Predicting Stroke Risk for Heart Surgery Patients: Comparison of Results with Logistic Regression vs. Machine Learning Techniques

Akhil Patel, Frank Shannon, and Nate Velarde

Abstract

Strokes are one of the most costly health conditions from both a financial and human standpoint. The stroke incidence rate per capita is twice as high for individuals aged 50 and above that have had heart surgery than for those who have not. The Society of Thoracic Surgery publishes Adult Cardiac Surgery Risk Models that, among other things, predict stroke risk for heart surgery patients. These models are based on logistic regression techniques on preoperative features selected by a panel of physicians. We modeled stroke risk using a variety of machine learning techniques on an extensive set of preoperative features as well as a set that included both preoperative and postoperative features. No models outperformed STS when utilizing solely preoperative data. However, we found several models that outperformed the STS model, as measured by the area under the receiver operating characteristic curve (c-statistic), when postoperative data was incorporated. This suggests a rationale for the increased use of postoperative data to improve perioperative care to reduce strokes after heart surgery.

Introduction

In the current era of heightened focus on the cost and performance of health care as well as the transparency of professional medical societies, the Society of Thoracic Surgery (STS) is one of the leaders in public reporting of institutional results. Formed 30 years ago, the STS Adult Cardiac Surgery Database was designed to collect outcome data of heart surgical cases so that mortality and complications would be reported in the context of patient risk factors, rather than raw numbers. This data collection and review process improved cardiac surgical outcomes by establishing performance benchmarks as well as a rigorous methodology for risk model formulation. The STS Adult Cardiac Surgery Database was composed of clinically granular and standardized data to characterize preoperative patient comorbidities.[1] These patient factors were followed by a description of specific operative parameters and culminated in a list of adverse outcomes including death. The variable number and definitions have varied over the years as the database was updated to include more predictive factors, newer surgical procedures, and updated quality metrics.

The statistical methodology for elaborating cardiac surgery outcome models evolved over the past 30 years. Built on years of surgical data from an ever increasing number of heart surgery programs in the US, the STS ACSD initially used Bayesian techniques for diagnostic evaluation and risk prediction. Logistic regression was introduced in 1998 for more robust risk modeling and has been used ever since because it achieved superior predictive accuracy with a parsimonious group of independent variables.[2] Composite risk models were initially created for the basic cardiac surgical procedures of isolated coronary artery bypass grafting (CABG) and isolated aortic valve replacement. The models for prediction of the six primary adverse outcomes of cardiac surgery (death, stroke, renal failure, sternal wound infection, prolonged ventilator support, and reoperation for bleeding) were refined every 6 to 8 years as surgical techniques evolved and outcomes improved. The overarching goal of STS risk modeling is to adjust for all preoperative factors that are significantly associated with adverse outcomes and vary across STS participant programs. Models for seven different cardiac surgical procedures were derived and recalibrated as the core procedures changed in difficulty. These procedures are isolated CABG (coronary artery bypass graft), isolated AVR (aortic valve replacement), isolated MVR (mitral valve replacement), isolated MVRrep (mitral valve repair) AVR + CABG, MVR + CABG and MVRrep + CABG.

The risk models for the primary procedures of isolated CABG, isolated valve replacement and valve replacement + CABG are used for the following reasons:

(1) patient counseling about the percent risk of adverse outcomes for a given procedure,
(2) benchmark best outcomes for each procedure,
(3) guide performance improvement initiatives,
(4) provide robust case-mix adjustment for estimating risk-adjusted outcomes that are then used in public reporting, regulatory compliance, and value-based reimbursement.

The substantial influence of a program's risk-adjusted performance on hospital funding, patient referral, and peer review requires transparency of the modeling process. Hierarchical logistic regression with hyperparameter tuning remains the modeling method of choice for the STS because it is more transparent than AI methods, computationally simpler, more likely to converge and not subject to the ills of an "overparameterized model" (unstable, noisy and overfit). Based on comparative ROC AUC tests for the latest STS Adult Cardiac Surgery Risk Models (2018), the logistic regression methodology performs well for all outcomes except for reoperation (range, 0.574 to 0.627) and stroke (range, 0.616 to 0.704). Of the two complications, stroke is more clinically significant than reoperation and second only to mortality as the most important adverse outcome of heart surgery.

We have chosen to determine whether or not we can improve on the performance of the 2018 STS risk model for stroke by using all variables in the Pre-Op phase in a variety of classification algorithms that included Logistic Regression, Decision Trees, and Neural Nets. This Pre-Op phase data consists of approximately 90 to 100 features that include demographic data, patient associated disease states, primary cardiac problems, laboratory results, and performance measurements. In the STS method, these features were ranked in causal importance for a specific complication by linear regression. The top 50 were then reviewed by a committee of 10 senior, academic cardiac surgeons who selected variables that they felt were most important in causing the complication. The selected variable list was further regressed until the best model by c-statistic criteria was produced. In contrast, our modelling steps are described below.

Methods

Dataset

The dataset for training and testing was obtained from the Michigan State Thoracic and Cardiovascular Society for use in our university-sponsored data science program. IRB approval for this project was previously obtained from the University of Michigan by the MSTCVS for quality improvement projects approved by the Research Committee. As per HIPPA requirements, the data was fully de-identified of all direct or indirect labels. A total of 42,746 STS ACSD records were obtained. The dataset included all cardiac surgery procedures that were risk-adjusted in the state of Michigan from July 1, 2011 to December 31, 2016. The procedure categories included isolated CABG, isolated AVR, isolated MVR, isolated MVRrep, AVR + CABG, MVR + CABG, and MVRrep + CABG. These records spanned STS database version 2.73 (7/1/2011 to 6/30/2014) and version 2.81 (7/1/2014 to 12/31/2016). This interval was chosen because it is the same time period used by the STS to revise their risk model that was published and promulgated in 2018.[3,4] The data columns for each patient was reduced to 409 data fields that were pertinent to our study. There were 21,903 cases in version 2.73 and 20,843 cases in version 2.81.

Version Mapping

Of the total 130 Pre-Op fields, there were 48 parent categorical fields that had been expanded in version 2.81. These parent fields had different names and had to be identified and mapped one-by-one for clarity. The PREOP parent and child data fields were harmonized by reducing the granular choices in 2.81 data to equivalent choices in 2.73. For data fields that categorized symptoms or disease states into different degrees of clinical severity, dummy variables were constructed for multi-level tree analysis. When completed, there were 72 PREOP fields (61 categorical and 11 numeric). These features were later expanded to 110 when multi-level categorical variables were split into their respective dummy variables.

The remaining 220 data fields were labeled POSTOP in our database. Of these, 67 were selected as surgical operation features. Of these 67 fields, 52 were categorical and 15 were numeric. The primary surgical operation types from version 2.81 were mapped to the same type in 2.73. The remaining fields were outcome details and therefore only included ‘stroke’ and ‘mortality’ for our study. The 67 operation features were expanded to 105 for multi-level categorical analysis.

Modeling

PREOP Features

The distinctive feature of the STS modeling method described above is the introduction of "humans-in-the-loop" to review the feature ranking produced by logistic regression and select variables that made the most "clinical sense".[5] Strictly speaking, this human layer in the selection process adds bias to the model that is respected by other physicians, especially cardiac surgeons, and adds credibility by not making it a strictly "machine process". Also, it helped create a parsimonious analytic process by manually removing surrogate and proxy variables that could be creating multicollinearity in the model. In the current climate of "data-driven processing", however, adding the human layer could be regarded as a step that removes features that could contribute to the performance of the model, rather than a sanity check. In our modeling at every level, we used all of the PREOP and POSTOP variables so as to avoid bias. Thus, our PREOP feature list is the same as the STS list but more complete (Table 3).

We modeled the PREOP data using a variety of classification algorithms that included Logistic Regression, Decision Trees and Neural Nets. We tried neural nets in a number of configurations, but there was insufficient data to produce a performance advantage in this dataset.[6] After tuning the hyperparameters and trying various transformations of the features (polynomial), Logistic Regression with LASSO and Random Forest Classifier exhibited the best discrimination for stroke risk, as measured by our primary evaluation metric, the c-statistic.

Based on the STS Pre-Op database features, the risk models for stroke generally performed about the same as the new STS model. When we examined feature importance measures of our PREOP feature set via (a) SelectKBest (chi-squared statistics for categorical features and ANOVA F-values for numerical), (b) non-zero coefficients from Logistic Regression with LASSO, and (c) Random Forest Classifier Feature Importances, we found a lot of overlap between the statistically significant/non-zero coefficient/high feature importance value features in our PREOP model and the features used by the STS model (Tables 4 and 5). On the basis of these results, it seems to be safe to posit that the Pre-Operative features in the STS database do not fully encompass the risk factors/physiological traits/-omic characteristics that can predict strokes that occur during cardiac surgical treatment.

Table 1: Comparison of c-statistic between STS and models trained on preoperative data

Model	CABG	Valve	CABG + Valve
MSTCVS data points (total)	28,945	8,500	5,295
MSTCVS data points (test)	5,789	1,700	1,059
STS (2018 model)	0.697	0.656	0.632
STS (train set)	0.700	0.619	0.671
STS (test set)	0.710	0.687	0.713
PREOP Logistic Regression	0.663	0.622	0.616
PREOP Random Forest	0.692	0.637	0.655

POSTOP Features

The next step in our modeling process was to determine if we could improve upon the baseline STS stroke model by adding POSTOP features to the model. Some of the features that we felt could be more important were details regarding the heart surgery procedure itself, like the duration the patient was on the heart lung machine, the blood pressure and blood count ranges on the heart lung machine, and oxygen levels measured in the blood during and after surgery. Other potential factors are the results of blood tests that measure kidney function and ultrasound examinations of the heart and major arteries in the chest. These features were included in the POSTOP data set.

Following the same data cleaning and variable mapping process that we performed for the PREOP variables, we added 67 primary POSTOP features (later expanded to 114 when multi-level categorical variables were split into their respective dummy variables).

We incorporated the POSTOP features into modeling process in two ways: (a) one-step procedure by which we simply added the POSTOP features to the PREOP features (designated PRE + POST in the Table 2) and (b) a two-step or STACKED process by which we ran the PREOP models (Logistic Regression and Random Forest Classifier), extracted the predicted probabilities of stroke for each observation and added these probability vectors as additional features to our POSTOP dataset.

We tried a multitude of classification algorithms and transformations to our POSTOP feature enhanced dataset, but found the best results (as well as the fastest compute times and better interpretability) from Logistic Regression with LASSO and Random Forest Classifier. As the results below indicate, our models with POSTOP features outperformed the baseline STS model on our primary model evaluation metric, the c-statistic. Feature Importance Analysis is tabulated for combined PRE + POST and STACKED PRE-POST methods (Tables 6 and 7).

Table 2: Comparison of c-statistic for STS models, models trained on solely preoperative data, and models utilizing both preoperative and postoperative data

Model	CABG	Valve	CABG + Valve
STS (train set)	0.700	0.619	0.713
STS (test set)	0.710	0.687	0.713
PREOP Logistic Regression	0.663	0.622	0.616
PREOP Random Forest	0.692	0.637	0.655
PRE + POST Logistic Regression	0.729	0.705	0.657
PRE + POST Stacked Log Regression	0.717	0.715	0.683
PRE + POST Random Forest	0.724	0.726	0.659

Discussion

The purpose of this study is to determine if there are any advantages to using machine learning techniques to improve prediction of stroke in a population of patients undergoing open heart surgery. The gold standard for risk prediction and modeling in cardiac surgery is the Society of Thoracic Surgery's Adult Cardiac Surgery Database that was launched in 1989. On its 30th anniversary, the STS ACSD has over 1,000 heart surgery hospitals participating in the database and a cumulative patient record repository of over six and a half million patients.[1] The prediction models for standard cardiac surgery cases are recalibrated annually and recalculated every 6 to 8 years. The features or risk factors that are used to determine surgical mortality and stroke risk have remained relatively constant for the past 20 years.

Despite the size of the database, the discriminating performance of the STS model for stroke prediction is disappointing in comparison to what we have grown accustomed to in our data-driven occupations. There are a number of potential reasons for this better than average (but not excellent) model: (a) the modeling has remained restricted to linear regression statistical techniques, (b) the modeling procedure involves experienced, cardiac surgical authorities to select risk factors from a ranked list that they think are clinically relevant - "human-in-the-loop" process for credibility, (c) strokes in the context of heart surgery are a dynamic process that is multifactorial, not entirely pre-determined by a patient's preoperative medical problems, a consequence of the operation itself and due to a factor that we have yet to identify.[7,8]

Our comparative study seems to address some of these possibilities and points to future approaches. Using the c-statistic as our performance metric, the STS 2018 risk model for stroke is better than our best AI/ML models (Logistic Regression and Random Forest) we could find using all of the preoperative features for the 3 standard cardiac surgery categories. The only wrinkle we added to our logistic regression model in comparison to STS was using LASSO technique for feature weighting. It seems valid to postulate in this circumstance that there are unspecified or uncaptured features that are not captured by the STS database and could improve our model performance. There are no reports in the medical literature that address the problem of modeling stroke occurrence after cardiac surgery or interventional cardiology procedures using machine learning. Consequently, we have found no evidence that the STS method using linear regression and expert surgeon risk factor selection could be improved. However, the second phase of our study is provocative because using data from what we label the “POSTOP” part of the STS database gives us better results than the STS standard for CABG and Valve surgery patients.

Including data from the operation and early post-op phase of care seems to significantly improve our model performance. As we can see, the c-statistic for either a stacked or combined dataset is significantly better with our feature importance weighted towards dynamic measurements like blood counts and brain oxygen measurements while on the heart lung machine, use of blood products during surgery, assessment of the aorta during surgery, and kidney function after surgery. All of these variables can be captured in real time with digital OR and ICU monitoring equipment. Use of concurrent data from digital monitoring systems in the ICU to predict mortality outcomes in critically ill ICU patients has been reported.[9] As Holmgren et al have reported, using ANN to process the SAPS (Simplified Acute Physiology Score) with a data foundation of 200,000+ ICU patients in Sweden has a predictive c-statistic for ICU mortality ranging from 0.882 to 0.910.[10] This mortality prediction can be updated daily, used to triage care and guide care improvements.

A leading cause of stroke following heart surgery is the development of atrial fibrillation (rapid heart beat, “palpitations”). AF occurs in up to 30% of patients within 5 days of surgery, is usually converted back to regular rhythm but causes blood clot formation within the heart that can then be pumped to the brain. Anticoagulation is the preventive treatment for clot formation and stroke, but that also has a complication of delayed bleeding around the heart. There is no algorithm to solve this clinical dilemma in post-op heart surgery patients, but continuous heart rhythm data from patients in atrial fibrillation using CNN, random forest, and LASSO regression has identified a “signature” pattern of atrial fibrillation that is predictive of stroke with a c-statistic of 0.702.[11] Use of this rhythm monitoring method in post-op heart surgery patients with atrial fibrillation could yield a reduction in stroke by recommending anticoagulation in patients with the “stroke signature” AF pattern.

The combination of preoperative features with perioperative variables to better predict stroke after heart surgery is a disruptive use of an existing database to improve outcomes. While the STS ACSD and its accompanying risk models are intended to be used as a method to compare surgical results from different hospitals, the data can also be used to prospectively guide care in patients who are predicted to have a higher risk of complications. The better risk score also makes it a more valid and fair measurement of hospital performance in our new era of value-based healthcare payments. Finally, it is a good project for hospitals that do open heart surgery and want to develop a machine and deep learning capabilities to improve care.

Conclusion

STS cardiac surgery data was obtained from the Michigan chapter of the Society to develop a better model for stroke prediction during heart surgery by using machine learning techniques. De-identified patient heart surgery records were obtained for all risk modeled operations done between July 1, 2011 to December 31, 2016. A total of 42,746 records were obtained with the data being cleaned and mapped to recent version of the database specifications. The records were divided into random training and testing cohorts with use of all preoperative features in a variety of classification algorithms that included logistic regression, decision trees, and neural nets. After tuning the hyperparameters and trying various feature transformations, logistic regression with LASSO and random forest classifier were found to exhibit the highest c-statistic. These c-statistics were not as good as those of the STS risk model so we decided to use POSTOP variables as well. These features were either added to the PREOP dataset for modeling or combined with the PREOP dataset in a stacked process by adding the probability vectors as additional features to the POSTOP dataset. Combined use of the PREOP and POSTOP datasets significantly improved the performance of our models and suggested a rationale for their use to improve perioperative care to reduce strokes after heart surgery.

Table 3: Risk factors for cerebral vascular accident (stroke)

CABG	Valve	CABG + Valve
ADP Usage/Timing of Stop	ADP Usage/Timing of Stop	Age
Age	Age	Aortic insufficiency
Alcohol consumption/week	Alcohol consumption/week	Arrhythmia and type
Aortic insufficiency	Aortic insufficiency	BSA
Arrhythmia and type	Aortic stenosis	Cardiac pressure
BSA	Arrhythmia and type	Chronic Lung Disease
Cancer	BMI	CVD and CVA
Chronic Lung Disease	BSA	CVD stenosis
CVD and CVA	Chronic Lung Disease	Diabetes and control method
CVD stenosis	CVD and CVA	Endocarditis
Diabetes and control method	Diabetes and control method	CHF class and timing
Ejection fraction	Endocarditis	Hematocrit
Family history of CAD	CHF class and timing	# diseased ventricles
Glycoprotein within 24 hours	Hematocrit	Payor
Heart failure class and timing	Illicit drug use	Platelet count
Hypertension	Immunosuppression therapy	PVD
Inotrope	Mitral insufficiency	Dialysis/last creatinine
MI history and timing	# previous cardiac surgeries	Status
Number of diseased vessels	Payor	Unresponsive neuro status
Payor	PCI and timing	
Platelet count	Platelet count	
Previous carotid surgery	Preop IABP	
Previous ICD	Prev other cardiac interventions	
Previous tricuspid valve op	Prev AO Valve procedures	
Proximal LAD	Previous CAB	
PVD	Race and ethnicity	
Race and ethnicity	Recent smokers	

Dialysis/last creatinine	Dialysis/last creatinine
Gender	Gender
Sleep apnea	Shock/ECMO/Impella
Surg status	Status
Steroid before surg	Steroid
Unresponsive neuro status	Syncope
White blood cell count	Time trend
	Tricuspid insufficiency
	Unresponsive neuro status
	White blood cell count

Table 4: PREOP Feature Importance – Ranked by SelectKBest p-value (Top 40)

indicates features with statistically significant test statistics: chi-square and F-value for categorical and numerical features, respectively
indicates features with non-zero coeffcients in LogisticRegression with Lasso model
STS=1, if the Feature is also a feature in the STS Model, else STS=0

				SelectKBest		LR with Lasso	
	Feature	feature_type	STS	kbest_statistic	pvalue	coef	abs_coef
1	age	numerical	1	66.29	4.03E-16	0.3104	0.3104
2	cva	categorical	1	64.07	1.20E-15	0.5224	0.5224
3	cvd	categorical	1	55.95	7.45E-14	0.1777	0.1777
4	heightcm	numerical	1	39.24	3.79E-10	(0.0768)	0.0768
5	classnyh_ANY_ACTIVITY	categorical	1	38.60	5.20E-10	0.3070	0.3070
6	hct	numerical	0	37.05	1.17E-09	0.0000	0.0000
7	totalbumin	numerical	0	35.97	2.02E-09	(0.1509)	0.1509
8	cvdstenlft_100%	categorical	0	33.53	7.03E-09	1.0960	1.0960
9	incidencREOP_FOURTH	categorical	1	30.20	3.89E-08	0.0000	0.0000
10	status_EMERGENCY	categorical	1	25.01	5.70E-07	0.6668	0.6668
11	cvdstenrt_100%	categorical	0	24.81	6.33E-07	0.4623	0.4623
12	weightkg	numerical	1	23.01	1.62E-06	(0.0735)	0.0735
13	pvd	categorical	1	22.75	1.84E-06	0.2441	0.2441
14	arrhyafib	categorical	1	21.53	3.49E-06	0.3610	0.3610
15	cvdtia	categorical	1	21.29	3.94E-06	0.2712	0.2712
16	chf	categorical	0	19.74	8.88E-06	0.0357	0.0357
17	cvdstenrt_80-99%	categorical	0	18.80	1.45E-05	0.1913	0.1913
18	pasys	numerical	0	16.83	4.10E-05	0.0664	0.0664
19	cvdcarsten_RIGHT	categorical	0	14.51	1.39E-04	0.2865	0.2865
20	medinotr	categorical	0	13.59	2.28E-04	0.0679	0.0679
21	prvalve	categorical	1	12.33	4.47E-04	0.4866	0.4866
22	vdaort	categorical	1	12.08	5.10E-04	0.2694	0.2694
23	hdef	numerical	1	11.84	5.81E-04	(0.1007)	0.1007
24	gender	categorical	1	11.82	5.85E-04	(0.3496)	0.3496
25	vdinsuft_MODERATE	categorical	0	11.14	8.47E-04	0.0000	0.0000
26	carshock24	categorical	1	10.97	9.24E-04	0.0000	0.0000
27	cvdcarsten_LEFT	categorical	0	9.47	2.09E-03	0.2234	0.2234
28	meldscr	numerical	0	9.11	2.55E-03	0.0133	0.0133
29	numdisv_2_CORONARIES	categorical	1	8.35	3.86E-03	(0.4040)	0.4040
30	vdinsufm_MODERATE	categorical	1	7.39	6.57E-03	0.0000	0.0000
31	numdisv_1_CORONARY	categorical	1	6.18	1.29E-02	(0.5896)	0.5896
32	cvdpcarsurg	categorical	0	6.05	1.39E-02	(0.1449)	0.1449
33	infendo	categorical	0	5.86	1.55E-02	0.0255	0.0255
34	cvdstenlft_80-99%	categorical	0	5.62	1.78E-02	0.0779	0.0779
35	a1clvl	numerical	0	5.60	1.79E-02	0.0572	0.0572
36	diabetes	categorical	1	5.36	2.06E-02	0.1256	0.1256
37	infendty	categorical	1	5.30	2.14E-02	0.0000	0.0000
38	vdinsufm_MILD	categorical	1	5.09	2.41E-02	0.0873	0.0873
39	numdisv_3_CORONARIES	categorical	1	4.81	2.84E-02	0.0136	0.0136
40	vdinsuft_MILD	categorical	0	4.51	3.36E-02	0.0000	0.0000

Table 5: PREOP Feature Importance – Ranked by Random Forest Feature Importance Score

indicates features with statistically significant test statistics: chi-square and F-value for categorical and numerical features, respectively
the higher the **feature_importance** value, the more important the feature
STS=1, if the Feature is also a feature in the STS Model, else STS=0

				SelectKBest		RandomForest
	Feature	feature_type	STS	kbest_statistic	pvalue	feature_importance
1	age	numerical	1	66.29	4.03E-16	0.0719
2	weightkg	numerical	1	23.01	1.62E-06	0.0600
3	heightcm	numerical	1	39.24	3.79E-10	0.0588
4	hct	numerical	0	37.05	1.17E-09	0.0585
5	bmi	numerical	0	0.74	3.91E-01	0.0581
6	a1clvl	numerical	0	5.60	1.79E-02	0.0488
7	hdef	numerical	0	11.84	5.81E-04	0.0479
8	pasys	numerical	0	16.83	4.10E-05	0.0471
9	totalbumin	numerical	0	35.97	2.02E-09	0.0454
10	meldscr	numerical	0	9.11	2.55E-03	0.0430
11	creatlst	numerical	1	1.97	1.61E-01	0.0407
12	surgdt_month	categorical	0	0.44	5.06E-01	0.0369
13	surgdt_DayOfWeek	categorical	0	9.16	2.48E-03	0.0297
14	mediastrad	categorical	0	3.96	4.65E-02	0.0209
15	ethnicity	categorical	0	0.34	5.61E-01	0.0199
16	raceasian	categorical	0	3.02	8.24E-02	0.0173
17	infendty	categorical	1	5.30	2.14E-02	0.0171
18	chrlungd	categorical	0	3.90	4.83E-02	0.0169
19	vdinsufm	categorical	1	24.21	8.65E-07	0.0146
20	hdefd	categorical	0	0.01	9.29E-01	0.0137
21	racenativeam	categorical	0	0.13	7.20E-01	0.0123
22	anginalclass	categorical	1	0.36	5.51E-01	0.0118
23	hypertn	categorical	1	1.35	2.45E-01	0.0118
24	cvdstenlft	categorical	0	54.96	1.23E-13	0.0104
25	priorhf	categorical	1	3.36	6.70E-02	0.0094
26	prcab	categorical	1	0.24	6.24E-01	0.0093
27	resusc24	categorical	1	1.43	2.32E-01	0.0091
28	chf	categorical	1	19.74	8.88E-06	0.0081
29	gender	categorical	1	11.82	5.85E-04	0.0076
30	ivdrugab	categorical	0	1.81	1.79E-01	0.0073
31	alcohol	categorical	0	1.54	2.15E-01	0.0072
32	incidencREOP	categorical	1	6.30	1.21E-02	0.0071
33	cvdcarsten	categorical	0	36.59	1.46E-09	0.0069
34	syncope	categorical	0	0.37	5.44E-01	0.0069
35	medinotr	categorical	0	13.59	2.28E-04	0.0067
36	prcvint	categorical	0	2.86	9.10E-02	0.0065
37	prvalve	categorical	1	12.33	4.47E-04	0.0063
38	vdstenm	categorical	0	0.65	4.20E-01	0.0060
39	cva	categorical	1	64.07	1.20E-15	0.0060
40	raceblack	categorical	0	2.88	8.94E-02	0.0058

Table 6: POSTOP Feature Importance – Ranked by SelectKBest p-value (Top 40)

indicates features with statistically significant test statistics: chi-square and F-value for categorical and numerical features, respectively
indicates features with non-zero coeffcients in LogisticRegression with Lasso model

			SelectKBest		LR with Lasso	
	Feature	feature_type	kbest_statistic	pvalue	coef	abs_coef
1	STS_predstro	numerical	310.71	3.10E-69	0.0000	0.0000
2	LR_PREOP_PROBA	numerical	244.45	6.55E-55	1.9920	1.9920
3	RF_PREOP_PROBA	numerical	227.53	3.01E-51	3.6658	3.6658
4	cperftyp_Retrograde	categorical	115.50	6.12E-27	0.0000	0.0000
5	postcreat	numerical	103.10	3.45E-24	0.1699	0.1699
6	lwsthct	numerical	63.43	1.71E-15	(0.0445)	0.0445
7	cotarrst	categorical	61.15	5.30E-15	0.3919	0.3919
8	ecmo	categorical	58.18	2.39E-14	0.4913	0.4913
9	ecmoind_Resp_Failure	categorical	53.82	2.20E-13	0.7313	0.7313
10	ibldprod	categorical	51.59	6.83E-13	0.0264	0.0264
11	ecmowhen_Post_OP	categorical	50.03	1.52E-12	0.4189	0.4189
12	perfustm	numerical	47.20	6.53E-12	0.3110	0.3110
13	IABP	categorical	41.90	9.60E-11	0.0000	0.0000
14	iabpind_Hemodyn_Instab	categorical	40.62	1.85E-10	0.4149	0.4149
15	ecmoind_Cadiac_Failure	categorical	35.68	2.33E-09	0.0000	0.0000
16	circarr	categorical	26.13	3.19E-07	0.0236	0.0236
17	ibdcryou	numerical	24.99	5.79E-07	0.0372	0.0372
18	iabpwhen_Intra_OP	categorical	21.70	3.20E-06	0.0306	0.0306
19	ibdffpu	numerical	20.96	4.71E-06	(0.0452)	0.0452
20	iabpind_Procedure_Support	categorical	19.40	1.06E-05	1.0920	1.0920
21	cotafib	categorical	19.34	1.09E-05	0.2537	0.2537
22	iabpwhen_Pre_OP	categorical	18.67	1.56E-05	0.0000	0.0000
23	urgntrsn_Infect_Dev	categorical	17.85	2.38E-05	0.3110	0.3110
24	cperfutil	categorical	17.85	2.38E-05	0.0000	0.0000
25	ecmowhen_Intra_OP	categorical	17.68	2.61E-05	0.0000	0.0000
26	cumulsatlft	numerical	17.05	3.66E-05	0.0473	0.0473
27	xclamptm	numerical	16.54	4.78E-05	(0.2023)	0.2023
28	lwsttemp	numerical	15.11	1.02E-04	(0.0424)	0.0424
29	opvalve	categorical	14.29	1.57E-04	0.2253	0.2253
30	unplproc_Yes_Complication	categorical	13.01	3.10E-04	0.1882	0.1882
31	cumulsatrt	numerical	11.85	5.76E-04	0.0029	0.0029
32	asmtaodx_atheroma>5mm	categorical	11.57	6.69E-04	0.4903	0.4903
33	ibdplatu	numerical	10.25	1.37E-03	(0.0368)	0.0368
34	emergrsn_Shock_Circ_Supp	categorical	9.91	1.65E-03	0.0357	0.0357
35	urgntrsn_AMI	categorical	9.21	2.41E-03	0.4064	0.4064
36	vsmv	categorical	9.04	2.64E-03	0.0097	0.0097
37	cathbasassistwhen_Pre_OP	categorical	8.84	2.94E-03	0.0207	0.0207
38	ibdrbcu	numerical	8.55	3.45E-03	(0.0722)	0.0722
39	iabpwhen_Post_OP	categorical	8.35	3.86E-03	(0.5865)	0.5865
40	mt30stat	categorical	7.09	7.74E-03	(1.2115)	1.2115

Table 7: POSTOP Feature Importance – Ranked by Random Forest Feature Importance Score

indicates features with statistically significant test statistics: chi-square and F-value for categorical and numerical features
the higher the **feature_importance** value, the more important the feature

			SelectKBest		RandomForest
	Feature	feature_type	kbest_statistic	pvalue	feature_importance
1	canartstoth	categorical	0.01	9.17E-01	0.1428
2	cperftime	numerical	0.85	3.56E-01	0.1349
3	ocarvsd	categorical	0.02	8.95E-01	0.1193
4	iabpwhen	categorical	74.62	5.71E-18	0.0894
5	cofirstind	categorical	0.23	6.30E-01	0.0793
6	ecmowhen	categorical	170.29	6.38E-39	0.0518
7	IABP	categorical	41.90	9.60E-11	0.0470
8	cotarrst	categorical	61.15	5.30E-15	0.0396
9	asmtaodx	categorical	5.25	2.20E-02	0.0352
10	perfustm	numerical	47.20	6.53E-12	0.0316
11	opcab	categorical	0.00	9.94E-01	0.0294
12	unplao	categorical	6.31	1.20E-02	0.0261
13	concalc	categorical	4.55	3.30E-02	0.0245
14	ibdrbcu	numerical	8.55	3.45E-03	0.0167
15	dhcatm	numerical	1.11	2.92E-01	0.0164
16	imedtran	categorical	0.32	5.69E-01	0.0147
17	ibldprod	categorical	51.59	6.83E-13	0.0141
18	lwsthct	numerical	63.43	1.71E-15	0.0136
19	opvalve	categorical	14.29	1.57E-04	0.0095
20	vsmv	categorical	9.04	2.64E-03	0.0090
21	RF_PREOP_PROBA	numerical	227.53	3.01E-51	0.0078
22	STS_predstro	numerical	310.71	3.10E-69	0.0056
23	iabpind	categorical	44.17	3.02E-11	0.0051
24	lwsttemp	numerical	15.11	1.02E-04	0.0048
25	canartstaort	categorical	0.14	7.07E-01	0.0041
26	cathbasassistind	categorical	0.01	9.17E-01	0.0038
27	unplvad	categorical	0.17	6.77E-01	0.0036
28	cumulsatlft	numerical	17.05	3.66E-05	0.0033
29	unplav	categorical	0.14	7.10E-01	0.0021
30	oponcard	categorical	0.42	5.19E-01	0.0021
31	cperftyp	categorical	55.87	7.73E-14	0.0018
32	unplproc	categorical	30.38	3.55E-08	0.0017
33	vsavpr	categorical	6.85	8.85E-03	0.0016
34	unplmv	categorical	0.71	3.98E-01	0.0016
35	aortoccl	categorical	0.08	7.83E-01	0.0016
36	mt30stat	categorical	7.09	7.74E-03	0.0009
37	circarr	categorical	26.13	3.19E-07	0.0008
38	cumulsatrt	numerical	11.85	5.76E-04	0.0007
39	opocard	categorical	2.70	1.00E-01	0.0005
40	ocarasd	categorical	5.68	1.71E-02	0.0005

References

1. Shahian, DM. Professional society leadership in health care quality: the Society of Thoracic Surgeons experience. Joint Commission J on Quality and Patient Safety 2019; 45:466-479
2. Clark RE. The STS Cardiac Surgery National Database: an update. Ann Thorac Surg. 1995: 59:1376-1380
3. Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery risk models: part 1 -- background, design considerations, and model development. Ann Thorac Surg 2018: 105: 1411-18
4. O'Brien SM, Feng L, Xia H, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: part 2 -- statistical methods and results. Ann Thorac Surg 2018; 105:1419-28
5. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics. 2016; 3:119-131
6. Lippman RP, Shahian DM, Coronary artery bypass risk prediction using neural networks. Ann Thorac Surg. 1997; 63:1635-43
7. Karim N, Epi MC, Reid CM, et al. Variable selection methods for multiple regressions influence the parsimony of risk prediction models for cardiac surgery. J Thorac Cardiovasc Surg. 2017; 153:1128-35
8. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of Clinical Epidemiology. 2019; 110: 12-22
9. Howitt SH, Grant SW, Riding DM et al. Risk models that use postoperative patient monitoring data to predict outcomes in adult cardiac surgery: a systematic review. Journal of Cardiothoracic and Vascular Anesthesia. 2017; 31: 1865-1877
10. Holmgren G, Andersson P, Jakobsson A and Frigyesi A. Artificial neural networks improve and simplify intensive care mortality prognostication: a national cohort study of 217,289 first-time intensive care unit admissions. Journal of Intensive Care. 2019; 7:44

11. Han L, Askari M, Altman RB, et al. Atrial fibrillation burden signature and near-term prediction of stroke. Circ Cardiovasc Qual Outcomes. 2019:12:e005595