

Predicting the Risk of Perioperative Mortality in Patients Undergoing Pancreaticoduodenectomy

A Novel Scoring System

Raghunandan Venkat, MD, MPH; Milo A. Puhan, MD, PhD; Richard D. Schulick, MD; John L. Cameron, MD; Frederic E. Eckhauser, MD; Michael A. Choti, MD, MBA; Martin A. Makary, MD, MPH; Timothy M. Pawlik, MD, MPH; Nita Ahuja, MD; Barish H. Edil, MD; Christopher L. Wolfgang, MD, PhD

Objective: To develop and validate a risk score to predict the 30- and 90-day mortality after a pancreaticoduodenectomy or total pancreatectomy on the basis of preoperative risk factors in a high-volume program.

Design: Data from a prospectively maintained institutional database were collected. In a random subset of 70% of patients (training cohort), multivariate logistic regression was used to develop a simple integer score, which was then validated in the remaining 30% of patients (validation cohort). Discrimination and calibration of the score were evaluated using area under the receiver operating characteristic curve and Hosmer-Lemeshow test, respectively.

Setting: Tertiary referral center.

Patients: The study comprised 1976 patients in a prospectively maintained institutional database who underwent pancreaticoduodenectomy or total pancreatectomy between 1998 and 2009.

Main Outcome Measures: The 30- and 90-day mortality.

Results: In the training cohort, age, male sex, preoperative serum albumin level, tumor size, total pancreatectomy, and a high Charlson index predicted 90-day mortality (area under the curve, 0.78; 95% CI, 0.71-0.85), whereas all these factors except Charlson index also predicted 30-day mortality (0.79; 0.68-0.89). On validation, the predicted and observed risks were not significantly different for 30-day (1.4% vs 1.0%; $P = .62$) and 90-day (3.8% vs 3.4%; $P = .87$) mortality. Both scores maintained good discrimination (for 30-day mortality, area under the curve, 0.74; 95% CI, 0.54-0.95; and for 90-day mortality, 0.73; 0.62-0.84).

Conclusions: The risk scores accurately predicted 30- and 90-day mortality after pancreatectomy. They may help identify and counsel high-risk patients, support and calculate net benefits of therapeutic decisions, and control for selection bias in observational studies as propensity scores.

Arch Surg. 2011;146(11):1277-1284

Author Affiliations: Department of Surgery, The Johns Hopkins University School of Medicine (Drs Venkat, Schulick, Cameron, Eckhauser, Choti, Makary, Pawlik, Ahuja, Edil, and Wolfgang), and the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health (Drs Venkat and Puhan), Baltimore, Maryland.

THE USE OF PANCREATICO-DUODENECTOMY for the resection of periampullary tumors and proximal pancreatic neoplasms has become routine in high-volume centers. Although the perioperative mortality of this procedure has markedly decreased during the past 3 decades, it remains significant in relation to other operations of the gastrointestinal tract.¹ With improvements in surgical techniques and critical care, studies^{2,3} from our institution have reported a substantial decrease in perioperative mortality from 30% in the 1970s to as low as 1% in the 2000s, although recent population-based studies⁴ have reported higher mortality rates in lower-volume centers, ranging from 3.5% to 8.3%. Despite this improvement in out-

come, a relatively wide range of mortality rates have been reported for this operation. It is likely that patient factors, such as significant comorbidities, account for at least a portion of this variability.

See Invited Critique at end of article

Risk scores may serve various purposes. Pancreatic cancer, which is the most common diagnosis that requires a pancreatectomy, is a disease of the elderly, who may also be experiencing multiple comorbidities. The risks and benefits must be carefully assessed before subjecting these subsets of patients, as well as patients with a higher risk of perioperative mortality, to such major operations. Knowledge of the risk of morbidity and mortality after the pro-

Table 1. Classification of Patients Based on Histologic Diagnosis

Category	Diagnosis
Benign cystic lesions	Macrocytic serous cystadenoma
	Mucinous cystic neoplasm (NOS)
	Mucinous cystic neoplasm with low-, moderate-, or high-grade dysplasia (carcinoma in situ)
	IPMN with low-, moderate-, or high-grade dysplasia (carcinoma in situ)
	Intraductal tubular neoplasms with low, moderate, or high grade dysplasia (carcinoma in situ)
Periampullary tumors	Invasive ductal adenocarcinoma
	Serous or mucinous cystic neoplasms with associated invasive carcinoma
	IPMN with associated invasive carcinoma
	Ampullary, duodenal, bile duct adenocarcinoma
Pancreatic neuroendocrine tumors	Ampullary, duodenal, bile duct adenoma
	Well-differentiated pancreatic endocrine neoplasms; poorly differentiated pancreatic endocrine neoplasms

Abbreviations: IPMN, intraductal papillary mucinous neoplasms; NOS, nonspecific.

cedure that is based on preoperative risk factors in each individual patient is critical in helping the patient to make his or her decision, as well as in satisfying the adequacy of an informed consent. Also, additional therapeutic options, such as salvage procedures, prolonged monitoring in the intensive care unit, or withholding surgery, may be considered in selected patients who may have a high risk of perioperative mortality, provided this information is available at the time of decision making. The net benefits of these therapeutic interventions, in terms of reduction of the risk of perioperative mortality, can also be calculated for individual patients using risk scores. They may also be used in clinical trials to identify certain risk groups that may be considered for stratification or exclusion from the trial to minimize confounding or for covariate adjustment and subgroup analysis in the trials. In observational studies, risk scores may serve the purpose of propensity scores to minimize selection bias.

There are studies⁵⁻¹⁰ that have built predictive models to calculate the risk of perioperative mortality and/or long-term survival; however, some of these models^{5,7,8} use operative data to predict the outcome. This makes the score of limited use in preoperative patient counseling and risk stratification. Other models^{6,9} that are based on national population-based data may be inappropriate to use in the patient population of a tertiary referral institution if the range of the patient profile differs from one institution to another. Also, limited studies have incorporated preoperative laboratory parameters, such as albumin and creatinine levels, or characteristics of the tumor, such as tumor size or histologic diagnosis, which may play an important role in predicting the perioperative mortality with greater accuracy and make the score more specific to patient or tumor characteristics.

Our objective was to develop and validate a simple and easily applicable score based on readily available preop-

erative parameters to predict perioperative mortality in patients scheduled for pancreatectomy.

METHODS

PATIENT SELECTION AND DATA ACQUISITION

The study population (n = 1976) consisted of adult patients (≥18 years old) admitted to The Johns Hopkins Hospital and who subsequently underwent total pancreatectomy or pancreaticoduodenectomy (classic or pylorus-preserving) from January 1, 1998, through June 30, 2009. We included patients from 1998 to 2009 to ensure a contemporary patient population and to account for a potential decreasing trend in perioperative mortality for pancreatectomy over time.¹¹ We excluded patients who underwent distal pancreatectomy because there was only 1 death within 90 days, out of a total of 209 cases, for this procedure (0.48% mortality). In patients who underwent subsequent procedures, only data from their primary surgery were included. We restricted our analysis to patients with a histologic diagnosis of benign cystic lesions, periampullary tumors, or neuroendocrine tumors (**Table 1**). All data were obtained from an institutional review board–approved pancreatectomy database that comprises prospectively collected patient data from January 1970 to June 2009 at The Johns Hopkins Hospital. The primary outcomes of 30- and 90-day mortality, defined as all-cause mortality occurring within 30 and 90 days of the date of surgery, respectively, were assessed from this institutional database and confirmed using the National Social Security Death Index. Information on patient comorbidities was available from the Hospital billing data, and we converted the *International Classification of Diseases, Ninth Revision (ICD-9)* codes obtained from this database for each patient admission to a Charlson comorbidity index.¹²

SELECTION OF PREDICTORS OF MORTALITY

The predictors of mortality were selected a priori on the basis of clinical usefulness and biological plausibility. To facilitate the use of the model in clinical practice and to ensure building stable models, we chose to restrict the maximum number of predictive parameters to 10 and included only those factors that were readily available preoperatively. Variables included were age, sex, race, histologic diagnosis, type of surgery, preoperative serum albumin and serum creatinine levels, tumor size, and Charlson index. The covariates age, Charlson index, and albumin level were modeled as continuous variables, whereas sex, tumor size, creatinine level, histologic diagnosis, and type of surgery were categorical. Data on tumor size were obtained from surgical pathological evaluation; although it was obtained postoperatively, we included it as a predictor because previous studies^{13,14} have shown a close agreement between tumor size at computed tomography examination and pathological evaluation.

STATISTICAL ANALYSIS

Standard approaches to developing and validating prediction scores were followed.¹⁵⁻¹⁷ We used split-sample cross-validation in which a random sample of 70% of the study population was used to develop the multivariate model (training data set) and the remaining 30% of the study population was used for validating the model (validation data set). In the training data set, we fitted separate multivariable logistic regression models using backward stepwise selection for 30- and 90-day perioperative mortality. We entered all variables that were associated with the outcomes in univariate analyses (using the χ^2 test

Table 2. Characteristics of the Study Population

Variable	No. (%)	30-Day Mortality		90-Day Mortality	
		No. (%)	P Value	No. (%)	P Value
Patients	1976 (100)	26 (1.3)		73 (3.7)	
Age, mean (SD), y	66 (12)	74 (10) ^a	<.001	74 (10) ^a	<.001
≤60	588 (29.8)	4 (0.7)	.03 ^b	7 (1.2)	<.001
61-70	596 (30.2)	5 (0.8)		14 (2.3)	
71-80	579 (29.3)	10 (1.7)		32 (5.5)	
>80	213 (10.8)	7 (3.3)		20 (9.4)	
Sex			.22		.02
Female	920 (46.6)	9 (1.0)		24 (2.6)	
Male	1056 (53.4)	17 (1.6)		49 (4.6)	
Race			.57 ^b		.24
White	1694 (85.7)	24 (1.4)		66 (3.9)	
Nonwhite	282 (14.3)	2 (0.7)		7 (2.5)	
Surgery type			.04 ^b		.02
Pancreaticoduodenectomy	1876 (94.9)	22 (1.2)		65 (3.5)	
Total pancreatectomy	100 (5.1)	4 (4.0)		8 (8.0)	
Histologic diagnosis			.24 ^b		.12 ^b
Benign cystic lesions	232 (11.7)	1 (0.4)		3 (1.3)	
Neuroendocrine tumors	123 (6.2)	3 (2.4)		5 (4.1)	
Periapillary tumors	1621 (82.0)	22 (1.4)		65 (4.0)	
Charlson comorbidity index, median (range)	6 (0-11)	6 (0-10)	.45 ^c	6 (0-10)	.008 ^c
0	207 (10.5)	1 (0.5)	.80 ^b	3 (1.4)	.01 ^b
1-3	406 (20.5)	5 (1.2)		10 (2.5)	
4-6	831 (42.1)	12 (1.4)		29 (3.5)	
≥7	532 (26.9)	8 (1.5)		31 (5.8)	
Tumor size, cm			.009		.008
<3	1033 (52.3)	7 (0.7)		27 (2.6)	
≥3	943 (47.7)	19 (2.0)		46 (4.9)	
Serum albumin level, mean (SD), g/dL	2.9 (0.7)	2.6 (0.7) ^a	.02	2.5 (0.7) ^a	<.001
≥3.5	312 (15.8)	2 (0.6)	.14 ^b	7 (2.2)	<.001
2.5-3.4	1163 (58.9)	13 (1.1)		31 (2.7)	
<2.5	501 (25.4)	11 (2.2)		35 (7.0)	
Serum creatinine level, mean (SD), mg/dL	0.7 (0.4)	0.8 (0.3) ^a	.19	0.8 (0.3) ^a	.30
<1.2	1871 (94.7)	23 (1.2)	.16 ^b	66 (3.5)	.10
≥1.2	105 (5.3)	3 (2.9)		7 (6.7)	

SI conversions: To convert serum albumin to grams per liter, multiply by 10; creatinine to micromoles per liter, multiply by 88.4.

^aMean (SD) for continuous variables with P values from unpaired t test.

^bFisher exact test.

^cWilcoxon rank sum test.

for categorical variables and the t test for continuous variables) with $P \leq .25$. Only variables that remained significant predictors of the outcome in the multivariate model at a significance level of $P \leq .25$ on the basis of the likelihood ratio test¹⁸ or that changed the model's discrimination by at least 1% were included in the final model. Clinically related variables were examined for interaction. All predictor variables were assessed for correlations.

Development and validation of the prediction model involved assessing the discrimination and calibration of the model. *Discrimination* is the ability to distinguish between those who survive and those who die and was estimated using the concordance index, also known as the C statistic or the area under the receiver operating characteristic curve (AUC). The AUCs of the different models were evaluated in the training and the validation data sets. The final model that was subsequently converted into the prediction score was thus developed based on a combination of AUC and likelihood ratio test. To reduce the variability of the error estimate and to control for overfitting, we used bootstrapping methods with 1000 resamples.¹⁹ Additional internal validation of the final model was done using the jackknifing method,¹⁹ in which we systematically estimated the outcome by leaving out 1 observation at a time from the total patient sample. The average 30- and 90-day mortality in these

subsets was then compared with that from the entire sample to estimate the bias of the latter. Discrimination and goodness of fit of the model were then assessed using this method.

We used shrinkage to limit the risks of the logistic regression model overestimating the associations of the predictors with the outcomes of 30- and 90-day mortality and to improve the accuracy of the model in other populations. The shrinkage factor represents a constant with which the regression coefficients of the predictor variables are multiplied. The magnitude of the factor depends on the number of predictors and the goodness of fit of the model and reflects the overoptimism of the multivariate model.²⁰

It is critical for prognostic prediction models to be validated in a separate set of patients before being used in clinical practice.¹⁵⁻¹⁷ This ensures that the predictions of the outcome based on the risk factors are reliable and accurate and thus can guide treatment decisions. To assess the calibration of our model, we used the Hosmer-Lemeshow goodness of fit test.²¹ Using this test, we compared the predicted and observed risks for 30- and 90-day mortality across all risk classes and assessed whether they differed significantly from one another.

Finally, to ensure its utility in clinical practice, the underlying regression equations were converted into a simplified integer score following an established approach used for the

Table 3. Characteristics of the Training and Validation Cohorts^a

Variable	Training Data Set (n=1383)			Validation Data Set (n=593)		
	No. (%)	30-Day Mortality	90-Day Mortality	No. (%)	30-Day Mortality	90-Day Mortality
Age, mean (SD), y	66 (12.0)			66 (11.0)		
≤60	424 (30.7)	3 (0.7)	6 (1.4)	164 (27.7)	1 (0.6)	1 (0.6)
61-70	398 (28.8)	2 (0.5)	10 (2.5)	198 (33.4)	3 (1.5)	4 (2.0)
71-80	404 (29.2)	10 (2.5)	23 (5.7)	175 (29.5)	0	9 (5.1)
>80	157 (11.4)	5 (3.2)	14 (8.9)	56 (9.4)	2 (3.6)	6 (10.7)
Sex						
Female	642 (46.4)	7 (1.1)	18 (2.8)	278 (46.9)	2 (0.7)	6 (2.2)
Male	741 (53.6)	13 (1.8)	35 (4.7)	315 (53.1)	4 (1.3)	14 (4.4)
Race						
White	1177 (85.1)	18 (1.5)	47 (4.0)	517 (87.2)	6 (1.2)	19 (3.7)
Nonwhite	206 (14.9)	2 (1.0)	6 (2.9)	76 (12.8)	0	1 (1.3)
Surgery type						
Pancreaticoduodenectomy	1305 (94.4)	17 (1.3)	46 (3.5)	571 (96.3)	5 (0.9)	19 (3.3)
Total pancreatectomy	78 (5.6)	3 (3.8)	7 (9.0)	22 (3.7)	1 (4.5)	1 (4.5)
Histologic diagnosis						
Benign cystic lesions	163 (11.8)	1 (0.6)	3 (1.8)	69 (11.6)	0	0
Neuroendocrine tumors	90 (6.5)	2 (2.2)	4 (4.4)	33 (5.6)	1 (3.0)	1 (3.0)
Periapillary tumors	1130 (81.7)	17 (1.5)	46 (4.1)	491 (82.8)	5 (1.0)	19 (3.9)
Charlson comorbidity index						
0	139 (10.0)	1 (0.7)	3 (2.2)	68 (11.5)	0	0
1-3	288 (20.8)	2 (0.7)	5 (1.7)	118 (19.9)	3 (2.5)	5 (4.2)
4-6	590 (42.7)	10 (1.7)	20 (3.4)	241 (40.6)	2 (0.8)	9 (3.7)
≥7	366 (26.5)	7 (1.9)	25 (6.8)	166 (28.0)	1 (0.6)	6 (3.6)
Tumor size, cm						
<3	731 (52.9)	6 (0.8)	19 (2.6)	302 (50.9)	1 (0.3)	8 (2.6)
≥3	652 (47.1)	14 (2.1)	34 (5.2)	291 (49.1)	5 (1.7)	12 (4.1)
Serum albumin level, mean (SD), g/dL	2.9 (0.7)			2.8 (0.6)		
≥3.5	229 (16.6)	1 (0.4)	4 (1.7)	83 (14.0)	1 (1.2)	3 (3.6)
2.5-3.4	806 (58.3)	10 (1.2)	22 (2.7)	357 (60.2)	3 (0.8)	9 (2.5)
<2.5	348 (25.2)	9 (2.6)	27 (7.8)	153 (25.8)	2 (1.3)	8 (5.2)
Serum creatinine level, mean (SD), mg/dL	0.7 (0.5)			0.7 (0.3)		
<1.2	1312 (94.9)	18 (1.4)	49 (3.7)	559 (94.3)	5 (0.9)	17 (3.0)
≥1.2	71 (5.1)	2 (2.8)	4 (5.6)	34 (5.7)	1 (2.9)	3 (8.8)

SI conversions: To convert serum albumin to grams per liter, multiply by 10; creatinine to micromoles per liter, multiply by 88.4.

^aData are given as number (percentage) unless otherwise indicated.

Framingham and the ADO (Age, Dyspnea, and airflow Obstruction) risk score.^{22,23} This method involved transformation of the regression coefficients into points that reflect their strengths of associations in which reference categories were assigned a value of zero. The theoretical range of points was calculated by summing up the points for each predictor. For each total score within the range of points, a specific risk for 30- and 90-day mortality, depending on the predictor variables for each outcome, was calculated using the regression equation. All the statistical analysis was performed using Intercooled Stata, version 11.0 (StataCorp, College Station, Texas).

RESULTS

DEMOGRAPHIC CHARACTERISTICS

Most of the patients had multiple comorbidities, with 75.0% (n=1482) of the patient population having a Charlson index of 3 or higher (**Table 2**). A randomly selected sample of 70% of the total study cohort were a part of the training data set (n=1383), and the remaining 30% were a part of the validation data set (n=593). The training and the validation cohorts of patients were seen to

be comparable in terms of demographic parameters (**Table 3**).

DEVELOPMENT OF THE PREDICTION MODEL

Of the 1383 patients randomly selected for building the prediction model, 20 patients (1.5%) and 53 patients (3.8%) had died within 30 and 90 days of surgery, respectively. Age, sex, tumor size, type of surgery, and preoperative serum albumin levels were predictors of 30-day mortality and age, sex, tumor size, Charlson index, type of surgery, and preoperative serum albumin levels were predictors of 90-day mortality (**Table 4**). Histologic diagnosis and creatinine levels were not included in the final model because neither were they significantly associated with the outcomes nor did they improve the predictive accuracy of the model. The Charlson index, although a predictor of 90-day mortality, failed to predict the 30-day perioperative mortality per our criteria. The AUCs of the 2 models in the training cohort were 0.79 (95% CI, 0.68-0.89) and 0.78 (0.71-0.85) for

Table 4. Multivariable Logistic Regression Model and Development of the Prediction Scores

Predictor, Category	30-Day Mortality				90-Day Mortality			
	Regression Coefficients $\beta^{a,b}$	Odds Ratio	P Value	Risk Score ^c	Regression Coefficients $\beta^{a,b}$	Odds Ratio	P Value	Risk Score ^c
Age, y			.002				.001	
≤ 60	Reference	Reference		0	Reference	Reference		0
61-70	1.098	2.99		4	0.995	2.70		4
71-80	1.708	5.52		6	1.548	4.70		6
≥ 81	2.258	9.56		7	2.048	7.75		7
Sex			.26				.053	
Female	Reference	Reference		0	Reference	Reference		0
Male	0.428	1.53		1	0.526	1.69		2
Tumor size, cm			.06				.053	
< 3	Reference	Reference		0	Reference	Reference		0
≥ 3	0.739	2.09		2	0.580	1.79		2
Serum albumin level, g/dL			.11				.004	
≥ 3.5	Reference	Reference		0	Reference	Reference		0
2.5-3.4	0.488	1.63		2	0.769	2.16		3
< 2.5	0.976	2.65		3	1.537	4.65		6
Type of surgery			.17				.22	
Pancreaticoduodenectomy	Reference	Reference		0	Reference	Reference		0
Total pancreatectomy	0.727	2.07		2	0.565	1.76		2
Charlson index	NA	NA	NA	NA			.07	
0					Reference	Reference		0
1-3					0.249	1.28		1
4-6					0.620	1.86		2
≥ 7					1.116	3.05		4

Abbreviation: NA, not applicable.

SI conversion: To convert serum albumin to grams per liter, multiply by 10.

^aCompared with reference category, that is, ≤ 60 y, female sex, tumor size < 3 cm, Charlson index of 0, albumin ≥ 3.5 g/dL, and type of surgery: pancreaticoduodenectomy, respectively. Constant (β_0) = -8.985 for 30-day mortality and -6.908 for 90-day mortality.

^bRegression coefficients multiplied by shrinkage factor: 0.79 (30-day mortality) and 0.86 (90-day mortality).

^cOne point was assigned for the increase in risk of mortality per 5-year increase in age, which is a coefficient (reference category) of 0.305 for 30-day mortality and 0.277 for 90-day mortality scores. Points were then rounded to the nearest integer.

30- and 90-day mortality, respectively, which indicated a good discrimination of patients who survived and died within the 30- and 90-day postoperative period. The point score derived from the regression model ranged from 0 to 15 for 30-day mortality, and from 0 to 23 for 90-day mortality, with a score of 0 corresponding to the lowest risk of both outcomes. A score of 15 corresponded to the highest risk of 30-day mortality (2.9%) and a score of 23 to the highest risk for 90-day mortality (26.7%) (**Figure 1** and **Figure 2** and **Table 5**).

VALIDATION OF THE PREDICTION MODEL

The validation population comprised the remaining 593 patients. In this cohort, the predicted risk for 30-day mortality was 1.4%, whereas the observed risk was 1.0%. The predicted and observed 90-day mortality was 3.8% and 3.4%, respectively. The Hosmer-Lemeshow test confirmed that there were no statistically significant differences between observed and expected 30-day ($P = .62$) and 90-day ($P = .87$) mortality across risk groups. The AUCs in the validation cohort were 0.74 (95% CI, 0.54-0.95) and 0.73 (0.62-0.84) for 30- and 90-day mortality, respectively, indicating a good discrimination in the validation cohort as well (**Figure 3** and **Figure 4**). We used the final model and validated it using the jackknifing method. The model showed a good predictive accuracy, consistent with the split-sample method of validation for 30-day mortality (AUC,

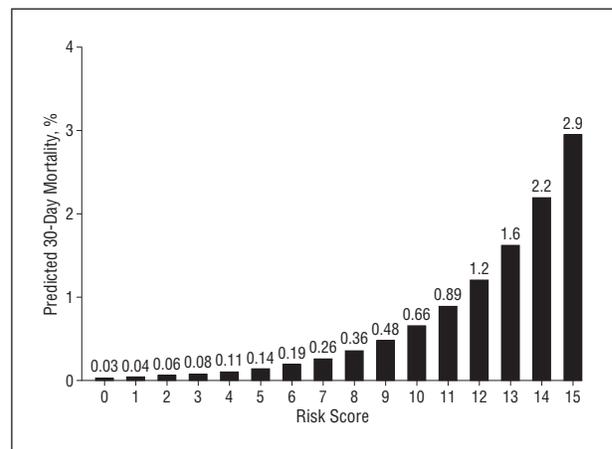


Figure 1. Integer risk score and corresponding predicted probability of 30-day mortality.

0.72; Hosmer-Lemeshow test, $P = .36$) and 90-day mortality (AUC, 0.75; Hosmer-Lemeshow test, $P = .09$).

COMMENT

We have developed and validated 2 integer scores to predict the perioperative mortality within 30 and 90 days after pancreatectomy for patients with benign and malignant tumors. The scores, based on 5 parameters for

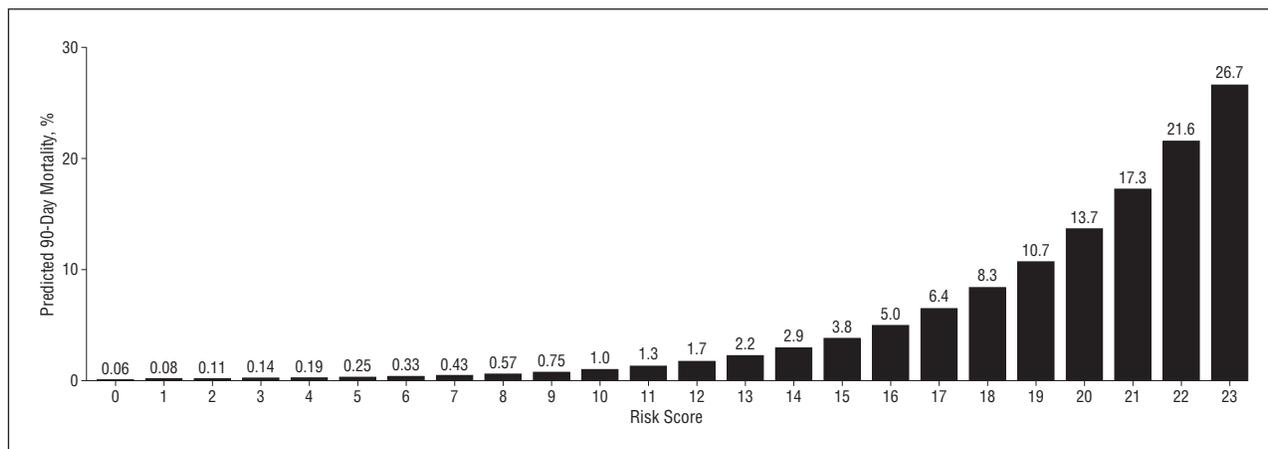


Figure 2. Integer risk score and corresponding predicted probability of 90-day mortality.

Table 5. Total Integer Risk Score and Corresponding Predicted Probability of 30- and 90-Day Mortality

Risk Score	Predicted Probability (95% CI)	
	30-Day Mortality	90-Day Mortality
0	0.03 (0.01-0.05)	0.06 (0.02-0.09)
1	0.04 (0.02-0.06)	0.08 (0.04-0.11)
2	0.06 (0.03-0.08)	0.11 (0.07-0.14)
3	0.08 (0.05-0.11)	0.14 (0.11-0.18)
4	0.11 (0.09-0.14)	0.19 (0.15-0.25)
5	0.14 (0.10-0.18)	0.25 (0.20-0.29)
6	0.19 (0.16-0.23)	0.33 (0.30-0.37)
7	0.26 (0.18-0.33)	0.43 (0.37-0.49)
8	0.36 (0.29-0.41)	0.57 (0.50-0.64)
9	0.48 (0.40-0.55)	0.75 (0.69-0.86)
10	0.66 (0.56-0.70)	1.00 (0.85-1.14)
11	0.89 (0.72-0.98)	1.30 (1.17-1.45)
12	1.20 (1.05-1.30)	1.70 (1.50-1.96)
13	1.60 (1.33-1.85)	2.20 (2.01-2.48)
14	2.20 (2.01-2.49)	2.90 (2.57-3.25)
15	2.90 (2.55-3.50)	3.80 (3.36-4.30)
16		5.00 (4.62-5.39)
17		6.40 (5.80-7.46)
18		8.30 (7.78-9.23)
19		10.70 (9.70-11.52)
20		13.70 (12.03-15.57)
21		17.30 (15.91-19.10)
22		21.60 (19.71-24.02)
23		26.70 (24.52-29.10)

30-day mortality and 6 parameters for 90-day mortality, showed good discrimination and calibration.

Based on these scores, the perioperative risk can be assessed. For instance, an 82-year-old male patient with a 3.5-cm tumor, a serum albumin level of 2.7 g/dL (to convert serum albumin to grams per liter, multiply by 10), and a Charlson index of 5 scheduled for a total pancreatectomy would have a predicted risk of 30-day post-operative mortality of 2.2% (risk score of 14; 95% CI, 2.01-2.49) irrespective of his Charlson index. On the other hand, his risk of 90-day perioperative mortality would be 8.3% (risk score of 18; 95% CI, 7.78-9.23). This may also suggest that the 90-day mortality may be a better benchmark to judge the outcome of the procedure than

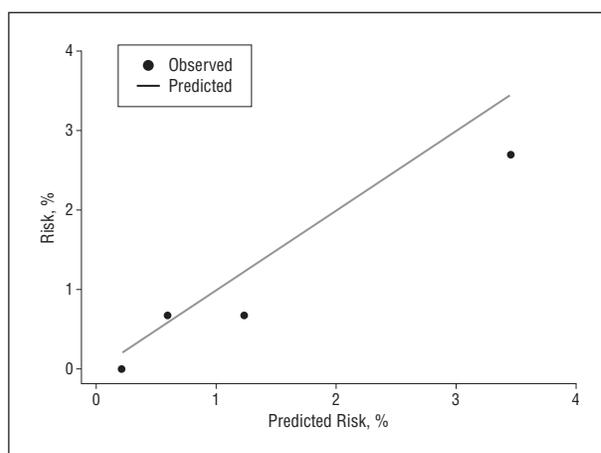


Figure 3. Calibration plot comparing observed and predicted probabilities of 30-day mortality (area under the receiver operating characteristic curve, 0.74; Hosmer-Lemeshow, $P=.62$) in the validation cohort.

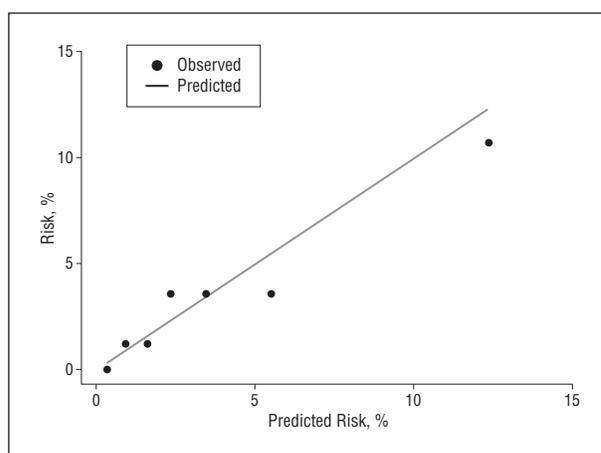


Figure 4. Calibration plot comparing observed and predicted probabilities of 90-day mortality (area under the receiver operating characteristic curve, 0.73; Hosmer-Lemeshow, $P=.87$) in the validation cohort.

the 30-day mortality, as observed by the differences in the risk of perioperative mortality. For the same patient, the risk for death in the 90-day perioperative period is more than 2-fold compared with the 30-day mortality. Thus, for a complete risk assessment, more emphasis

Table 6. Characteristics of Studies That Predict Perioperative Mortality After Pancreatic Resection

Source	Patient Selection and Data Source	Outcome	Validation Technique (C Statistic in Validation Cohort)	Predictive Factors
Present study	Prospective data from a tertiary referral center, n = 1976, benign and malignant tumors	30- and 90-d mortality	70%-30% random split-sample cross-validation (0.74, 0.73) ^a ; jackknife (0.72, 0.75) ^a	30-d mortality: age, sex, tumor size, type of surgery, and serum albumin level; 90-d mortality: age, sex, tumor size, Charlson comorbidity index, type of surgery, and serum albumin level
Hill et al, ⁶ 2010	NIS, n = 7143, only malignant tumors	In-hospital death	80%-20% random split-sample cross-validation (0.74)	Age, sex, Charlson comorbidity index, type of surgery, and hospital volume
Kelly et al, ¹⁰ 2011	ACS-NSQIP, n = 2322, only distal pancreatectomy	30-d mortality	80%-20% random split-sample cross-validation (0.79)	Esophageal varices, neurologic disease, dependent functional status, recent weight loss, alkaline phosphatase level, and blood urea nitrogen level
Are et al, ⁹ 2009	NIS, n = 5481, only malignant tumors	In-patient mortality	Temporal split-sample cross-validation (0.76)	Age, sex, type of surgery, renal failure, neurologic disease, hypothyroid, CHF, liver disease, elective surgery, HTN, hospital type/size, arrhythmia, COPD

Abbreviations: ACS-NSQIP, American College of Surgeons National Surgical Quality Improvement Program; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; HTN, hypertension; NIS, Nationwide Inpatient Sample.

^aC statistic for 30- and 90-d mortality, respectively.

should be given to the 90-day mortality rate rather than the 30-day mortality rate.

Risk scores have the utility of combining information available at a stage of clinical evaluation where critical decisions may have to be made. The main goal of our scores is the preoperative evaluation of postoperative mortality. Each patient's risk for perioperative mortality can be calculated before surgery so that interventions, such as improvements in patient care, may be directed in high-risk patients during the entire perioperative period. In addition, patients who have a higher risk of mortality on the basis of the score may be provided this information, thus helping them better understand the risks of going through their surgery as well as satisfying the adequacy of an informed consent. This makes our score more useful in preoperative risk stratification and patient counseling compared with the surgical Apgar and the POSSUM (Physiologic and Operative Severity Score for the enUmeration of Mortality and Morbidity) scores,^{5,7} which also include intraoperative factors. Although intraoperative parameters, such as blood loss or vein resection, may still be considered to update the preoperative risk prediction and to further support the decision making for the postoperative management, such a score would again require development and validation. We have also incorporated important independent predictors of mortality, such as preoperative serum albumin level and tumor size, that may enhance the predictive accuracy of our score compared with similar studies in the past^{6,9} (**Table 6**).

Strengths of this study include a large sample size that allowed us to develop and validate the prediction score. We have used advanced statistical techniques, such as bootstrapping and shrinkage, which may increase the applicability and validity of the score in other related settings. We have used parameters of accuracy, such as discrimination and calibration. Assessment of these parameters is indispensable before the use of a risk score in clinical practice. We have also used 2 separate validation techniques, split-sample cross-validation and jackknifing methods, to

assess the performance of the model. Our study uses institutional data rather than administrative data and hence would not be associated with the potential risks and limitations of using claims and registry data.²⁴

A potential limitation of our study may be that the patient cohort of a single center was used for the development of the model. Because it was a high-volume tertiary referral hospital, the characteristics of the patients may be different from those observed in lower-volume centers. The results of the score should thus be interpreted with caution in centers that may not receive a similar patient population. However, this does not mean that our scores would not perform adequately in other populations, but they may need some simple recalibration of, for example, the risk of mortality. We have used the established correlations between postoperative pathologic findings and preoperative imaging techniques^{13,14} to obtain data on tumor size and used it as a proxy for assessing the strength of the association between tumor size and perioperative mortality. Further validation with data on tumor size using preoperative computed tomography may be warranted to test the performance of the scores and to update them, if required.

Our scores also need to be validated and updated in other populations before they are widely used. The range of the patient profiles and the incidence of perioperative mortality may be different in our patient population compared with other institutions. This has been the case for the widely used Framingham risk score and the ADO risk score, which have been updated for populations differing in the incidence of the outcome.^{22,25} A multi-institutional validation would further strengthen the external validity of the score. It would also be interesting to study the impact of other preoperative, intraoperative, and postoperative risk factors on the risk of mortality as well as morbidity and to update future prediction models to incorporate these parameters. This would help target specific interventions based on the risk stratification to prevent these outcomes as well as to adequately counsel the patients at each stage of therapy.

In conclusion, we have developed and validated a risk score to accurately predict the 30- and 90-day perioperative mortality in patients undergoing pancreatotomy. The risk score will be useful in identifying patients at high risk for perioperative mortality on the basis of simple and easily obtained preoperative risk factors. This information will help in supporting important therapeutic decisions as well as assist in the realization of interventions to improve patient care in high-risk individuals and to calculate their net benefits. The risk scores may be used for risk stratification in clinical trials and, as propensity scores, may be useful in controlling for selection bias in observational studies. The score will also be beneficial in making individual patients better understand the risks of surgery based on their preoperative characteristics and augment the adequacy of an informed consent.

Accepted for Publication: May 16, 2011.

Correspondence: Christopher L. Wolfgang, MD, PhD, Department of Surgery, Johns Hopkins Hospital, 600 N Wolfe St, Osler 624, Baltimore, MD 21287 (cwolfga2@jhmi.edu).

Author Contributions: *Study concept and design:* Venkat, Puhon, Schulick, Cameron, Eckhauser, Makary, Ahuja, Edil, and Wolfgang. *Acquisition of data:* Venkat, Choti, Pawlik, and Wolfgang. *Analysis and interpretation of data:* Venkat, Puhon, Pawlik, and Wolfgang. *Drafting of the manuscript:* Venkat, Puhon, Schulick, Eckhauser, Makary, Ahuja, Edil, and Wolfgang. *Critical revision of the manuscript for important intellectual content:* Venkat, Puhon, Cameron, Eckhauser, Choti, Pawlik, Ahuja, Edil, and Wolfgang. *Statistical analysis:* Venkat, Puhon, Pawlik, and Wolfgang. *Obtained funding:* Wolfgang. *Administrative, technical, and material support:* Venkat, Schulick, Makary, and Wolfgang. *Study supervision:* Venkat, Puhon, Schulick, Cameron, Eckhauser, Choti, Makary, Pawlik, Ahuja, Edil, and Wolfgang. *Financial Disclosure:* None reported.

REFERENCES

1. Bilimoria KY, Bentrem DJ, Feinglass JM, et al. Directing surgical quality improvement initiatives: comparison of perioperative mortality and long-term survival for cancer surgery. *J Clin Oncol*. 2008;26(28):4626-4633.
2. Winter JM, Cameron JL, Campbell KA, et al. 1423 Pancreaticoduodenectomies for pancreatic cancer: a single-institution experience. *J Gastrointest Surg*. 2006;10(9):1199-1210.
3. Cameron JL, Riall TS, Coleman J, Belcher KA. One thousand consecutive pancreaticoduodenectomies. *Ann Surg*. 2006;244(1):10-15.
4. Simons JP, Shah SA, Ng SC, Whalen GF, Tseng JF. National complication rates after pancreatotomy: beyond mere mortality. *J Gastrointest Surg*. 2009;13(10):1798-1805.
5. Pratt W, Joseph S, Callery MP, Vollmer CM Jr. POSSUM accurately predicts morbidity for pancreatic resection. *Surgery*. 2008;143(1):8-19.
6. Hill JS, Zhou Z, Simons JP, et al. A simple risk score to predict in-hospital mortality after pancreatic resection for cancer. *Ann Surg Oncol*. 2010;17(7):1802-1807.
7. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg*. 2007;204(2):201-208.
8. Brennan MF, Kattan MW, Klimstra D, Conlon K. Prognostic nomogram for patients undergoing resection for adenocarcinoma of the pancreas. *Ann Surg*. 2004;240(2):293-298.
9. Are C, Afuh C, Ravipati L, Sasson A, Ullrich F, Smith L. Preoperative nomogram to predict risk of perioperative mortality following pancreatic resections for malignancy. *J Gastrointest Surg*. 2009;13(12):2152-2162.
10. Kelly KJ, Greenblatt DY, Wan Y, et al. Risk stratification for distal pancreatotomy utilizing ACS-NSQIP: preoperative factors predict morbidity and mortality. *J Gastrointest Surg*. 2011;15(2):250-261.
11. McPhee JT, Hill JS, Whalen GF, et al. Perioperative mortality for pancreatotomy: a national perspective. *Ann Surg*. 2007;246(2):246-253.
12. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
13. Bluemke DA, Cameron JL, Hruban RH, et al. Potentially resectable pancreatic adenocarcinoma: spiral CT assessment with surgical and pathologic correlation. *Radiology*. 1995;197(2):381-385.
14. Zeman RK, Cooper C, Zeiberg AS, et al. TNM staging of pancreatic carcinoma using helical CT. *AJR Am J Roentgenol*. 1997;169(2):459-464.
15. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. *JAMA*. 1997;277(6):488-494.
16. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.
17. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586.
18. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604. doi:10.1136/bmj.b604.
19. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. *Ann Stat*. 1986;14(4):1261-1295.
20. Copas JB. Using regression models for prediction: shrinkage and regression to the mean. *Stat Methods Med Res*. 1997;6(2):167-183.
21. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982;115(1):92-106.
22. Puhon MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet*. 2009;374(9691):704-711.
23. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631-1660.
24. Nathan H, Pawlik TM. Limitations of claims and registry data in surgical oncology research. *Ann Surg Oncol*. 2008;15(2):415-423.
25. Borden WB, Davidson MH. Updating the assessment of cardiac risk: beyond Framingham. *Rev Cardiovasc Med*. 2009;10(2):63-71.

INVITED CRITIQUE

Personalizing Surgical Risk

“To Be or Not to Be” Should Not Be the Question

Predictive models to calculate individual surgical mortality are a critical element of “personalized surgery.” Risk modeling is particularly suited to technically challenging operations with a narrow therapeutic index, such as radical pancreaticoduodenectomy

for pancreatic cancer, and will transform surgery from a “practice” to reproducible performance.

The current report by Venkat et al¹ objectifies an emotionally charged “go, no go” decision for patients and surgeons dealing with pancreatic cancer and prom-