JAMA Network Open

Invited Commentary | Anesthesiology

Linking Preoperative and Intraoperative Data for Risk Prediction
More Answers or Just More Data?

Seshadri C. Mudumbai, MD, MS; Parisa Rashidi, PhD

Elsewhere in *JAMA Network Open*, Xue and colleagues[1] share the results of their cohort study involving 111 888 surgeries at a large academic medical center. The research team deployed machine learning (ML) models to predict the risk of postoperative complications related to pneumonia, acute kidney injury, deep vein thrombosis, delirium, and pulmonary embolism. Studies indicate that more than 10% of surgical patients may experience a major postoperative complication (eg, heart attack, infection, and blood clots), with the incidence of postoperative complications varying by type of surgery. The use of ML approaches that take advantage of the rich storehouses of electronic health record (EHR) data and support perioperative clinical decision-making have been talked about for years; however, real-life examples are still few. Therefore, we applaud this group of investigators for pursuing this important and timely topic.

Substantial evidence exists to support the prevention and early management of postoperative complications. The occurrence of a 30-day postoperative complication is more important than preoperative risk and intraoperative factors in determining survival after major surgery. For example, within 30 days of major surgery, between 14% and 30% of patients who develop a postoperative pulmonary complication (eg, pneumonia) will die vs 0.2% to 0.3% without the complication.[2] However, the causes of a postoperative complication are multifactorial and relate not only to the patient's health conditions but also to potential risks introduced by the surgery itself and anesthetic management. These complex relationships have long been recognized, and in an effort to improve the quality of surgical care, various metrics and risk assessment tools have been developed.

While the original intent of the American Society of Anesthesiologists physical status score was to stratify severity of illness prior to surgery, more recently the physical status score has been applied as a simple means to estimate outcomes. The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) surgical risk calculator has a web-based interface and primarily uses markers of preoperative status (eg, age, comorbidities) and type of surgery.[3] Underlying both tools is a reliance on large, national administrative data sets or registries for their validation. In the case of ACS NSQIP, clinicians are needed to assemble data on predictors and outcomes from medical chart reviews. Given the growing, ready availability of EHR data, physicians and hospitals are increasingly searching for ways to automate risk stratification within clinical workflows and provide institution-specific risk assessments for their unique set of patients.

Intraoperative data typically capture key aspects of the procedure, including duration, blood loss, and surgeon, but also elements of anesthetic management, such as hemodynamics, administered medications, and ventilator settings. Of particular interest is how intraoperative management decisions and interventions can worsen postoperative complications in the context of preexisting diseases. Using preoperative and intraoperative data together with an ensemble of ML algorithms, the authors[1] noted highest area under the receiver operating characteristic curve (AUROC; a measure of predictive accuracy, with 1.000 indicating theoretically perfect accuracy) for predicting the risk of postoperative complication associated with pneumonia (0.905), acute kidney injury (0.848), deep venous thrombosis (0.881), pulmonary embolism (0.831), and delirium (0.762). These results help to address an ongoing question within the surgical literature: whether the predictive performance of ML algorithms is enhanced with both intraoperative and preoperative data. The authors[1] identified a series of predictors, including hematocrit and intraoperative mean

+ Related article

Author affiliations and article information are listed at the end of this article.

arterial blood pressure, to aid risk stratification. The causal frameworks tethering predictors and outcomes, and to what degree the predictors are actionable, remain open questions not only for this study but other ML-based clinical studies.[4]

This retrospective study also does not offer practical details on how this predictive model would be deployed as a real-time clinical decision support system to decrease postoperative complications. An additional implementation trial of a decision support system combined with outcomes analysis will be necessary to address these questions. Because hospitals and health care systems are increasingly poised to use their EHR data, these types of pragmatic trials will determine how risk prediction tools can be usefully incorporated into existing informatics and clinical workflows.

Working with a data set developed from a cohort of real-world surgical patients, the authors[1] carefully defined a group of potentially modifiable outcomes. Then, the authors selected 5 ML models to develop their predictions: logistic regression, support vector machine, random forest, gradient boosting tree, and deep neural network. Readers will want to assess the data preprocessing stage of the analytic pipeline, particularly surrounding missing data, including its extent and pattern. Missingness is a common issue in EHR data with a potential impact on ML feature selection, model training, validation, and, importantly, interpretability.[5]

Logistic regression falls under the umbrella of interpretable models. In other words, a human can understand the cause of the decisions. While current risk scoring systems may offer limited predictive performance, their reliance on interpretable models nevertheless allows physicians to examine how the algorithm arrived at its prediction. On the other hand, model-agnostic methods, such as support vector machines, random forests, gradient boosting trees, and deep neural networks, can outperform traditional models like logistic regression. However, these so-called black box approaches do not lend themselves to direct interpretability.[6] As a result, clinicians may be reluctant to accept recommendations when applied to treatment decisions. This study[1] used an approach called Shapley Additive Explanations along with visualizations to help clinicians understand how and which variables were selected to be important by the models. Future studies could benefit from model personalization by including information on surgeons' previous performance in an aggregate and privacy-preserving manner in relation to case mix, as suggested by Bihorac et al.[7] Furthermore, evaluation and mitigation of bias and evaluating fairness should be an integral part of perioperative models to address potential disparities in surgical care. While it may be uncertain how clinicians will accept recommendations arising from ML-based risk models, we can be confident that these types of studies will only increase in the years ahead.

ARTICLE INFORMATION

Published: March 30, 2021. doi:10.1001/jamanetworkopen.2021.2547

Corresponding Author: Seshadri C. Mudumbai, MD, MS, Anesthesiology and Perioperative Care Service, Veterans Affairs Palo Alto Health Care System, 3801 Miranda Ave, 112A, Palo Alto, CA 94304 (mudumbai@stanford.edu).

Author Affiliations: Anesthesiology and Perioperative Care Service, Veterans Affairs Palo Alto Health Care System, Palo Alto, California (Mudumbai); Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, California (Mudumbai); Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System, Menlo Park, California (Mudumbai); Biomedical Engineering, University of Florida, Gainesville (Rashidi); Electrical and Computer Engineering, University of Florida, Gainesville (Rashidi).

Conflict of Interest/Disclosures: Dr Rashidi reported receiving grants from the National Institutes of Health and the National Science Foundation during the conduct of the study and outside the submitted work and having patent 16616534 pending. No other disclosures were reported.

Disclaimer: The views expressed do not necessarily reflect those of the US Department of Veterans Affairs or other institutions.

REFERENCES

1. Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open*. 2021;4(3):e212240. doi:10.1001/jamanetworkopen.2021.2240

2. Miskovic A, Lumb AB. Postoperative pulmonary complications. *Br J Anaesth*. 2017;118(3):317-334. doi:10.1093/bja/aex002

3. ACS NSQIP. Welcome to the ACS NSQIP surgical risk calculator. Accessed January 25, 2021. http://riskcalculator.facs.org/RiskCalculator/

4. Liu Y, Chen PC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. *JAMA*. 2019;322(18):1806-1816. doi:10.1001/jama.2019.16489

5. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes*. 2020;13(10):e006556. doi:10.1161/CIRCOUTCOMES.120.006556

6. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep*. 2019;49(1):15-21. doi:10.1002/hast.973

7. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg*. 2019;269(4):652-662. doi:10.1097/SLA.0000000000002706