2772-9931
https://doi.org/10.1016/j.atssr.2024.02.005

Coronary: Short Report

Intraoperative Features Improve Model Risk Predictions After Coronary Artery Bypass Grafting

Willa Potosnak, BS,[1] Keith A. Dufendach, MD,[1,2] Chirag Nagpal, PhD,[1]
David J. Kaczorowski, MD,[2] Pyongsoo Yoon, MD,[2] Johannes Bonatti, MD,[2]
James K. Miller, PhD,[1] and Artur W. Dubrawski, PhD[1]

ABSTRACT

BACKGROUND Intraoperative physiologic parameters could offer predictive utility in evaluating risk of adverse postoperative events yet are not included in current standard risk models. This study examined whether the inclusion of continuous intraoperative data improved machine learning model predictions for multiple outcomes after coronary artery bypass grafting, including 30-day mortality, renal failure, reoperation, prolonged ventilation, and combined morbidity and mortality (MM).

METHODS The Society of Thoracic Surgeons (STS) database features and risk scores were combined with retrospectively gathered continuous intraoperative data from patients. Risk models were developed for each outcome by training a logistic regression classifier on intraoperative data using 5-fold cross-validation. STS risk scores were included as offset terms in the models.

RESULTS Compared with the STS Risk Calculator, models developed using a combination of the intraoperative features and the STS preoperative risk score had improved mean area under the receiver operating characteristic curve for prolonged ventilation (0.750 [95% CI, 0.690-0.809] vs 0.800 [95% CI, 0.750-0.851]) and MM (0.695 [95% CI, 0.644-0.746] vs 0.724 [95% CI, 0.673-0.775]). Additionally, models developed using intraoperative features had improved calibration, measured with Brier score, for prolonged ventilation (0.060 [95% CI, 0.050-0.070] vs 0.055 [95% CI, 0.045-0.065]) and MM (0.092 [95% CI, 0.081-0.103] vs 0.087 [95% CI, 0.075-0.098]).

CONCLUSIONS The inclusion of time series intraoperative data in risk models may improve early postoperative care by identifying patients who require closer monitoring postoperatively.

(Ann Thorac Surg Short Reports 2024;2:336-340)

Clinical risk adjudication before cardiac surgery is routinely carried out for appropriate treatment decisions to mitigate adverse postoperative outcomes. Machine learning (ML) has proven useful in developing models that identify patients at risk of adverse outcomes after coronary artery bypass grafting (CABG).[1] Previous work demonstrated that the inclusion of continuous intraoperative parameters derived from time series data can further improve

IN SHORT

- Inclusion of interoperative data with preoperative data in machine learning risk models improves risk predictions for prolonged ventilation and combined morbidity and mortality.
- These machine learning models may improve postoperative care by identifying those patients at highest risk for complications.

Accepted for publication Feb 26, 2024.

[1]Auton Lab, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania; and [2]Division of Cardiac Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Address correspondence to Dr Dufendach, Division of Cardiac Surgery, University of Pittsburgh Medical Center, 200 Lothrop St, Pittsburgh, PA 15213; email: dufendachka@upmc.edu.

renal failure risk predictions,[2,3] a finding suggesting that time series intraoperative data may be used to improve predictions for other postoperative outcomes, such as mortality and prolonged ventilation. However, research that develops models for other outcomes using time series data is scarce in the current literature. Furthermore, the current literature that develops models using continuous intraoperative parameters does not directly compare performance against established baseline models used in practice, such as The Society of Thoracic Surgeons (STS) Risk Calculator.[1] As such, current understanding of the potential additive utility of intraoperative models in clinical practice is unclear. The aim of this study was to determine whether ML models developed using continuous parameters derived from time series intraoperative data offer improved risk predictions over the baseline STS Risk Calculator for multiple outcomes after CABG.

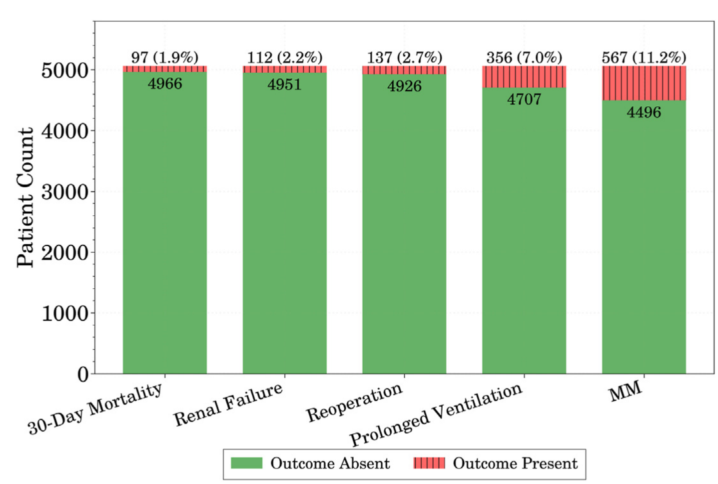

FIGURE 1 Count (percentage) of patients who experienced adverse outcomes after coronary artery bypass grafting. (MM, combined morbidity and mortality)

MATERIAL AND METHODS

Data were retrospectively gathered from a large tertiary care center in the United States and included preoperative patient characteristics and risk scores generated from the STS Risk Calculator. Additionally, static and continuous time series intraoperative data from the anesthesia record were gathered for 5623 patients who underwent isolated on-pump CABG between 2011 and 2021. A total of 328 patients were excluded because of incomplete intraoperative data, which had an insignificant impact on the incidence rates of adverse outcomes in the final cohort (*P* value $>$ .05). A total of 34.6% of patients underwent elective CABG (n = 1753). At least 1 internal mammary artery graft was used in 96.3% of patients (unknown in 1.1%; n = 58). The mean number of vein graft targets was 2.2 (range, 0-6). For additional cohort data, see Supplemental Table 1. Static intraoperative data included blood product use, perfusion duration, and cross-clamp duration, among others (Supplemental Table 2). Physiologic intraoperative data included hemodynamic parameters such as blood pressure, heart rate, and ventilator settings such as fraction of inspired oxygen content or tidal volume (Supplemental Table 3). All medications and fluid administration were also reported. Data preprocessing was performed on the basis of on-pump CABG procedure surgery phases, as outlined in the Supplemental Methods and Supplemental Figure 1.

Risk models were developed for 5 postoperative outcomes using a binary logistic regression classifier. Individual models were developed for each outcome, including 30-day mortality, renal failure, reoperation, prolonged ventilation, and combined morbidity and mortality (MM) defined by STS database criteria.[1] Outcome incidence rates are shown in Figure 1. Separate ML models were trained using only intraoperative data or intraoperative data with the STS risk score as an offset term using the GLMNET R software package, as outlined in the Supplemental Methods.[4,5] Importantly, when included as an offset term in the model, the STS risk score bounds model performance to be either equal or superior to that of the STS Risk Calculator. Model performance was measured using mean area under the receiver operating characteristic curve (AUC-ROC) and Brier score computed from bootstrap samples (n = 500).

Institutional Review Board approval from the University of Pittsburgh Medical Center was obtained for this research. Patient consent statements were not required for this research. All patient data were anonymized to remove personally identifiable information.

RESULTS

Compared with the STS Risk Calculator, models developed using a combination of the intraoperative features and the STS preoperative risk score had improved mean ROC-AUC for prolonged ventilation (0.750 vs 0.800) and MM (0.695 vs 0.724), as shown in Figures 2A and 2B and the Table. Additionally, calibration curves for the intraoperative model with the STS risk score shown in Figures 2C and 2D have a slope of approximately 1.00, which denotes improved calibration over the STS Risk Calculator. Improved calibration is also demonstrated by Brier scores for prolonged ventilation (0.060 vs 0.055) and MM (0.092 vs 0.087), as shown in the Table. For the 30-day mortality and renal failure outcomes, the intraoperative

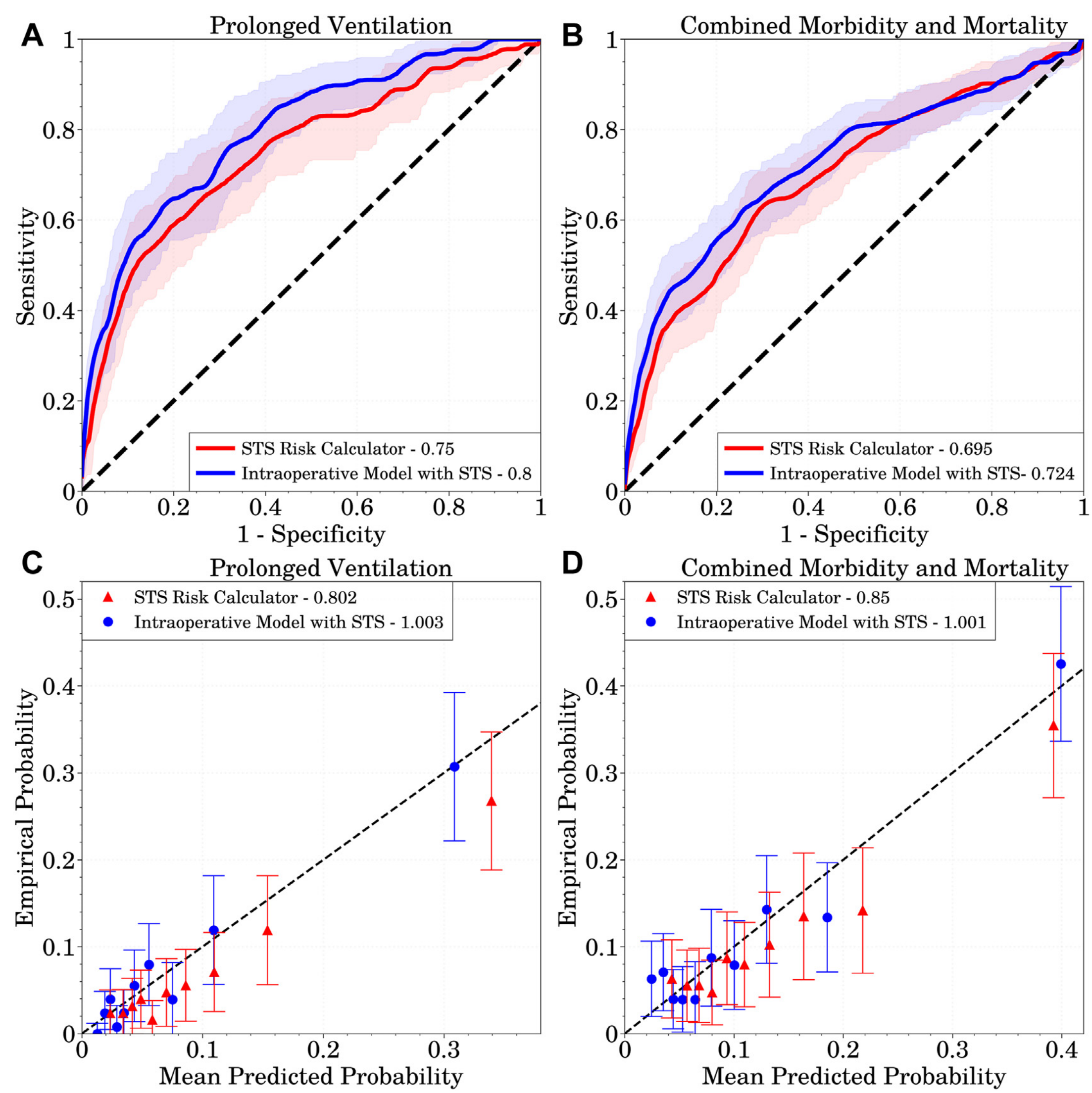

FIGURE 2 Model performance was measured using the area under the receiver operating characteristic curve. The inclusion of intraoperative data improved the area under the receiver operating characteristic curve for (A) prolonged ventilation and (B) combined morbidity and mortality outcomes. Model calibration, or similarity of risk scores to empirical probabilities (observed event rate), was assessed using calibration curves. The intraoperative model with The Society of Thoracic Surgeons (STS) risk score demonstrated improved model calibration compared with the STS Risk Calculator for (C) prolonged ventilation and (D) combined morbidity and mortality models, particularly for higher-risk patients, as shown by the calculated slope in the figure.

models with the STS risk score offset had performance similar to that of the STS Risk Calculator. Notably, models trained using solely intraoperative data without the STS risk score generally had worse performance on the basis of the ROC-AUC, with the exception of the reoperation outcome, which had an improved ROC-AUC over the STS Risk Calculator, yet at a higher Brier score (Table).

We assessed the ability of the intraoperative model with the STS risk score offset to reclassify patients into low-, intermediate-, and high-risk groups on the basis of terciles computed from STS risk scores.[6] With the intraoperative model, more patients were downgraded to lower-risk groups compared with the STS model, as shown in Figure 3. Furthermore, Figure 3 shows that the intraoperative model with the STS risk score offset can more effectively classify patients into risk groups than the STS Risk Calculator, as evidenced by the outcome rates of each risk strata. The observed outcome rates for the high- and intermediate-risk groups of the intraoperative model are higher than that of the STS Risk Calculator. Additionally, the observed outcome rate for

TABLE Area Under the Receiver Operating Characteristic Curve and Brier Score (95% CI) for The STS Risk Calculator, Intraoperative Data Models, and Intraoperative Data Models With The STS Risk Score Offset

		Outcomes, Mean (95% CI)				
Metric	Data Set	30-Day Mortality	Renal Failure	Reoperation	Prolonged Ventilation	MM
ROC-AUC	STS	0.783 (0.687-0.879)	0.767 (0.664-0.871)	0.535 (0.434-0.636)	0.750 (0.690-0.809)	0.695 (0.644-0.746)
	Intraoperative	0.662 (0.547-0.777)	0.644 (0.5310.757)	0.676 (0.575-0.776)	0.778 (0.726-0.831)	0.615 (0.562-0.667)
	Intraoperative with STS offset	0.792 (0.7-0.883)	0.769 (0.668-0.87)	0.535 (0.434-0.636)	0.800 (0.750-0.851)	0.724 (0.673-0.775)
Brier score	STS	0.018 (0.010-0.025)	0.021 (0.014-0.027)	0.028 (0.02-0.037)	0.060 (0.050-0.070)	0.092 (0.081-0.103)
	Intraoperative	0.188 (0.182-0.194)	0.136 (0.131-0.141)	0.158 (0.155-0.160)	0.148 (0.142-0.153)	0.194 (0.191-0.197)
	Intraoperative with STS offset	0.018 (0.011-0.025)	0.020 (0.014-0.027)	0.028 (0.019-0.036)	0.055 (0.045-0.065)	0.087 (0.075-0.098)

MM, major morbidity and mortality; ROC-AUC, area under the receiver operating characteristic curve; STS, The Society of Thoracic Surgeons.

the low-risk group of the intraoperative model is lower than that of the STS Risk Calculator. Statistics of important model features computed for risk groups are shown in Supplemental Tables 4 and 5.

Intraoperative features that contribute to model predictions for prolonged ventilation and MM are shown in Supplemental Figures 2 and 3. Important derived features for prolonged ventilation and MM on the basis of model weights include difference in mean heart rate values between prebypass and postbypass (0.026 and 0.023) status, mean postbypass arterial oxygen saturation (Spo_2) (−0.041 and −0.048), and epinephrine volume administered during the preincision phase (0.013 and 0.007), among others.

COMMENT

Risk predictions from intraoperative data models with the STS risk score offset outperform the STS Risk Calculator for prolonged ventilation and MM. Model performance for these outcomes demonstrates that careful inclusion of time series intraoperative data with preoperative patient characteristics improves model risk predictions. Given that ML models are regularly used in clinical practice before surgery, use of a supplemental model with combined preoperative and intraoperative data may be feasibly incorporated to identify high-risk patients and mitigate adverse outcomes. The STS models developed for isolated CABG surgery consider

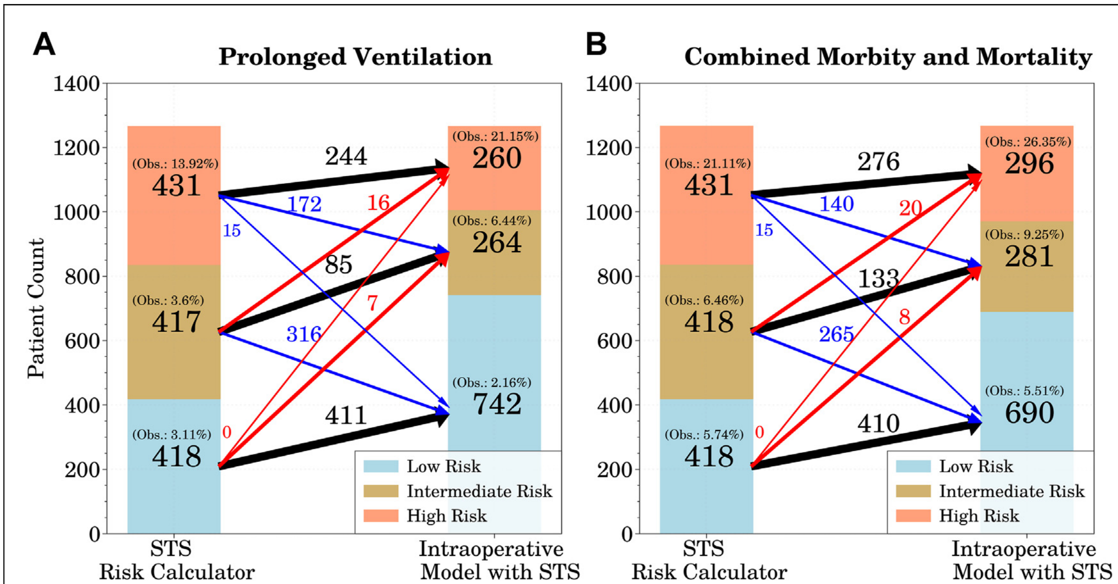

FIGURE 3 The intraoperative model with The Society of Thoracic Surgeons (STS) risk score offset reclassified patients into low-risk (prolonged ventilation: ≤0.047; combined morbidity and mortality: ≤0.078), intermediate-risk (prolonged ventilation: >0.047 and ≤0.088; combined morbidity and mortality: >0.078 and ≤ 0.135), and high-risk (prolonged ventilation: >0.088; combined morbidity and mortality: >0.135) categories on the basis of terciles computed from STS risk score. For (A) prolonged ventilation and (B) combined morbidity and mortality outcomes, the intraoperative model generally reclassified patients into lower-risk categories, with exception of several patients who were reclassified from lower-risk categories to higher-risk categories. (Obs., observed; STS, The Society of Thoracic Surgeons.)

only 4 intraoperative parameters: intraaortic balloon pump use, catheter-based assist device use, extracorporeal membrane oxygenation use, and need for resuscitation. Other intraoperative features include medication administration and physiologic measurements such as heart rate and blood pressure. These features may also serve as physiologic targets for intraoperative management, thereby allowing for reduced postoperative risk.

Only 4 studies have developed preoperative and intraoperative data models and directly compared model performance to measure the relative utility of intraoperative data in predicting outcomes.[2,7-9] Other than 1 study that developed a preoperative model on the basis of the STS Risk Calculator,[9] none of the 4 studies directly compared performance with the STS Risk Calculator. Additionally, only 2 studies considered intraoperative time series data.[2,3] One study, by Tseng and colleagues,[3] used a combination of preoperative and time series intraoperative data to predict postoperative acute kidney injury. Time series features included in modeling were limited to hemodynamic parameters such as heart rate and blood pressure measurements but did not include ventilator parameters or medication administration. Another study, by Adhikari and colleagues,[2] included a larger subset of time series hemodynamic features, as well as two binary operative medication features. Both studies focused solely on predicting postoperative renal failure, with other postoperative outcomes not considered. Our study demonstrates that time series intraoperative data also improve predictions for other postoperative outcomes.

Limitations to this study include that risk models were developed and assessed using data with a relatively small number of patients (on-pump CABG procedure) compared with the data used to develop the STS Risk Calculator. Additionally, there is a potential for provider bias, including medication administration and hemodynamic targets, which is a challenge to identify and remove in ML models. A standardized intraoperative data collection strategy is needed to reduce missing data and increase reliability of physiologic targets discovered through this work.

The Supplemental Material can be viewed in the online version of this article [https://doi.org/10.1016/j.atssr.2024.02.005] on http://www.annalsthoracicsurgery.org.

FUNDING SOURCES

This research was partially funded with gifts through the Center for Machine Learning and Health at Carnegie Mellon University and UPMC Enterprises. This research was supported by National Institutes of Health grant R01EB029751. The authors have no other funding sources to disclose.

DISCLOSURES

The authors have no conflicts of interest to disclose.

REFERENCES

1. O'Brien SM, Feng L, He H, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: part 2-statistical methods and results. *Ann Thorac Surg*. 2018;105:1419-1428.

2. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, et al. Improved predictive models for acute kidney injury with idea: intraoperative data embedded analytics. *PloS One*. 2009;14:e0214904.

3. Tseng P, Chen Y, Wang C, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020;24:478.

4. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;1:1-22.

5. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.

6. Kumar A, Sato K, Narayanswami J, et al. Current Society of Thoracic Surgeons model reclassifies mortality risk in patients undergoing transcatheter aortic valve replacement. *Circulation*. 2018;11:e006664.

7. Mori M, Durant TJS, Huang C, et al. Toward dynamic risk prediction of outcomes after coronary artery bypass graft: improving risk prediction with intraoperative events using gradient boosting. *Circ Cardiovasc Qual Outcomes*. 2021;14:e007363.

8. Kern H, Redlich U, Hotz H, et al. Risk factors for prolonged ventilation after cardiac surgery using APACHE II, SAPS II, and TISS: comparison of three different models. *Intensive Care Med*. 2001;27:407-415.

9. Saleh HZ, Shaw M, Al-Rawi O, et al. Outcomes and predictors of prolonged ventilation in patients undergoing elective coronary surgery. *Interact Cardiovasc Thorac Surg*. 2012;15:51-56.