Medicine®

Risk factor stratification for postoperative delirium
A retrospective database study

Susana Vacas, MD, PhD[a],* , Tristan Grogan, MS[b] , Drew Cheng, MD[a], Ira Hofer, MD[a,c,d]

Abstract

Postoperative delirium (POD) is a disturbing reality for patients and their families. Absence of easy-to-use and accurate risk scores prompted us to retrospectively extract data from the electronic health records (EHR) to identify clinical factors associated with POD. We seek to create a multivariate nomogram to predict the risk of POD based upon the most significant clinical factors.

Methods: The EHR of 32734 patients >18 years of age undergoing surgery and had POD assessment were reviewed. Patient characteristics and study variables were summarized between delirium groups. We constructed univariate logistic regression models for POD using each study variable to estimate odds ratios (OR) and constructed a multivariable logistic regression model with stepwise variable selection. In order to create a clinically useful/implementable tool we created a nomogram to predict risk of delirium.

Results: Overall, we found a rate of POD of 3.7% across our study population. The Model achieved an AUC of the ROC curve of 0.83 (95% CI 0.82–0.84). We found that age, increased American Society of Anesthesiologists (ASA) score (ASA 3–4 OR 2.81, CI 1.49–5.28, $P < .001$), depression (OR 1.28, CI 1.12–1.47, $P < .001$), postoperative benzodiazepine use (OR 3.52, CI 3.06–4.06, $P < .001$) and urgent cases (Urgent OR 3.51, CI 2.92–4.21, $P < .001$; Emergent OR 3.99, CI 3.21–4.96, $P < .001$; Critically Emergent OR 5.30, CI 3.53–7.96, $P < .001$) were associated with POD.

Discussion: We were able to distinguish the contribution of individual risk factors to the development of POD. We created a clinically useful easy-to-use tool that has the potential to accurately identify those at high-risk of delirium, a first step to prevent POD.

Abbreviations: ASA = American Society of Anesthesiologists, BMI = body mass index, CAM = confusion assessment method, EHR = electronic health records, POD = postoperative delirium, OR = odds ratios.

Keywords: postoperative delirium, risk factors, electronic health records, postoperative neurologic outcomes, perioperative neurocognitive disorders, surgery, anesthesia

1. Introduction

Each year approximately 52 million Americans undergo surgical procedures,[1] characterized by immense physiologic stress. Despite a fundamental goal of returning patients to optimal health, harmful cognitive changes, such as postoperative delirium (POD), are a common and well-established occurrence after surgery.[2] POD is the most common surgical complication in older adults[3] and is estimated to cost $32.9 billion per year in the US alone.[4–6] POD doubles a patient's risk of post-discharge institutionalization, enhances all-cause mortality by 10% to 20% for every 48 hours of delirium,[7] and is linked to long-term cognitive decline, including a 8-fold increased risk of dementia.[8–10] It is estimated that 30% to 40% of POD cases (and ultimately long-lasting deficits) are preventable.[2,11,12]

Currently, perioperative prevention strategies are resource and personnel intensive and general prevention for all surgical patients is not cost-effective and probably unnecessary. A more realistic solution is risk stratification, whereby patients who are at high risk for POD can be targeted for preventative measures. Recognizing which patients are at heightened risk for developing

Anesthesiology and Medicine, Icahn School of Medicine at Mount Sinai, NY, USA, [d] Department of Medicine, Division of Data Driven Medicine (D3M), Icahn School of Medicine at Mount Sinai.

*Correspondence: Susana Vacas, Department of Anesthesiology and Perioperative Medicine, David Geffen School of Medicine, University of California Los Angeles, 757 Westwood Plaza, Suite 2331, Los Angeles, CA, 90095, USA (e-mail: svacas@mednet.ucla.edu).

POD continues to be an active area of research. Although several risk factors have been identified to be predictors of POD, most studies are limited by small sample size and/or are targeted for specific procedures or patient groups.[13–18] By 2014, a systemic review and metanalysis identified 37 different risk prediction models for POD, but most were not externally validated.[15,19] Approximately 80% of these tools focused on cardiac and orthopedic surgery,[19,20] while some were limited to a single type of pathology.[13] Unfortunately, many commonly-used POD risk stratification instruments show no superior advantage in application and performance than random chance,[14,15,21] which may be indicative of the small-sized studies on which these instruments were developed.

Over the past several years, the introduction of electronic health records (EHR) has increased the feasibility of obtaining a number of data points across a large patient population. In particular, our group has demonstrated the ability to extract complex medical phenotypes from a variety of EHR data.[22,23] The ability to extract these data from the EHR allows for the possibility of expanding on current work on POD risk stratification to include larger populations and a wider variety of surgical procedures.

In this manuscript we seek to leverage our expertise in complex data extraction from the EHR to retrospectively identify clinical factors associated with the risk of POD. As a secondary outcome we then seek to create a multivariate nomogram to predict the risk of POD based upon the most significant clinical factors.

2. Methods

2.1. Data extraction

The University of California Los Angeles (UCLA) Institutional Review Board approved and waived the requirement for informed consent (IRB#15-000518) for this single-center retrospective observational study. All patients aged 18 years and older undergoing surgery or procedures under anesthesia between April 2013 and July 2020 at the UCLA Medical Center and affiliated surgery centers were considered eligible for inclusion. We excluded from the analysis patients aged less than 18 years old, and cases that spent less than one night in the hospital.

All data for this study were extracted from the perioperative data warehouse, a custom-built robust data warehouse containing all patients who have undergone surgery at the UCLA Health since the implementation of our EHR (EPIC Systems, Madison, WI) in March 2013. We have previously described the creation of the perioperative data warehouse, which has a 3-stage design.[24,25] Briefly, in the first stage, data are extracted from EPIC's Clarity database into 29 tables organized around 3 distinct concepts: patients, surgical procedures, and health system encounters. These data are then cleaned and used to populate a series of 4000 distinct measures and metrics such as procedure duration, readmissions, admission International Classification of Diseases (ICD) codes, and postoperative outcomes.

2.2. Definition of outcome measure: postoperative delirium

We defined the primary outcome as POD using the confusion assessment method (CAM) score.[24] The CAM is a commonly used tool that identifies the presence/absence of POD and is consistently applied and documented in the EHR of surgical patients at UCLA.[24] It consists of a series of yes/no questions meeting the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) criteria for delirium: acute onset and/or fluctuating course during the day of disturbance of attention, awareness or cognition not explained by another preexisting or established neurocognitive disorder.[25,26] The responses to the individual metrics are stored in the EHR at the time of

documentation. These data were extracted from the EHR and the individual metrics grouped at the level of the calendar day. For that calendar day, a positive response to any of the CAM questions was taken as the patient having POD on that day. For our study, a patient was considered to have POD if they had a single positive CAM score in the seven calendar days following surgery. Because not all elements were always documented contemporaneously, in the event that only some elements were charted the most recent values for the other elements were used to compute the score.[25,26] Patients who were never screened for delirium were excluded from the study.

2.3. Perioperative characteristics

For the purposes of this study, preoperative patient characteristics were chosen based on current literature and included age, sex, height, weight, American Society of Anesthesiologists (ASA) status, surgical service, procedure type, inpatient versus outpatient status, elective versus emergent/critically emergent procedure, metabolic equivalents of task, SPICES score and comorbidities such as congestive heart failure, peripheral vascular disease, pulmonary history including Chronic Obstructive pulmonary Disease, kidney and endocrine disease, obstructive sleep apnea, neuropsychiatric history including anxiety and depression, history of tobacco or alcohol use, as well as imaging (echocardiogram) and laboratory values (glomerular filtration rate).

Intraoperative characteristics included type of anesthetic, duration of anesthesia, administration of benzodiazepines and intraoperative hemodynamic monitoring data, such as time spent with mean arterial pressure <55 or <65 mm Hg.

2.4. Statistical analysis

Patient characteristics and study variables were summarized between delirium groups using frequency (%) or mean (SD) and compared between groups using the chi-square test or t test as appropriate. We then constructed univariate logistic regression models for delirium using each study variable to estimate odds ratios (OR) with 95% confidence intervals. Next, we constructed a multivariable logistic regression model with stepwise variable selection to get a parsimonious model. Finally, in order to create a more clinically useful/implementable tool we created a nomogram from this model to predict risk of delirium both visually and as a point system score. The overall discrimination/prognostic ability of the model was assessed using the area under the ROC curve (AUC). P values < .05 were considered statistically significant and statistical analyses were performed using IBM SPSS V26 (Armonk, NY) and R V4.1.0 (www.r-project.org, Vienna, AU). Missing data was excluded from the analysis and reported in the tables.

3. Results

3.1. Patient demographics and univariate analysis

Supplementary Table S1, http://links.lww.com/MD/H645 shows the overall demographics of the cohort and the rates of the covariates studied. We identified 32,734 patients who met inclusion criteria. The overall incidence of POD was 3.7%. The cohort consisted of 7315 (22.3%) patients aged 65 to 75 years and 5401 (16.5%) patients age >75 years old. The majority (20,199 [62.7%]) of patients were ASA physical status class 3 or 4 and the most common surgical services were Orthopedics (7200), General Surgery (6126), and Neurosurgery (3345). Most cases (19,950 [60.9%]) were elective.

Overall, in univariate analysis a wide variety of factors were directly associated with POD. Of note, patient age, ASA physical

status, female sex, surgical service and case urgency were associated with POD (see Supplementary Table S1, http://links.lww.com/MD/H645). In addition, body mass index (BMI), depression and postoperative benzodiazepine use) were associated with POD; however, premedication with benzodiazepines was not associated with POD, nor was anesthesia type. Study variables can be found in Supplementary Table S1, http://links.lww.com/MD/H645.

3.2. Multivariate model

Table 1 shows the odds ratios for the features included in the multivariate analysis of POD. Overall, the Model achieved an AUC of the ROC curve of 0.83 (95% CI 0.82–0.84). Age above 45 was consistently associated with POD, with age greater than 75 years having a OR 8.26, 95% CI 6.39-10.68 of POD. In addition, increased ASA physical status score (ASA 3–4 OR 2.81, 95% CI 1.49–5.28, $P < .001$), history of depression (OR 1.28, 95% CI 1.12–1.47, $P < .001$), postoperative benzodiazepine use (OR 3.52, 95% CI 3.06–4.06, $P < .001$) and more urgent cases (Urgent OR 3.51, 95% CI 2.92–4.21, $P < .001$; Emergent OR 3.99, 95% CI 3.21–4.96, $P < .001$; Critically Emergent OR

5.30, 95% CI 3.53–7.96, $P < .001$) were associated with POD. Of interest premedication with benzodiazepines was protective of POD.

3.3. Creation of a score for POD risk stratification

In order to facilitate prospective risk stratification Figure 1 contains a nomogram that can be used to convert the odds ratios from the multivariate model into a "score" for POD prediction. The points associated with the various sections of the nomogram can be found in Supplementary excel file: 1 (https://docs.google.com/spreadsheets/d/1U4fzoNTsIZT4kbAd3cIpfhPktx0FJNru/edit?usp=sharing&ouid=101929998507796207436&rtpof=true&sd=true).

4. Discussion

This is (to our knowledge) the largest retrospective cohort analysis of patients studied for POD to date. We examined nearly 33,000 patients aged >18 years old with an overall incidence of POD of 3.7%. The final multivariate model, containing nine features achieved an AUC of 0.85 (0.8–0.86) and was thus highly discriminating for POD. While many of our findings are similar to those of previous work in smaller studies on more focused populations, we report several differences. Like others, we found that POD was highly associated with advancing age, increased patient co-morbidities (ASA physical status), male gender, depression, high BMI, emergent cases, and invasive surgical procedures. However, we found no differences regarding type of anesthesia or if patients where premedicated with midazolam. The surgical services most likely where patients were more likely to develop POD where orthopedics, liver transplant and neurosurgery.

Preoperative identification of high-risk patients for POD is essential to allocate time and resources to implement targeted preventive strategies. As noted, this manuscript is not the first to perform a retrospective analysis of risk factors for POD.[27–29] Medical literature contains many risk scores designs, each attempting to distill the likelihood of one of a variety of different outcomes.[13–18] However, to be easily calculated, were based on a limited number of highly predictive features. These scores evaluate disease status in broad strokes[30–33] and often lack precision at the patient level—making them better suited to risk adjustment for populations than individualized risk prediction.[34] Most of these studies were on smaller populations or focused on specific types of surgeries, or patient groups. For example, one important risk score tool for POD based on a large population (approximately 10,000 patients), recently developed by Whitlock et al[14] accounts for precipitating factors solely by classifying surgical risk as moderate or high, and their risk score is unweighted and additive (i.e., summing the points for each risk factor), while clinical risk comes from the complex interplay of multiple factors. Our study differs in several ways. Firstly, rather than providing broad surgical groupings (low, medium, and high risk) we stratified by surgical services while incorporating urgency as its own feature. Secondly, by directly harnessing EHR data we were able to incorporate a wider range of clinical co-morbidities hypothesized to be associated with POD including depression and BMI. Of note, the ability to increase the amount of data included in our model resulted in an AUC substantially higher than those reported previously. This gives further credence for the use of advanced modeling techniques, such as machine learning, that has been used to predict other conditions.[35–38]

In developing our model, we attempted to extract a wide variety of relevant data from the EHR to include as potential co-variates. Nonetheless, not all information was necessarily readily available. Important predictors, such as detailed information on cognition and sensory testing was not available. Further, previous studies (including several by our group) have

Table 1

Multivariate odds ratios for postoperative delirium and *p* values for each category.

Demographic/clinical characteristics	Multivariable OR (95%CI)	p value
Age group		<.001
18–45 (ref)	--	
45–65	2.39 (1.87–3.06)	
65–75	3.26 (2.51–4.22)	
>75	8.26 (6.39–10.68)	
ASA score group		<.001
1–2 (ref)	--	
3–4	2.86 (2.29–3.58)	
5	2.81 (1.49–5.28)	
Male sex (ref = female)	1.29 (1.13–1.48)	<.001
Hx Depression	1.28 (1.12–1.47)	<.001
Perioperative midazolam	0.75 (0.65–0.87)	<.001
Postoperative benzodiazepine	3.52 (3.06–4.06)	<.001
Primary surgical service		<.001
Obstetrics and gynecology (ref)	--	
Cardiac surgery	1.15 (0.48–2.76)	
Gastroenterology	1.01 (0.43–2.39)	
General surgery	1.41 (0.61–3.25)	
Liver transplant	3.10 (1.30–7.43)	
Neurosurgery	4.50 (1.96–10.34)	
Orthopaedics	2.86 (1.25–6.52)	
Other	1.35 (0.55–3.28)	
Otolaryngology	1.20 (0.50–2.90)	
Plastic surgery	2.13 (0.82–5.53)	
Surgical oncology	1.17 (0.44–3.07)	
Thoracic surgery	1.85 (0.74–4.62)	
Urology	1.37 (0.58–3.25)	
Vascular surgery	1.66 (0.67–4.10)	
BMI categories		.001
18.5–29.99 (ref)	--	
<18.5	1.45 (1.15–1.84)	
30–39.99	0.93 (0.78–1.10)	
40+	0.66 (0.44–0.99)	
Case_type		<.001
Elective (ref)	--	
Critically emergent	5.30 (3.53–7.96)	
Emergent	3.99 (3.21–4.96)	
Inpatient	3.49 (2.89–4.23)	
Transplant case	1.88 (1.25–2.83)	
Urgent	3.51 (2.92–4.21)	
AUC (95%CI)	0.85 (0.84–0.86)	

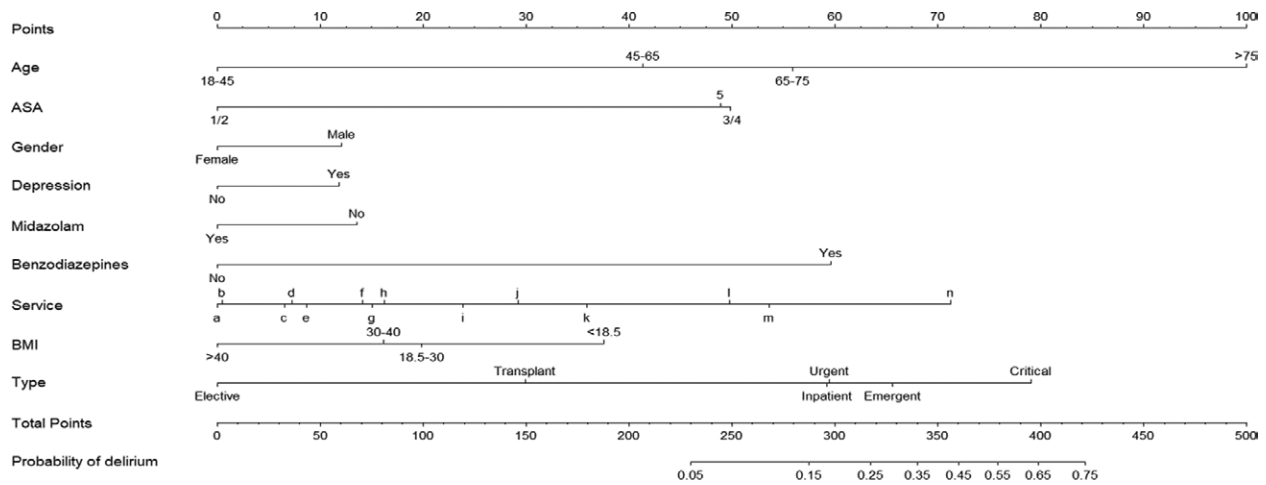ASA = American Society of Anesthesiologists, BMI = body mass index.

Figure 1. Nomogram point to predict individual risk for postoperative delirium (service legend a) obstetrics and gynecology, b) gastroenterology, c) cardiac surgery, d) surgical oncology, e) otolaryngology, f) other, g) urology, h) general surgery, i) vascular surgery, j) thoracic surgery, k) plastic surgery, l) orthopaedics, m) liver transplant, n) Neurosurgery).

repeatedly shown that EHR data is sparely populated and thus potentially inaccurate.[23] We attempted to account for this by leveraging various types of EHR data to make more complete clinical phenotypes, however for patients less known to our system we may not have been sufficiently sensitive.

As important as the features that were associated with POD, are findings of features that were not associated with POD. In particular the use of midazolam as a preoperative anxiolytic has been the topic of much discussion for several years. In our cohort, this was actually protective of POD. However, it is critical to call out the limits of associations in a retrospectively analysis such as this one. Given the awareness of theoretically associations between preoperative benzodiazepine use and POD, it is possible that there was significant selection bias in which patients were given midazolam preoperatively thereby affecting the results. Thus, these results must be interpreted carefully given the inability to address all possible confounders.

Importantly, we only included those patients who were screened for POD at least once during their hospitalization. Although, all patients admitted to the hospital should undergo delirium screening, it does not always happen. It is highly probable that patients at higher risk of delirium were screened more frequently introducing a selection bias to our cohort. In addition, previous work has shown that hyperactive delirium is more likely to be recognized by providers than hypoactive delirium. Our healthcare system has trained staff to recognize signs and symptoms of delirium. At least, for those patients that developed hyperactive delirium, it is highly likely that this information was introduced in the chart. Nonetheless, the exact rates reported in this manuscript may not be completely accurate and it is possible that the exact risk calculated based on our nomogram is not completely generalizable to patients who were not screened for POD. However, these broader trends and associations are likely correct and thus informative in creating a risk score.

POD is a multifactorial problem and until real-time clinical implementation of machine learning models become accurate and feasible, POD prediction that relies on weighted risk factors with an easy-to-use tool has the tremendous potential to improve predictive capacity and deployment of targeted interventions.

Author contributions

Susana Vacas: This author helped in study conception and design, data acquisition, data validation and interpretation and with manuscript drafting and revision. The author has approved the final version of the manuscript.

References

[1] Centers for disease control and prevention - national hospital discharge survey, 2010 - number of all-listed procedures for discharges from short-stay hospitals, by procedure category and age: United States. [Access date February, 2020].
[2] Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet (London, England). 2014;383:911–22.
[3] Postoperative delirium in older adults: best practice statement from the American geriatrics society. J Am Coll Surg. 2015;220:136–48.e1.
[4] Leslie DL, Inouye SK. The importance of delirium: economic and societal costs. J Am Geriatr Soc. 2011;59(Suppl 2):S241–3.
[5] Leslie DL, Marcantonio ER, Zhang Y, et al. One-year health care costs associated with delirium in the elderly population. Arch Intern Med. 2008;168:27–32.
[6] Gou RY, Hshieh TT, Marcantonio ER, et al. One-year medicare costs associated with delirium in older patients undergoing major elective surgery. JAMA Surg. 2021;156:430–42.
[7] Whitlock EL, Vannucci A, Avidan MS. Postoperative delirium. Minerva Anestesiol. 2011;77:448–56.
[8] Davis DH, Muniz Terrera G, Keage H, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. Brain. 2012;135(Pt 9):2809–16.
[9] Aranake-Chrisinger A, Avidan MS. Postoperative delirium portends descent to dementia. Br J Anaesth. 2017;119:285–8.

[10] Sprung J, Roberts RO, Weingarten TN, et al. Postoperative delirium in elderly patients is associated with subsequent cognitive impairment. Br J Anaesth. 2017;119:316–23.

[11] Inouye SK, Bogardus BS, Jr, Charpentier PA, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med. 1999;340:669–76.

[12] Marcantonio ER, Flacker JM, Wright RJ, et al. Reducing delirium after hip fracture: a randomized trial. J Am Geriatr Soc. 2001;49:516–22.

[13] Choi NY, Kim EH, Baek CH, et al. Development of a nomogram for predicting the probability of postoperative delirium in patients undergoing free flap reconstruction for head and neck cancer. Eur J Surg Oncol. 2017;43:683–8.

[14] Whitlock EL, Braehler MR, Kaplan JA, et al. Derivation, validation, sustained performance, and clinical impact of an electronic medical record-based perioperative delirium risk stratification tool. Anesth Analg. 2020;131:1901–10.

[15] Jansen CJ, Absalom AR, de Bock GH, et al. Performance and agreement of risk stratification instruments for postoperative delirium in persons aged 50 years or older. PLoS One. 2014;9:e113946e113946.

[16] Marcantonio ER, Goldman L, Mangione CM, et al. A clinical prediction rule for delirium after elective noncardiac surgery. Jama. 1994;271:134–9.

[17] Litaker D, Locala J, Franco K, et al. Preoperative risk factors for postoperative delirium. Gen Hosp Psych. 2001;23:84–7.

[18] Kim EM, Li G, Kim M. Development of a risk score to predict postoperative delirium in patients with hip fracture. Anesth Analg. 2020;130:79–86.

[19] van Meenen LC, van Meenen DM, de Rooij SE, et al. Risk prediction models for postoperative delirium: a systematic review and meta-analysis. J Am Geriatr Soc. 2014;62:2383–90.

[20] Rudolph JL, Jones RN, Levkoff SE, et al. Derivation and validation of a preoperative prediction rule for delirium after cardiac surgery. Circulation. 2009;119:229–36.

[21] Lindroth H, Bratzke L, Purvis S, et al. Systematic review of prediction models for delirium in the older adult inpatient. BMJ Open. 2018;8:e019223.

[22] Hofer IS, Cheng D, Grogan T. A retrospective analysis demonstrates that a failure to document key comorbid diseases in the anesthesia preoperative evaluation associates with increased length of stay and mortality. Anesth Analg. 2021;133:698–706.

[23] Hofer IS, Cheng D, Grogan T, et al. Automated assessment of existing patient's revised cardiac risk index using algorithmic software. Anesth Analg. 2019;128:909–16.

[24] Inouye SK, van Dyke CH, Alessi CA, et al. Clarifying confusion: the confusion assessment method. Ann Intern Med. 1990;113:941–8.

[25] Evered L, Silbert B, Knopman DS, et al. Recommendations for the nomenclature of cognitive change associated with anaesthesia and surgery-2018. Anesthesia Analgesia. 2018;127:1189–95.

[26] Association AP. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition: DSM-5. American Psychiatric Association. 2013.

[27] Ansaloni L, Catena F, Chattat R, et al. Risk factors and incidence of postoperative delirium in elderly patients after elective and emergency surgery. Br J Surg. 2010;97:273–80.

[28] Benoit AG, Campbell MI, Tanner JR, et al. Risk factors and prevalence of perioperative cognitive dysfunction in abdominal ancurysm patients. J Vasc Surg. 2005;42:884–90.

[29] Moller JT, Cluitmans P, Rasmussen LS, et al. Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International study of post-operative cognitive dysfunction. Lancet. 1998;351:857–61.

[30] Le Manach Y, Collins G, Rodseth R, et al. Preoperative score to predict postoperative mortality (POSPOM): derivation and validation. Anesthesiology. 2016;124:570–9.

[31] Sobol JB, Wunsch H. Triage of high-risk surgical patients for intensive care. Review. Crit Care. 2011;15:217.

[32] Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European society of intensive care medicine. Intensive Care Med. 1996;22:707–10.

[33] Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest. 1991;100:1619–36.

[34] Hofer IS, Halperin E, Cannesson M. Opening the black box: understanding the science behind big data and predictive analytics. Anesth Analg. 2018;127:1139–43.

[35] Hill BL, Brown R, Gabel E, et al. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. Br J Anaesth. 2019;123:877–86.

[36] Hofer IS, Burns M, Kendale S, et al. Realistically integrating machine learning into clinical practice: a road map of opportunities, challenges, and a potential future. Anesth Analg. 2020;130:1115–8.

[37] Hofer IS, Lee C, Gabel E, et al. Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set. NPJ Digit Med. 2020;3:58.

[38] Lee CK, Hofer I, Gabel E, et al. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. Anesthesiology. 2018;129:649–62.