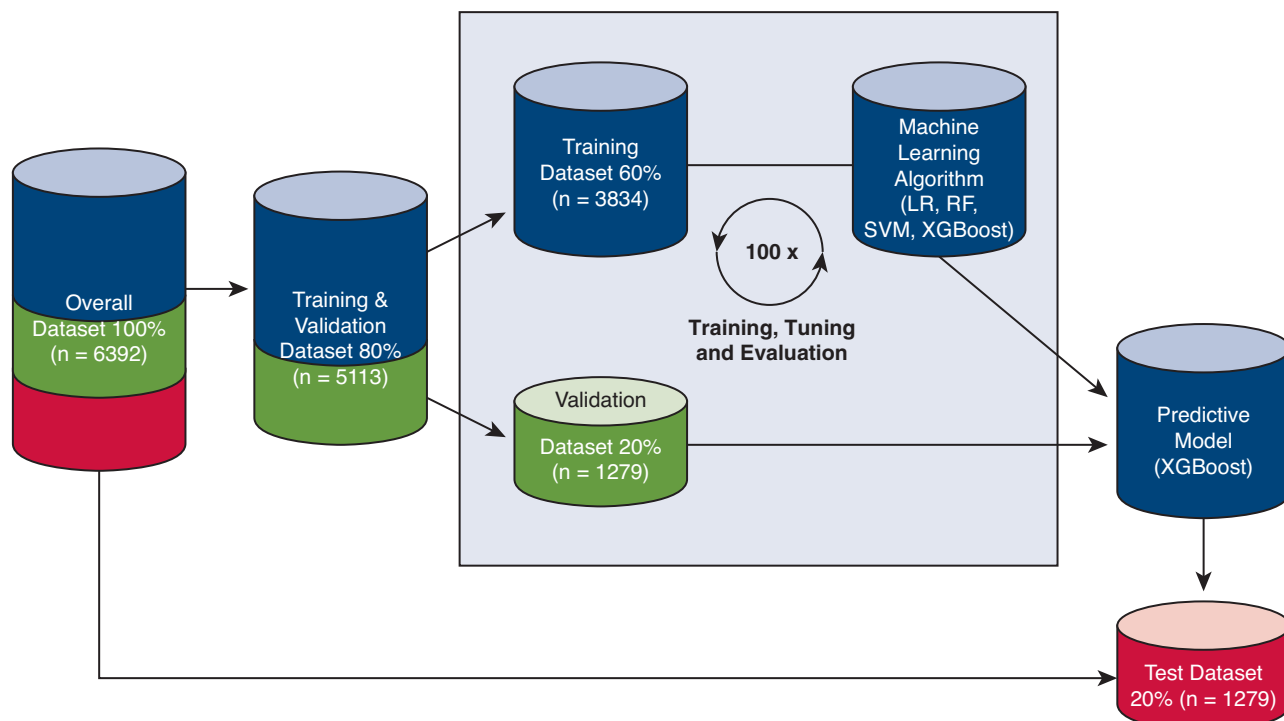
were then analyzed as to whether imputation would improve performance, and if so, at which cutoff. The final training set with these features included was used for the subsequent model development steps.

### Determination of Model Thresholds

Machine learning–based classification models predict a probability that each observation in the test set is labeled a “class.” For the conversion of probabilities to binary outcome predictions, this threshold is typically 0.5, with probabilities 0.5 or greater labeled 1 class (alive) and probabilities less than 0.5 labeled the other class (deceased). However, for a highly skewed mortality outcome, such a default threshold may result in poor prediction performance, because most patients would be assigned to the “deceased” classes.<sup>17</sup> For all of our candidate predictive models, this threshold was determined in a data-driven manner. Specifically, across all thresholds applicable to the probabilities produced by a candidate model on the corresponding validation set, we determined the final threshold as that which maximized the F-measure value for the minority (deceased) class to facilitate the model most likely to capture this more clinically relevant class.

### Feature Selection

To determine if a model trained on a smaller feature set would perform better, we tested the recursive feature elimination (RFE) algorithm<sup>18</sup> in a setup analogous to the missing value imputation. The imputed training set was split in a 75:25 ratio into development and validation sets. By using RFE with the same 4 classification algorithms on the development set, candidate models were built with top  $x\%$  of all features, with  $x$  decreasing from 100% to 5% in decrements of 5%. These candidates were evaluated on the validation set using the 6 evaluation metrics. This development-validation process was repeated 100 times, and the evaluated metrics were averaged to obtain the full range of results at different feature percentages. These results were used to define the most effective number of features to include in the final predictive model.



**FIGURE 2.** The machine learning–based risk prediction model development for postoperative mortality in cardiac surgery patients using only preoperative features. The overall dataset ( $n = 6392$  patients) was divided in an 80:20 ratio into training ( $n = 5113$ ) and test ( $n = 1279$ ) datasets. During model development, the training set was repeatedly randomly split in a 75:25 ratio into development ( $n = 3834$ ) and validation ( $n = 1279$ ) datasets. These datasets were used to train, tune, and evaluate several candidate mortality prediction models. Four classification algorithms (LR, RF, SVM, XGBoost) were trained on the development set to develop these candidates, which were then evaluated on the validation set. The best-performing algorithm (XGBoost) was used to train the final predictive model on the full training set (development set + validation set) and subsequently evaluated on the held-out test set. *LR*, Logistic regression; *RF*, random Forest; *SVM*, Support Vector Machine; *XGBoost*, eXtreme Gradient Boosting.

## Development and Evaluation of the Final Mortality Predictive Model

These steps yielded an optimally performing set of features with an acceptable level of missing values, and among them, the best-performing fraction of features that would be included in the final model. These analyses also identified which of the classification algorithms yielded the best performing final model. The final mortality prediction model was then built on the entire original training set (development set + validation set) using the chosen classification algorithm applied to the selected features and then evaluated on the holdout test set using the 6 evaluation metrics.

## Index Case-specific Model Performance and Comparison With Society of Thoracic Surgeons

The STS risk scores were generated at the time of surgery and stored in a prospective institutional database for the 7 index procedures: coronary artery bypass grafting (CABG), aortic valve replacement (AVR), mitral valve repair (MVRRepa), mitral valve replacement (MVR), AVR + CABG, MVRRepa + CABG, and MVR + CABG. For each index procedure, the risk scores (probabilities) for the patients in our training set with known alive/death status were used to determine the best-performing classification threshold using the same method as the candidate predictive models. The performance of the index-specific STS scores were calculated on the respective subsets of the test set using the same evaluation metrics. After the development and evaluation of the final model, we also determined its performance for the 7 index procedures and compared with the

performance of the corresponding STS risk models. The patients in the test set were identified as belonging to one of the index cases or as a non-STS index procedure. The final machine learning–based mortality prediction model was then evaluated individually on 5 STS index case cohorts (CABG, AVR, MVRRepa, AVR + CABG, and MVRRepa + CABG) in the test set. The MVR and MVR + CABG cohorts were not considered because of few events. Furthermore, STS scores were only available for 4487 patients, of whom 3562 and 925 were included in the training and test sets, respectively. Thus, the described evaluation and comparison were restricted to these patients.

## Statistical Analysis

For the overall cohort, qualitative features are denoted as frequencies and percentages, and quantitative features as medians and interquartile ranges. Baseline characteristics were compared between groups using the Student *t* test and chi-square test for continuous and categorical variables, respectively. All analyses were performed with publicly available software packages as noted in [Appendix 3](#). [Appendix 4](#) provides the details for hyperparameter tuning for the XGBoost model.

## RESULTS

### Study Cohort Characteristics

The overall patient cohort consisted of 6392 patients ([Figure E1](#)). For each patient, observations included 3883 preoperative features and 1 outcome (postoperative

mortality). [Table 1](#) provides summary statistics of clinically relevant multi-modal features for the overall patient cohort, training set, and test set. The median age was 64.7 years, and 4072 (63.7%) of the patients were male. The majority of patients identified as White ( $n = 3198$ , ~50.0%). The number of patients with a history of cardiac surgery was 885 (13.8%). There were 193 patients (3.0%) who died. There were 928 unique combinations of surgical procedures with the most frequent being CABG ( $n = 1585$ ), AVR ( $n = 528$ ), and MVRRepa plus tricuspid valve repair ( $n = 437$ ) ([Table E1](#)). There were also substantial non-STS index cases, including aortic root replacements, ascending aorta replacements, left ventricular assist devices, and cardiac transplants.

### Development and Performance of the Mortality Prediction Model

[Figure 2](#) provides the multi-step machine learning–based methodology used to develop and validate the mortality risk model. First, the overall cohort was randomly split in an 80:20 ratio into training ( $n = 5113$ ) and test ( $n = 1279$ ) sets to develop and validate the model, respectively. The distributions of the predictor features and the mortality outcome were similar between both sets ([Table 1](#)).

Within the training set, we assessed the level of missing values that could be reliably imputed. The results ([Figures 3 and E2](#)) demonstrated that including any imputed missing values would not lead to a performance improvement. Thus, we proceeded with only the 336 features with no missing values in the training set ([Table E2](#)). Next, RFE assessed if selecting a subset of the features yielded a better performing model. The results ([Figures E3 and E4](#)) demonstrated model performance plateauing with including approximately 10% of the 336 features, whereas a smaller subset of features did not improve performance. Therefore, no additional features were eliminated during the final model development as to be more inclusive and avoid overfitting.

In both of these steps, the XGBoost algorithm produced the best performing candidate model. Accordingly, we trained the final mortality prediction model by applying the XGBoost algorithm to the 336 features with no missing values in the original training set. The median of the optimized thresholds (0.261) of the XGBoost models was used as the threshold to convert the probabilities produced by the final model into deceased and alive labels. The final model was then evaluated on the held out test set. The results demonstrated that the model performed well across all metrics ([Table 2](#)).

Next, the information gain method<sup>19</sup> was used to identify the important features for the XGBoost model. Information gain represents the relative contribution of the corresponding feature to the model with higher values implying more importance to the model for generating predictions. [Figure 4](#) shows the 15 most important features

in the final model. These multi-modal features included preoperative medications (beta-blockers), comorbidities (hemodialysis and atrial fibrillation), administrative codes (acute kidney failure), and clinical acuity (intra-aortic balloon pump).

### Index Case Evaluation and Comparison With Society of Thoracic Surgeons Risk Scores

Finally, we sought to determine how the performance of the XGBoost model compared with the STS risk scores for index surgery types in the test set. [Table E3](#) describes the cohort stratified by index cardiac surgery procedure, as well as the number of patients who were alive and deceased.

[Table 3](#) provides a performance comparison of the XGBoost model with each of the STS models for their respective index case types within the test set. The XGBoost model outperformed the STS model for each of the 5 specific index procedure types with sufficient data. [Figure 5](#) provides confusion matrices detailing the correct/incorrect predictions made by the XGBoost and the respective STS risk models evaluated. In patients undergoing CABG, the XGBoost model correctly predicted all events except 1 false-positive, whereas the STS risk score had 5 false-positives and 2 false-negatives. For the AVR cohort, the XGBoost model correctly predicted all cases except for 1 false-negative, whereas the STS model had 21 false-positives. In particular, for all index surgery types, the STS models correctly predicted fewer deaths (true-positives), and their evaluation metrics for this important class were relatively low ([Table 3](#)).

## DISCUSSION

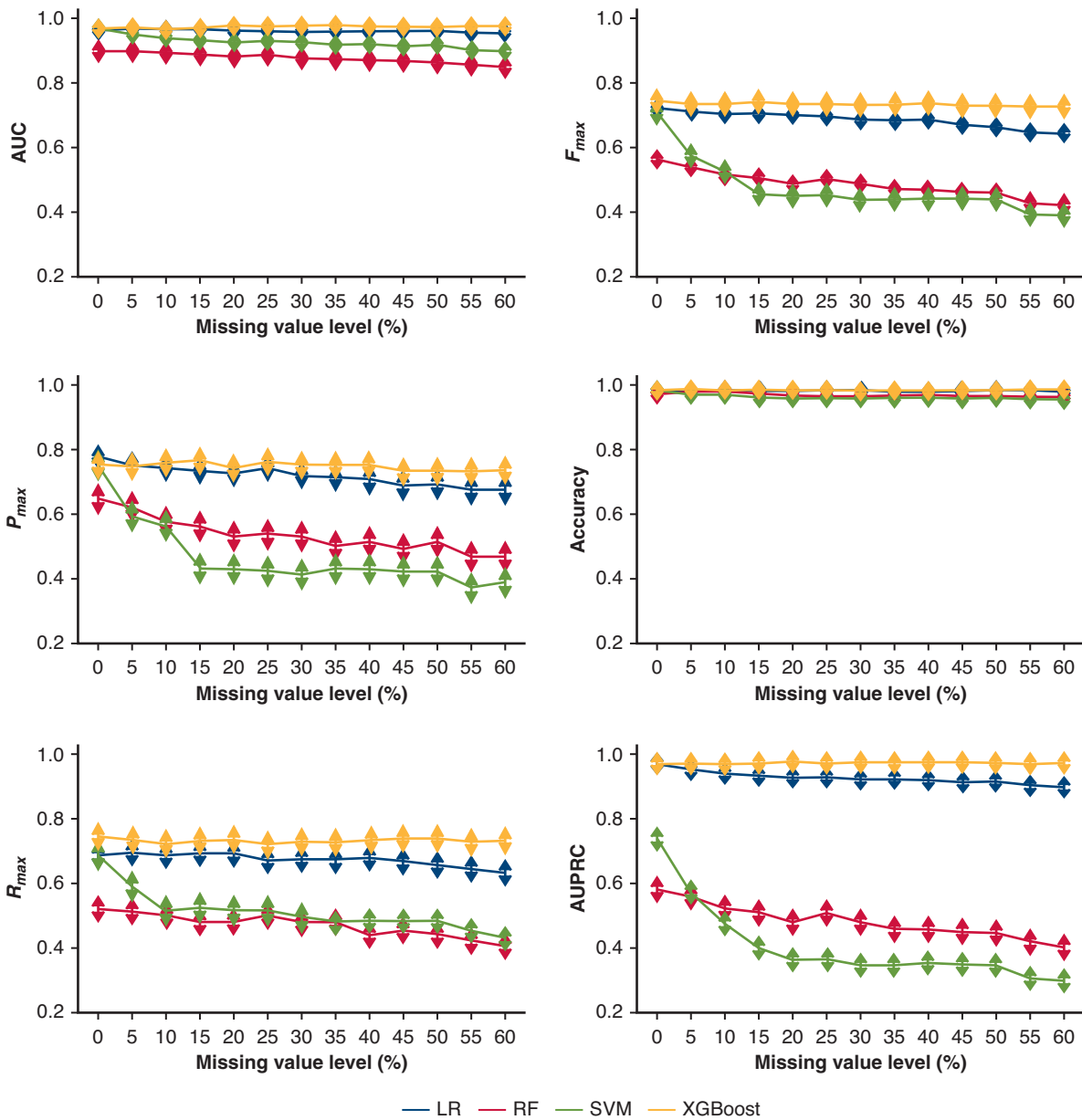
In this study, we rigorously developed an institution-specific, machine learning–based mortality risk prediction model for patients undergoing cardiac surgery using routinely collected EHR. The final XGBoost model outperformed the commonly used STS risk scores for mortality for all index case types with sufficient data. Although the STS models are the gold standard for mortality risk assessment in cardiac surgery, they may be inadequately designed for several surgical procedures, as well as combinations of surgical procedures that are performed routinely. For potential wider applicability, our model was developed from and evaluated on a range of cardiac surgery patients undergoing various combinations of procedures ([Table E1](#)). Furthermore, the cases included were from all adult cardiac surgeons over the study period to reflect the variability and complexity of a real-world, high-volume academic practice.

Our model was developed from the noisy data routinely collected in the EHR during clinical encounters using a fully data-driven methodology. The process by which the multi-modal feature set was developed was rigorous and labor intensive, and necessitated expert clinical oversight for data quality, organization, transparency, and reproducibility.

**TABLE 1. Statistical characteristics of the overall cohort and the split training and test sets used to develop and evaluate the mortality risk prediction model, respectively**

Feature	Overall (n = 6392)	#Missing n (%)	Training (n = 5113)	#Missing n (%)	Test (n = 1279)	#Missing n (%)	P value*
Age, y	64.7 (55.4-73.3)	0 (0.0%)	64.8 (55.6-73.5)	0 (0.0%)	64.7 (54.9-72.7)	0 (0.0%)	.920
Sex, male	4072 (63.7%)	0 (0.0%)	3259 (63.7%)	0 (0.0%)	813 (63.6%)	0 (0.0%)	.933
Weight, kg	77.0 (66.0-89.7)	0 (0.0%)	77.3 (66.0-90.0)	0 (0.0%)	77.0 (66.0-89.0)	0 (0.0%)	.565
BMI, kg/m <sup>2</sup>	26.9 (23.8-30.6)	118 (1.8%)	26.9 (23.9-30.7)	101 (2.0%)	26.6 (23.7-30.5)	17 (1.3%)	.176
Race		646 (10.1%)		513 (10.0%)		133 (10.4%)	.165
White	3198 (50.0%)		2553 (49.9%)		645 (50.4%)		
African American	679 (10.6%)		538 (10.5%)		141 (11.0%)		
Asian	321 (5.0%)		265 (5.2%)		56 (4.4%)		
Hispanic or Latino	61 (1.0%)		54 (1.1%)		7 (0.5%)		
Native American	17 (0.3%)		11 (0.2%)		6 (0.5%)		
Pacific Islander	14 (0.2%)		9 (0.2%)		5 (0.4%)		
Other	1456 (22.8%)		1170 (22.8%)		286 (22.4%)		
Diabetes	2217 (34.7%)	0 (0.0%)	1778 (34.8%)	0 (0.0%)	439 (34.3%)	0 (0.0%)	.787
HTN	4383 (68.6%)	0 (0.0%)	3509 (68.6%)	0 (0.0%)	874 (68.3%)	0 (0.0%)	.866
PVD	443 (6.9%)	0 (0.0%)	348 (6.8%)	0 (0.0%)	95 (7.4%)	0 (0.0%)	.471
Dialysis	236 (3.7%)	0 (0.0%)	182 (3.6%)	0 (0.0%)	54 (4.2%)	0 (0.0%)	.298
CVA	168 (2.6%)	0 (0.0%)	136 (2.7%)	0 (0.0%)	32 (2.5%)	0 (0.0%)	.827
Preoperative mechanical ventilation	76 (1.2%)	0 (0.0%)	62 (1.2%)	0 (0.0%)	14 (1.1%)	0 (0.0%)	.838
Pulmonary hypertension	1063 (16.6%)	0 (0.0%)	843 (16.5%)	0 (0.0%)	220 (17.2%)	0 (0.0%)	.568
Atrial fibrillation	1409 (22.0%)	0 (0.0%)	1138 (22.3%)	0 (0.0%)	271 (21.2%)	0 (0.0%)	.431
History of tobacco use	3162 (49.5%)	213 (3.3%)	2549 (49.9%)	175 (3.4%)	613 (48.0%)	38 (3.0%)	.171
Chest radiation	132 (2.1%)	0 (0.0%)	105 (2.1%)	0 (0.0%)	27 (2.1%)	0 (0.0%)	.985
Preoperative admission	3560 (55.7%)	0 (0.0%)	2856 (55.9%)	0 (0.0%)	704 (55.0%)	0 (0.0%)	.622
White blood cell, k/ $\mu$ L	7.1 (5.8-8.7)	756 (11.8%)	7.1 (5.8-8.7)	587 (11.5%)	7.1 (5.9-8.7)	169 (13.2%)	.651
Hematocrit, %	37.9 (33.9-41.3)	756 (11.8%)	37.9 (33.8-41.3)	587 (11.5%)	38.0 (33.9-41.3)	169 (13.2%)	.849
Platelets $\times 10^3$ , k/ $\mu$ L	199.0 (162.0-238.0)	758 (11.9%)	199.0 (163.0-238.0)	588 (11.5%)	196.0 (161.0-240.0)	170 (13.3%)	.293
Creatinine, mg/dL	1.0 (0.9-1.3)	844 (13.2%)	1.0 (0.9-1.3)	646 (12.6%)	1.0 (0.9-1.3)	198 (15.5%)	.379
Glomerular filtration rate	60.0 (52.8-60.0)	850 (13.3%)	60.0 (53.0-60.0)	650 (12.7%)	60.0 (52.2-60.0)	200 (15.6%)	.441
ASA Status		0		0		0	
1	2 (0.1%)	0 (0.0%)	1 (0.1%)	0 (0.0%)	1 (0.1%)	0 (0.0%)	
2	21 (0.3%)		18 (0.4%)		3 (0.2%)		
3	1844 (28.8%)		1454 (28.4%)		390 (30.5%)		
4	4399 (68.8%)		3539 (69.2%)		860 (67.2%)		
5	126 (2.0%)		101 (2.0%)		25 (2.0%)		.458
Reoperation	885 (13.8%)	0 (0.0%)	693 (13.6%)	0 (0.0%)	192 (15.0%)	0 (0.0%)	.192
Ejection fraction, %	59.0 (46.0-62.0)	2742 (42.9%)	60.0 (46.0-62.0)	2186 (42.8%)	59.0 (46.5-62.0)	556 (43.5%)	.630
Mean arterial pressure, mm Hg	89.8 (81.5-98.0)	486 (7.6%)	89.9 (81.8-98.0)	387 (7.6%)	89.4 (81.2-98.3)	99 (7.7%)	.289
Heart rate, bpm	72.0 (63.0-81.0)	415 (6.5%)	72.0 (63.0-81.0)	331 (6.5%)	72.0 (63.5-82.0)	84 (6.6%)	.522
Preoperative hospital LOS, d	1.0 (0.0-2.0)	0 (0.0%)	1.0 (0.0-2.0)	0 (0.0%)	1.0 (0.0-3.0)	0 (0.0%)	.519
Death	193 (3.0%)	0 (0.0%)	154 (3.0%)	0 (0.0%)	39 (3.0%)	0 (0.0%)	.983

Counts and percentages are presented for categorical features, and median values and interquartile ranges for continuous features. Also shown are *P* values denoting the statistical significance of the difference between the distributions of values of individual features between the training and test sets. *BMI*, Body mass index; *HTN*, hypertension; *PVD*, peripheral vascular disease; *CVA*, cerebrovascular accident; *ASA*, American Society of Anesthesiologists; *LOS*, length of stay. \*Comparisons between the values of the individual features in the training and test sets were carried out using the Student *t* test for continuous features and the chi-square test for categorical features.



**FIGURE 3.** Results of a data-driven assessment of the acceptable level of missing values in features that can be reliably included and imputed during model development for the deceased class. These levels varied from 0% to 60% in increments of 5% for each feature (X-axis). At each level, the mean-mode imputation method (mean for continuous features, mode for categorical features) was applied to the training dataset. Four classifiers (LR, RF, SVM, and XGBoost) were then trained on the imputed training set and evaluated on the validation set in terms of 6 metrics (AUROC, Accuracy,  $F_{max}$ ,  $R_{max}$ ,  $P_{max}$ , and AUPRC). Each algorithm was run 100 times for each missing data level, and the average of the performance was plotted for each metric for the deceased class. *AUROC*, Area under the ROC curve; *LR*, logistic regression; *RF*, random Forest; *SVM*, Support Vector Machine; *XGBoost*, eXtreme Gradient Boosting;  $F_{max}$ , maximum value of F-measure across all prediction score thresholds;  $P_{max}$ , value of precision at  $F_{max}$ ;  $R_{max}$ , value of recall at  $F_{max}$ ; *AUPRC*, Area under the precision-recall curve; *AUROC*, area under the receiver operating characteristic curve.

As EHRs evolve, automatic text mining of a priori defined discrete structured data points or natural language processing of unstructured clinical data can be leveraged to automatically select multi-modal risk features important in determining outcomes. These efforts when combined with data-driven analytics could theoretically populate

institution-specific risk prediction models that continuously update in real time to aid decision making.

XGBoost is an effective prediction algorithm that builds an ensemble of decision trees by iteratively focusing on harder to predict subsets of the training data.<sup>20</sup> Because of its systematic optimization-based design, this algorithm

**TABLE 2. Performance of the final XGBoost-based model for mortality risk prediction on the held out test set as measured in terms of 6 different evaluation metrics (F-measure, Precision, Recall, Accuracy, AUROC, and AUPRC)**

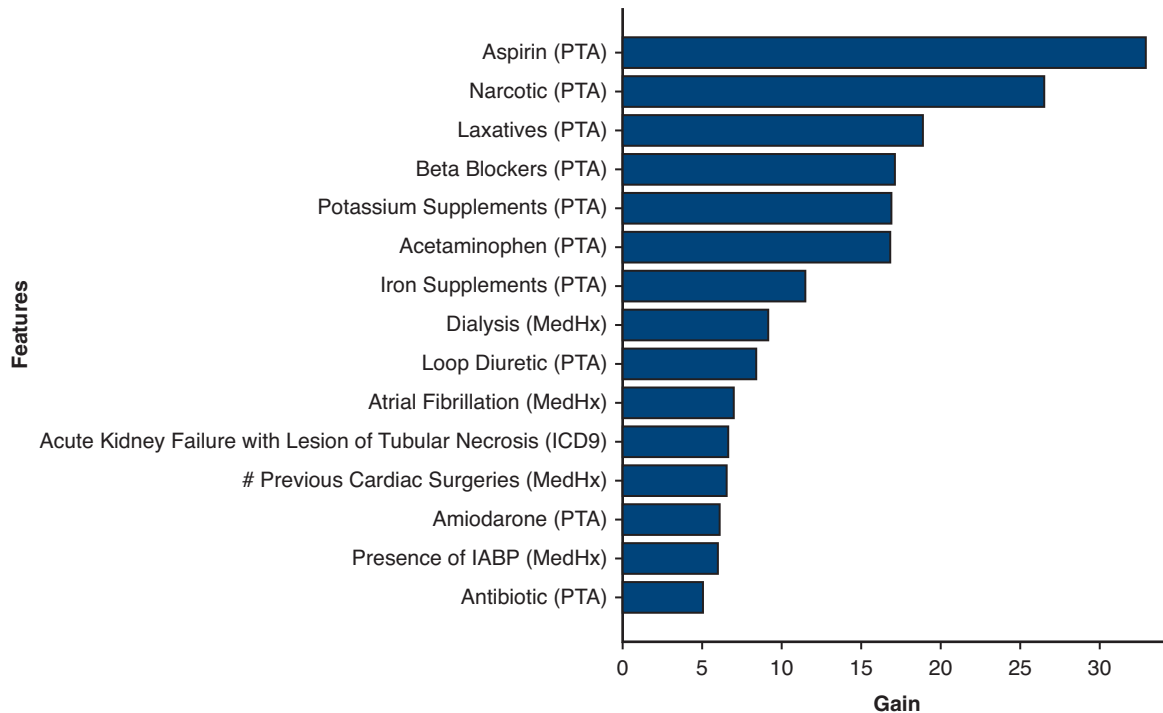
Metrics	XGBoost model's performance	Confidence interval (95%)
F-measure	0.775	0.689-0.857
Precision (positive predictive value)	0.756	0.646-0.861
Recall (sensitivity)	0.795	0.686-0.900
Accuracy	0.986	0.980-0.991
AUROC	0.978	0.964-0.989
AUPRC	0.804	0.709-0.890

The model was based on only preoperative features that had no missing values. *XGBoost*, eXtreme Gradient Boosting; *AUROC*, area under the receiver operating characteristic curve; *AUPRC*, area under the precision-recall curve.

has shown superior performance for predictive modeling involving structured data, findings consistent with our observations. Another noteworthy aspect of XGBoost is its ability to discern nonlearning relationships among features and individual feature importance through the information gain method.<sup>19</sup> Of the 15 most important features from

the XGBoost model (Figure 3), many have previously been associated with mortality risk including atrial fibrillation<sup>21</sup> and renal failure requiring hemodialysis,<sup>22</sup> as well as the presence of an intra-aortic balloon pump.<sup>23</sup> Previous cardiac surgery has also been shown to increase the risk of mortality when patients undergo reoperative surgery.<sup>24</sup> Several of the important features in the XGBoost model also appear in the STS risk model, including atrial fibrillation, acute kidney failure, hemodialysis, reoperation, and intra-aortic balloon pump, providing further validity. Of note, these 15 features were from a variety of EHR data types and sources, demonstrating the importance of multi-modal data to inform patient-level risk prediction.

Prediction models based on machine learning algorithms have been generated across clinical medicine with some demonstrating improved results over their standard of care counterparts.<sup>25-31</sup> In cardiac surgery, a number of studies have used machine learning-based classifiers to model operative risk,<sup>32-41</sup> with some demonstrating improved performance.<sup>33,37,40,41</sup> However, most limit the assessed patient cohort to highly selected surgical procedures, leaving unaddressed more complex combinations of procedures. Additionally, several studies only incorporated features that were structured for national registries, potentially



**FIGURE 4.** The 15 most important features in the XGBoost-based mortality risk prediction model. Information gain (X-axis) represents the relative contribution of the corresponding feature to the model. A higher gain value for one feature when compared with another implies that the former is more important to the model for generating predictions. This method was used to identify the most important features of the model, of which the top 15 are shown. *PTA*, Medications before admission; *MedHx*, medical history; *ICD-9*, International Classification of Diseases, Ninth Revision; *IABP*, intra-aortic balloon pump; *XGBoost*, eXtreme Gradient Boosting.



**TABLE 3. Performance comparison of the XGBoost-based model and the Society of Thoracic Surgeons mortality risk prediction models for 5 index cardiac surgery types**

Index procedure	F-measure (deceased class)		F-measure (alive class)		AUROC		AUPRC (deceased class)	
	XGBoost	STS	XGBoost	STS	XGBoost	STS	XGBoost	STS
CABG	0.800	0.000	0.998	0.989	1.000	0.580	1.000	0.017
AVR	0.667	0.087	0.995	0.880	1.000	0.888	1.000	0.042
MVRepa	0.750	0.000	0.997	0.990	0.967	0.914	0.676	0.219
AVR + CABG	0.667	0.400	0.985	0.954	0.939	0.894	0.633	0.196
MVRepa + CABG	1.000	0.000	1.000	0.990	1.000	0.755	1.000	0.038

Results are not shown for the MVR and MVR + CABG surgery types as their training or test sets had zero deaths. F-measure is the harmonic mean of the class-specific Precision and Recall measures, and ranges from 0 to 1, with a higher value indicating superior classification performance. *AUROC*, Area under the receiver operating characteristic curve; *AUPRC*, area under the precision-recall curve; *XGBoost*, eXtreme Gradient Boosting; *STS*, Society of Thoracic Surgeons; *CABG*, coronary artery bypass grafting; *AVR*, aortic valve replacement; *MVRepa*, mitral valve repair; *MVR*, mitral valve replacement.

missing features not collected or more granular center-specific features that may contribute to residual unexplained risk.<sup>6</sup>

Allyn and colleagues<sup>42</sup> retrospectively studied 6520 cardiac surgery patients from a single institution to predict postoperative mortality. The study included patients undergoing elective cardiac surgery with cardiopulmonary bypass, excluding high-risk emergency and noncardiopulmonary bypass cases. They evaluated EuroSCORE I and II, an LR model using EuroSCORE II covariates, 4 different machine learning models, and an ensemble machine learning model. The ensemble model outperformed the EuroSCORE II as assessed by the AUROC score, an evaluation metric with limited accuracy when applied to datasets with significant outcome class imbalance.<sup>43</sup> This study only included EuroSCORE features, thereby potentially missing important multi-modal and institution-specific features. In contrast, our model was developed using a multi-modal feature set, and its performance was assessed in terms of 6 evaluation metrics more standardly reported in machine learning-based classifications.

Two studies from the same authors assessed machine learning algorithms for predicting outcomes for STS index cases.<sup>44,45</sup> The first study included 11,190 patients from a single institution undergoing only STS index cases. XGBoost was used to develop a risk model for operative mortality,<sup>44</sup> demonstrating improved performance over the STS models. However, this study was limited by including only features captured for the STS models rather than multi-modal, center-specific features. The second study assessed the performance of an XGBoost mortality risk model for a national cohort of STS index patients undergoing isolated surgical AVR.<sup>45</sup> Although the number of patients included was large ( $n = 243,142$ ), the study was also limited by the same fixed STS feature set and lacked a fully separate test set, which is essential for assessing a predictive model's validity.

Our study methodology has the potential to be replicated in different institutions using patient- and center-specific

data. Individual cardiac surgery practices benefit from tracking and comparing their outcomes with the national benchmarks derived from population-based risk scores such as the STS. However, as our study demonstrates, there is also benefit from incorporating hospital- and patient-specific data to provide an accurate prediction for an individual patient. This personalized risk prediction may help streamline patient selection and treatment options. Enhanced risk prediction may also enable the identification of unique center-specific features that can be incorporated into quality improvement initiatives.

### Study Limitations

Because this is a retrospective, single-institution study, the methodology should be validated in a prospective manner before being used to inform decision making. Additionally, the relatively small number of outcome events may affect the stability of some evaluation metrics. Relatedly and despite a large sample size for a single institution, we lacked sufficient outcome events to assess model performance for 2 STS index types. Moreover, with a large number of features with missing data, this may have resulted in important features being excluded from the final feature set that potentially contribute to residual uncharacterized risk. Last, the optimization parameters inherent to the different machine learning algorithms used may have further improved performance, but this was not attempted to avoid overfitting.

### CONCLUSIONS

Using a rigorous, data-driven, machine learning methodology applied to routinely collected, multi-modal EHR data on a diverse cardiac surgery patient cohort from a single institution, we developed an XGBoost-based mortality risk prediction model. This model demonstrated improved performance over the STS models for index cases. Our results show that data-driven, institution-specific models developed with machine learning algorithms may improve

	XGBoost Model	STS Model
<b>CABG</b>		
<b>AVR</b>		
<b>MVRepa</b>		
<b>CABG + AVR</b>		
<b>CABG + MVRepa</b>		

**FIGURE 5.** Confusion matrices enumerating correct and erroneous predictions from the XGBoost and STS models for 5 index cardiac surgery types. Shown in these matrices are the numbers of true-negatives (patients who were alive and predicted to be so), false-negatives (patients who were dead but predicted to be alive), false-positives (patients who were alive but were predicted to be dead), and true-positives (patients who were dead and predicted to be so) for each of the index types. The MVR and MVR + CABG types are not shown because their training or test cases had zero deaths. Patients who did not have an STS score calculated within the institutional database were not included in the confusion matrices, which is why the numbers of patients in these matrices are different than those in [Table E1](#). *XGBoost*, eXtreme Gradient Boosting; *STS*, Society of Thoracic Surgeons; *CABG*, coronary artery bypass grafting; *TN*, true-negative; *FP*, false-positive; *FN*, false-negative; *TP*, true-positive; *AVR*, aortic valve replacement; *MVRepa*, mitral valve repair; *MVR*, mitral valve replacement.

risk prediction in cardiac surgery and provide complementary insights to traditional population-based models.

### Conflict of Interest Statement

The Icahn School of Medicine at Mount Sinai receives royalty payments from Edwards Lifesciences and Medtronic for intellectual property related to Dr Adams' involvement in the development of 2 MVrepa rings and 1 tricuspid valve repair ring. Dr Adams is the National Co-Principal Investigator of the CoreValve United States Pivotal Trial, which is supported by Medtronic. Ravi Iyengar had a consulting agreement with Tectonic Therapeutics in 2021. All other authors reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

The authors thank Yan Chak Li for help with the Graphical Abstract and [www.flaticon.com](http://www.flaticon.com), and creators RaftelDesign, xnim-rodx, Freepik, and juicy\_fish for several icons included in the abstract.

### References

- Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009;88:S2-22.
- O'Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac Surg.* 2009;88:S23-42.
- Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3—valve plus coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009;88:S43-62.
- Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg.* 1999;16:9-13.
- Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardiothorac Surg.* 2012;41:734-44; discussion 744-5.
- Raza S, Sabik JF III, Rajeswaran J, Idrees JJ, Trezzi M, Riaz H, et al. Enhancing the value of population-based risk scores for institutional-level use. *Ann Thorac Surg.* 2016;102:70-7.
- Nowicki ER. What is the future of mortality prediction models in heart valve surgery? *Ann Thorac Surg.* 2005;80:396-8.
- Chan V, Ahrari A, Ruel M, Elmestekawy E, Hynes M, Mesana TG. Perioperative deaths after mitral valve operations may be overestimated by contemporary risk models. *Ann Thorac Surg.* 2014;98:605-10; discussion 610.
- Kennedy JL, LaPar DJ, Kern JA, Kron IL, Bergin JD, Kamath S, et al. Does the Society of Thoracic Surgeons risk score accurately predict operative mortality for patients with pulmonary hypertension? *J Thorac Cardiovasc Surg.* 2013;146:631-7.
- Alnajjar A, Chatterjee S, Chou BP, Khabsa M, Rippstein M, Lee VV, et al. Current surgical risk scores overestimate risk in minimally invasive aortic valve replacement. *Innovations.* 2021;16:43-51.
- Iturra SA, Suri RM, Greason KL, Stulak JM, Burkhart HM, Dearani JA, et al. Outcomes of surgical aortic valve replacement in moderate risk patients: implications for determination of equipoise in the transcatheter era. *J Thorac Cardiovasc Surg.* 2014;147:127-32.
- Vassileva CM, Aranki S, Brennan JM, Kaneko T, He M, Gammie JS, et al. Evaluation of the Society of Thoracic Surgeons online risk calculator for assessment of risk in patients presenting for aortic valve replacement after prior coronary artery bypass graft: an analysis using the STS adult cardiac surgery database. *Ann Thorac Surg.* 2015;100:2109-15; discussion 2115-6.
- Barili F, Pacini D, Grossi C, Di Bartolomeo R, Alamanni F, Parolari A. Reliability of new scores in predicting perioperative mortality after mitral valve surgery. *J Thorac Cardiovasc Surg.* 2014;147:1008-12.
- Healthcare Cost and Utilization Project (HCUP); 2013. Accessed November 1, 2019. <https://www.ahrq.gov/data/hcup/index.html>
- R Core Team. *R: A language and environment for statistical computing.* Accessed May 8, 2023. <https://www.R-project.org/>
- Lever J, Krzywinski M, Altman N. Classification evaluation. *Nat Methods.* 2016;13:603-4.
- Ishwaran H, Blackstone EH. Commentary: dabblers: beware of hidden dangers in machine-learning comparisons. *J Thorac Cardiovasc Surg.* 2022;164:2088-90.
- Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn.* 2002;46:389-422.
- Brownlee J. *XGBoost With Python: Gradient Boosted Trees with XGBoost and Scikit-Learn.* Machine Learning Mastery; 2016.
- Chen T, Guestrin C. *XGBoost: a scalable tree boosting system.* Paper presented at: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California;* August 2016.
- Quader MA, McCarthy PM, Gillinov AM, Alster JM, Cosgrove DM III, Lytle BW, et al. Does preoperative atrial fibrillation reduce survival after coronary artery bypass grafting? *Ann Thorac Surg.* 2004;77:1514-22; discussion 1522-4.
- Aljohani S, Alqahtani F, Almufata A, Boobes K, Modi S, Alkhouli M. Trends and outcomes of aortic valve replacement in patients with end-stage renal disease on hemodialysis. *Am J Cardiol.* 2017;120:1626-32.
- Christenson JP, Simonet F, Schmuziger M. The effect of preoperative intra-aortic balloon pump support in high risk patients requiring myocardial revascularization. *J Cardiovasc Surg.* 1997;38:397-402.
- Elbadawi A, Hamed M, Elgendy IY, Omer MA, Ogunbayo GO, Megaly M, et al. Outcomes of reoperative coronary artery bypass graft surgery in the United States. *J Am Heart Assoc.* 2020;9:e016282.
- Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920-30.
- Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. *Proc IEEE Inst Electr Electron Eng.* 2016;104:444-66.
- Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med.* 2016;44:368-74.
- Kessler RC, van Loo HM, Wardenar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry.* 2016;21:1366-71.
- Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med.* 2016;23:269-78.
- Varghese B, Chen F, Hwang D, Palmer SL, De Castro Abreu AL, Ukimura O, et al. Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images. *Sci Rep.* 2019;9:1570.
- Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health.* 2020;2:e516-25.
- Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Comput Biomed Res.* 1993;26:220-9.
- Nilsson J, Ohlsson M, Thulin L, Högglund P, Nashef SA, Brandt J. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. *J Thorac Cardiovasc Surg.* 2006;132:12-9.
- Rowan M, Ryan T, Hegarty F, O'Hare N. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. *Artif Intell Med.* 2007;40:211-21.
- Peng SY, Peng SK. Predicting adverse outcomes of cardiac surgery with the application of artificial neural networks. *Anaesthesia.* 2008;63:705-13.
- Loghmanpour NA, Druzdzel MJ, Antaki JF. Cardiac Health Risk Stratification System (CHRiSS): a Bayesian-based decision support system for left ventricular assist device (LVAD) therapy. *PLoS One.* 2014;9:e111264.
- Loghmanpour NA, Kanwar MK, Druzdzel MJ, Benza RL, Murali S, Antaki JF. A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality. *ASAIO J.* 2015;61:313-23.

38. LaFaro RJ, Pothula S, Kubal KP, Inchiosa ME, Pothula VM, Yuan SC, et al. Neural network prediction of ICU length of stay following cardiac surgery based on pre-incision variables. *PLoS One*. 2015;10:e0145395.
39. Smedira NG, Blackstone EH, Ehrlinger J, Thuita L, Pierce CD, Moazami N, et al. Current risks of HeartMate II pump thrombosis: non-parametric analysis of Interagency Registry for Mechanically Assisted Circulatory support data. *J Heart Lung Transplant*. 2015;34:1527-34.
40. Delen D, Oztekin A, Kong ZJ. A machine learning-based approach to prognostic analysis of thoracic transplantations. *Artif Intell Med*. 2010;49:33-42.
41. Oztekin A, Delen D, Kong ZJ. Predicting the graft survival for heart-lung transplantation patients: an integrated data mining methodology. *Int J Med Inform*. 2009;78:e84-96.
42. Allyn J, Allou N, Augustin P, Philip I, Martinet O, Belghiti M, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS One*. 2017;12:e0169772.
43. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10:e0118432.
44. Kilic A, Goyal A, Miller JK, Gjekmarkaj E, Tam WL, Gleason TG, et al. Predictive utility of a machine learning algorithm in estimating mortality risk in cardiac surgery. *Ann Thorac Surg*. 2020;109:1811-9.
45. Kilic A, Goyal A, Miller JK, Gleason TG, Dubrawski A. Performance of a machine learning algorithm in predicting outcomes of aortic valve replacement. *Ann Thorac Surg*. 2020;111:503-10.

**Key Words:** cardiac surgery, clinical outcomes, electronic health records, machine learning, prediction modeling, risk prediction

## APPENDIX 1. EHR DATA IDENTIFICATION, ABSTRACTION, AND CLEANING

A collection of multi-modal clinical features was identified and abstracted through feeds from enterprise data warehouse, the MSDW, additional source systems including the EHR (Epic, Epic Systems Corporation), and from custom internal cardiology imaging databases (cardiac catheterization and echocardiography). The following process provides a detailed look at the data identification, abstraction, and cleaning process that were performed to enable an analyzable dataset that included routine granular clinical data for all patients who underwent cardiac surgery.

Where applicable, each feature was obtained with a time stamp, and features with repeated measurements over time were chronologically organized. All observations for all features were uploaded into open-source R software (R Foundation),<sup>15</sup> where they were systematically cleaned and assessed for clinical relevance. If appropriate, features were transformed to more clinically relevant interpretations and subsequently recorded to enable transparency and reproducibility. Metadata for individual features were assessed to identify potential erroneous or outlier values, as well as to determine the frequency of unknown or missing values. Each feature and its associated range of values were systematically verified by a trained cardiothoracic surgeon (A.J.W.).

The description of each surgical procedural was manually abstracted and recorded from individual operative reports after harmonizing with the appropriate patient observation in the database. A new file was created that characterized the specific granular components of each operative procedure for each patient. This allowed identification of the surgical detail in a binary manner for each observation in a single location that also enabled quality checks and reproducibility. After correctly labeling and confirming all included surgical procedures, this yielded a total of 6392 open surgeries as unique patient observations for the study cohort.

### Generation of Patient Masterlist

Patients who had undergone cardiac surgery at the Mount Sinai Hospital during the period of June 1, 2011, to June 1, 2016, were identified using HCUP clinical classification software codes. By using enterprise-wide data warehouse sources, demographic information was queried for all the matching entries in the cohort. After reviewing, the data were cleaned from a semi-structured format to a structured format with more uniformity. Additionally, it was determined that multiple observations that were captured using the HCUP codes were not actual cardiac surgeries or performed by a cardiac surgeon and therefore would need to be removed. All entries with a noncardiac surgery primary

surgeon were then removed, and the remaining surgeons were standardized in their identified values. All noncardiac surgery cases were also identified and removed.

We then checked for redundant observations in EPIC. For patients who had multiple surgeries (observations) during the same visit (same MRN and VISIT ID but different case name), we included only the first surgery during that visit and excluded any subsequent same hospitalization surgeries. After identifying feeds with demographic information from the MSDW, an inclusive list of demographic variables for each of the patients included in the cohort was harmonized to the master list. These variables were then cleaned, formalized, and standardized within the master list.

To determine the specific procedures that patients underwent and because CPT codes do not always accurately capture the granular level of detail needed for specific procedural identification, we undertook a manual (expert cardiothoracic surgeon) chart review of the operative notes to ascertain specific procedural details as detailed below: A file was created with the 6400 observations (only later became 6392 after this manual process eliminated 8 more patients from the cohort) along with the case name, MRN, VISIT ID, service date, and the 5 CPT columns from the master list obtained along with the demographic data with the corresponding values harmonized to each specific observation. Additional surgical procedure features were then created using clinical expertise corresponding to the different procedural components that could occur in surgery. Using the service date (date of cardiac surgery) and MRN as matching criteria, a manual data abstraction was taken for each of the 6392 patients. Each of the values for Preoperative Diagnosis, Postoperative Diagnosis and Surgeon Operative Description was copied verbatim from their respective operative note for each corresponding observation. Any discrepancies were noted in separate columns (labeled Notes 1 and Notes 2) for later reference if needed. The different procedural components were then filled out for the newly created procedure spreadsheet based on the Surgeon Operative Description. Following this, a manual entry of the corrected CPT codes (CPT1New–CPT8New) was done by A.J.W. (cardiothoracic surgeon) that corresponded with his interpretation of what CPT codes should be coded for that respective observation's procedure. A new column was created called Incorrect. This column took the codes from cpt1-5 and compared with CPT1New–CPT8New, and if the codes differed for each entry a 1 was recorded in the Incorrect column. If the codes matched between the 2 columns, then a 0 was inserted in that column for that respective observation. This allowed us to understand if the codes that were used for billing purposes were appropriately representative of the surgery that was performed.

The format of this spreadsheet facilitated transparency to the interpretation of the surgeries performed but also provided an easier and more consolidated way of looking at the data from each patient without having to go back into each patient's chart. This allowed identification of the surgical detail of each observation in this spreadsheet and permit repeated interpretation if needed. After all 6392 observations were appropriately labeled, 2 to 3 manual reviews that were separated in time were undertaken to fix incorrect entries. This process allowed for capture of errors due to the manual abstraction. After the final review, these new surgical procedure details were harmonized with the master list. Next, the format for dates and times was standardized and additional important clinical variables that involved lengths of time were created such as preoperative length of stay and hospital length of stay. Any additional redundant features were removed. Following merging with the master list, we were left with a final master list document that could be used for further harmonization.

### Processing of Administrative Features

We obtained data from the MSDW of all the International Classification of Diseases, 9th Revision (ICD-9) and International Classification of Diseases, 10th Revision (ICD-10) codes labeled as secondary diagnoses that were recorded for each observation matched to their VISIT-ID. These codes were captured at the time of the patient encounter for billing purposes (system data were extracted from Eagle). Each observation in this list (110,005 individual observations in the data feed) comes with an MRN, VISIT ID, and service date that match those found in the final master list. The aim was to identify all of the diagnosis codes present on admission that indicated what medical problems the patients had before being admitted. Additionally, the medical problems that developed at the hospital acquired were also cleaned and placed into a separate file. Issues that arose when reviewing these raw data:

Needed to first obtain all codes for all possible ICD-9 and ICD-10 codes and be able to make sure that the codes and descriptions matched. If they did not match or if a description did not have a code associated, it was removed.

Removed Duplicate Entries and Entries Not Applicable to a Present on Admission Code

Needed to determine if all 6392 patients had observations in this file and if any observations in this file were not for the 6392 patients in our cohort.

Needed to convert all of the ICD-10 codes to ICD-9 codes. This involved direct conversion when applicable but also some codes were identified as needing manually converted by expert opinion due to either coding for multiple different conversions or not coding for any particular ICD-9 or ICD-10 code.

Separated out the ICD-9 codes associated with the present on admission and those who were not or unable to be determined.

Created a new file where each diagnosis code was a feature. For each feature, the presence of that feature received a 1 in the corresponding observation and a 0 if the feature was not present.

Created a frequency table for all of the new features to determine the percentage of patients who had each individual feature.

The 2 Secondary Diagnosis final master list (Present on Admission and Hospital Acquired) documents were generated in the same format as the finalmasterlist document to facilitate database merging by VISIT ID. This resulted in creation of the following 4 files:

```
finalmasterlist_DiagnosisSecondaryPresentOnAdmission
finalmasterlist_DiagnosisSecondaryHospitalAcquired
frequencytable_DiagnosisSecondaryPresentOnAdmission
frequencytable_DiagnosisSecondaryHospitalAcquired.
```

### Processing of Admitting and Principal Diagnosis

We received a data feed from the MSDW of all the ICD-9 and ICD-10 codes labeled as admitting or principal diagnosis that were recorded for each observation matched to their VISIT-ID. These codes were captured at the time of the patient encounter for billing purposes (system data extracted from was Eagle). Each observation in this list (123,006 individual observations in the data feed organized as 10 features) comes with an MRN, VISIT ID, and service date that match those found in the finalmasterlist. The aim was to identify the admitting and principal diagnosis codes that indicated what medical problems the patient was being admitted to the hospital for (surrogate for chief symptom because this was unable to be reliably obtained). Additionally, the codes that were listed as Reason for Visit or labeled as H (not present at admission) were removed. Issues that arose when reviewing the raw data:

Needed to first obtain all codes for all possible ICD-9 and ICD-10 codes and be able to make sure that the codes and descriptions matched. If they did not match or if a description did not have a code associated, it was removed.

Removed duplicate entries and entries not applicable to a present on admission code for either Admitting or Principal Diagnosis as well as remove all entries that were listed as Reason for Visit.

Needed to determine if all 6392 patients had observations in this file and if any observations in this file were not for the 6392 patients in our cohort.

Needed to convert all of the ICD-10 codes to ICD-9 codes. This involved direct conversion when applicable, but also some codes were identified as needing manually

converted by expert opinion due to either coding for multiple different conversions or not coding for any particular ICD-9 or ICD-10 code.

Separated out the ICD-9 codes associated with the present on admission and those who were not or unable to be determined.

For an individual observation (1-6392), it was possible to have more than 1 principal diagnosis code with the corresponding description and hospital acquired value (only P in this case). Therefore, we created a new file with the 6392 individual observations as matched by VISIT ID and the columns of features were as follows.

VISIT ID

Case name

MRN

Service date

Admitting Diagnosis ICD-9 Code

Admitting Diagnosis Description

Admitting Hospital Acquired (value = P)

Principal Diagnosis ICD-9 Code 1

Principal Diagnosis Description 1

Principal Hospital Acquired 1 (value = P)

Principal Diagnosis ICD-9 Code 2

Principal Diagnosis Description 2

Principal Hospital Acquired 2 (value = P)

Principal Diagnosis ICD-9 Code 3

Principal Diagnosis Description 3

Principal Hospital Acquired 3 (value = P)

For those observations that did not have more than 1 principal diagnosis ICD-9 code, then the values of the feature for that corresponding observation were labeled as not available (ie, missing).

Created a frequency table for all of the new features to determine the percentage of patients who had each individual feature for both admitting and principal diagnosis.

The final file generated was in the same format as the finalmasterlist document to facilitate later merging of the databases by VISIT ID. Along with the frequency distribution file, this resulted in creation of the following files:

finalmasterlist\_DiagnosisAdmittingPrincipal

frequencytable\_DiagnosisAdmittingPrincipal

### Processing of Intraoperative Time Features

This data feed was obtained from the intraoperative anesthesia record (CompuRecord) used at Mount Sinai and had 4 features with 279,219 observations. Each of the observations was an event that occurred in the operating room during a specific patient's procedure (Case Name). There were 94 different event types as values for the feature named Label and these were time stamped, indicating a chronological series of events that transpired in the operating room. The database had the following 4 features:

Case Name

Time stamp (exact date, hour, and minute) that corresponded with the Label feature

ItemID that corresponds with feature 4

Label (an event that occurred in the operating room)

The values for the feature Label (feature 4) were recorded by an anesthesiologist in real-time so this required vetting of input errors and lack of standardization of inputs. This was done by looking at the presence (or absence) of each value for Label as it occurred for each patient. These intraoperative events were then used to create clinically relevant features that characterized important time intervals in the surgery. A new file was created that would record the various features created from Events and the file was named "Intraoperative Times." This database input the 6392 individual observations from masterlist. After cross-referencing to the 6392 observations from finalmasterlist, all entries that did not correspond by Case Name were eliminated from Events.

The following features were created for the database:

Intraoperative Times

MRN

VISIT ID

Case Name (matched by this identifier)

Service Date

Procedure Start Time (date/time)

Procedure Finish Time (date/time)

Time of Surgery (min)

Time of Surgery Missing (0 or 1)

Total Cardiopulmonary Bypass Time (min)

First Cardiopulmonary Bypass Time (min)

Subsequent Cardiopulmonary Bypass Time (min)

Number of Cardiopulmonary Bypass Runs (integers)

Cardiopulmonary Bypass Time Missing (0 or 1)

Total Aortic Crossclamp Time (min)

Number of Crossclamps Performed (integer)

Crossclamp Time Missing (0 or 1)

Circulatory Arrest Time (min)

Circulatory Arrest Time Missing (0 or 1)

Where applicable for an above feature, the distribution of results for a created column was checked to make sure there were no missense values present.

### Processing of Preoperative In-Hospital Medications Administered

This raw file contains information pertaining to observations for medications administered to a patient who had been admitted to the hospital before surgery. The file included 10 features with 615,504 observations with 10 features as outlined below:

MRN

VISIT ID

Service Date

Material Name—specific name of medication with concentration of medication where applicable (1890 different values for this feature)

Pharmacy Class—the overall class of drugs that the medication belonged to (267 different values for this feature)

Pharmacy Subclass—the subclass within the Pharmacy Class that the medication belonged into (422 different values for this feature)

Date Given—a time and date stamp for when the medication was administered

Units—numerical dose administered

Route—how the medication was administered (67 different values)

Action—a specific action taken for each observation of a drug (Values included ê Others, Restarted, Given)

Careful review of the range of values as well as the format of values for each of the columns was performed. Date entries without a complete time/date stamp were removed so as to only include administered medications that were confirmed to be given before the operation. Any observation that occurred after the feature from the Intraoperative Times sheet labeled “Patient In OR” was eliminated. This reduced the number of observations from 615,504 to 610,152 for 12 features (with addition of Case Name and Patient In OR Time (labeled Time)). A new file was created ê finalmasterlist\_PreoperativeInHospitalMedAdministration that had the 6392 specific observations from finalmasterlist as denoted by MRN, VISIT ID, Case Name, and Service Date. Because patients who were admitted preoperatively were admitted for varying lengths of time before surgery, we looked at the frequency distribution of these times to determine a clinically relevant period where medications had been administered before surgery to include in the file (only included observations of medications administered within 1 week preoperatively).

Next, the following features were created due to their unique clinical significance:

Parenteral Nutrition Within 1 wk

Inotropic Support Within 48 h

Heparin Infusion Within 24 h

Nesiritide Infusion Within 48 h

Beta-Blocker Within 24 h

The Pharmacy Subclass was used as the feature set that the patients Preoperative In-hospital Medication Administration would be derived from. This was determined after reviewing the mostly unstructured and numerous unique features of Material Name as well as only 267 unique values for the Pharmacy Class. The Pharmacy Subclass had 422 unique features initially; however, after cleaning and imputing values from material name where Class and Subclass had none these results changed. A file consisting of a frequency table was created for each of the Pharmacy Subclass Values with the overall frequency at the top and the rows with the individual Material Names and their frequencies within each Pharmacy Subclass. The individual

Pharmacy Subclass features and their corresponding values for each patient (0 = medication not administered and 1 = medication administered) and filled in the respective observations in the final file for Preoperative Administration of In-hospital Medications.

This file included for the 6392 observations:

4 features—MRN, VISIT ID, Case Name, Service Date

5 clinically relevant features—Parenteral Nutrition within 1 wk, inotropes within 48 h, heparin infusion within 24 hours, nesiritide within 48 h and beta-blockers within 24 h

414 features of individual pharmacy subclasses of medications that may or may not have been administered to a patient

### Processing of Medications Present on Preoperative Admission

This file contained information pertaining to observations of specific outpatient medications used by patients before being admitted to the hospital. The raw file had 8 features and 69,145 observations. The original raw data file had the following features:

MRN

VISIT ID

Service Date

Description—specific name of medication with dosage where it was entered (2682 different values)

Generic Name—the generic name associated to the drug in the feature Description (1004 different values)

Pharmacy Subclass—the subclass within the Pharmacy Class that the medication belonged into (562 different values)

Pharmacy Class—overall class of drugs that the medication belonged (357 different values) for this feature

Therapeutic Class—a less-specific drug classification system (51 different values)

Specific issues to this file included:

How to deal with combination drugs (eg, amlodipine/benazepril)

Drugs used for different purposes but had the same Class or Subclass (eg, Sildenafil listed as Viagra and classified as an Erectile Dysfunction Drug and Pulmonary Hypertension Drug)

A manual review of these medications was undertaken with an expert informed classification system developed (available by request). The designations were informed whenever possible by the previously described pharmacy Class or Subclass. Additionally, many patients were on multiple medications from the same subclass or class. If a generic value did not exist, the description then was checked, and it was appropriately assigned a generic value that was in line with what other observations with similar descriptions that did have a generic value assigned.



A file was created with the name `Finalmasterlist_PTA-Meds`. The first 4 features were as with all the above-mentioned databases:

MRN  
VISIT ID  
Case Name  
Service Date

Then, 3 features were created for each specific category `ptaXXX`—whether or not the patient was taking this medication (values 0 or 1)

`ptaXXXValue`—the specific generic value for that observation from the original file after cleaning

if more than 1, then both were listed separated by a comma

`ptaXXXDescription`—the specific description value for that observation from the original file after cleaning

if more than 1, then both were listed separated by a comma.

Only the first of the 3 features for each category of medications was included in the analysis, but this formatting enabled to both review and check for accuracy and provide transparency with retainment of detailed. The file created was `finalmasterlist_PTAMeds` and included

4 features—MRN, VISIT ID, Case Name, Service Date

348 features that characterize 116 classes of medications with only 116 features (the column with 0 or 1) included in the final analysis

### Processing of Family History Features

This file contained the family history information for the observations included in the overall patient cohort and consisted of 7 features and 17,164 observations including the following:

MRN  
VISIT ID  
Service Date

Relative With Problem—specific value for which relative had the family history (32 different values including Missing Values, Family History Negative For, and NonContributory)

Problem Description—specific value for the medical problem that the family history was positive for (165 different values)

Some of these “unique” features were really the same as others but labeled slightly different.

Comments—a feature that provided free form text describing either the problem description or provided the problem description (1166 different values)

There was considerable messy data in this file, and the following steps were taken to generate a standardized feature set for family history for the data provided.

Deleted any entry with Family History Negative For or Noncontributory.

The value “no known problems” was kept as being negative for family history.

For those with Other as the Problem Description, if there was a value in comments, then this value was put in the Problem Description for that specific entry.

The exact family member with the medical problem was unable to be reliably captured; therefore, this was eliminated and instead just noted if there was presence or absence for a specific observation.

The remaining values were categorized and detailed steps of this classification of the various Family History features can be provided upon request.

A file was created `Finalmasterlist_FamilyHistory` with the following features:

MRN  
VISIT ID  
Case Name  
Service Date

Family History Reported—whether or not a patient had any value recorded for their family history and if they did not then the value was 0 and if they did it was 1.

If the family history was not reported, then the rest of the features listed below were determined to be missing.

Only if family history was known to be negative was a value of 0 given.

Two additional features were created for each specific family history category:

`famhxXXX`—whether or not the patient had a positive family history for this category (values 0 or 1)

`famhxXXX_descript`—what the specific value for that category was from the original file

if patients had more than 1 in the same category then both were listed separated by a semicolon.

The file created `finalmasterlist_FamilyHistory` that included

4 features—MRN, VISIT ID, Case Name, Service Date  
1 feature—`famhxreported` (0 or 1)

30 features of which 15 were binary for a specific class of Descriptions in Family History and 15 were the detailed descriptions for what those were (only the 15 binary features (the column with 0 or 1) were included in the final analysis)

### Processing of Preoperative Vital Signs

This file contained information pertaining to vital sign values that were for patients before the operating room. The file had 7 features and 1,366,940 observations including:

MRN  
VISIT ID  
Service Date

Vital Sign—specific value for the type of vital sign recorded (38 different values)

Measured Date Time—an exact date and time for when the Vital sign was recorded

Format provides a date, an hour and minutes  
 Measured Value—the measured vital sign value  
 Unit of Measure—units of measure for the value of the vital sign (9 different values)

Dates and times were standardized to the format used in other files. The date/time allowed identification of the most recent vital sign proximal to surgery and entries without a full time/date stamp were removed. Additionally, any entry that occurred after the feature from the Intraoperative Times sheet labeled “Patient In OR” was eliminated. We also used this Intraoperative feature to localize the most recent set of vitals for a patient before entering the operating room. The file `Finalmasterlist_PreoperativeVitalSigns` was created and included

MRN  
 VISIT ID  
 Case Name  
 Service Date  
 Surgery Time  
 Respirations (Breaths Per Minute)  
 Respirations (Measured Date Time)  
 Pulse (Beats Per Minute)  
 Pulse (Measured Date Time)  
 Temperature (degrees Fahrenheit)  
 Temperature (Measured Date Time)  
 Diastolic Blood Pressure (mm Hg)  
 Diastolic Blood Pressure (Measured Date Time)  
 Systolic Blood Pressure (mm Hg)  
 Systolic Blood Pressure (Measured Date Time)  
 Pain Scale (Pain Score)  
 Pain Scale (Measured Date Time)  
 O2 Saturation (Percent)  
 O2 Saturation (Measured Date Time)  
 Mean Arterial Pressure (mm Hg)—calculated from systolic blood pressure and diastolic blood pressure  
 Pulse Pressure (mm Hg)—calculated from systolic blood pressure and diastolic blood pressure  
 The final file included  
 4 features—MRN, VISIT ID, Case Name, Service Date  
 Surgery Time (later eliminated due to redundancy)  
 7 Vital Signs Features (respirations, pulse, temperature, diastolic blood pressure, systolic blood pressure, pain scale, O2 saturation)  
 7 features that had the measured date and time for their respective vital sign  
 2 features that were created (mean arterial pressure and pulse pressure)

### Processing of Preoperative Surgical History

This file contains information pertaining to a patient’s specific surgical history. The file had 8 features and 23,297 observations including:

MRN  
 VISIT ID

Service Date  
 Surgical Procedure—where the individual observation came from (Surgical History of Surgical Procedure)  
 Surgery Date—if a date or date/time provided for when the surgical procedure took place  
 Context Procedure Code—proved to be an unusable column that was an internal procedure code that MSDW used without an association dictionary  
 Procedure Description—specific value of a surgery that had been performed before this hospitalization (1852 different values)  
 Context Name—which EHR the observation had been obtained from (10 different values)  
 Values: EPIC, HSM, Rad Mod Site 3, Carefusion Lab, Tamtron EAP, IP Outpatient Charges, Cardiology Charges, Surgery Charges, Dialysis Charges, Labor, and Delivery Charges

Upon review, the file needed structured features created and this entailed manually classifying the 1852 unique Procedure descriptions into defined categories. The list of unique Procedure Descriptions was listed into a file and categories of surgical procedures were created so that the 1852 unique entries would be classified under 1 of these categories (available upon request). Any entry associated with procedures that did not make sense or did not correspond to 1 of the 6392 patients in finalmasterlist then it was removed. This reduced the number of observations in Preoperative Surgical History from 23,297 to 14,634. The file `Finalmasterlist_SurgicalHistory` was created. The first 4 features were in line with all the previous files and the rest of the features were a binary feature (did this patient have this type of surgery, 0 or 1) and if they did have that type of surgery the second feature was its value (eg, Dental Surgery: 1, Dental Surgery value: history of dental surgery). Only 4839 patients of 6392 patients had information regarding their surgical history and therefore the patients who did not have anything listed for needed to be accounted for. An additional feature was created called Surgical History Present and if there was no surgical history then there was a 0 and the rest of the values within the file corresponding to that observation were labeled as missing data. If there was information regarding surgical history then it was labeled as a 1.

The final file `finalmasterlist_SurgicalHistory` and included

4 features—MRN, VISIT ID, Case Name, Service Date  
 Surgery History Present (0 or 1)  
 49 features of Categories of surgical procedures (binary 0 or 1)  
 49 features that have the detailed value that went into the designated categories

### Processing of Echocardiogram

These files contained the data from the echocardiograms performed at Mount Sinai and their corresponding reports generated from the Mount Sinai Cardiology

Echocardiography Database. These files were related by their Unit Number (MRN), Report Number (unique identifying report number for each echocardiogram), and Test Date (date of echocardiogram exam) to match the various entries from each file to provide a full report for an individual exam. These files provided all of the entries for all patients who had echocardiograms at Mount Sinai both preoperatively and postoperatively. Each file was systematically evaluated, and all the features were listed out and manually reviewed and cleaned (full list of files with subsequent echo values provided available on request).

The report numbers were first matched between the various echo files. A master echo file was then created for all the features for all the observations. Because some features were the same name but had a different meaning, we renamed the individual features with the sheet name they came from in front followed by an underscore followed by the feature name (eg, TricuspidValve\_Etiology). Duplicated features and features with 100% missing values were removed. Features with dates were standardized in the format %d-%m-%Y. Entries whose MRN did not match an MRN from the finalmasterlist (6392 patients) were removed. This resulted in 620 features with 14,428 observations.

**For stress echocardiogram.** We then created a file (finalmasterlist\_EchoStress) that had the most recent STRESS ECHOCARDIOGRAM (Stress Echo) before undergoing surgery. This provided us with 1 observation that was categorized as a Stress Echo and was the most proximal to the date of surgery (to give as much of a recent characterization of the information provided by a Stress Echo before surgery). This provided the most proximal observation of a stress echocardiogram performed preoperatively to the date of surgery for any patient with a stress echocardiogram. If a patient did not have a stress echocardiogram before surgery then missing value was input in its place. This resulted in 623 features with 6392 observations (corresponding to the same format as the 6392 observations in finalmasterlist). A total of 315 of the 6392 patients (4.9%) had a preoperative stress echocardiogram.

**For transthoracic echocardiogram.** We then created a file (finalmasterlist\_EchoColor) that had the most recent TRANSTHORACIC ECHOCARDIOGRAM before undergoing surgery. This provided us with 1 observation that was categorized as a Transthoracic Echo and was the most proximal to the date of surgery (to give as much of a recent characterization of the information provided by a Transthoracic Echo before surgery). This provided the most proximal observation of a transthoracic echocardiogram performed preoperatively to the date of surgery for any patient with echocardiogram. If a patient did not have a transthoracic echocardiogram before surgery then missing value was input in its place. The file has 623 features with 6392 observations (corresponding to the same format as the 6392

observations in finalmasterlist). A total of 2698 of the 6392 patients (42.2%) had a preoperative transthoracic echocardiogram.

**For transesophageal echocardiogram.** We then created a file (finalmasterlist\_EchoTrans) that had the most recent TRANSESOPHAGEAL ECHOCARDIOGRAM before undergoing surgery. This provided us with 1 observation that was categorized as a Transesophageal Echo and was the most proximal to the date of surgery (to give as much of a recent characterization of the information provided by a Transesophageal Echo before surgery). This provided the most proximal observation of a transesophageal echocardiogram performed preoperatively to the date of surgery for any patient with echocardiogram. If a patient did not have a transesophageal echocardiogram before surgery, then missing value was input in its place. The file had 623 features with 6392 observations (corresponding to the same format as the 6392 observations in finalmasterlist). A total of 666 of the 6392 patients (10.4%) had a preoperative transesophageal echocardiogram.

**For last echocardiogram before discharge.** We then created a file (finalmasterlist\_EchoPriorDischarge) that had the last echocardiogram before the date of discharge of the hospitalization the patient underwent surgery (had to fall in time interval between OR time and date of discharge). This provided us with 1 observation that was categorized as the last echocardiogram (regardless of type—Stress, transthoracic, transesophageal) that the patient had performed before being discharged. If a patient had multiple echocardiograms in the postoperative period, only the 1 performed closest to the date of discharge was included (this provided as much of a recent characterization of the echocardiogram information before discharge). This provided the most proximal observation of an echocardiogram performed before discharge in a patient who underwent cardiac surgery. If a patient did not have an echocardiogram in the time interval between the service date and the date of discharge, then missing value was input in its place. The file had 623 features with 6392 observations (corresponding to the same format as the 6392 observations in finalmasterlist). A total of 4795 of the 6392 (75.0%) patients had an echocardiogram between the service date and before discharge.

**For most recent echocardiogram available in system.** We then created a file (finalmasterlist\_EchoRecentPostOp) that had the results for the most recent echocardiogram available at the time of the data acquire from the Cardiology department. This took the most recent echocardiogram that was performed after surgery until the date of data acquire (the date of the exam could occur after the date of discharge). This provided us with 1 observation that was categorized as the most recent echocardiogram (regardless of type—Stress, transthoracic, transesophageal) that the patient had performed after having surgery even if it was after

discharge. This provided the most recent observation of an echocardiogram performed after a patient had surgery until the data acquire (ie, the postoperative echo farthest out from surgery). If a patient did not have an echocardiogram performed at all after surgery, then a missing value was inserted for features in that observation. The file had 623 features with 6392 observations (corresponding to the same format as the 6392 observations in finalmasterlist). 4936 of the 6392 (77.2%) patients had an echocardiogram between the service date and the data acquire (January 31, 2017).

**For overall preoperative echo.** We then created a file (finalmasterlist\_EchoDataTotalPreop) that had the most recent preoperative echocardiogram before undergoing surgery regardless of what type of echo it was (Stress, Transesophageal, Transthoracic). This provided us with 1 observation for the most proximal echocardiogram regardless of type before having surgery (a combination of Stress, Transesophageal, Transthoracic but only taking the 1 observation out of all 3 that was closest to the day of service). This provided the most proximal observation (regardless of type of echocardiogram) performed preoperatively to the date of surgery for any patient with an echocardiogram. If a patient did not have an echocardiogram before surgery, then missing value was input in its place. The file had 623 features with 6392 observations (corresponding to the same format as the 6392 observations in finalmasterlist). A total of 2982 of the 6392 (46.7%) patients had a preoperative echocardiogram of any type.

The final list of features available in each of these files is available upon request.

### Processing of Preoperative Cardiac Catheterization

This file contains information pertaining to a patient's preoperative cardiac catheterization if 1 was available. The file had 355 features and 7460 observations. The unstructured features are available upon request. Careful review of the range of values as well as the format of values for each of the columns was undertaken to first understand the data. All dates and times were standardized to the format used in other files. The date stamp allowed us to determine when the catheterization took place in relation to the surgery date. Further review resulted in deletion of redundant features or personal identifying information or features with no data.

Cross-referencing to finalmasterlist determined that 3650 patients had a catheterization out of the 6392 total patients. There were 3854 catheters that were matching to observations from finalmasterlist. However, some patients had multiple catheterizations that were farther back in time. Therefore, we wanted to take the catheterization observation most proximal to the date of service. Only the most recent catheterization observation was then isolated and put in a file called mostrecentcath and included the following:

MRN

VISIT ID

Case Name

Service Date

TESTDATE

REPORT NUMBER

And 344 features related to cardiac catheterization

### Processing of Intraoperative Fluid Outputs

This file was obtained from intraoperative anesthesia records (CompuRecord) and had 6 features with 20,368 observations. Each of the observations is a value for an output that was recorded in the operating room during a specific patient's procedure (Case Name). There are 6 different categories that are possible values for the feature of Label, and these are time stamped. The file had 6 features:

Case Name

Time stamp (exact date hour and minute) that corresponds with feature 4

ItemID that corresponds with feature 4

Label

Urine; E.B.L.; CSF; Ascites; Pump-Remains; N/G

Value—an amount associated with the specific label feature

Units—units of measurement of the output

The values for the feature Label (feature 4) were recorded by an anesthesiologist in real-time so this required vetting of input errors and lack of standardization of inputs. This was done by looking at the presence (or absence) of each label as it occurred for each patient. These intraoperative outputs were used to create specific features that characterized important overall outputs in the surgery.

A new file was created that would contain the features created from Outputs and was called Intraoperative Outputs. The following features were created for the file finalmasterlist\_IntraoperativeOutputs.

VISIT ID

MRN

Case Name

Service Date

CSF Total Output—cerebral spinal fluid total output recorded for a case

Numeric range

Urine Total Output—urine total output recorded for a case

Numeric range

### Processing of Intraoperative Blood Product Use

This file was obtained from the intraoperative anesthesia records (CompuRecord) and had 6 features with 18,718 observations. Each of the observations was a value for a blood/blood product transfusion that was recorded in the operating room during a specific patient's procedure (Case Name). There are 13 different categories that are possible values

for the feature of Label, and these are time stamped. The file had 6 features:

- Case Name
- Time stamp (exact date hour and minute) that corresponds with feature 4
- ItemID that corresponds with feature 4
- Label
  - RBCs; Cell Saver; Cryoprecipitate; Platelets; FFP; RBCs-Autol; Platelets-SD; RBCs-Washed; Whole Blood; Whole Blood-DD; RBCs-DD; PRBC; Whole Blood-blb bank.
  - Value—an amount associated with the specific label feature
  - Units—units of measurement of the output
- The values for the feature Label (feature 4) were recorded by an anesthesiologist in real-time so this required vetting of input errors and lack of standardization of inputs. This was done by looking at the presence (or absence) of each label as it occurred for each patient. These intraoperative blood and blood products were used to create specific features that characterize important usage in the surgery. A new file was created that would contain the various features created from Intraoperative Blood Products and was called final-masterlist\_IntraoperativeBloodProducts. Blood and blood product transfusions that occurred after Procedure Finish Time were eliminated. Any entries with values less than 25 mL was removed because these were likely to be a recording error. The final file created included the following:
  - VISIT ID
  - MRN
  - Case Name
  - Service Date
  - Cell Saver—did a patient have cell saver transfusion given to them (value 0, 1)
  - Cell Saver Total mL (numeric range)
  - Cryoprecipitate—did a patient have cryoprecipitate transfusion given to them (value 0, 1)
  - Cryoprecipitate Units (numeric range)
  - Cryoprecipitate Total mL (numeric range)
  - FFP—did a patient have FFP transfusion given to them (value 0, 1)
  - FFP Units (numeric range)
  - FFP Total mL (numeric range)
  - Platelets—did a patient have platelets transfusion given to them (value 0, 1)
  - Platelets Units (numeric range)
  - Platelets Total mL (numeric range)
  - Autologous RBCs—did a patient have autologous RBCs transfusion given to them (value 0, 1)
  - Autologous RBCs Units (numeric range)
  - Autologous RBCs Total mL (numeric range)
  - RBCs—did a patient have RBCs transfusion given to them (value 0, 1)

- RBCs Units (numeric range)
- RBCs Total mL (numeric range)
- Whole Blood—did a patient have whole blood transfusion given to them (value 0, 1)
- Whole Blood Units (numeric range)
- Whole Blood total mL (numeric range)
- Intraoperative transfusion—did a patient receive platelets, FFP, cryoprecipitate, or RBCs (value 0, 1)

### Processing of Preoperative Laboratory Values

This file contained information pertaining to observations of laboratory features that were routinely check preoperatively in patients who would be having open surgery. The file had 11 features and 2,789,525 observations and included the following:

- MRN
- VISIT ID
- Service Date
- Order Code—the internal code used by Mount Sinai to identify different orders (678 different values)
- Result Date Time—an exact date and time for when the laboratory result was drawn
  - Format provides a date, an hour, and minutes
- Result Status—the current status of the value of the result (6 different values)
  - Final Result; Preliminary Result; Result Canceled; Correct Result; Result Not Available; Other Result
- Result Code—the internal code used by Mount Sinai to identify different specific results (1222 different values)
  - If an order code was for a panel then the result code only corresponded to the specific feature in that panel.
- This also provided the singular laboratory tests that were important to abstract and create features from.
  - Abnormal Flag—if the laboratory value assessed was abnormal compared with the normal reference range used (7 unique values)
    - NA; Below low normal; Above high normal; Abnormal (applies to non-numeric results); Above upper panic limits; Below lower panic limits; Critical or Absurd (alphanumeric only)
  - Reference range—reference range internally used to determine if a specific laboratory value was normal versus abnormal.
  - Qualitative and Quantitative values for this feature as it depended on the specific laboratory value feature it related to.
    - Value—the value for the specific laboratory test that was performed specific to an individual patient.
    - Qualitative and Quantitative values for this feature as it depended on the specific laboratory value feature it related to.
      - Unit of Measure—the reference unit of measure used for a specific laboratory test.

Qualitative and Quantitative values for this feature as it depended on the specific laboratory value feature it related to.

Each of these features was explored to determine if any data were corrupt or did not make sense. Any erroneous entries or entries that were not completed or deemed to not have an appropriate value for that corresponding feature were removed. To keep proximity to the date of surgery, we eliminated all observations that were more than 1 week before the Procedural Start Time and only included the laboratory values in that interval (<1 week before the surgery start). This resulted in 1175 unique Result Codes (#7 above). Because there were many unique features, we had to determine what all of the unique laboratory test features were that could be created and if some laboratory tests were the same but labeled in a different way. Laboratory tests that were the same but labeled slightly different were identified and their labels were unified under 1 harmonized feature label.

We then created a file that provided the Order Code, Result Code, No Observations with that Lab test result (of 6392), Units, and Reference Range. All of the unique Result codes with their corresponding information for the columns of Order Code, Number of Observations, Units, and Reference Ranges were included. If more than 1 unit or Reference Range corresponded to that specific Result Code then they were listed in their feature column along with a semi colon between them. A full list of the unique Result Codes was created with their frequency of appearance (of 6392). This query returned 1081 individual laboratory test features of which many had only rare values. Therefore, we decided to eliminate all laboratory tests that had a non-missing value filled in for an observation less than 1% of all observations (frequency of values needed to be  $\geq 64$ ). The rest of the entries corresponding to the infrequent observations were removed from the dataset. This brought the number of features down from 1081 to 174. We then identified the most proximal observation for a specific laboratory result code to the date of surgery. This was performed for all of the unique Result Codes so that each patient would not have more than 1 entry of the same laboratory test.

The features were then grouped by clinically relevant groups (eg, all ABG lab values in the same group of features, all CBC lab values in the same group of features). The groups were as follows:

- General labs
- Coagulation profile labs
- ABG labs
- CBC labs
- VBG labs
- Urinlytes labs
- Urinalysis labs
- Urine microscopy labs
- Urine dipstick labs

A file was created named finalmasterlist\_Preoperative-Labs and included

- MRN
- VISIT ID
- Case Name
- Service Date

And then the following 3 feature format was made to represent each of the laboratory features that were cleaned (feature X, Y, Z...):

Laboratory Feature X

Resulted in 174 unique features that each needed to be individually vetted and cleaned to make sure the results were standardized

Laboratory Feature X Date and Time

The time and date that the laboratory feature X resulted  
Laboratory Feature X AB

This reflected whether or not the measured value for Feature X was considered to be abnormal when compared to the reference range used at the time the laboratory value resulted.

One additional feature was created from 2 previous features stratified by race. Because GFR was reported for both Non-African Americans and African Americans as 2 separate features, we created a new column Final GFR that took the appropriate value from the Non-African American GFR feature or the African American GFR feature depending on race denoted from finalmasterlist. This would then have a column with only the GFR applicable to that patient as noted by their race.

Full list of generated laboratory value features is available upon request.

### Processing of Preoperative Cardiac Catheterization Anatomy

This file contained information pertaining to a patient's preoperative cardiac catheterization specifically in relation to the coronary vessel anatomy. The file had 10 features and 45,899 observations including the following:

Report Number

Vessel—specifically what vessel each observation was referring to

Obstruction—defined the stenosis present

Morphology—descriptive term

Morphology 1—descriptive term

Morphology 2—descriptive term

Distal Vessel 1—described the vessel after the stenosis

Distal Vessel 2—described the vessel after the stenosis

Distal Vessel 3—described the vessel after the stenosis

Careful review of the range of values as well as the format of values for each of the features was undertaken to first understand the data. The Report Number allowed us to correlate it to the administrative data such as date of catheterization as previously described in Most Recent Cath. Further cleaning resulted in deletion of redundant

features or personal identifying information or features with no data at all. Cross-referencing to finalmasterlist it was determined that 3650 patients had a catheterization of the 6392 patients. There were 3854 catheterizations that were matching to observations from finalmasterlist. However, some patients had multiple catheterization that were farther back in time. Therefore we wanted to take the catheterization observation most proximal to the date of service. Only the observations with Report Numbers corresponding to those present in finalmasterlist\_MostRecentCath were used.

The file finalmasterlist\_CatheterizationAnatomy was created.

The number of unique Vessels were 6 (RCA; Left Main; LAD; LCx; NA; Ramus Intermedius). The number of unique Segments were 28 (Proximal; NA; OM1; Distal; Mid; High Lateral; D1; OM2; RPDA; RPL1; D2; AV Continuation; Ostial; D1 Ostial; OM1 Ostial; LPL2; LPDA; LPL1; Acute Marginal; D3; D2 Ostial; RPDA Ostial; RPL2; OM3; 1st Septal; LPDA Ostial; Separate Ostia; LAV).

We created a vector of column names that corresponded to the unique combinations present between the feature of Vessels and the feature of Segments. Also collapsed the 3 Morphology features down into 1 feature with the values separated by semi-colons. The same was done for the 3 separate Distal Vessel features. Then a new feature (feature 1) was created that was the combination of the unique values of Vessels and Segments, and a second feature (feature 2) was created that provided the Morphology related to the first feature, and a third feature (feature 3) was created that provided the Distal Vessel values related to the first feature. The value in the feature Obstruction was inserted into the value for feature 1 that corresponded to its specific observation.

Therefore, to provide an example: a patient with report number XXXXX with an observation with the following information.

Vessel—LAD

Segment—D1

Obstruction—No Obstruction

Morphology—Bifurcation

Morphology 1—Aneurysmal

Morphology 2—(blank)

Distal Vessel 1—Moderate size, moderate disease

Distal Vessel 2-branching vessel

Distal Vessel 3-(blank)

This would read as:

Feature 1—LAD-D1 with a value of (No obstruction)

Feature 2—LAD-D1-Morphology with a value of (Bifurcation; Aneurysmal)

Feature 3—LAD-D1-Distal Vessel with a value of (Moderate size, moderate disease; branching vessel)

This was then done for all of the unique combinations of Vessels and Segments and resulted in 411 features for 6392 observations that matched finalmasterlist. The file included

VISIT ID

MRN

Case Name

Service Date

Test Date

Report Number

Then 135 unique combinations of Vessel and Feature laid out in the format above.

Feature 1

Feature 2

Feature 3

### Generation of Preoperative Medical History

A unique file was created that had a hand curated number of features that corresponded to clinically impactful as well as previously used features from the STS Risk models. This resulted in the file finalmasterlist\_PreoperativeMedicalHistory being created, and it included 34 features for 6392 observations. The 34 features were as follows:

MRN

VISIT ID

Case Name

Service Date

Diabetes

Diabetes Control

Hypertension

Peripheral Arterial Disease

Preoperative Mechanical Ventilation

Immunocompromise

Dialysis

Cerebrovascular Disease

Cerebrovascular Accident

Inotropes Within 48 h.

IABP

Liver Disease

Liver Disease Type

Pulmonary hypertension

Atrial fibrillation

Number Diseased coronary arteries

NYHA Class

Left Main Stenosis

Ejection fraction

Aortic insufficiency

Aortic stenosis

Mitral insufficiency

Mitral stenosis

Tricuspid insufficiency

Status at admission

Cardiogenic shock

Radiation

Tobacco history

Endocarditis

Meld Score

A thorough explanation can be provided upon request as to how these features and their corresponding values was generated. Additional source code can be found in the PreopMedicalHistory.R file used to create this database for any further clarification.

## APPENDIX 2. EXPLANATION OF MACHINE LEARNING ALGORITHMS AND EVALUATION METRICS

### Machine Learning Algorithms

There were 4 different classification algorithms used in this study (LR, random forest, support vector machine, and extreme gradient boosting). LR and support vector machine are linear classification algorithms that infer a hyperplane separating the 2 classes under consideration. LR identifies this hyperplane by optimizing the probability of a data point's class label based on its features, and support vector machines identifies the hyperplane that maximizes the separation between the classes. In contrast, random forest and extreme gradient boosting are ensemble algorithms that build a collection of decision tree classifiers, and aggregate the predictions from the trees as the final outcome. In particular, random forest learns the decision trees independently of each other, while extreme gradient boosting learns them sequentially to maximize the ensemble's collective classification performance.

### Evaluation Metrics

Multiple measures exist to evaluate classifier performance (Figure E5). The most common of these is accuracy, which is the fraction of the total number of patients the model correctly classifies and is calculated as  $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$ . However, this measure can be misleading in cases of severe class imbalance, because uninformed classification of all the patients to the majority class can yield an artificially high value of Accuracy.

The AUROC is another commonly used measure, because it evaluates classifier performance over all prediction score thresholds. Although more effective than Accuracy, AUROC still faces challenges in cases of severe class imbalance like ours.

Precision, Recall and F-measure are class-specific evaluation measures that are designed for unbalanced classes (Figure E5). Precision measures how many predictions for a given class are actually correct and is also referred to as the positive predictive value for the positive class, and vice versa. Recall measures how many of the true patients of a given class are correctly classified and is also referred

to as the sensitivity and specificity for the positive and negative classes, respectively. A precision-recall curve shows how the 2 measures vary with respect to each other for a given classifier. AUPRC provides a global evaluation of the model and is more informative than AUROC when evaluating classifiers with unbalanced datasets.

The F-measure is the harmonic mean of precision and recall that represents a more reliable aggregate evaluation metric for unbalanced class distributions. F-measure has a range of 0 to 1, with higher values supporting improved classification performance. This measure is calculated as  $2 \times ((\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}))$ . Because of the relative strengths and weaknesses of the described measures, we report in this study Accuracy and AUROC, but focus on F-measure, precision, recall, and AUPRC as the main evaluation metrics.

## APPENDIX 3. SOFTWARE USED

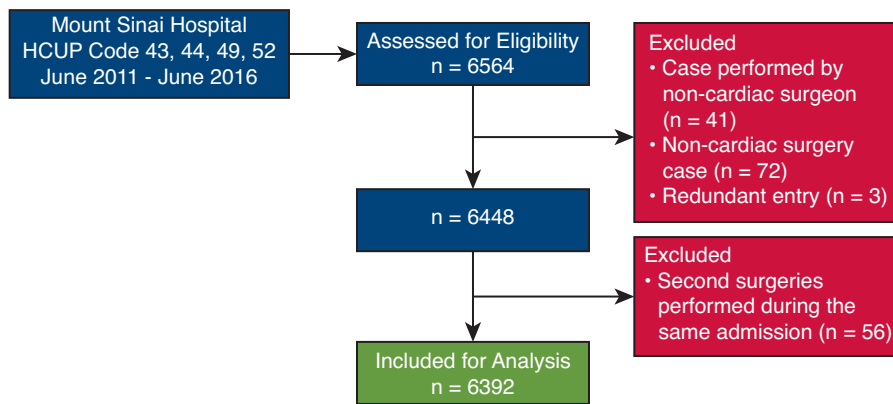
All analyses were performed with the following publicly available software packages. Data preprocessing, model development and evaluation, and data visualization were performed with Jupyter Notebook, Anaconda software, and the Python programming language using the NumPy, Pandas, SciPy, Seaborn, Matplotlib, and Sklearn packages.

Specifically, initial data preprocessing was performed in R (R Foundation, Vienna, Austria) and preprocessed data were imported using Pandas version 1.1.4 and NumPy version 1.18.1 for normalization and imputation. All model code was in Python version 3.7.6 and models were run in Anaconda version 4.9.0 using Jupyter Notebook with XGB version 0.90 for the XGBoost classification model and Sklearn version 0.22.1 for the other machine learning models. Scipy version 1.4.1 was used to calculate measures of central tendency and confidence intervals. Figure plotting was performed using Matplotlib version 3.1.3 and Seaborn version 0.10.0.

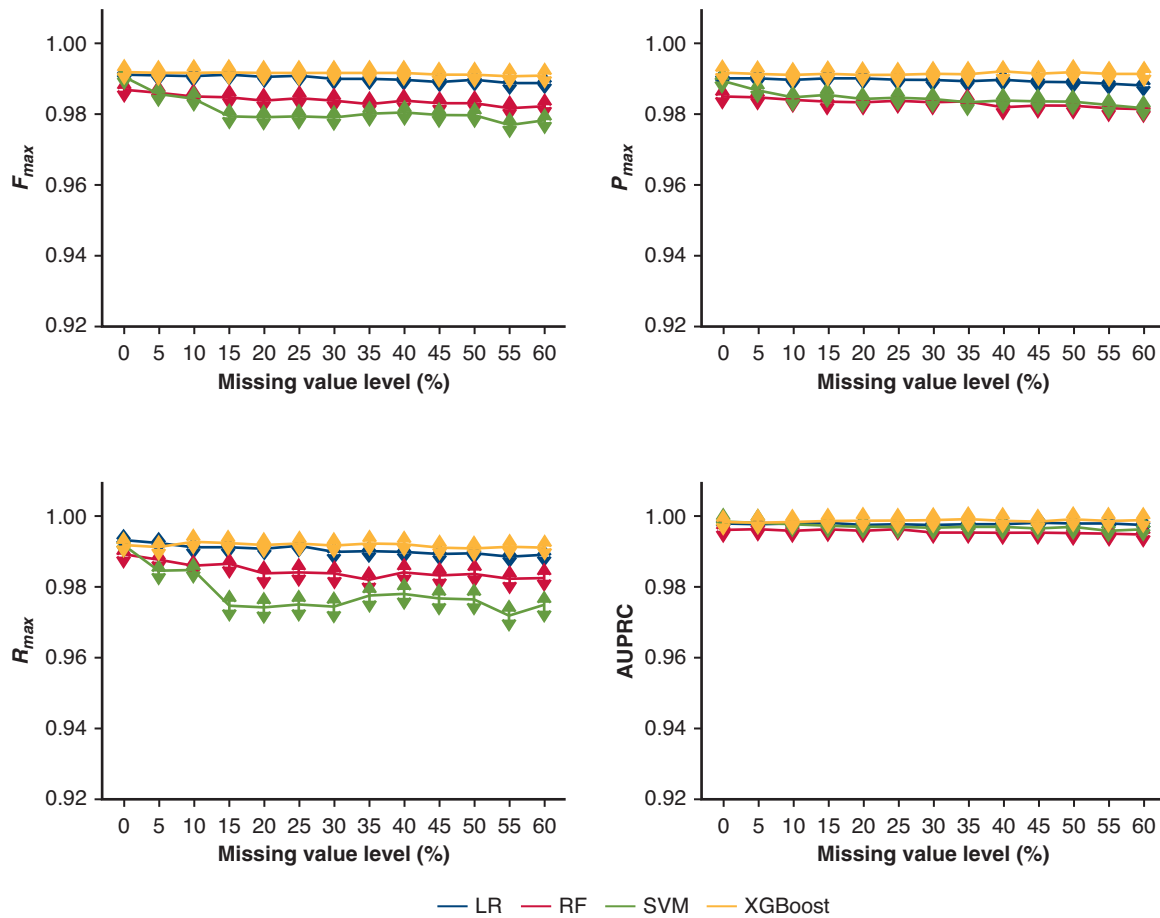
## APPENDIX 4. XGBoost HYPERPARAMETERS

Of note, hyperparameter tuning was not used in the model building process. For the XGBoost model, the standard parameters were applied as follows: XGBClassifier base\_score = 0.5, booster = 'gbtree,' colsample\_bylevel = 1, colsample\_bynode = 1, colsample\_bytrees = 1, gamma = 0, learning\_rate = 0.1, max\_delta\_step = 0, max\_depth = 3, min\_child\_weight = 1, missing = None, n\_estimators = 100, n\_jobs = 1, nthread = None, objective = 'binary:logistic,' random\_state = 0, reg\_alpha = 0, reg\_lambda = 1, scale\_pos\_weight = 1, seed = None, silent = None, subsample = 1, verbosity = 1.

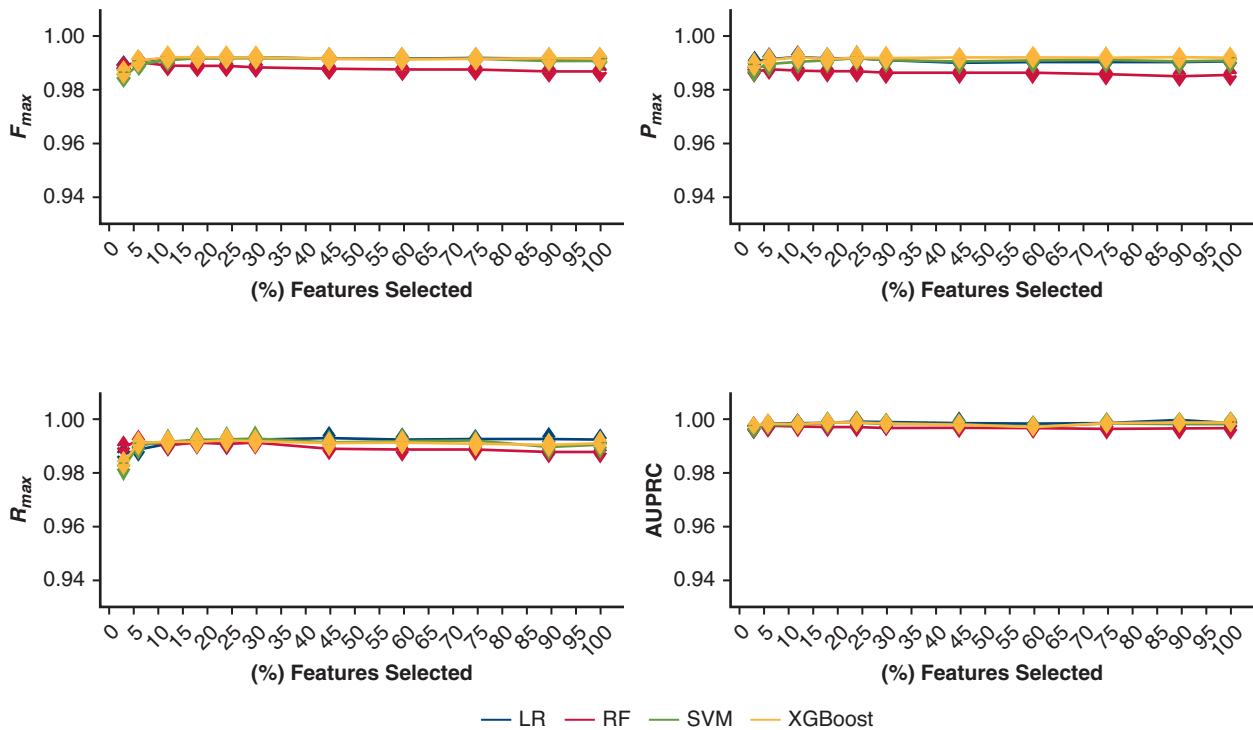




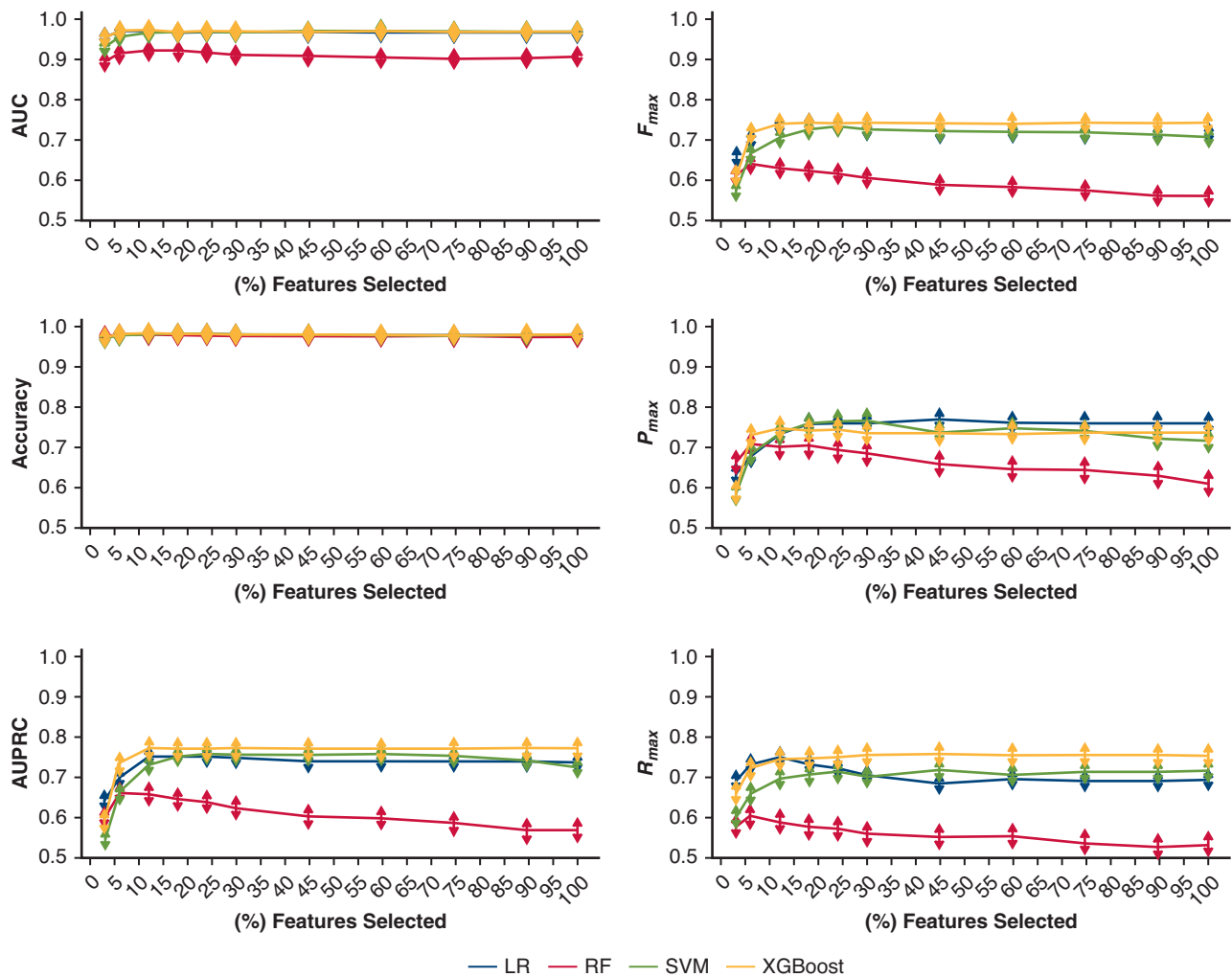
**FIGURE E1.** CONSORT diagram demonstrating patient inclusion and exclusion criteria for the overall study cohort. *HCUP*, Healthcare Cost and Utilization Project.



**FIGURE E2.** Results of a data-driven assessment of the acceptable level of missing values in features that can be reliably included and imputed during model development for the alive class. These levels varied from 0% to 60% in increments of 5%. At each level, the mean-mode imputation method (mean for continuous features, mode for categorical features) was applied to the training dataset. Four classifiers (LR, RF, SVM, XGBoost) were then trained on the imputed training set, and evaluated on the 20% validation set in terms of 6 metrics (AUROC, Accuracy,  $F_{max}$ ,  $R_{max}$ ,  $P_{max}$ , AUPRC).  $F_{max}$  is the highest value of F-measure across all classification score thresholds, and  $R_{max}$  and  $P_{max}$  are the corresponding values of precision and recall, respectively, at that point. Each algorithm was run 100 times for each missing data level, and the average of the performance was plotted for each metric for the alive class (AUROC and accuracy are not presented because they are the same for both classes).  $F_{max}$ , Maximum value of F-measure across all prediction thresholds; *LR*, logistic regression; *RF*, Random Forest; *SVM*, Support Vector Machine; *XGBoost*, eXtreme Gradient Boosting;  $P_{max}$ , value of precision at  $F_{max}$ ;  $R_{max}$ , value of recall at  $F_{max}$ ; *AUPRC*, Area under the precision-recall curve; *AUROC*, area under the receiver operating characteristic curve.



**FIGURE E3.** Results of feature selection using RFE during model development for the alive class. RFE is a feature selection method that fits a predictive model to the training data and removes the weakest features until a prespecified number or percentage of features is reached. Various percentages from 0% to 100% of the overall number of features (100% = 336 preoperative features with no missing values) were assessed. Four different classifiers (LR, RF, SVM, XGBoost) were then trained using the specified percentage of features in the training set and evaluated on the validation set in terms of 6 metrics (AUROC, Accuracy, Fmax, Rmax, Pmax, AUPRC). Fmax is the highest value of Fmeasure across all classification score thresholds, and Rmax and Pmax are the corresponding values of precision and recall, respectively, at that point. Each algorithm was run 100 times for each percentage of features, and the average of the performance was plotted for each metric for the alive class (AUROC and accuracy are not presented because these are the same for both classes). Fmax, Maximum value of F-measure across all prediction thresholds; LR, logistic regression; RF, Random Forest; SVM, Support Vector Machine; XGBoost, eXtreme Gradient Boosting; Pmax, value of precision at Fmax; Rmax, value of recall at Fmax; AUPRC, Area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; XGBoost, eXtreme Gradient Boosting.



**FIGURE E4.** Results of feature selection using RFE during model development for the deceased class. RFE is a feature selection method that fits a predictive model to the training data and removes the least predictive features until a prespecified number or percentage of features is reached. Various percentages from 0% to 100% of the overall number of features (100% = 336 preoperative features with no missing values) were assessed (X-axis). Four different classifiers (LR, RF, SVM, and XGBoost) were then trained using the specified percentage of features in the training set and evaluated on the validation set in terms of 6 metrics (AUROC, Accuracy,  $F_{max}$ ,  $R_{max}$ ,  $P_{max}$ , and AUPRC; Y, axis of the corresponding subplots). Each algorithm was run 100 times for each percentage of features, and the average of the performance was plotted for each metric for the deceased class. *AUROC*, Area under the ROC curve; *LR*, logistic regression; *RF*, random Forest; *SVM*, Support Vector Machine; *XGBoost*, eXtreme Gradient Boosting;  $F_{max}$ , maximum value of F-measure across all prediction score thresholds;  $P_{max}$ , value of precision at  $F_{max}$ ;  $R_{max}$ , value of recall at  $F_{max}$ ; *AUPRC*, Area under the precision-recall curve; *AUROC*, area under the receiver operating characteristic curve.

		PREDICTED CLASS	
		Class = +	Class = -
ACTUAL CLASS	Class = +	a (TP)	b (FN)
	Class = -	c (FP)	d (TN)

$$\text{Precision } + (P_+) = \text{PPV} = \frac{a}{a + c}$$

$$\text{Precision } - (P_-) = \text{NPV} = \frac{d}{b + d}$$

$$\text{Recall } + (R_+) = \text{Sensitivity} = \frac{a}{a + b}$$

$$\text{Recall } - (R_-) = \text{Specificity} = \frac{d}{c + d}$$

$$F\text{-measure } + (F_+) = \frac{2P_+R_+}{P_+ + R_+} = \frac{2a}{2a + b + c}$$

$$F\text{-measure } - (F_-) = \frac{2P_-R_-}{P_- + R_-} = \frac{2d}{2d + b + c}$$

**FIGURE E5.** Evaluation measures for classifiers that generate a positive (+) or negative (−) prediction for a given individual. The relationships between Sensitivity, Specificity, Positive/Negative Predictive Values, Precision, Recall and F-measure, the main evaluation measures used in our work, are summarized here. F-measure, which is a harmonic (conservative) mean of Precision and Recall that is computed separately for each class, provides a comprehensive and reliable assessment of model performance when classes are imbalanced. *FN*, False-negative; *FP*, false-positive; *TN*, true-negative; *TP*, true-positive.

**TABLE E1. Twenty-five most frequent surgical procedure combinations performed in the overall study cohort**

Surgical procedure performed	Frequency, n
CABG	1585
AVR	528
MVRepa (complex) + TVRepa (ring)	437
CABG + AVR	207
MVRepa (complex)	206
MVRepa (complex) + TVRepa (ring) + LAA closure + Maze	135
MVRepa (ring) + TVRepa (ring)	75
MVRepa (complex) + TVRepa (ring) + PFO closure	74
MVRepa (complex) + TVRepa (ring) + LAA closure	73
AVR + ascending aortic replacement	69
MVR + TVRepa (ring)	65
MVRepa (complex) + TVRepa (ring) + Maze	64
MVRepa (complex) + TVRepa (no ring)	62
Aortic root replacement (Bio-Bentall) + ascending aortic replacement	59
MVRepa (ring) + TVRepa (ring) + LAA closure + Maze	56
Heart transplant	55
Ascending aortic replacement	55
MVRepa (complex) + TVRepa (ring) + CABG	52
Aortic root replacement (Bio-Bentall)	49
MVR + TVRepa (ring) + LAA closure + Maze	48
MVRepa (ring) + CABG	45
MVR	42
MVRepa (ring) + TVRepa (ring) + Maze	38
Ross	37
Intracorporeal LVAD (less invasive)	35

Some index cases did not have an associated STS risk score calculated at the time of the surgery, thereby explaining the discrepancy in frequency of specific procedures with Figure 4 and Table 3. CABG, Coronary artery bypass grafting; AVR, Aortic valve replacement; MVRepa, mitral valve repair; TVRepa, tricuspid valve repair; LAA, left atrial appendage; PFO, patent foramen ovale; LVAD, left ventricular assist device; MVR, mitral valve replacement.

TABLE E2. Complete list of 336 preoperative features included in the final model

Administrative
(V4511)-RENAL DIALYSIS STATUS
(4242)-TRICUSPID VALVE DISORDERS SPECIFIED AS NONRHEUMATIC_AP
(42820)-UNSPECIFIED SYSTOLIC HEART FAILURE_AP
(4264)-RIGHT BUNDLE BRANCH BLOCK_AP
(44021)-ATHEROSCLEROSIS OF NATIVE ARTERIES OF THE EXTREMITIES WITH INTERMITTENT CLAUDICATION
(2762)-ACIDOSIS_AP
(V1046)-PERSONAL HISTORY OF MALIGNANT NEOPLASM OF PROSTATE
(25080)-DIABETES WITH OTHER SPECIFIED MANIFESTATIONS TYPE II OR UNSPECIFIED TYPE NOT STATED AS UNCONTROLLED_AP
(V153)-PERSONAL HISTORY OF IRRADIATION PRESENTING HAZARDS TO HEALTH
(2724)-OTHER AND UNSPECIFIED HYPERLIPIDEMIA_AP
(25070)-DIABETES WITH PERIPHERAL CIRCULATORY DISORDERS TYPE II OR UNSPECIFIED TYPE NOT STATED AS UNCONTROLLED_AP
(V4589)-OTHER POSTSURGICAL STATUS
(49390)-ASTHMA UNSPECIFIED_AP
(V142)-PERSONAL HISTORY OF ALLERGY TO SULFONAMIDES
(2738)-OTHER DISORDERS OF PLASMA PROTEIN METABOLISM_AP
(412)-OLD MYOCARDIAL INFARCTION_AP
(V145)-PERSONAL HISTORY OF ALLERGY TO NARCOTIC AGENT
(496)-CHRONIC AIRWAY OBSTRUCTION NOT ELSEWHERE CLASSIFIED_AP
(42842)-CHRONIC COMBINED SYSTOLIC AND DIASTOLIC HEART FAILURE_AP
(3899)-UNSPECIFIED HEARING LOSS_AP
(V4579)-OTHER ACQUIRED ABSENCE OF ORGAN
(78659)-OTHER CHEST PAIN_AP
(V4582)-PERCUTANEOUS TRANSLUMINAL CORONARY ANGIOPLASTY STATUS_AP
(30000)-ANXIETY STATE UNSPECIFIED_AP
(V4501)-CARDIAC PACEMAKER IN SITU
(V141)-PERSONAL HISTORY OF ALLERGY TO OTHER ANTIBIOTIC AGENT
(79029)-OTHER ABNORMAL GLUCOSE_AP
(71690)-UNSPECIFIED ARTHROPATHY SITE UNSPECIFIED
(E8781)-SURGICAL OPERATION WITH IMPLANT OF ARTIFICIAL INTERNAL DEVICE CAUSING ABNORMAL PATIENT REACTION OR LATER COMPLICATION WITHOUT MISADVENTURE AT TIME OF OPERATION_AP
(V433)-HEART VALVE REPLACED BY OTHER MEANS
(2749)-GOUT UNSPECIFIED_AP
(42989)-OTHER ILL-DEFINED HEART DISEASES_AP
(V180)-FAMILY HISTORY OF DIABETES MELLITUS
(2753)-DISORDERS OF PHOSPHORUS METABOLISM_AP
(V4983)-AWAITING ORGAN TRANSPLANT STATUS_AP
(25000)-DIABETES MELLITUS WITHOUT MENTION OF COMPLICATION TYPE II OR UNSPECIFIED TYPE NOT STATED AS UNCONTROLLED_AP
(V8741)-PERSONAL HISTORY OF ANTINEOPLASTIC CHEMOTHERAPY
(2851)-ACUTE POSTHEMORRHAGIC ANEMIA_AP
(42832)-CHRONIC DIASTOLIC HEART FAILURE_AP
(42789)-OTHER SPECIFIED CARDIAC DYSRHYTHMIAS_AP
(32723)-OBSTRUCTIVE SLEEP APNEA (ADULT) (PEDIATRIC)_AP
(V140)-PERSONAL HISTORY OF ALLERGY TO PENICILLIN
(4111)-INTERMEDIATE CORONARY SYNDROME_AP
(42833)-ACUTE ON CHRONIC DIASTOLIC HEART FAILURE_AP
(41400)-CORONARY ATHEROSCLEROSIS OF UNSPECIFIED TYPE OF VESSEL NATIVE OR GRAFT_AP
(4240)-MITRAL VALVE DISORDERS_AP
(5533)-DIAPHRAGMATIC HERNIA WITHOUT OBSTRUCTION OR GANGRENE_AP
(V5866)-LONG-TERM (CURRENT) USE OF ASPIRIN
(V1508)-PERSONAL HISTORY OF ALLERGY TO RADIOGRAPHIC DYE
(2449)-UNSPECIFIED ACQUIRED HYPOTHYROIDISM_AP
(4293)-CARDIOMEGALY_AP
(3970)-DISEASES OF TRICUSPID VALVE_AP

(Continued)

TABLE E2. Continued

(4239)-UNSPECIFIED DISEASE OF PERICARDIUM_AP
(V1581)-PERSONAL HISTORY OF NONCOMPLIANCE WITH MEDICAL TREATMENT PRESENTING HAZARDS TO HEALTH
(V146)-PERSONAL HISTORY OF ALLERGY TO ANALGESIC AGENT
(5849)-ACUTE KIDNEY FAILURE UNSPECIFIED_AP
(4412)-THORACIC ANEURYSM WITHOUT RUPTURE
(40391)-HYPERTENSIVE CHRONIC KIDNEY DISEASE UNSPECIFIED WITH CHRONIC KIDNEY DISEASE STAGE V OR END STAGE RENAL DISEASE_AP
(51852)-OTHER PULMONARY INSUFFICIENCY NOT ELSEWHERE CLASSIFIED FOLLOWING TRAUMA AND SURGERY_AP
(4295)-RUPTURE OF CHORDAE TENDINEAE_AP
(311)-DEPRESSIVE DISORDER NOT ELSEWHERE CLASSIFIED_AP
(27800)-OBESITY UNSPECIFIED_AP
(4148)-OTHER SPECIFIED FORMS OF CHRONIC ISCHEMIC HEART DISEASE_AP
(V5867)-LONG-TERM (CURRENT) USE OF INSULIN_AP
(25002)-DIABETES MELLITUS WITHOUT MENTION OF COMPLICATION TYPE II OR UNSPECIFIED TYPE UNCONTROLLED_AP
(4263)-OTHER LEFT BUNDLE BRANCH BLOCK_AP
(5119)-UNSPECIFIED PLEURAL EFFUSION_AP
(7464)-CONGENITAL INSUFFICIENCY OF AORTIC VALVE
(5180)-PULMONARY COLLAPSE_AP
(V462)-DEPENDENCE ON SUPPLEMENTAL OXYGEN
(42731)-ATRIAL FIBRILLATION_AP
(56400)-UNSPECIFIED CONSTIPATION_AP
(V4502)-AUTOMATIC IMPLANTABLE CARDIAC DEFIBRILLATOR IN SITU
(E8497)-ACCIDENTS OCCURRING IN RESIDENTIAL INSTITUTION_AP
(60000)-HYPERTROPHY (BENIGN) OF PROSTATE WITHOUT URINARY OBSTRUCTION AND OTHER LOWER URINARY TRACT SYMPTOMS (LUTS)
(42823)-ACUTE ON CHRONIC SYSTOLIC HEART FAILURE_AP
(4260)-ATRIOVENTRICULAR BLOCK COMPLETE_AP
(4400)-ATHEROSCLEROSIS OF AORTA_AP
(4280)-CONGESTIVE HEART FAILURE UNSPECIFIED_AP
(2768)-HYPOPOTASSEMIA_AP
(60001)-HYPERTROPHY (BENIGN) OF PROSTATE WITH URINARY OBSTRUCTION AND OTHER LOWER URINARY TRACT SYMPTOMS (LUTS)_AP
(78551)-CARDIOGENIC SHOCK_AP
(49320)-CHRONIC OBSTRUCTIVE ASTHMA UNSPECIFIED
(4142)-CHRONIC TOTAL OCCLUSION OF CORONARY ARTERY_AP
(E8798)-OTHER SPECIFIED PROCEDURES AS THE CAUSE OF ABNORMAL REACTION OF PATIENT OR OF LATER COMPLICATION WITHOUT MISADVENTURE AT TIME OF PROCEDURE_AP
(V4364)-HIP JOINT REPLACEMENT_AP
(5845)-ACUTE KIDNEY FAILURE WITH LESION OF TUBULAR NECROSIS_AP
(E8790)-CARDIAC CATHETERIZATION AS THE CAUSE OF ABNORMAL REACTION OF PATIENT OR OF LATER COMPLICATION WITHOUT MISADVENTURE AT TIME OF PROCEDURE_AP
(3963)-MITRAL VALVE INSUFFICIENCY AND AORTIC VALVE INSUFFICIENCY
(5854)-CHRONIC KIDNEY DISEASE STAGE IV (SEVERE)_AP
(2720)-PURE HYPERCHOLESTEROLEMIA_AP
(2631)-MALNUTRITION OF MILD DEGREE_AP
(2752)-DISORDERS OF MAGNESIUM METABOLISM_AP
(4589)-HYPOTENSION UNSPECIFIED_AP
(V1083)-PERSONAL HISTORY OF OTHER MALIGNANT NEOPLASM OF SKIN
(4210)-ACUTE AND SUBACUTE BACTERIAL ENDOCARDITIS_AP
(V5863)-LONG-TERM (CURRENT) USE OF ANTIPLATELETS/ANTITHROMBOTICS
(3659)-UNSPECIFIED GLAUCOMA
(2809)-IRON DEFICIENCY ANEMIA UNSPECIFIED_AP
(3572)-POLYNEUROPATHY IN DIABETES
(E8788)-OTHER SPECIFIED SURGICAL OPERATIONS AND PROCEDURES CAUSING ABNORMAL PATIENT REACTION OR LATER COMPLICATION WITHOUT MISADVENTURE AT TIME OF OPERATION_AP
(27652)-HYPOVOLEMIA_AP

(Continued)



TABLE E2. Continued

(41401)-CORONARY ATHEROSCLEROSIS OF NATIVE CORONARY ARTERY_AP
(9971)-CARDIAC COMPLICATIONS NOT ELSEWHERE CLASSIFIED_AP
(42611)-FIRST DEGREE ATRIOVENTRICULAR BLOCK_AP
(43889)-OTHER LATE EFFECTS OF CEREBROVASCULAR DISEASE
(07070)-UNSPECIFIED VIRAL HEPATITIS C WITHOUT HEPATIC COMA
(99593)-SYSTEMIC INFLAMMATORY RESPONSE SYNDROME DUE TO NONINFECTIOUS PROCESS WITHOUT ACUTE ORGAN DYSFUNCTION_AP
(4168)-OTHER CHRONIC PULMONARY HEART DISEASES_AP
(V4561)-CATARACT EXTRACTION STATUS
(V148)-PERSONAL HISTORY OF ALLERGY TO OTHER SPECIFIED MEDICINAL AGENTS
(V1749)-FAMILY HISTORY OF OTHER CARDIOVASCULAR DISEASES_AP
(4011)-BENIGN ESSENTIAL HYPERTENSION_AP
(V1255)-PERSONAL HISTORY OF PULMONARY EMBOLISM_AP
(4019)-UNSPECIFIED ESSENTIAL HYPERTENSION_AP
(V1251)-PERSONAL HISTORY OF VENOUS THROMBOSIS AND EMBOLISM_AP
(5856)-END STAGE RENAL DISEASE_AP
(44020)-ATHEROSCLEROSIS OF NATIVE ARTERIES OF THE EXTREMITIES UNSPECIFIED_AP
(4139)-OTHER AND UNSPECIFIED ANGINA PECTORIS_AP
(262)-OTHER SEVERE PROTEIN-CALORIE MALNUTRITION_AP
(V5861)-LONG-TERM (CURRENT) USE OF ANTICOAGULANTS
(V8801)-ACQUIRED ABSENCE OF BOTH CERVIX AND UTERUS
(30500)-NONDEPENDENT ALCOHOL ABUSE UNSPECIFIED DRINKING BEHAVIOR_AP
(2875)-THROMBOCYTOPENIA UNSPECIFIED_AP
(7140)-RHEUMATOID ARTHRITIS
(5853)-CHRONIC KIDNEY DISEASE STAGE III (MODERATE)_AP
(5859)-CHRONIC KIDNEY DISEASE UNSPECIFIED_AP
(7469)-UNSPECIFIED CONGENITAL ANOMALY OF HEART
(4254)-OTHER PRIMARY CARDIOMYOPATHIES_AP
(V1254)-PERSONAL HISTORY OF TRANSIENT ISCHEMIC ATTACK (TIA) AND CEREBRAL INFARCTION WITHOUT RESIDUAL DEFICITS_AP
(4928)-OTHER EMPHYSEMA_AP
(V4581)-POSTSURGICAL AORTOCORONARY BYPASS STATUS
(V1582)-PERSONAL HISTORY OF TOBACCO USE_AP
(5990)-URINARY TRACT INFECTION SITE NOT SPECIFIED_AP
(27801)-MORBID OBESITY_AP
(78079)-OTHER MALAISE AND FATIGUE_AP
(33818)-OTHER ACUTE POSTOPERATIVE PAIN_AP
(27669)-OTHER FLUID OVERLOAD_AP
(V4365)-KNEE JOINT REPLACEMENT
(2761)-HYPOSMOLALITY AND/OR HYPONATREMIA_AP
(2859)-ANEMIA UNSPECIFIED_AP
(2639)-UNSPECIFIED PROTEIN-CALORIE MALNUTRITION_AP
(4439)-PERIPHERAL VASCULAR DISEASE UNSPECIFIED_AP
(V5869)-LONG-TERM (CURRENT) USE OF OTHER MEDICATIONS
(28521)-ANEMIA IN CHRONIC KIDNEY DISEASE_AP
(53081)-ESOPHAGEAL REFLUX_AP
(78605)-SHORTNESS OF BREATH
(V070)-NEED FOR ISOLATION_AP
(28860)-LEUKOCYTOSIS UNSPECIFIED_AP
(42843)-ACUTE ON CHRONIC COMBINED SYSTOLIC AND DIASTOLIC HEART FAILURE_AP
(4241)-AORTIC VALVE DISORDERS_AP
(99801)-POSTOPERATIVE SHOCK CARDIOGENIC_AP
(V707)-EXAMINATION OF PARTICIPANT IN CLINICAL TRIAL
(V173)-FAMILY HISTORY OF ISCHEMIC HEART DISEASE
(73300)-OSTEOPOROSIS UNSPECIFIED
(56210)-DIVERTICULOSIS OF COLON (WITHOUT HEMORRHAGE)

(Continued)

TABLE E2. Continued

(3051)-NONDEPENDENT TOBACCO USE DISORDER  
 (99672)-OTHER COMPLICATIONS DUE TO OTHER CARDIAC DEVICE IMPLANT AND GRAFT\_AP  
 (7455)-OSTIUM SECUNDUM TYPE ATRIAL SEPTAL DEFECT\_AP  
 (71590)-OSTEOARTHRISIS UNSPECIFIED WHETHER GENERALIZED OR LOCALIZED INVOLVING UNSPECIFIED SITE  
 (V103)-PERSONAL HISTORY OF MALIGNANT NEOPLASM OF BREAST  
 (V1504)-PERSONAL HISTORY OF ALLERGY TO SEAFOOD\_AP  
 (42822)-CHRONIC SYSTOLIC HEART FAILURE\_AP  
 (40390)-HYPERTENSIVE CHRONIC KIDNEY DISEASE UNSPECIFIED WITH CHRONIC KIDNEY DISEASE STAGE I THROUGH STAGE IV OR UNSPECIFIED\_AP  
 (42732)-ATRIAL FLUTTER\_AP  
 (34690)-MIGRAINE UNSPECIFIED WITHOUT MENTION OF INTRACTABLE MIGRAINE WITHOUT MENTION OF STATUS MIGRAINOSUS\_AP  
 (34590)-EPILEPSY UNSPECIFIED WITHOUT INTRACTABLE EPILEPSY\_AP  
 (E8782)-SURGICAL OPERATION WITH ANASTOMOSIS BYPASS OR GRAFT WITH NATURAL OR ARTIFICIAL TISSUES USED AS IMPLANT CAUSING ABNORMAL PATIENT REACTION OR LATER COMPLICATION WITHOUT MISADVENTURE AT TIME OF OPERATION\_AP  
 (42821)-ACUTE SYSTOLIC HEART FAILURE\_AP  
 (78650)-UNSPECIFIED CHEST PAIN\_AP

**Demographic**

AGECALCYEARS  
 WEIGHT\_KG  
 GENDER  
 VIP

**Clinical**

PREOPERATIVE\_LENGTH\_OF\_STAY  
 LIVER\_DISEASE  
 PREVIOUS\_CARDIAC\_SURGERIES\_TOTAL\_NO  
 HYPERTENSION  
 ASA\_STATUS  
 ATRIAL\_FIBRILATION  
 CARDIOGENIC\_SHOCK  
 INOTROPES\_WI\_48 h  
 PERIPHERAL\_ARTERIAL\_DISEASE  
 DIABETES  
 RADIATION  
 ADMITTED\_PREOP  
 IMMUNOCOMPROMISE  
 TYPE\_A\_DISSECTION  
 Nesiritide\_Inf\_WI\_48 h  
 PREOPERATIVE\_MECHANICAL\_VENTILATION  
 IABP  
 Beta=Blocker\_WI\_24\_h  
 SURGICAL\_HISTORY\_PRESENT  
 CALCIFIED\_AORTA  
 FAMHXREPORTED  
 Heparinin\_Inf\_WI\_24 h  
 MAC  
 DIALYSIS

**Preoperative Before Admission Medications**

PTA\_ANTILIPIDABSORPTION  
 PTA\_LIDOCAINEPATCH  
 PTA\_MINERALSELECTROLYTESPOTASSIUM  
 PTA\_LMWHEPARIN  
 PTA\_ALDOSTERONERECEPTORANTAGONIST

(Continued)

TABLE E2. Continued

PTA\_DIABETESORAL  
 PTA\_ALLERGYANTIHISTAMINE  
 PTA\_PLATELETINHIBITORS  
 PTA\_DIURETICTHIAZIDE  
 PTA\_CLONIDINE  
 PTA\_AMIODARONEDRONEDARONE  
 PTA\_IMMUNOSUPPRESSIVE  
 PTA\_CCBDIHYDROPYRIDINE  
 PTA\_MINERALSELECTROLYTESFOLICACID  
 PTA\_DIABETESINSULIN  
 PTA\_ACEINHIBITOR  
 PTA\_ANTIDEPRESSANT  
 PTA\_GOUTAGENT  
 PTA\_ANTIPSYCHOTIC  
 PTA\_GLUCOCORTICOIDTOPICAL  
 PTA\_ASTHMACOPD  
 PTA\_ANTIANGINAL  
 PTA\_ERYTHROPOIETINS  
 PTA\_ALLERGYNASALSTEROID  
 PTA\_THYROIDHORMONE  
 PTA\_BENZODIAZEPINE  
 PTA\_ANTIEMETIC  
 PTA\_BONERESORPTIONINHIBITORS  
 PTA\_OPTHALMIC  
 PTA\_ALTERNATIVETHERAPY  
 PTA\_MINERALSELECTROLYTESMAGNESIUM  
 PTA\_ANTICONVULSANT  
 PTA\_DIGOXIN  
 PTA\_GABAANALOG  
 PTA\_ANALGESICNARCOTIC  
 PTA\_SMOKINGDETERRENTS  
 PTA\_ACETAMINOPHEN  
 PTA\_BPHAGENT  
 PTA\_DIURETICLOOP  
 PTA\_ANTIREFLUXAGENTH2BLOCKER  
 PTA\_CCBNONDIHYDROPYRIDINE  
 PTA\_ASPIRIN  
 PTA\_ANTIPLIPIDDIET  
 PTA\_MINERALSELECTROLYTESIRON  
 PTA\_HYDRALAZINE  
 PTA\_ANTIBIOTIC  
 PTA\_ANTIPLIPIDFIBRICACID  
 PTA\_ANTIREFLUXAGENTPPI  
 PTA\_LAXATIVE  
 PTA\_ANTIPLIPIDSTATIN  
 PTA\_ARB  
 PTA\_BETA-BLOCKERS  
 PTA\_NSAID  
 PTA\_DIRECTFACTORXAINHIBITOR  
 PTA\_ANTIPLATULENT  
 PTA\_SLEEPAID  
 PTA\_PHOSPHATEBINDER

**Preoperative In Hospital Medications**

Analgesic Narcotic Oxycodone Combinations  
 General Anesthetic-Parenteral, Benzodiazepines

(Continued)

**TABLE E2. Continued**


---

Minerals and Electrolytes-Iron
Vaccine Bacterial-Gram-Positive Cocci
Electrolyte Depleters-Ion Exchange Resin
Nesiritide IV Infusion
Vitamins-Folic Acid and Derivatives
Sedative-Hypnotic-Antihistamines
Gastric Acid Secretion Reducers-Histamine H2-Receptor Antagonists
Human Insulins-Short Acting
VitaminsK, Phytonadione and Derivatives
Antihyperlipidemic-Selective Cholesterol Absorption Inhibitor
Antihyperglycemic-Sulfonylurea Derivatives
Minerals and Electrolytes-Potassium for Injection
Cardiac Inotropes-Phosphodiesterase Inhibitors
Antacid-Simethicone Combinations
Dextrose Solutions, Concentrated
Asthma Therapy-Leukotriene Receptor Antagonists
Diuretic-Loop
Asthma/COPD-Anticholinergic Agents, Inhaled Short Acting
Gastric Acid Secretion Reducing Agents-Proton Pump Inhibitors (PPIs)
Anticonvulsant-GABA Analogs
Laxative-Stimulant
Digitalis Glycosides
Minerals and Electrolytes-Magnesium
Glucocorticoids
Antianxiety Agent-Benzodiazepines
Heparins
Asthma/COPD Therapy-Beta Adrenergic-Glucocorticoid Combinations
Platelet Aggregation Inhibitors-Thienopyridine Agents
Phosphate Binders
Antianginal-Coronary Vasodilators (Nitrates)
Minerals and Electrolytes-Potassium, Oral
Diuretic-Aldosterone Receptor Antagonist, Nonselective
Beta-Blockers Cardiac Selective
Laxative-Saline and Osmotic
Sodium Chloride, Parenteral
Hyperuricemia Therapy-Xanthine Oxidase Inhibitors
Minerals and Electrolytes-Calcium Replacement
Antiarrhythmic-Class III
Platelet Aggregation Inhibitors-Salicylates
Multivitamins
Prostatic Hypertrophy Agent-alpha-1-Adrenoceptor Antagonists
Nesiritide IV in D5W
Sedative-Hypnotic-GABA-Receptor Modulators
Insulin Analogs-Long Acting
Asthma/COPD Therapy-Beta 2-Adrenergic Agents, Inhaled, Short Acting
Cardiovascular Sympathomimetics
Minerals and Electrolytes-Parenteral Electrolyte Combinations
Diuretic-Thiazides and Related
Anticoagulants-Coumarin
Analgesic Narcotic Agonists
Glycopeptide Antibiotics
Salicylate Analgesics
Dermatological-Antibacterial Other
Thyroid Hormones-Synthetic T4 (Thyroxine)
Analgesic or Antipyretic Non-Narcotic
Laxative-Surfactant

---

*(Continued)*

**TABLE E2. Continued**

Ophthalmic-Intraocular Pressure Reducing Agents, Prostaglandin Analogs  
 Amino Acids, Single Ingredient, Oral (noninjectable)  
 Vitamin D Derivatives  
 Prostatic Hypertrophy Agent-Type II 5-alpha Reductase Inhibitors  
 Cephalosporin Antibiotics-3rd Generation  
 Calcium Channel Blockers-Benzothiazepines  
 Smoking Deterrents-Nicotine-Type  
 Low Molecular Weight Heparins  
 Antidepressant-Selective Serotonin Reuptake Inhibitors (SSRIs)  
 Contrast Media-Ultrasound Agents  
 Antihyperlipidemic-HMG CoA Reductase Inhibitors (statins)  
 Fluoroquinolone Antibiotics  
 Antianginal and Anti-ischemic Agents, Nonhemodynamic  
 Insulin Analogs-Rapid Acting  
 Antiemetic-Selective Serotonin 5-HT3 Antagonists  
 General Anesthetic Adjuncts-Narcotic  
 Cephalosporin Antibiotics-4th Generation  
 Direct Acting Vasodilators  
 ACE Inhibitors  
 Alpha-Beta Blockers  
 Angiotensin II Receptor Blockers (ARBs)  
 Antidiuretic and Vasopressor Hormones Calcium Channel Blockers-Dihydropyridines

For the administrative features, the numerical Current Procedural Terminology (CPT) code with its corresponding feature is provided.

**TABLE E3. The number of patients in the full cohort, non-STS index case cohort, and STS index case cohort, along with the specific STS index case types, in the overall, training, and test sets**

Surgery type	Total patients, n	Alive, n	Deceased, n
Overall data set	6392	6199	193
Training	5113	4959	154
Test	1279	1240	39
Non-STS index procedures	1905	1781	124
Training	1551	1448	103
Test	354	333	21
STS index procedures	4487	4418	69
Training	3562	3511	51
Test	925	907	18
CABG	1527	1510	17
Training	1215	1200	15
Test	312	310	2
AVR	497	494	3
Training	398	396	2
Test	99	98	1
MVR	199	195	4
Training	147	143	4
Test	52	52	0
MVRepa	1399	1382	17
Training	1107	1095	12
Test	292	287	5
CABG + AVR	196	191	5
Training	161	158	3
Test	35	33	2
CABG + MVR	24	23	1
Training	20	20	0
Test	4	3	1
CABG + MVRepa	248	245	3
Training	198	196	2
Test	50	49	1

Additionally, the number of patients who lived or died after cardiac surgery are provided. *STS*, Society of Thoracic Surgeons; *CABG*, coronary artery bypass grafting; *AVR*, aortic valve replacement; *MVR*, mitral valve replacement; *MVRepa*, mitral valve repair.