POTTER-ICU: An artificial intelligence smartphone-accessible tool to predict the need for intensive care after emergency surgery

Anthony Gebran, MD\textsuperscript{a,b}, Annita Vapsi, BS, MBAn\textsuperscript{c}, Lydia R. Maurer, MD, MPH\textsuperscript{a,b}, Mohamad El Moheb, MD\textsuperscript{a,b}, Leon Naar, MD\textsuperscript{a,b}, Sumiran S. Thakur, BS, MBAn\textsuperscript{a}, Robert Sinyard, MD, MBA\textsuperscript{a}, Dania Daye, MD, PhD\textsuperscript{b,d}, George C. Velmahos, MD, PhD\textsuperscript{a}, Dimitris Bertsimas, PhD\textsuperscript{c}, Haytham M.A. Kaafarani, MD, MPH\textsuperscript{a,b,}\textsuperscript{*}

\textsuperscript{a} Division of Trauma, Emergency Surgery, and Surgical Critical Care, Department of Surgery, Massachusetts General Hospital, Boston, MA
\textsuperscript{b} Center for Outcomes and Patient Safety in Surgery (COMPASS), Massachusetts General Hospital, Boston, MA
\textsuperscript{c} Massachusetts Institute of Technology, Cambridge, MA
\textsuperscript{d} Division of Interventional Radiology, Massachusetts General Hospital, Boston, MA

\textbf{Article info}

\textbf{Article history:}
Accepted 15 March 2022
Available online xxx

\textbf{Abstract}

\textbf{Background:} Delays in admitting high-risk emergency surgery patients to the intensive care unit result in worse outcomes and increased health care costs. We aimed to use interpretable artificial intelligence technology to create a preoperative predictor for postoperative intensive care unit need in emergency surgery patients.

\textbf{Methods:} A novel, interpretable artificial intelligence technology called optimal classification trees was leveraged in an 80:20 train:test split of adult emergency surgery patients in the 2007–2017 American College of Surgeons National Surgical Quality Improvement Program database. Demographics, comorbidities, and laboratory values were used to develop, train, and then validate optimal classification tree algorithms to predict the need for postoperative intensive care unit admission. The latter was defined as postoperative death or the development of 1 or more postoperative complications warranting critical care (eg, unplanned intubation, ventilator requirement \textgeq 48 hours, cardiac arrest requiring cardiopulmonary resuscitation, and septic shock). An interactive and user-friendly application was created. C statistics were used to measure performance.

\textbf{Results:} A total of 464,861 patients were included. The mean age was 55 years, 48\% were male, and 11\% developed severe postoperative complications warranting critical care. The Predictive Optimal Trees in Emergency Surgery Risk Intensive Care Unit application was created as the user-friendly interface of the complex optimal classification tree algorithms. The number of questions (ie, tree depths) needed to predict intensive care unit admission ranged from 2 to 11. The Predictive Optimal Trees in Emergency Surgery Risk Intensive Care Unit application had excellent discrimination for predicting the need for intensive care unit admission (C statistics: 0.89 train, 0.88 test).

\textbf{Conclusion:} We recommend the Predictive Optimal Trees in Emergency Surgery Risk Intensive Care Unit application as an accurate, artificial intelligence–based tool for predicting severe complications warranting intensive care unit admission after emergency surgery. The Predictive Optimal Trees in Emergency Surgery Risk Intensive Care Unit application can prove useful to triage patients to the intensive care unit and to potentially decrease failure to rescue in emergency surgery patients.

© 2022 Elsevier Inc. All rights reserved.

\textbf{Introduction}

The burden of disease requiring emergency surgery (ES) is substantial and continues to rise. Between 2001 and 2010, ES admissions accounted for 7.1\% of all hospitalizations and an annual cost of $28.37 billion.\textsuperscript{1–3} More importantly, when compared to patients undergoing elective surgery, ES patients have a
significantly increased risk of postoperative mortality and serious complications requiring life- and organ-sustaining treatment and thus more often require intensive care unit (ICU) admission. Failure to recognize the high-risk ES patients or delay in their admission to the ICU can result in even higher mortality and morbidity in this vulnerable patient population. From a capacity perspective, admitting a patient unnecessarily to the ICU is also problematic to most hospitals, especially those with scarce resources or capacity challenges. As such, a strong triage tool for ES patients not only can prevent failure to rescue ES patients but can simultaneously help preserve resources.

Several risk prediction systems are available to predict postoperative ICU admission in select patient populations, including the ES patient. Most of these models are inherently limited since they assume that the interaction between variables is linear and additive when, in reality, the presence or absence of a variable influences the significance and impact of other variables. Newer risk prediction tools, particularly those that use artificial intelligence (AI) and, specifically, machine learning (ML) methodologies, have the ability to reboots themselves with each variable, and as such capture the nonlinear complex interactions of different variables. Our group has previously developed and validated an AI-based risk calculator, the Predictive Optimal Trees in Emergency Surgery Risk (POTTER), which leverages a novel ML methodology called optimal classification trees (OCT) to accurately predict the postoperative outcomes of ES patients. POTTER outperforms all existing risk calculators in the field (C statistics for predicting mortality in the emergency general surgery patient is 0.92), is available as a user-friendly smartphone application, and has been downloaded through the android and iPhone platforms for use by thousands of surgeons worldwide.

In this study, we sought to build on our previous experience with POTTER and use this ML-based OCT technology to create and validate a preoperative predictor for postoperative ICU need among ES patients.

Methods

Patient population: Derivation and validation cohorts

All patients who underwent ES in the 2007–2017 American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database were included. ES is identified in the ACS-NSQIP database by the “EMERGENCY” variable as reported by the institution’s surgeon and anesthesiologist. We excluded patients who: (1) did not undergo general anesthesia, (2) had outpatient surgery, or (3) were already on mechanical ventilation preoperatively (as they will definitely need ICU postoperatively). The patient cohort was then divided in an 80:20 derivation-validation split to train and validate the model, respectively, using the OCT methodology described herein.

Data variables

Data variables included demographics (age, sex, race, ethnicity), comorbidities (diabetes, current smoker, dyspnea on exertion/ at rest, chronic obstructive pulmonary disease, ascites, congestive heart failure, hypertension, chronic renal failure, dialysis dependence, wound infection, recent weight loss, bleeding disorder, and steroid use), functional status, preoperative transfusion requirement, and laboratory values (sodium, blood urea nitrogen, creatinine, albumin, bilirubin, serum glutamic-oxaloacetic transaminase, alkaline phosphatase, white blood cell count, hematocrit, platelet count, partial thromboplastin time, and international normalized ratio). Patients who were septic were identified using the “PRSEPIIS” variable available in the ACS-NSQIP and subcategorized into (1) no sepsis, (2) Systemic Inflammatory Response Syndrome or sepsis, (3) septic shock, as the sepsis-3 new definitions eliminated the Systemic Inflammatory Response Syndrome terminology. Since all variables had less than 10% missing data, missing data were imputed using the ML method Optimal Impute, which demonstrated superior predictive performance for missing values when compared to more traditional methods of imputation.

Outcome: Defining “need for ICU”

To define our main outcome, need for ICU admission, all the ACS-NSQIP defined postoperative complications were reviewed by a panel of 6 board-certified surgical critical care experts to determine the ones that would most likely require management in the ICU. This methodology has been previously described by our team. Accordingly, we created the composite variable “ICUneed,” defined as postoperative death or the development of 1 or more postoperative complications warranting critical care (unplanned intubation, ventilator requirement for more than 48 hours, cardiac arrest requiring cardiopulmonary resuscitation, and septic shock). The panel did not include acute renal failure in the definition, as the presence of this complication alone is not generally a reason for ICU admission, but rather its concurrent presentation with another severe complication (eg, hyperkalemia).

Machine-learning OCT

The OCT methodology used to develop, train, and validate the predictive algorithm has been previously described in detail. Briefly, the OCT method accounts for nonlinear interactions among variables by reboots itself with each variable. This methodology tests all input variables and selects the most impactful ones for each tree to maximize performance. This allows the model to choose the variables as opposed to having them “forced” into the algorithm, and attributes specific cutoffs for the nodes (eg, laboratory value) after testing all possible scenarios. OCT does this by determining the splits based on knowledge of their down and upstream impact, which differentiates it from more classically “greedy” methods such as classification and regression trees, which follow a top-down approach for optimization. In addition to offering higher accuracy, OCT provides an easily interpretable model that builds predictions on a series of splits using a small number of high-importance variables. In that aspect, OCT is superior to other more opaque “black box” ML methods such as neural networks or gradient boosted decision-trees. For this analysis, a comprehensive tree was created to predict “ICUneed” after ES.

Measuring performance and calibration

We used the C statistic methodology to assess the model discrimination in both the 80% derivation and the 20% validation cohorts. Receiver operator characteristic curves were developed, and C statistics were calculated for “ICUneed.”

Interactive POTTER-ICU smartphone application

A user-friendly smartphone application was created as an interactive interface for direct use by providers. The application incorporates information from the OCT and presents a different number of questions depending on the individual patient’s characteristics as manifested in the trees. The application reports the predicted percentage for “ICUneed” and the number of patients used to develop those estimates.
A total of 464,861 patients were included (371,889 derivation; 92,972 validation). The median age was 56 years (interquartile range 40–77), and 57% of patients were male (56%). The overall cohort included 49,437 patients (11%) with severe COPD. The most common surgical specialty was General Surgery (78.1%), followed by Orthopedics (8.3%), Vascular Surgery (7.5%), and Cardiac Surgery (0.6%). A total of 4,647 patients (1%) had a history of severe COPD. The most common comorbidities were hypertension (41.0%) and diabetes (7.8%). The median BMI was 25.1 (IQR: 19.0–33.0), and the median serum creatinine was 0.9 (IQR: 0.7–1.1) mg/dL. The median INR was 1.1 (IQR: 1.0–1.2), and the median platelet count was 236,000 (IQR: 190,000–292,000) cells/µL. The median albumin was 3.8 (IQR: 3.3–4.2) g/dL, and the median hemoglobin was 15.0 (IQR: 13.5–16.0) g/dL. The median duration of surgery was 217 minutes (IQR: 140–303 minutes). The median number of units of blood products transfused was 1.0 (IQR: 0–2). The median hospital length of stay was 6 days (IQR: 3–11 days). A total of 46,353 patients (10%) had a diagnosis of sepsis, and 36,589 patients (8%) had a diagnosis of pneumonia. The overall incidence of surgical site infection was 5.4%, and the overall incidence of postoperative death was 0.3%.

**Ethical oversight**

This study was approved by the Mass General Brigham Institutional Review Board.
range [IQR] 38–71), 223,617 (48.1%) were male, and the most common comorbidities were hypertension (41.0%), diabetes (14.3%), and bleeding disorders (10.0%). The majority of patients underwent general surgery (77.8%), followed by orthopedic surgery (8.3%) and vascular surgery (7.5%). The overall 30-day postoperative mortality was 4.8%, and 11% of patients developed severe postoperative complications warranting critical care. Table I and II summarize the patient characteristics and outcomes of the derivation and validation cohorts, respectively.

A comprehensive OCT algorithm was developed to predict severe complications that require ICU admission. Figure 1 illustrates the complex and comprehensive structure of the overall tree, with a closer look at one example of the branch point structure. In this leaf, the initial branch point is based on whether the patient is septic or not. For patients in preoperative septic shock, the tree offers a final prediction (referred to as terminal node), with a probability of ICU admission (P[ICU]) of 63.67%. However, for patients who are not in preoperative septic shock, the next question is about functional status. For patients totally dependent at baseline, the tree offers a final prediction (P[ICU]) of 38.34%. For those who are independent or partially dependent, the tree goes on to ask about the patient’s age, followed by additional questions.

Based on the OCT trees, the interactive and user-friendly POTTER-ICU application was created (Figure 2). The number of questions (ie, node depths) needed to predict ICU admission ranged from as little as 2 questions to a maximum of 11 questions. Based on each answer given by the user, the application supplies the next question in accordance with the structure of the OCT until ultimately it yields the prediction.

**POTTER-ICU performance**

POTTER-ICU accurately predicted the need for ICU admission (C statistics: 0.89 train, 0.88 test). The receiver operator characteristic curve for the test cohort is displayed in Figure 3.

**Discussion**

POTTER-ICU is a user-friendly, accurate, ML-based, and interpretable risk prediction algorithm to predict severe complications after ES requiring ICU admission. POTTER-ICU can be easily and quickly used by bedside surgeons and anesthesiologists to predict preoperatively the need for postoperative critical care in ES patients. Allocating these patients to the appropriate level of care is essential to decrease the rates of failure to rescue while preserving hospital resources. This becomes especially important in resource-limited health care settings where critical care abilities are restricted by the number of available ICU beds, staffing, and equipment.
POTTER-ICU provides a predicted probability for requiring ICU admission postoperatively; however, it does not provide an explicit recommendation for a threshold at which clinicians should opt for ICU admission. Instead, as hospitals have different resources, we recommend that each hospital sets its own cutoff for ICU admission based on available critical care resources. For example, a health care facility with only 2 critical care beds and a step-down intermediate care unit can decide to set a high cutoff for the allocation of an ICU bed for an ES patient. On the other hand, a center with much larger ICU capacity and resources can elect to be more conservative and set a lower probability cutoff. As such, it can avoid under-triaging patients who are at a slightly increased risk of developing severe complications postoperatively.

Several studies have previously attempted to create and test ICU risk prediction models for elective and emergency surgery. Chan et al designed the Combined Assessment of Risk Encountered in Surgery (CARES) surgical risk calculator to predict the need for ICU stay based on routinely collected preoperative variables. Through the use of multivariable logistic regression analyses, they identified and assigned ranks to individual model parameters and attained a C statistic of 0.84. The Surgical Apgar score (SAS) is another scoring system that was presented as a tool to guide the decision to admit a patient directly to the ICU after surgery. However, SAS uses intraoperative information on hemodynamics and blood loss and cannot be calculated preoperatively. This undermines its usability as a triage tool because it offers little time between its derivation and the readiness of the patient to leave the operating room. For the ES patient in particular, Kongkaewpaisan et al tested the Emergency Surgery Score (ESS) as a tool to predict ICU need after ES. ESS gradually and accurately predicted critical care need with a C statistic of 0.90. However, all these models can be cumbersome to use by bedside and are still inherently limited as they assume a linear interaction between variables, whereas, in reality, the presence or absence of a variable can significantly affect the degree of impact of the next one. Consider the theoretical example of a 76-year-old man who requires ES for mesenteric ischemia. Preoperative septic shock may determine what other factors become important in influencing his outcome. If the patient is not in septic shock, whether the patient has renal disease or not may become important in predicting admission to the ICU. In contrast, if the patient is in septic shock preoperatively, the impact of renal failure will likely be different. Thus, nonlinear models better represent the complexity of real-life health care and are the next frontier of risk prediction in surgery.

There is some precedent to using ML-based methodology to predict ICU need after surgery. Chiew et al developed a highly...
accurate algorithm to predict the need for ICU stay after surgery.18 However, these algorithms are not specific to ES, and they rely on a “black box” methodology, in that the end-user does not understand the sequence of decisions the algorithms used to predict a certain outcome.19–21 POTTER-ICU is not only accurate and user-friendly, but it is also interpretable, offering a window into the patient characteristics driving the final prediction.

Our study has some limitations. First, we defined ICU need based on critical care surgeons’ consensus using only the available variables in the ACS-NSQIP database; other indications for ICU admission, such as severe electrolyte imbalances, were not included. Second, due to the retrospective nature of the study, we could not assess whether the patients who required ICU admission in our model did, in fact, receive critical care postoperatively. Third, as is the case with all ML models, the performance of our algorithm relies on the quality of the data used to train the trees. POTTER-ICU was created using the well-validated ACS-NSQIP database, which provides a large volume of data but remains limited by the number of variables that are collected. Our cohort is also skewed toward White (85.5%), and non-Hispanic (90.2%) patients. Fourth, POTTER-ICU is designed to offer preoperative prediction based on patient characteristics and laboratory values and therefore does not account for intraoperative surgical or anesthesia complications, or other features of a patient’s clinical course that might influence ICU need. POTTER-ICU was specifically built in this way to allow clinicians to predict and plan early on whether an ES patient will require an ICU level of care after the operation. Fifth, we used the ACS-NSQIP definition of ES, which is slightly different from the American Association for the Surgery of Trauma emergency general surgery definition, and as such could have resulted in different results.22 Finally, all our data were collected before the COVID-19 pandemic. Infection with COVID-19 might require additional considerations for ICU admission that were not accounted for in our model.

In conclusion, we therefore recommend POTTER-ICU as an accurate, user-friendly ML-based tool for predicting severe complications warranting ICU admission after ES. POTTER-ICU can prove useful as an adjunct to clinical judgment to triage patients to the ICU and to potentially decrease failure to rescue in ES patients. In resource-limited settings, POTTER-ICU may serve as a useful tool for appropriate triage to ensure prompt rescue of patients at highest risk of developing severe complications while avoiding unwarranted ICU admissions.

Funding/Support

CRICO.

Conflict of interest/Disclosure

Dr Bertsimas is a founder of Interpretable AI. Dr Kaafarani is a member of the scientific board of Alexandria health and earns royalties from UpToDate.

References