



Machine learning predicts unpredicted deaths with high accuracy following hepatopancreatic surgery

Kota Sahara^{1,2}, Anghela Z. Paredes¹, Diamantis I. Tsilimigras¹, Kazunari Sasaki³, Amika Moro¹, J. Madison Hyer¹, Rittal Mehta¹, Syeda A. Farooq¹, Lu Wu¹, Itaru Endo², Timothy M. Pawlik¹

¹Division of Surgical Oncology, The Ohio State University Wexner Medical Center and James Comprehensive Cancer Center, Columbus, OH, USA; ²Gastroenterological Surgery Division, Yokohama City University School of Medicine, Yokohama, Japan; ³Department of General Surgery, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, OH, USA

Contributions: (I) Conception and design: K Sahara, AZ Paredes, DI Tsilimigras, K Sasaki; (II) Administrative support: I Endo, TM Pawlik; (III) Provision of study material or patients: K Sahara, AZ Paredes, DI Tsilimigras, A Moro; (IV) Collection and assembly of data: K Sahara, SA Farooq, L Wu; (V) Data analysis and interpretation: K Sahara, K Sasaki, JM Hyer, R Mehta; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Timothy M. Pawlik, MD, MPH, PhD, FACS, FRACS (Hon.). Professor and Chair, Department of Surgery, The Urban Meyer III and Shelley Meyer Chair for Cancer Research, The Ohio State University, Wexner Medical Center, 395 W. 12th Ave., Suite 670, Columbus, OH, USA. Email: Tim.Pawlik@osumc.edu.

Background: Machine learning to predict morbidity and mortality—especially in a population traditionally considered low risk—has not been previously examined. We sought to characterize the incidence of death among patients with a low estimated morbidity and mortality risk based on the National Surgical Quality Improvement Program (NSQIP) estimated probability (EP), as well as develop a machine learning model to identify individuals at risk for “unpredicted death” (UD) among patients undergoing hepatopancreatic (HP) procedures.

Methods: The NSQIP database was used to identify patients who underwent elective HP surgery between 2012–2017. The risk of morbidity and mortality was stratified into three tiers (low, intermediate, or high estimated) using a k-means clustering method with bin sorting. A machine learning classification tree and multivariable regression analyses were used to predict 30-day mortality with a 10-fold cross validation. C statistics were used to compare model performance.

Results: Among 63,507 patients who underwent an HP procedure, median patient age was 63 (IQR: 54–71) years. Patients underwent either pancreatectomy (n=38,209, 60.2%) or hepatic resection (n=25,298, 39.8%). Patients were stratified into three tiers of predicted morbidity and mortality risk based on the NSQIP EP: low (n=36,923, 58.1%), intermediate (n=23,609, 37.2%) and high risk (n=2,975, 4.7%). Among 36,923 patients with low estimated risk of morbidity and mortality, 237 patients (0.6%) experienced a UD. According to the classification tree analysis, age was the most important factor to predict UD (importance 16.9) followed by preoperative albumin level (importance: 10.8), disseminated cancer (importance: 6.5), preoperative platelet count (importance: 6.5), and sex (importance 5.9). Among patients deemed to be low risk, the c-statistic for the machine learning derived prediction model was 0.807 compared with an AUC of only 0.662 for the NSQIP EP.

Conclusions: A prognostic model derived using machine learning methodology performed better than the NSQIP EP in predicting 30-day UD among low risk patients undergoing HP surgery.

Keywords: Mortality; unpredicted; machine learning; National Surgical Quality Improvement Program (NSQIP)

Submitted Sep 27, 2019. Accepted for publication Nov 12, 2019.

doi: 10.21037/hbsn.2019.11.30

View this article at: <http://dx.doi.org/10.21037/hbsn.2019.11.30>

Introduction

Over the past several decades, operative mortality following high-risk surgery has steadily declined (1). The decline in mortality has been due in part to improved patient selection, consolidation of high-risk operations at high volume centers and improved medical management. Specifically, following many complex surgical interventions, the incidence of 30-day mortality now ranges from 0.9% to 6.7% with most variation attributed to patient level factors and procedure complexity (2,3). To this point, hepatopancreatic (HP) operations remain among the most complex set of general surgical procedures with associated rates of postoperative complications and mortality as high as 40–50% and 2–6%, respectively (4–7). Given the risk of morbidity and mortality for these types of complex operations, as well as the aged population in which many of these procedures are performed, patient selection is critical (8).

In this context, prognostic models have been increasingly proposed as a means to optimize patient selection, as well as stratify risk among patients to improve shared decision-making and balance risks versus benefits of surgery (9). To this end, the American College of Surgeons (ACS) developed the National Surgical Quality Improvement Program (NSQIP) Surgical Risk Calculator (SRC) to help inform providers and patients of the likelihood of an adverse outcome following surgery (10,11). Prior studies have noted prognostic gaps within the SRC. As such, disease-specific and population-specific prognostic models have been developed expanding on variables included in the NSQIP SRC (12,13). For example, Min *et al.* developed the preoperative “VESPA” tool that incorporated five preoperative activities of daily living, in addition to other measures, to identify elderly patients at risk for postoperative surgical and geriatric complications (14). In a separate study, Hyder *et al.* utilized preoperative lab values to develop a risk score with high sensitivity and specificity to predict 90-day mortality among a cohort of patients undergoing liver resection (15). These and other tools have largely been created, however, based on assumptions about the distribution of the data and their relative independence from other factors. This assumption may be problematic as prognostic factors can often behave in a synergistic—rather than independent—manner, leading to greater than expected increases in the risk of certain outcomes (16). As such, previous prediction models that have employed traditional statistical methods may not have accounted for more complicated and nuanced relationships among data

variables to predict outcomes.

To this end, machine learning methods have been increasingly adopted and have demonstrated high sensitivity to predict certain outcomes including cancer recurrence (17), overall survival (18), and hospital readmission (19). Machine learning to predict morbidity and mortality—especially in a population traditionally considered low risk—has not been previously examined. As such, the objective of the current study was to characterize the incidence of death among patients with a low estimated morbidity and mortality risk based on the NSQIP estimated probability (EP) in the ACS NSQIP dataset. In particular, we sought to develop a machine learning model to identify individuals at risk for “unpredicted death” (UD) among patients undergoing HP procedures.

Methods

Data source

ACS NSQIP is the premier surgical quality and outcomes assessment program, which provides reliable and valid surgical outcome measures for over 5.5 million cases in both the inpatient and outpatient setting from over 700 NSQIP participating facilities (20). The NSQIP sampling approach and clinical abstraction methods have been previously reported (21). Briefly, the program aggregates detailed information on patient demographics, preoperative risk factors and laboratory values, intraoperative variables and postoperative outcomes using standardized definitions.

Study population

Patients who underwent an elective hepatectomy and pancreatectomy for a benign or malignant indication between 2012 and 2017 were identified using current procedural terminology (CPT) codes (Table S1). Patients who were younger than 18 years old, as well as individuals who underwent emergency surgery were excluded from the analytic cohort. Information in the dataset included preoperative comorbidities and perioperative clinical variables, as well as 30-day postoperative complications and mortality.

The ACS NSQIP database includes the EP of 30-day morbidity (MORBPROB) and 30-day mortality (MORTPROB) (22,23). The probabilities were developed for all cases based on a logistic regression analysis using the patient’s preoperative characteristics as the independent or predictive variables (24). The risk of morbidity and

mortality was stratified into three tiers (low, intermediate, or high estimated risk) using a k-means clustering method with bin sorting by median in order to compute cluster seed (25). An UD was defined as a death within 30-days of surgery in a patient with low risk for a morbidity or mortality based on the NSQIP EP.

Data analysis

Descriptive statistics were presented as median and interquartile range (IQR) and frequency (%) for continuous and categorical variables, respectively. Categorical variables were compared using chi-square tests and Fisher exact tests, where appropriate. Continuous variables were compared using Wilcoxon rank-sum test. Demographics, patient characteristics, 30-day postoperative complications were compared among patients both categorized and not categorized as a UD.

A classification tree was constructed as a predictive tool to assist in the prediction of UD by stratifying patients into different risk groups on the basis of preoperative patient characteristics and laboratory values. Classification trees are a nonparametric predictive modeling technique commonly used in machine learning to predict a binary outcome (26). To prune the tree and minimize overfitting, a 10-fold cross validation method was utilized (27). The performance of the classification tree and the NSQIP EP was measured by the C-statistic, also known as the area under the curve (AUC). AUC comparisons were made using the DeLong test, a nonparametric method that exploits the mathematical equivalence of the AUC to the Mann-Whitney U-statistic (28). Two distinct methods were used to determine factors associated with UD: multivariable logistic regression analysis and a machine learning method. All factors associated with a UD on bivariate analysis were considered in the full multivariable model. As previously reported, effect sizes of factors associated with UD from the multivariable analysis were measured using LogWorth values [wherein LogWorth represents $-\log_{10}$ (P value), such that $P=0.01$ is equivalent to a LogWorth of 2.0] (29). The relative importance of preoperative variables was calculated to identify factors that were noted to contribute to the decision-tree model algorithm (18). Additional analyses were conducted with the machine learning algorithm that utilized factors stratified into hepatectomy and pancreatectomy variables while incorporating procedure specific variables for hepatectomy (e.g., cirrhosis, biliary stent placement, minimally invasive approach), as well as for

pancreatectomy (e.g., vascular resection, pancreatic gland texture). For the subgroup analysis, the Procedure Targeted module of the ACS NSQIP dataset from 2014 to 2017 were utilized to incorporate the procedure specific variables for hepatectomy (e.g., cirrhosis, biliary stent placement, minimally invasive approach), as well as for pancreatectomy (e.g., vascular resection, pancreatic gland texture). Statistical significance was assessed at $\alpha=0.05$. All analyses were performed using SPSS version 25 (IBM Corporation, Armonk, NY, USA) and JMP statistical package version 14 (SAS Institute Inc., Cary, NC, USA). The study was exempt from review by the Ohio State University Institutional Review board because the NSQIP database was available to all participating institutions and contained no identifiable protected health information.

Results

Study population

A total of 63,507 patients who underwent an HP procedure met inclusion criteria (Figure S1). Median patient age was 63 (IQR: 54–71) years. Approximately half of the cohort was female ($n=31,879$, 50.2%); the majority were white ($n=44,588$, 70.2%) and functionally independent prior to surgery ($n=62,904$, 99.3%). Most patients underwent a pancreatectomy ($n=38,209$, 60.2%), whereas a smaller cohort had a hepatic resection ($n=25,298$, 39.8%). Overall, patients were stratified into three tiers of predicted morbidity and mortality risk based on the NSQIP EP: low ($n=36,923$, 58.1%), intermediate ($n=23,609$, 37.2%) and high risk ($n=2,975$, 4.7%) (Figure 1A). Among 36,923 patients with low estimated risk of morbidity and mortality, only 237 patients (0.6%) experienced a UD following HP surgery (Figure 1B), whereas the vast majority did not die within 30-days of surgery ($n=36,686$, 99.4%).

Patient characteristics and perioperative outcomes relative to UD

Among patients with a low risk of morbidity and mortality, patients who experienced a UD were more likely to be older [median age, 64 (IQR: 58–72) vs. 59 (IQR: 49–68) year, $P<0.001$] and male ($n=140$, 59.1% vs. $n=16,686$, 45.5%; $P<0.001$). Additionally, a greater proportion of patients with a postoperative UD were in the higher ASA classification category ($P<0.001$). Patients with a UD also had a greater incidence of comorbidities and preoperative conditions

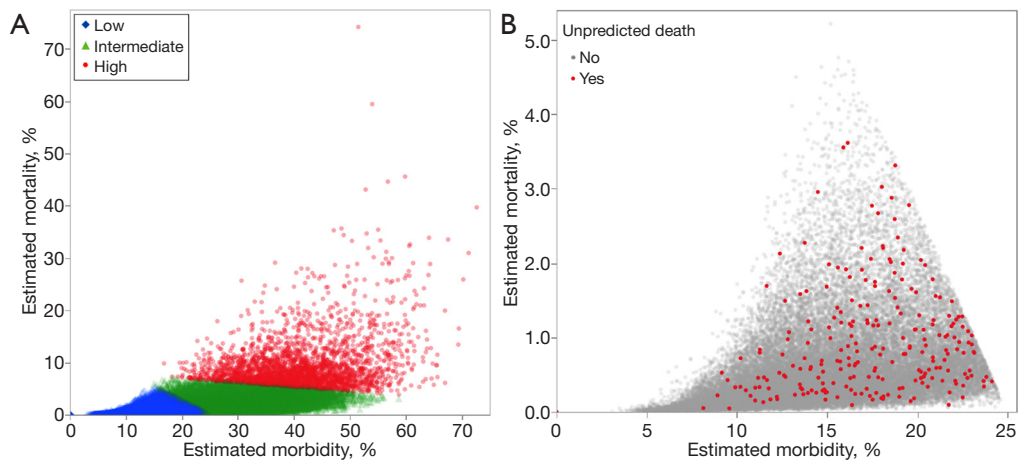


Figure 1 Scatter plots using cluster analysis with estimated morbidity and mortality: (A) in the entire cohort; (B) among patients with low estimated morbidity and mortality relative to the occurrence of an unpredicted death.

such as diabetes, hypertension, steroid use for a chronic condition, and weight loss >10% prior to surgery (all $P < 0.05$). In contrast, other characteristics such as race, body mass index (BMI), concurrent chronic obstructive pulmonary disease (COPD) and disseminated cancer were comparable among patients who had and did not experience a UD (all $P > 0.05$) (Table 1).

Median time from surgery to a UD was 10 (IQR 6–18) days (Table 2). Perhaps not surprising, patients who had a UD experienced more complications [median number, 2 (IQR: 1–3) *vs.* 0 (IQR: 0–0); $P < 0.001$]. The most common adverse outcome among patients with a UD was reoperation ($n=72$, 30.4% *vs.* $n=1,040$, 2.8%; $P < 0.001$), followed by renal failure ($n=55$, 23.2% *vs.* $n=77$, 0.2%; $P < 0.001$) and organ space surgical site infection (SSI) ($n=43$, 18.1% *vs.* $n=3,019$, 8.2%; $P < 0.001$). Similarly, patients experiencing a UD had a higher incidence of pneumonia, myocardial infarction, renal insufficiency, pulmonary embolism, and cerebrovascular accident (CVA)/stroke following HP surgery (all $P < 0.001$); in contrast, the incidence of sepsis, urinary tract infection (UTI), superficial SSI, deep SSI, and wound disruption was comparable (all $P > 0.05$).

Classification tree analysis

The classification tree predicting the occurrence of UD within 30-day of HP surgery is depicted in Figure 2A. An example of a decision-tree that estimated the risk of a UD following HP surgery is shown in Figure 2B, with the example tree limited to display four decision nodes. The

root node represented the entire cohort ($n=36,923$). The first split was secondary to the age variable (53 years), as determined by the computer algorithm that deemed this age as the optimal cutoff to split the node into two more homogenous sub-nodes. Specifically, if a given patient was 53-year-old or older, the algorithm led to the left branch of the tree with an associated risk of a UD of 0.8% (1 in 125). As the nodes continued to split into more homogenous groups, the risk was re-calculated by the machine learning algorithm.

Predicting UD: model accuracy

Using a multivariable logistic regression analysis, age was the strongest factor associated with UD (LogWorth 4.618), followed by history of steroid use (LogWorth 2.858), pre-operative albumin level (LogWorth 2.654), pre-operative total bilirubin level (LogWorth 2.176), and sex (LogWorth 1.611) (Figure 3A). According to the classification tree analysis, age was similarly noted to be one of the most important factors to predict UD (importance: 16.9) (Figure 3B). In the classification tree analysis, preoperative albumin level (importance: 10.8) was noted to be the second most important factor, followed by the presence of disseminated cancer (importance: 6.5), preoperative platelet count (importance: 6.5), and sex (importance: 5.9) (Figure 3B). Among patients deemed to be low risk, the c-statistic for the machine learning derived prediction model was 0.807 and better compared with an AUC of only 0.662 for the NSQIP EP ($P < 0.001$) (Figure 4).

Table 1 Patients characteristics

Variable	Unpredicted death		P
	No, N=36,686	Yes, N=237	
Age, median [IQR]	59 [49–68]	64 [58–72]	<0.001
Male	16,686 (45.5%)	140 (59.1%)	<0.001
Race			0.71
White	25,221 (68.7%)	158 (66.7%)	
AA	3,216 (8.8%)	21 (8.9%)	
Asian	2,053 (5.6%)	14 (5.9%)	
Hispanic	1,299 (3.5%)	6 (2.5%)	
Other/unknown	4,897 (13.3%)	38 (16.0%)	
Functional status			0.73
Independent	36,503 (99.5%)	236 (99.6%)	
Partially dependent	93 (0.3%)	0	
Totally dependent	5 (0.0%)	0	
Unknown	85 (0.2%)	1 (0.4%)	
ASA classification			<0.001
1	673 (1.8%)	1 (0.4%)	
2	13,971 (38.1%)	61 (25.7%)	
3	21,283 (58.0%)	164 (69.2%)	
4	700 (1.9%)	11 (4.6%)	
Unknown	59 (0.2%)	0	
BMI, median [IQR]	27.2 [23.9–31.2]	28.2 [23.7–31.9]	0.31
Procedure			1.00
Hepatectomy	18,948 (51.6%)	122 (51.5%)	
Pancreatectomy	17,738 (48.4%)	115 (48.5%)	
Diabetes			0.001
No	30,774 (83.9%)	178 (75.1%)	
Non-insulin	3,897 (10.6%)	42 (17.7%)	
Insulin	2,015 (5.5%)	17 (7.2%)	
Current smoker			0.81
No	31,432 (85.7%)	213 (89.9%)	
Yes	5,254 (14.3%)	24 (10.1%)	
Dyspnea			0.34
No	35,540 (96.9%)	227 (95.8%)	
Yes	1,146 (3.1%)	10 (4.2%)	
Ventilator dependent			1.00
No	36,684 (100.0%)	237 (100.0%)	
Yes	2 (0.0%)	0	

Table 1 (continued)**Table 1** (continued)

Variable	Unpredicted death		P
	No, N=36,686	Yes, N=237	
Severe COPD			0.42
No	36,125 (98.5%)	232 (97.9%)	
Yes	561 (1.5%)	5 (2.1%)	
Ascites			1.00
No	36,606 (99.8%)	237 (100.0%)	
Yes	80 (0.2%)	0	
Congestive heart failure			1.00
No	36,651 (99.9%)	237 (100.0%)	
Yes	35 (0.1%)	0	
Hypertension			<0.001
No	21,664 (59.1%)	110 (46.4%)	
Yes	15,022 (40.9%)	127 (53.6%)	
Acute renal failure			1.00
No	36,676 (100.0%)	237 (100.0%)	
Yes	10 (0.0%)	0	
Currently on dialysis			0.33
No	36,625 (99.8%)	236 (99.6%)	
Yes	61 (0.2%)	1 (0.4%)	
Disseminated cancer			0.17
No	28,274 (77.1%)	192 (81.0%)	
Yes	8,412 (22.9%)	45 (19.0%)	
Steroid use for chronic condition			0.017
No	35,801 (97.6%)	225 (94.9%)	
Yes	885 (2.4%)	12 (5.1%)	
Body weight loss >10%			0.012
No	35,207 (96.0%)	219 (92.4%)	
Yes	1,479 (4.0%)	18 (7.6%)	
Bleeding disorder			0.050
No	35,989 (98.1%)	228 (96.2%)	
Yes	697 (1.9%)	9 (3.8%)	
Transfusion within 72 h			0.33
No	36,626 (99.8%)	236 (99.6%)	
Yes	60 (0.2%)	1 (0.4%)	

IQR, interquartile range; AA, African American; ASA, American Society of Anesthesiologists; BMI, body mass index; COPD, chronic obstructive pulmonary disease.

Table 2 Postoperative outcomes relative to the occurrence of unpredicted death

Variable	Unpredicted death		P
	No, N=36,686	Yes, N=237	
Number of complications, median [IQR]	0 [0-0]	2 [1-3]	<0.001
Days from surgery to death, median [IQR]	-	10 [6-18]	-
Return to OR	1,040 (2.8%)	72 (30.4%)	<0.001
Renal failure	77 (0.2%)	55 (23.2%)	<0.001
Organ space SSI	3,019 (8.2%)	43 (18.1%)	<0.001
Pneumonia	807 (2.2%)	41 (17.3%)	<0.001
Myocardial infarction	169 (0.5%)	20 (8.4%)	<0.001
Renal insufficiency	136 (0.4%)	14 (5.9%)	<0.001
Pulmonary embolism	394 (1.1%)	13 (5.5%)	<0.001
Sepsis	1,534 (4.2%)	12 (5.1%)	0.51
CVA/stroke	53 (0.1%)	5 (2.1%)	<0.001
UTI	824 (2.2%)	5 (2.1%)	1.00
Superficial SSI	1,325 (3.6%)	4 (1.7%)	0.16
Deep SSI	285 (0.8%)	3 (1.3%)	0.44
Wound disruption	182 (0.5%)	1 (0.4%)	1.00

IQR, interquartile range; OR, operating room; SSI, surgical site infection; CVA, cerebrovascular accident; UTI, urinary tract infection.

Subgroup analysis of patients undergoing hepatectomy and pancreatectomy

Additional analyses were conducted with the machine learning algorithm that utilized factors stratified into hepatectomy and pancreatectomy variables while incorporating procedure specific variables for hepatectomy (e.g., cirrhosis, biliary stent placement, minimally invasive approach), as well as for pancreatectomy (e.g., vascular resection, pancreatic gland texture). Among 36,923 patients with low estimated risk of morbidity and mortality, 11,477 and 10,911 patients who underwent hepatectomy and pancreatectomy, respectively, had data in the Procedure Targeted module of the ACS NSQIP dataset from 2014 to 2017. Within these cohorts of patients, 70 (0.6%) and 71 (0.7%) patients experienced a UD after hepatectomy and pancreatectomy, respectively. According to the classification tree analysis, preoperative albumin level was the most

important factors to predict UD (importance: 16.9) followed by age (importance: 13.9), BMI (importance: 9.6), and preoperative biliary stent placement (importance: 8.9) among patients undergoing hepatectomy (Figure S2A). Meanwhile, age was noted to be the most important factor to predict UD (importance: 11.0), followed by preoperative albumin level (importance: 7.3), vascular resection (importance: 7.2), and bleeding disorder (importance: 6.9) (Figure S2B) among patients who underwent pancreatectomy. Of note, the c-statistic for the machine learning derived prediction model increased to 0.942 and 0.880 for hepatectomy and pancreatectomy, respectively, after incorporating procedure specific variables.

Discussion

Despite improvements in overall mortality following high-risk surgery, HP surgery remains one of the most complex set of operations with a persistent notable risk of morbidity and mortality (1,4-7,30). Prognostic tools have been created to identify individuals at risk for peri-operative complications and death, yet have largely focused on identification of patients at the highest risk of morbidity and mortality. In contrast, the current study specifically sought to delineate outcomes among patients deemed low risk. While most patients were classified as “low” risk by traditional parameters, death among this cohort of patients was not uncommon. In fact, roughly 1 in 150 patients who were deemed “low” risk by the NSQIP EP died with 30-days of an HP procedure. Estimating risk among low risk patients may be of particular interest as the likelihood of a complication may not be as anticipated by the provider or patient. In turn, any deviation from an expected “textbook” clinical course may be accompanied with decision-related regret (31,32). As such, the current study was important because it specifically sought to identify individuals who were at low risk of morbidity and mortality following HP surgery, yet died within 30-days of an HP operation. The approach to developing a prognostic model to predict UD among low risk patients was novel in that it was based on machine learning methodology. Of note, the prognostic model derived using machine learning methodology outperformed the NSQIP EP to accurately predict individuals most likely to experience a UD. In addition, there was discordance between the traditional logistic regression model and the machine learning model relative to which factors were associated with the risk of UD. Specifically, while both methods identified age as the

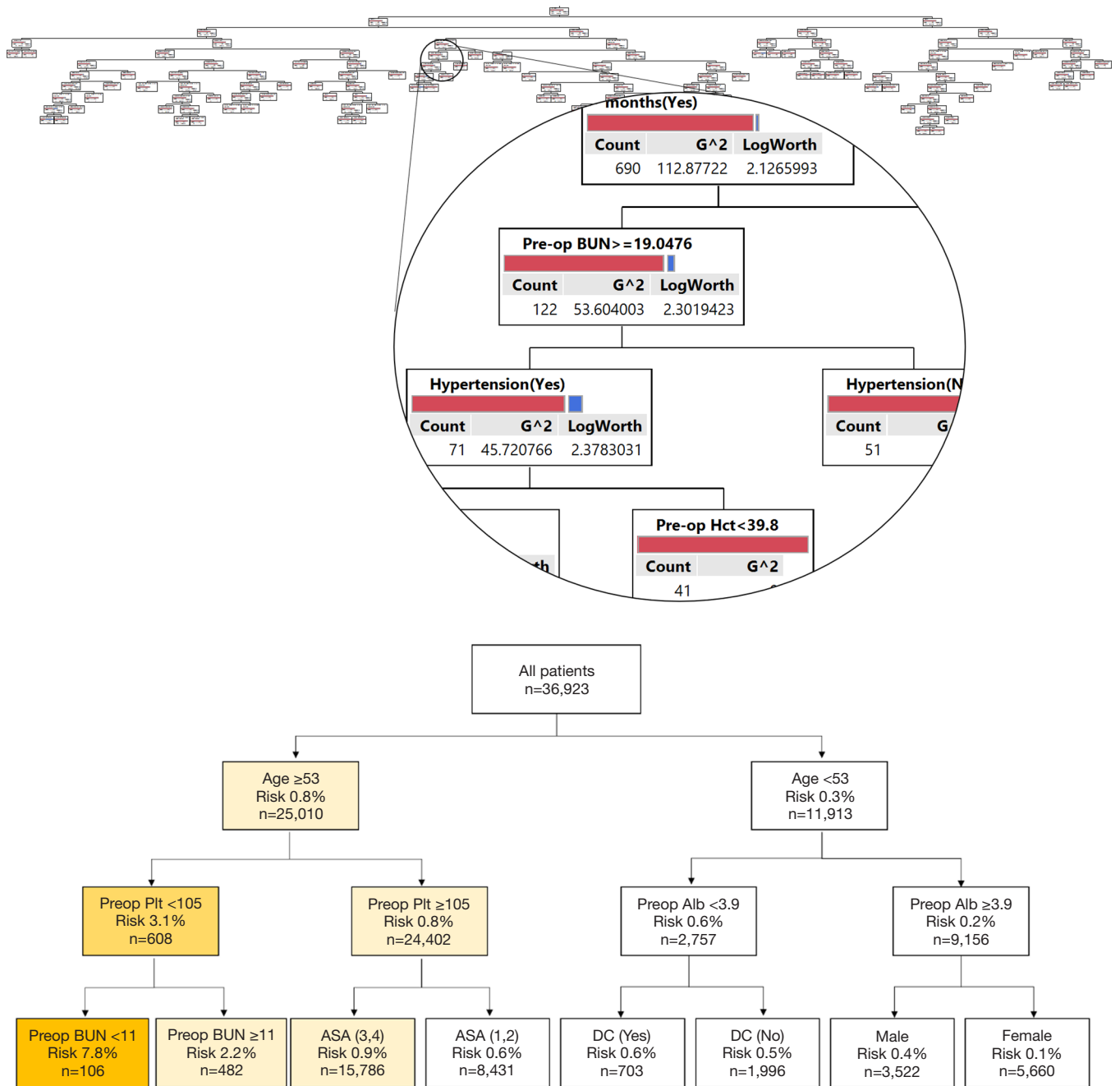


Figure 2 The classification tree models to predict UDs. (A) The final classification tree model; (B) an illustrative example of a segment of a classification-tree. UD, unpredicted death.

most important factor associated with UD, as well as noted pre-operative albumin level and patient sex to be important, machine learning also identified pre-operative platelet count and disseminated cancer to be other important factors to predict UD. In addition, subgroup that incorporated

procedure-specific variables demonstrated an increase in the c-index for the machine learning derived prediction model. Taken together, the data suggest that machine learning techniques may be better suited to build prognostic models, especially those events that may be relatively rare such as

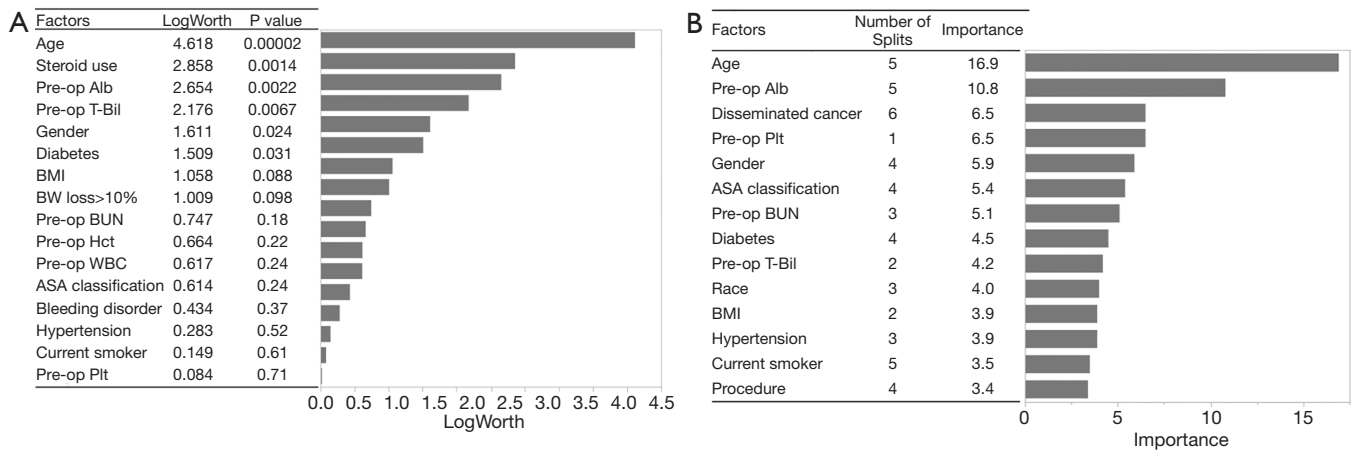


Figure 3 Relative effect of each factor to predict 30-day mortality based on: (A) the multivariable logistic regression model; (B) the classification tree model.

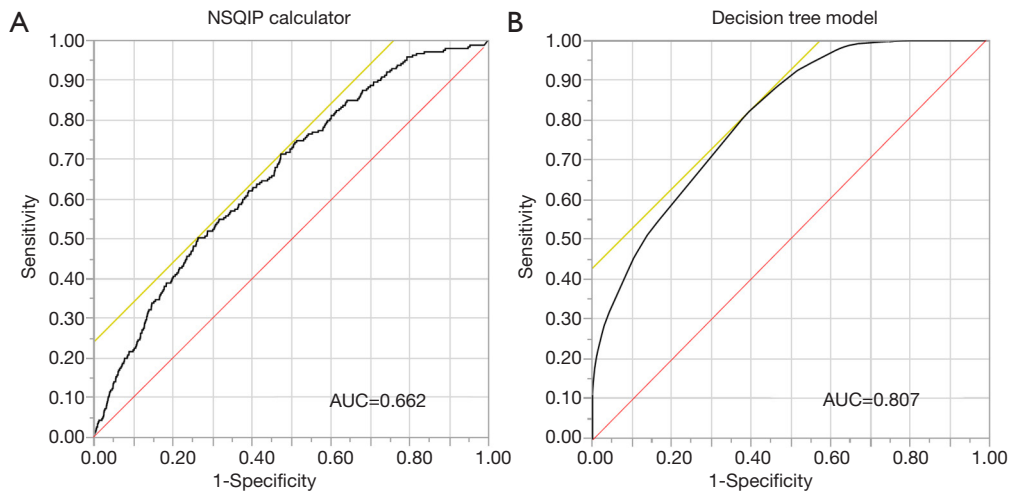


Figure 4 Receiver operating characteristic (ROC) curves to predict 30-day mortality among patients with low estimated morbidity and mortality by: (A) the NSQIP estimated probability; (B) the classification tree model. AUC, area under the curve.

death within a low-risk cohort.

Machine learning methodology has been utilized in many aspects of modern life and recently has been increasingly used in the medical setting to predict various clinical outcomes (33). For example, Gulshan *et al.* developed an algorithm base on machine learning to detect diabetic retinopathy in macula-centered retinal fundus images (34). The use of machine based learning algorithms facilitates the incorporation of “big” data, as well as the avoidance of a priori bias regarding which factors to include in the prediction model. As an example, Karadaghy and colleagues reported that a prediction model derived from

machine learning algorithms that incorporated various social, demographic, clinical and pathologic features more accurately predicted 5-year overall survival versus the traditional Tumor, Node, Metastasis staging scheme (18). Machine based learning algorithms may also superior to “expert” opinion or other types of human based prognostic models. For example, Ehteshami Bejnordi *et al.* noted that deep learning algorithms performed superior to a panel of expert pathologists to diagnose and detect lymph node metastases among women with breast cancer (35). In a separate study, Ally and colleagues reported that a machine learning model was more accurate in predicting mortality

after elective cardiac surgery versus the established EuroScore risk model (36). In the current study, using machine learning classification tree analysis, we were able to predict UD more accurately than the NSQIP EP (AUC 0.807 vs. 0.662).

Among patients undergoing HP surgery, certain patient and clinical factors can be associated with increased risk of poor outcomes and perioperative death. For example, Mayo *et al.* reported that older age, multiple medical comorbidities and larger extent of resection were associated with an increased risk of 30- and 90-day mortality (37). In a separate study, McPhee and colleagues noted that age, comorbidities such as renal failure, peripheral vascular disease, and liver disease were each associated with increased likelihood of in-hospital death following pancreatectomy (38). Interestingly, in the current study, machine learning algorithm identified age as the most important factor associated with risk of UD among low risk patients undergoing an HP procedure. Other factor included potentially modifiable factors such as preoperative albumin level and platelet count, as well as fixed variables such as sex and stage of disease. Identification and optimization of modifiable factors (e.g., optimization of preoperative nutritional status, etc.) may help improve outcomes and lessen peri-operative morbidity (39-41). In addition, information on unmodifiable factors (e.g., sex, disseminated cancer, etc.) associated with UD may help to counsel patients during the informed consent process to ensure that even low risk patients understand the chances of morbidity and mortality associated with HP surgery (9).

While most tools only assess preoperative factors when stratifying patients relative to risk, data from the current study strongly suggest that other perioperative factors strongly impact risk of death. Perhaps not surprisingly, the incidence of post-operative complications was higher among patients who experienced a UD (Table 2). In addition, one-third of patients who experienced a UD had undergone a re-operation within 30 days of initial surgery. Complications such as hemorrhage and anastomotic leak are the most often indications for reoperation following HP procedures, and early reoperation following HP surgeries can dramatically increase risk of mortality (42). Collectively, the data highlight the importance of risk re-stratification throughout the phases of surgical care. To this end, Marubashi *et al.* has proposed a “real-time” prognostic model to estimate risk of morbidity and mortality following transplantation based on preoperative variables, preoperative and intraoperative variables (43).

Several limitations should be considered when interpreting the results of the current study. Due to its retrospective

design, the current study was subject to information bias. To limit any inaccuracies and inconsistencies with data abstraction from the electronic health record, the ACS NSQIP dataset was specifically chosen due to its high-quality data aggregation methods (44). Furthermore, the generalizability of the current findings may be limited to patients receiving surgical care within one of the 718 ACS NSQIP participating hospitals. As opposed to logistic regression analysis while classification trees assigns a static risk value for each leaf (group of patients) in the tree, the construction of the tree and thus, group assignment of patients, can allow the same variable to be used multiple times within the same tree allowing for the same variable to partition patients in different ways (i.e., values) depending on the branch. As such, variables do not have their own point estimates; rather, they are used to identify clustering of patients with similar clinical presentations and are uniquely different in their outcome compared with other groups of patients. Despite the ability of classification trees to partition the data into smaller more homogenous groups, machine learning may be subject to over-fitting. As such, similar to previous studies, a 10-fold cross validation method was employed to limit over-fitting (45,46). This validation technique has been previously reported to maximize the use of the dataset for all processing stages (validation and testing) while maximizing overall model performance (47). Lastly, although the ability of classification trees was compared with the NSQIP EP, the probability has some limits in the field of HP surgery. Indeed, previous study from our group reported that the c-statistic of the NSQIP risk calculator for mortality was 0.752 and 0.633 among patients undergoing hepatectomy and pancreatectomy (13,48), respectively, which are lower compared with colorectal surgery (49). In that context, use of the machine learning method may be desired especially in the field of HP surgery.

In conclusion, using a machine learning classification tree algorithm, a prognostic model was developed that had better accuracy than the NSQIP EP to predict UD among individuals at low risk for a morbidity and mortality following HP surgery. The machine-based algorithm identified both modifiable factors (e.g., optimization of preoperative nutritional status, platelet count) and unmodifiable factors (e.g., sex, disseminated cancer) that were associated with UD. The data highlight the utility of machine learning methodology to develop prognostic tools for rare events such as UD among low risk patients. Such data may help target factors to optimize, as well as provide information to providers and patients, in the perioperative period.

Acknowledgments

Funding: None.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/hbsn.2019.11.30>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was approved by the Institutional Research Review Committee.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med* 2011;364:2128-37.
2. Endo I, Kumamoto T, Matsuyama R. Postoperative complications and mortality: Are they unavoidable? *Ann Gastroenterol Surg* 2017;1:160-3.
3. Hyder O, Dodson RM, Nathan H, et al. Influence of patient, physician, and hospital factors on 30-day readmission following pancreatoduodenectomy in the United States. *JAMA Surg* 2013;148:1095-102.
4. Tamirisa NP, Parmar AD, Vargas GM, et al. Relative Contributions of Complications and Failure to Rescue on Mortality in Older Patients Undergoing Pancreatectomy. *Ann Surg* 2016;263:385-91.
5. Merath K, Chen Q, Bagante F, et al. Synergistic Effects of Perioperative Complications on 30-Day Mortality Following Hepatopancreatic Surgery. *J Gastrointest Surg* 2018;22:1715-23.
6. Kutlu OC, Lee JE, Katz MH, et al. Open Pancreaticoduodenectomy Case Volume Predicts Outcome of Laparoscopic Approach: A Population-based Analysis. *Ann Surg* 2018;267:552-60.
7. Chen Q, Beal EW, Kimbrough CW, et al. Perioperative complications and the cost of rescue or failure to rescue in hepato-pancreato-biliary surgery. *HPB (Oxford)* 2018;20:854-64.
8. Sahara K, Paredes AZ, Tsilimigras DI, et al. Impact of Liver Cirrhosis on Perioperative Outcomes Among Elderly Patients Undergoing Hepatectomy: the Effect of Minimally Invasive Surgery. *J Gastrointest Surg* 2019;23:2346-53.
9. Childers R, Lipsett PA, Pawlik TM. Informed consent and the surgeon. *J Am Coll Surg* 2009;208:627-34.
10. Mitka M. Data-Based Risk Calculators Becoming More Sophisticated—and More Popular. *JAMA* 2009;302:730-1.
11. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg* 2013;217:833-42.e423.
12. Hyder O, Marques H, Pulitano C, et al. A nomogram to predict long-term survival after resection for intrahepatic cholangiocarcinoma: an Eastern and Western experience. *JAMA Surg* 2014;149:432-8.
13. Sahara K, Paredes AZ, Merath K, et al. Evaluation of the ACS NSQIP Surgical Risk Calculator in Elderly Patients Undergoing Hepatectomy for Hepatocellular Carcinoma. *J Gastrointest Surg* 2020;24:551-9.
14. Min L, Hall K, Finlayson E, et al. Estimating Risk of Postsurgical General and Geriatric Complications Using the VESPA Preoperative Tool. *JAMA Surg* 2017;152:1126-33.
15. Hyder O, Pulitano C, Firoozmand A, et al. A risk model to predict 90-day mortality among patients undergoing hepatic resection. *J Am Coll Surg* 2013;216:1049-56.
16. Merath K, Chen Q, Bagante F, et al. Synergistic Effects of Perioperative Complications on 30-Day Mortality Following Hepatopancreatic Surgery. *J Gastrointest Surg* 2018;22:1715-23.
17. Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. *JAMA Netw Open* 2019;2:e193721.
18. Karadaghy OA, Shew M, New J, et al. Development and Assessment of a Machine Learning Model to Help Predict Survival Among Patients With Oral Squamous Cell Carcinoma. *JAMA Otolaryngol Head Neck Surg* 2019;145:1115-20.
19. Elfiky AA, Pany MJ, Parikh RB, et al. Development and Application of a Machine Learning Approach to Assess Short-term Mortality Risk Among Patients With Cancer Starting Chemotherapy. *JAMA Netw Open* 2018;1:e180926.
20. Raval MV, Pawlik TM. Practical Guide to Surgical Data Sets: National Surgical Quality Improvement Program (NSQIP) and Pediatric NSQIP. *JAMA Surg* 2018;153:764-5.

21. Sellers MM, Merkow RP, Halverson A, et al. Validation of new readmission data in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg* 2013;216:420-7.
22. Merath K, Hyer JM, Mehta R, et al. Use of Machine Learning for Prediction of Patient Risk of Postoperative Complications After Liver, Pancreatic, and Colorectal Surgery. *J Gastrointest Surg* 2020;24:1843-51.
23. Abraham CR, Werter CR, Ata A, et al. Predictors of Hospital Readmission after Bariatric Surgery. *J Am Coll Surg* 2015;221:220-7.
24. Program ANSQL. ACS NSQIP 2016 PUF USER GUIDE 2016. Available online: <https://www.facs.org/quality-programs/acs-nsqip/participant-use>. Accessed Sep 6 2019.
25. Tan PN, Steinbach M, Karpatne A, et al. Introduction to data mining. 2019.
26. Breiman L. Classification and Regression Trees. New York: Routledge, 1984.
27. Wong T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition* 2015;48:2839-46.
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
29. Group MDAHNCWSW. Fatigue following radiation therapy in nasopharyngeal cancer survivors: A dosimetric analysis incorporating patient report and observer rating. *Radiother Oncol* 2019;133:35-42.
30. Merath K, Hyer JM, Mehta R, et al. Use of perioperative epidural analgesia among Medicare patients undergoing hepatic and pancreatic surgery. *HPB (Oxford)* 2019;21:1064-71.
31. Merath K, Chen Q, Bagante F, et al. A Multi-Institutional International Analysis of Textbook Outcomes Among Patients Undergoing Curative-Intent Resection of Intrahepatic Cholangiocarcinoma. *JAMA Surg* 2019;154:e190571.
32. Winner M, Wilson A, Ronnekleiv-Kelly S, et al. A Singular Hope: How the Discussion Around Cancer Surgery Sometimes Fails. *Ann Surg Oncol* 2017;24:31-7.
33. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA* 2018;319:1317-8.
34. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016;316:2402-10.
35. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 2017;318:2199-210.
36. Allyn J, Allou N, Augustin P, et al. A Comparison of a Machine Learning Model with EuroSCORE II in Predicting Mortality after Elective Cardiac Surgery: A Decision Curve Analysis. *PLoS One* 2017;12:e0169772.
37. Mayo SC, Shore AD, Nathan H, et al. Refining the definition of perioperative mortality following hepatectomy using death within 90 days as the standard criterion. *HPB (Oxford)* 2011;13:473-82.
38. McPhee JT, Hill JS, Whalen GF, et al. Perioperative mortality for pancreatectomy: a national perspective. *Ann Surg* 2007;246:246-53.
39. Minnella EM, Awasthi R, Loiselle SE, et al. Effect of Exercise and Nutrition Prehabilitation on Functional Capacity in Esophagogastric Cancer Surgery: A Randomized Clinical Trial. *JAMA Surg* 2018;153:1081-9.
40. Feldman LS, Carli F. From Preoperative Assessment to Preoperative Optimization of Frailty. *JAMA Surg* 2018;153:e180213.
41. Fearon KC, Jenkins JT, Carli F, et al. Patient optimization for gastrointestinal cancer surgery. *Br J Surg* 2013;100:15-27.
42. Lyu HG, Sharma G, Brovman EY, et al. Unplanned reoperation after hepatectomy: an analysis of risk factors and outcomes. *HPB (Oxford)* 2018;20:591-6.
43. Marubashi S, Ichihara N, Kakeji Y, et al. "Real-time" risk models of postoperative morbidity and mortality for liver transplants. *Ann Gastroenterol Surg* 2019;3:75-95.
44. Program ACoSNSQI. ACS NSQIP Participant Use Data File. 2017. Available online: <https://www.facs.org/quality-programs/acs-nsqip/participant-use>. Accessed Jun 15 2019.
45. Somnay YR, Craven M, McCoy KL, et al. Improving diagnostic recognition of primary hyperparathyroidism with machine learning. *Surgery* 2017;161:1113-21.
46. Goto T, Camargo CA Jr, Faridi MK, et al. Machine Learning-Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. *JAMA Netw Open* 2019;2:e186937.
47. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013;7:21.
48. Dave A, Beal EW, Lopez-Aguilar AG, et al. Evaluating the ACS NSQIP Risk Calculator in Primary Pancreatic Neuroendocrine Tumor: Results from the US Neuroendocrine Tumor Study Group. *J Gastrointest Surg* 2019;23:2225-31.
49. Hu WH, Chen HH, Lee KC, et al. Assessment of the Addition of Hypoalbuminemia to ACS-NSQIP Surgical Risk Calculator in Colorectal Cancer. *Medicine* 2016;95:e2999.

Cite this article as: Sahara K, Paredes AZ, Tsilimigras DI, Sasaki K, Moro A, Hyer JM, Mehta R, Farooq SA, Wu L, Endo I, Pawlik TM. Machine learning predicts unpredicted deaths with high accuracy following hepatopancreatic surgery. *HepatoBiliary Surg Nutr* 2021;10(1):20-30. doi: 10.21037/hbsn.2019.11.30