Medicine®

OBSERVATIONAL STUDY

OPEN

Comparison of National Operative Mortality in Gastroenterological Surgery Using Web-based Prospective Data Entry Systems

Takayuki Anazawa, MD, PhD, Jennifer L. Paruch, MD, Hiroaki Miyata, PhD, Mitsukazu Gotoh, MD, PhD, Clifford Y. Ko, MD, MS, MSHS, Mark E. Cohen, PhD, Norimichi Hirahara, PhD, Lynn Zhou, Hiroyuki Konno, MD, PhD, Go Wakabayashi, MD, PhD, Kenichi Sugihara, MD, PhD, and Masaki Mori, MD, PhD

Abstract: International collaboration is important in healthcare quality evaluation; however, few international comparisons of general surgery outcomes have been accomplished. Furthermore, predictive model application for risk stratification has not been internationally evaluated. The National Clinical Database (NCD) in Japan was developed in collaboration with the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), with a goal of creating a standardized surgery database for quality improvement. The study aimed to compare the consistency and impact of risk factors of 3 major gastroenterological surgical procedures in Japan and the United States (US) using web-based prospective data entry systems: right hemicolectomy (RH), low anterior resection (LAR), and pancreaticoduodenectomy (PD).

Data from NCD and ACS-NSQIP, collected over 2 years, were examined. Logistic regression models were used for predicting 30-day mortality for both countries. Models were exchanged and evaluated to determine whether the models built for one population were accurate for the other population.

We obtained data for 113,980 patients; 50,501 (Japan: 34,638; US: 15,863), 42,770 (Japan: 35,445; US: 7325), and 20,709 (Japan: 15,527; US: 5182) underwent RH, LAR, and, PD, respectively. Thirty-day mortality rates for RH were 0.76% (Japan) and 1.88% (US); rates for LAR were 0.43% versus 1.08%; and rates for PD were 1.35% versus 2.57%. Patient background, comorbidities, and practice style were different between Japan and the US. In the models, the odds ratio for each variable was similar between NCD and ACS-NSQIP. Local risk models could predict mortality using local data, but could not accurately predict mortality using data from other countries.

We demonstrated the feasibility and efficacy of the international collaborative research between Japan and the US, but found that local risk models remain essential for quality improvement.

(*Medicine* 94(49):e2194)

Abbreviations: ACS-NSQIP = American College of Surgeons National Surgical Quality Improvement Program, ASA = American Society of Anesthesiologists, AST = Aspartate aminotransferase, BMI = body mass index, CKD = chronic kidney disease, COPD = chronic obstructive pulmonary disease, LAR = low anterior resection, NCD = National Clinical Database, PD = pancreaticoduodenectomy, PT-INR = prothrombin time–international normalized ratio, RH = right hemicolectomy, SIRS = systemic inflammatory response syndrome, US = United States.

INTRODUCTION

Improving the quality of surgical procedures is dependent on the collection of accurate data. The National Clinical Database (NCD) in Japan was developed in collaboration with the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), with a shared goal of creating a standardized surgery database for quality improvement. NCD and ACS-NSQIP have developed systems using standardized variable definitions to collect data on risk factors and outcomes after surgery. These databases collect prospective rather than retrospective data. Both use web-based data collection software, contributing to effective quality improvement, via benchmarking and risk-adjusted feedback reports to hospitals; this enables the identification of specific problems and works towards their improvement. The ACS initiated ACS-NSQIP in 2006 and demonstrated improved surgical outcomes among participating private sector hospitals.[1] More than 500 hospitals participated in ACS-NSQIP. NCD in Japan, which was launched in 2010, is a nationwide prospective registry linked to the surgical board certification system. NCD systematically collects accurate data on structures, processes, and outcomes, to develop a standardized surgery database for quality improvement and healthcare quality evaluation.[2] NCD contains the records of >1,200,000 surgical cases collected in 2011, with approximately 4000 institutions participating in 2013.

One of the important advantages of NCD and ACS-NSQIP is the ability to benchmark and compare risk-adjusted outcomes. This ability allows fair comparisons to be made along with collaborative learning. International collaboration is

Editor: Maria Kapritsou.
Received: August 3, 2015; revised: November 5, 2015; accepted: November 7, 2015.
From the Member of Database Committee of Japanese Society of Gastroenterological Surgery (TA, GW); Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL (JLP, CYK, MEC, LZ); National Clinical Database, Tokyo, Japan (HM, MG, NH, HK); and Japanese Society of Gastroenterological Surgery, Tokyo, Japan (HM, MG, HK, KS, MM).
Correspondence: Mitsukazu Gotoh, The Japanese Society of Gastroenterological Surgery, 104-0041 Chuouku Sintomi 1-14-1 Central East Building 5F, Tokyo, Japan (e-mail: mgotoh@fmu.ac.jp).
Funding: This study is partially supported from the Ministry of Health, Labour and Welfare, Japan.
Conflict of interest disclosures: The authors report no conflicting financial interests.

ISSN: 0025-7974
DOI: 10.1097/MD.0000000000002194

important in healthcare quality evaluation and produces meaningful results; however, few international comparisons of general surgery outcomes using clinical registry data have been accomplished. There is a lack of data regarding outcomes of Japanese patients undergoing gastroenterological surgery and comparison with the United States (US). Furthermore, predictive model application for clinical risk stratification has not been internationally evaluated. Differences in the prevalence of patient comorbidities and their association with outcomes remain unknown. The purpose of this study was to compare patient characteristics; procedure details; operative outcomes; the consistency and impact of risk factors for 3 major gastroenterological surgery procedures: right hemicolectomy (RH), low anterior resection (LAR), and pancreaticoduodenectomy (PD) in Japan and US; and to examine whether risk prediction models built for one population were accurate for the other population. To the best of our knowledge, this is the first study to use large, high-quality data from different patient populations.

METHODS

Study Design and Outcomes

Patient cohorts and risk subcategories for RH, LAR, and PD were selected from both NCD/Japan and ACS-NSQIP/US data. Univariate analysis of each selected predictors for the 3 procedures was conducted for both datasets. Subsequent multivariate models were separately constructed using data from NCD and ACS-NSQIP. Finally, risk models were exchanged and evaluated to determine whether risk prediction models built for one population were accurate for the other population.

Data from NCD and ACS-NSQIP collected over 2 years (2011–2012) were examined. The primary outcome measure of this study was 30-day mortality. Thirty-day mortality was defined as death within 30 days after the operation date regardless of whether the patient had been discharged from initial admission.

Data Acquisition and Patient Selection

The NCD project was approved on November 2010 by the Japan Surgical Society Ethics Committee. The developmental history and current status of the NCD, including sampling strategy, data abstraction procedures, variables collected, outcomes, and structure, are described elsewhere.[2,3] NCD recruits individuals to approve the inputted data from members of various departments in charge of annual cases, as well as data entry officers through a web-based data management system to assure the traceability of the data. NCD conducted onsite audits using source data randomly for mortality, and the results were found to be accurate. Currently, NCD is planning to perform onsite and remote audit for verifying the accuracy of existing data for morbidities. The ACS-NSQIP program and dataset have been described elsewhere.[4,5] Data are abstracted by trained Surgical Clinical Reviewers using standardized definitions, including patient demographics, comorbidities, laboratory values, operative variables, and complications.

Patients who underwent RH, LAR, and PD were identified using Current Procedural Terminology (CPT) (US) and NCD codes (Japan). Both NCD and ACS-NSQIP contain patient cohorts limited to malignant tumor patients only. Any records with entry denied by patients were excluded from this analysis. Records with missing information regarding age, sex, or status at 30 days postoperation were also excluded. For the PD procedure, we excluded cases with simultaneous major hepatectomy.

All variables, definitions, and inclusion criteria in NCD/Japan are accessible on the NCD website (http://www.ncd.or.jp/). Descriptions of the qualifications, auditing of data collection personnel, case inclusion criteria, sampling data collection strategy, and variable and outcome definitions in ACS-NSQIP/US are available online in ACS-NSQIP user guide.[6]

Variables

Two sets of predictive variables were constructed from the NCD/Japan and ACS NSQIP/US data fields. Patient demographic variables considered were age and sex. General factors considered were as follows: preoperative functional status (independent, partially dependent, or totally dependent); American Society of Anesthesiologists (ASA) class; dyspnea (none, moderate exertion, or at rest); emergency cases; and body mass index (BMI: normal, underweight, overweight, and 3 categories of obesity). ASA class was not considered in further multivariate analysis because the criterion to determine class was inconsistent between countries. Comorbidities included were diabetes (oral medication or insulin-dependent); a history of chronic obstructive pulmonary disease (COPD); hypertension requiring medication; congestive heart failure; bleeding disorders; sepsis (systemic inflammatory response syndrome, sepsis, and septic shock); disseminated cancer; chronic kidney disease (CKD) stage; and weight loss (>10% in previous 6 mo). Length of hospital stay was also compared.

Preoperative laboratory variables examined included albumin, white blood count, prothrombin time-international normalized ratio (PT-INR), total bilirubin, and aspartate aminotransferase (AST). Missing laboratory data continued as separate categories. It should be noted that missing values are virtually nonexistent for predictors, except for laboratory variables, where clinical issues have a substantial impact on the ordering of tests.

Statistical Analysis

Raw frequencies and chi-square tests were used to assess differences in the distribution of general factors, comorbidities, and laboratory values, as well as their association with 30-day mortality. Because of the low number of deaths, the risk models were developed with a limited number of variables.[7] To identify these variables, we first used a logistic regression technique with forward selection to identify the most significant predictor variables. Sharing the same SAS code, we generated 3 models (1 for each surgical group) in each country, with lists of the top predictors (data not shown). We used these lists to select a common set of predictors to be used in the final risk models. For the final risk models, logistic regression techniques with forced selection were used to develop models that predict 30-day mortality using a set of relevant comparably defined risk factor variables in both countries. Model fit was assessed using Hosmer–Lemeshow goodness-of-fit statistic for calibration and c-statistic for discrimination.[8,9] The c-statistic allows model discrimination to be measured, with 1.0 indicating perfect discrimination and 0.5 being no better than chance. These models were then used to predict mortality using data from the other dataset (ie, the NCD model was used to predict mortality using the ACS-NSQIP data and vice versa). Observed and expected mortality rates were compared. All data manipulation and analysis were performed with SAS version 9.3 (SAS Institute Inc.).

RESULTS

Risk Profiles and Outcomes

During the study period, a total of 50,501 patients underwent RH (Japan, 34,638; US, 15,863), 42,770 patients underwent LAR (Japan, 35,445; US, 7325), and 20,709 patients underwent PD (Japan, 15,527; US, 5182). Thirty-day unadjusted mortality rates for RH were 0.76% in Japan and 1.88% in US; mortality rates for LAR were 0.43% in Japan and 1.08% in US; and mortality rates for PD were 1.35% in Japan and 2.57% in US. The risk profiles and outcomes of each procedure from both databases are described in Table 1 (RH), Table 2 (LAR), and Table 3 (PD). The ACS-NSQIP population for each procedure tended to be younger. When we looked at the 30-day mortality associated with age, we observed that in both countries, mortality increases as age increases; however, the effect was more pronounced in the ACS-NSQIP data. Laparoscopy was conducted in 36.6% of the Japanese and 56.9% of the US RHs, and 42.9% of the Japanese and 44.2% of the US LARs. Notably, the percentage of patients with a high BMI substantially differed between cohorts. The ACS-NSQIP cohort had a significantly shorter length of hospital stay. The prevalence of patients with CKD differed between Japan and US. Univariate analysis revealed the patient risk factors that were significant predictors of mortality after RH, LAR, and PD (Tables 1–3).

The Risk Models for Mortalities

The final logistic model for 30-day mortalities of each procedure, along with odds ratios (ORs) and 95% confidence intervals (CIs), is presented in Table 4. For RH, 14 significant risk factors for 30-day mortality were identified in Japan, and, in contrast, 17 significant risk factors were identified in US. The c-statistic was calculated to evaluate model performance. The c-statistic was 0.857 for the Japan model and 0.840 for the US model, indicating adequate discrimination. The Hosmer–Lemeshow statistic was 11.243 ($P = 0.19$) for the Japan model and 5.8660 ($P = 0.66$) for the US model, indicating both models adequately assigned risk. For LAR, 12 significant risk factors for 30-day mortality were identified in Japan; in contrast, 10 significant risk factors were identified in US. The c-statistic was 0.782 for the Japan model and 0.822 for the US model, indicating adequate discrimination. The Hosmer–Lemeshow statistic was 5.2355 ($P = 0.63$) for the Japan model and 10.9464 ($P = 0.20$) for the US model, indicating both models adequately assigned risk. For PD, 9 significant risk factors for 30-day mortality were identified in Japan; in contrast, 11 significant risk factors were identified in US. The c-statistic was 0.684 for the Japan model and 0.719 for the US model, indicating good discrimination. The Hosmer–Lemeshow statistic was 9.908 ($P = 0.27$) for the Japan model and 8.6192 ($P = 0.38$) for the US model, indicating both models adequately assigned risk. ORs for each variable were similar between countries.

Exchange Each Risk Model

Models were exchanged between countries and were used to create new models with forced selection to evaluate model transferability (Table 5). For RH, the c-statistic was 0.789 based on US data using the Japan model formula and 0.828 based on Japanese data using the US model formula, indicating adequate but decreased discrimination. The Hosmer–Lemeshow statistic was 171.01 ($P < 0.001$) based on US data using the Japan model formula and 955.23 ($P < 0.001$) based on Japanese data using the US model, indicating neither model adequately assigned risk. For LAR, the c-statistic was 0.786 based on US data using the Japan model formula and 0.778 based on Japanese data using the US model formula, indicating good but decreased discrimination. The Hosmer–Lemeshow statistic was 49.54 ($P < 0.001$) based on US data using the Japan model formula and 145.37 ($P < 0.001$) based on Japanese data using the US model, indicating neither model adequately assigned risk. For PD, the c-statistic was 0.674 based on US data using the Japan model formula and 0.540 based on Japanese data using the US model formula, indicating inadequate discrimination. The Hosmer–Lemeshow statistic was 8.8173 ($P = 0.36$) based on US data using the Japan model formula and 366.22 ($P < 0.001$) based on Japanese data using the US model. In all three procedures, we ran each risk model using the other country's data to assess the discrimination of each model.

Observed and Expected Mortality

Both NCD and ACS-NSQIP models were able to predict the number of deaths in the Japan dataset accurately. However, we decreased accuracy when using models from one country's dataset to predict the number of deaths in the other; we sound the ACS-NSQIP model overpredicted deaths in the NCD dataset, whereas the NCD model underpredicted deaths in the ACS-NSQIP dataset (Table 5). Figures 1 and 2 show the 30-day mortality model calibrations and observed event rates versus predicted rates. In measures of calibration (Hosmer–Lemeshow plots), the *y* axis gives the predicted number of deaths, and the *x* axis gives the actual number of deaths observed, that is, a perfect straight line would be a perfect model. Risk models based on local data accurately predicted mortality rates; however, risk models based on the other country's data could not accurately predict mortality rates.

DISCUSSION

Our study is the first international comparison of nationwide operative mortality in gastroenterological surgery using similar web-based prospective data entry systems, with collaboration between the NCD/Japan and the ACS-NSQIP/US. Although some international comparative studies provided variations in mortality rate between countries,[10,11] these studies were under the restriction of the inherent differences in the data collection methods between the datasets. Also, these studies did not examine whether risk prediction models built for one population were accurate for the other population. ACS-NSQIP participation have not offered a clear mechanism for quality improvement[12,13]; however, these are undeniably considered the highest clinical quality standards for evaluating risk-adjusted surgical outcomes. Both use rigorous, standardized data collection methods. Preoperative variables are clearly defined with the same definitions used in both databases and the same defined data collection methodology, including a strict follow-up period for outcomes.[2,14] By comparing the 2 datasets, we found differences in the following: descriptive data including preoperative patient variables; definitions and interpretation of ASA classifications; missing data from preoperative blood tests; duration of hospital stay after surgery; and 30-day mortality. We then created risk models based on local data for each country to predict mortality after each procedure. We found that although the exchanged models had adequate discrimination for mortality after each procedure, the models failed to yield adequate calibration between countries. This finding clearly indicated that risk models based on local data remain essential for quality assessment and improvement.

TABLE 1. Univariate Analysis for 30-Day Mortality of Right Hemicolectomy

	US/NSQIP (N = 15,863) (Died = 299, Died [%] = 1.88%)			Japan/NCD (N = 34,638) (Died = 264, Died [%] = 0.76%)		
	Total (N [%])	Died (n [%])	*P*	Total (N [%])	Died (n [%])	*P*
Age			<0.001			<0.001
<60	3903 (24.6)	14 (0.4)		3756 (10.8)	13 (0.3)	
60–70	4242 (26.7)	50 (1.2)		8453 (24.4)	33 (0.4)	
70–80	4408 (27.8)	86 (2.0)		12825 (37.0)	73 (0.6)	
80–90	2851(18.0)	114 (4.0)		8616 (24.9)	110 (1.3)	
≥90	459 (2.9)	35 (7.6)		988 (2.9)	35 (3.5)	
Sex			0.113			0.108
Men	7398 (46.6)	153 (2.1)		17596 (50.8)	121 (0.7)	
Women	8465 (53.4)	146 (1.7)		17042 (49.2)	143 (0.8)	
Diabetes			0.015			0.73
Insulin	999 (6.3)	29 (2.9)		1171 (3.4)	10 (0.9)	
Noninsulin	14864 (93.7)	270 (1.8)		33467 (96.6)	254 (0.8)	
COPD			<0.001			<0.001
No	14823 (93.4)	258 (1.7)		33618 (97.1)	244 (0.7)	
Yes	1040 (6.6)	41 (3.9)		1020 (2.9)	20 (2.0)	
Hypertension			<0.001			0.041
No	6371 (40.2)	85 (1.3)		21706 (62.7)	149 (0.7)	
Yes	9492 (59.8)	214 (2.3)		12932 (37.3)	115 (0.9)	
Congestive heart failure			<0.001			<0.001
No	15676 (98.8)	281 (1.8)		34147 (98.6)	241 (0.7)	
Yes	187 (1.2)	18 (9.6)		491 (1.4)	23 (4.7)	
Bleeding disorder			<0.001			<0.001
No	15207 (95.9)	260 (1.7)		33282 (96.1)	228 (0.7)	
Yes	656 (4.1)	39 (6.0)		1356 (3.9)	36 (2.7)	
Emergency status			<0.001			<0.001
No	15153 (95.5)	247 (1.6)		33027 (95.3)	201 (0.6)	
Yes	710 (4.5)	52 (7.3)		1611 (4.7)	63 (3.9)	
Functional status			<0.001			<0.001
Independent	15299 (96.4)	247 (1.6)		31345 (90.5)	138 (0.4)	
Partially dependent	489 (3.1)	41 (8.4)		2536 (7.3)	78 (3.1)	
Totally dependent	75 (0.5)	11 (14.7)		757 (2.2)	48 (6.3)	
ASA class			<0.001			<0.001
1-No disturb	333 (2.1)	0 (0.0)		10660 (30.8)	24 (0.2)	
2-Mild disturb	6205 (39.1)	26 (0.4)		19454 (56.2)	104 (0.5)	
3-Severe disturb	8349 (52.6)	191 (2.3)		4278 (12.4)	107 (2.5)	
4-Life threat	965 (6.1)	78 (8.1)		189 (0.5)	21 (11.1)	
5-Moribund	11 (0.1)	4 (36.4)		57 (0.2)	8 (14.0)	
Dyspnea			<0.001			<0.001
None	14030 (88.4)	231 (1.7)		33707 (97.3)	225 (0.7)	
Moderate	1691 (10.7)	55 (3.3)		789 (2.3)	25 (3.2)	
At rest	142 (0.9)	13 (9.2)		142 (0.4)	14 (9.9)	
Sepsis			<0.001			<0.001
None	15371 (96.9)	251 (1.6)		34323 (99.1)	222 (0.6)	
SIRS/sepsis/septic shock	492 (3.1)	48 (9.8)		315 (0.9)	42 (13.3)	
Weight loss			<0.001			<0.001
No	15129 (95.4)	259 (1.7)		32654 (94.3)	219 (0.7)	
Yes	734 (4.6)	40 (5.5)		1984 (5.7)	45 (2.3)	
Disseminated cancer			<0.001			<0.001
No	14770 (93.1)	235 (1.6)		32539 (93.9)	200 (0.6)	
Yes	1093 (6.9)	64 (5.9)		2099 (6.1)	64 (3.0)	
BMI			<0.001			<0.001
Underweight	390 (2.5)	20 (5.1)		5294 (15.3)	69 (1.3)	
Normal	4718 (29.7)	115 (2.4)		22845 (66.0)	156 (0.7)	
Overweight	5465 (34.5)	85 (1.6)		5640 (16.3)	32 (0.6)	
Obese 1	3077 (19.4)	49 (1.6)		688 (2.0)	5 (0.7)	
Obese 2/3	2213 (13.9)	30 (1.4)		170 (0.5)	2 (1.2)	

	US/NSQIP (N = 15,863) (Died = 299, Died [%] = 1.88%)			Japan/NCD (N = 34,638) (Died = 264, Died [%] = 0.76%)		
	Total (N [%])	Died (n [%])	*P*	Total (N [%])	Died (n [%])	*P*
Steroid			<0.001			0.006
No	15432 (97.3)	281 (1.8)		34276 (99.0)	256 (0.7)	
Yes	431 (2.7)	18 (4.2)		362 (1.0)	8 (2.2)	
Surgical approach			<0.001			<0.001
Lap	9031 (56.9)	82 (0.9)		12664 (36.6)	39 (0.3)	
Open	6832 (43.1)	217 (4.0)		21974 (63.4)	225 (1.0)	
Albumin (g/dL)			<0.001			<0.001
≥3.5	7994 (50.4)	80 (1.0)		23102 (66.7)	69 (0.3)	
2.8–3.5	2095 (13.2)	89 (4.3)		6831 (19.7)	78 (1.1)	
≤2.8	769 (4.9)	76 (9.9)		2824 (8.2)	103 (3.6)	
Missing	5005 (31.5)	54 (1.1)		1881 (5.4)	14 (0.7)	
PT-INR			<0.001			<0.001
≤1.25	7520 (47.4)	163 (2.2)		30576 (88.3)	169 (0.6)	
>1.25	804 (5.1)	52 (6.5)		1474 (4.3)	61 (4.1)	
Missing	7539 (47.5)	84 (1.1)		2588 (7.5)	34 (1.3)	
Chronic kidney stage			<0.001			<0.001
Stage 1 (GFR ≥90)	3928 (24.8)	77 (2.0)		3100 (8.9)	27 (0.9)	
Stage 2 (GFR 60–89)	7598 (47.9)	104 (1.4)		13287 (38.4)	64 (0.5)	
Stage 3 (GFR 30–59)	3017 (19.0)	85 (2.8)		15661 (45.2)	120 (0.8)	
Stage 4 (GFR 15–29)	244 (1.5)	16 (6.6)		1229 (3.5)	29 (2.4)	
Stage 5 (GFR ≤15 or dialysis)	145 (0.9)	12 (8.3)		499 (1.4)	14 (2.8)	
Missing	931 (5.9)	5 (0.5)		862 (2.5)	10 (1.2)	
Platelets (×1000/μL)			<0.001			<0.001
>120	14743 (92.9)	269 (1.8)		30332 (87.6)	165 (0.5)	
≤120	361 (2.3)	25 (6.9)		3843 (11.1)	95 (2.5)	
Missing	759 (4.8)	5 (0.7)		463 (1.3)	4 (0.9)	
Total bilirubin (mg/dL)			<0.001			<0.001
≤2.0	10847(68.4)	241 (2.2)		33607 (97.0)	248 (0.7)	
>2.0	107 (0.7)	6 (5.6)		281 (0.8)	11 (3.9)	
Missing	4909 (30.9)	52 (1.1)		750 (2.2)	5 (0.7)	
AST (U/L)			<0.001			<0.001
≤100	10762(67.9)	242 (2.3)		33914 (97.9)	236 (0.7)	
>100	82 (0.5)	4 (4.9)		212 (0.6)	24 (11.3)	
Missing	5019 (31.6)	53 (1.1)		512 (1.5)	4 (0.8)	
WBC (×1000/μL)			<0.001			<0.001
≥3.5 to ≤9.0	11818 (74.5)	171 (1.5)		28358 (81.8)	144 (0.5)	
<3.5 or >9.0	3289 (20.7)	121 (3.7)		5530 (16.0)	115 (2.1)	
Missing	756 (4.8)	7 (0.9)		750 (2.2)	5 (0.7)	
		Median (IQR)			Median (IQR)	
Length of stay (d)						
Total		5 (4–7)			14 (10–20)	
Survived		5 (4–7)			14 (10–20)	
Died		8 (5–13)			15 (6.25–22)	

Data are expressed as mean ± standard deviation or frequency(%).

T test/Wilcoxon Mann–Whitney test applied for continuous variables and chi-square/Fisher exact test applied for categorical variables.

ASA = American Society of Anesthesiologists, AST = aspartate aminotransferase, BMI = body mass index, COPD = chronic obstructive pulmonary disease, GFR = glomerular filtration rate, IQR = interquartile range, NCD = National Clinical Database, NSQIP = National Surgical Quality Improvement Program, PT-INR = prothrombin time–international normalized ratio, SIRS = systemic inflammatory response syndrome, WBC = white blood cell.

More than 2000 hospitals performing gastrointestinal (GI) tract surgery joined NCD, and 95% of surgical cases (949,824 cases in 2011 and 2012) were collected in this database, making NCD a nationally representative sample.[15] Meanwhile, data submitted to ACS-NSQIP from participating hospitals include a considerable proportion of cases (442,149 cases from 315 sites in 2011) sufficient to provide benchmark support to individual hospitals. We identified a number of differences in risk factor prevalence between datasets. Mortality rates reported in this study differed slightly, with lower unadjusted mortality rates for all 3 procedures in Japan than in US. The duration of hospital stay also differed, being longer in Japan compared to the US for all 3 procedures. The Japanese patients were older and had a higher prevalence of CKD. In contrast, US patients were younger and substantially more obese.

TABLE 2. Univariate Analysis for 30-Day Mortality of Low Anterior Resection

	US/NSQIP (N = 7325) (Died = 79, Died [%] = 1.08%)			Japan/NCD (N = 35,445) (Died = 154, Died [%] = 0.43%)		
	Total (N [%])	Died (n [%])	*P*	Total (N [%])	Died (n [%])	*P*
Age			<0.001			<0.001
<60	3172 (43.3)	10 (0.3)		8329 (23.5)	9 (0.1)	
60–70	1944 (26.6)	19 (1.0)		12300 (34.7)	38 (0.3)	
70–80	1409 (19.2)	25 (1.8)		10541 (29.8)	59 (0.6)	
80–90	720 (9.8)	22 (3.1)		4010 (11.3)	40 (1.0)	
≥90	80 (1.1)	3 (3.8)		265 (0.7)	8 (3.0)	
Sex			0.030			0.015
Men	4126 (56.3)	54 (1.3)		23140 (65.3)	115 (0.5)	
Women	3199 (43.7)	25 (0.8)		12305 (34.7)	39 (0.3)	
Diabetes			0.395			0.109
Insulin	321 (4.4)	5 (1.6)		1194 (3.4)	9 (0.8)	
Noninsulin	7004 (95.6)	74 (1.1)		34251 (96.6)	145 (0.4)	
COPD			<0.001			0.008
No	6994 (95.5)	66 (0.9)		34512 (97.4)	144 (0.4)	
Yes	331 (4.5)	13 (3.9)		933 (2.6)	10 (1.1)	
Hypertension			<0.001			<0.001
No	3775 (51.5)	24 (0.6)		24061 (67.9)	78 (0.3)	
Yes	3550 (48.5)	55 (1.6)		11384 (32.1)	76 (0.7)	
Congestive heart failure			<0.001			0.001
No	7282 (99.4)	75 (1.0)		35190 (99.3)	148 (0.4)	
Yes	43 (0.6)	4 (9.3)		255 (0.7)	6 (2.4)	
Bleeding disorder			<0.001			<0.001
No	7123 (97.2)	70 (1.0)		34441 (97.2)	139 (0.4)	
Yes	202 (2.8)	9 (4.5)		1004 (2.8)	15 (1.5)	
Emergency status			<0.001			0.006
No	7216 (98.5)	72 (1.0)		35077 (99.0)	148 (0.4)	
Yes	109 (1.5)	7 (6.4)		368 (1.0)	6 (1.6)	
Functional status			<0.001			<0.001
Independent	7200 (98.3)	69 (1.0)		33982 (95.9)	124 (0.4)	
Partially dependent	111 (1.5)	9 (8.1)		1242 (3.5)	20 (1.6)	
Totally dependent	14 (0.2)	1 (7.1)		221 (0.6)	10 (4.5)	
ASA class			<0.001			<0.001
1-No disturb	201 (2.7)	0 (0.0)		14205 (40.1)	21 (0.1)	
2-Mild disturb	3402 (46.4)	15 (0.4)		18307 (51.6)	86 (0.5)	
3-Severe disturb	3471 (47.4)	50 (1.4)		2848 (8.0)	42 (1.5)	
4-Life threat	250 (3.4)	14 (5.6)		58 (0.2)	5 (8.6)	
5-Moribund	1 (0.1)	0 (0.0)		27 (0.1)	0 (0.0)	
Dyspnea			<0.001			<0.001
None	6799 (92.8)	62 (0.9)		34909 (98.5)	145 (0.4)	
Moderate	487 (6.7)	11 (2.3)		464 (1.3)	7 (1.5)	
At rest	39 (0.5)	6 (15.4)		72 (0.2)	2 (2.8)	
Sepsis			<0.001			<0.001
None	7232 (98.7)	71 (1.0)		35356 (99.7)	149 (0.4)	
SIRS/sepsis/septic shock	93 (1.3)	8 (8.6)		89 (0.3)	5 (5.6)	
Weight loss			<0.001			<0.001
No	6996 (95.5)	65 (0.9)		34254 (96.6)	137 (0.4)	
Yes	329 (4.5)	14 (4.3)		1191 (3.4)	17 (1.4)	
Disseminated cancer			<0.001			<0.001
No	6690 (91.3)	61 (0.9)		33895 (95.6)	131 (0.4)	
Yes	635 (8.7)	18 (2.8)		1550 (4.4)	23 (1.5)	
BMI			0.849			0.005
Underweight	161 (2.2)	2 (1.2)		4223 (11.9)	29 (0.7)	
Normal	2266 (31.0)	28 (1.2)		23889 (67.4)	94 (0.4)	
Overweight	2520 (34.4)	26 (1.0)		6433 (18.2)	22 (0.3)	
Obese 1	1479 (20.2)	16 (1.1)		755 (2.1)	9 (1.2)	
Obese 2/3	899 (12.2)	7 (0.8)		145 (0.4)	0 (0.0)	

	US/NSQIP (N = 7325) (Died = 79, Died [%] = 1.08%)			Japan/NCD (N = 35,445) (Died = 154, Died [%] = 0.43%)		
	Total (N [%])	**Died (n [%])**	***P***	**Total (N [%])**	**Died (n [%])**	***P***
Steroid			0.231			0.019
No	7171 (97.9)	76 (1.1)		35214 (99.3)	150 (0.4)	
Yes	154 (2.1)	3 (2.0)		231 (0.7)	4 (1.7)	
Surgical approach			0.007			<0.001
Lap	3238 (44.2)	23 (0.7)		15208 (42.9)	42 (0.3)	
Open	4087 (55.8)	56 (1.4)		20237 (57.1)	112 (0.6)	
Albumin			<0.001			<0.001
≥3.5	4437 (60.6)	28 (0.6)		28807 (81.3)	91 (0.3)	
2.8–3.5	631 (8.6)	15 (2.4)		3804 (10.7)	45 (1.2)	
≤2.8	190 (2.6)	16 (8.4)		946 (2.7)	13 (1.4)	
Missing	2067 (28.2)	20 (1.0)		1888 (5.3)	5 (0.3)	
INR			<0.001			<0.001
≤1.25	3704 (50.6)	42 (1.1)		32069 (90.5)	122 (0.4)	
>1.25	167 (2.3)	10 (6.0)		872 (2.5)	18 (2.1)	
Missing	3454 (47.1)	27 (0.8)		2504 (7.0)	14 (0.6)	
Chronic kidney stage			0.002			<0.001
Stage 1 (GFR ≥90)	2441 (33.3)	26 (1.1)		2808 (7.9)	11 (0.4)	
Stage 2 (GFR 60–89)	3575 (48.8)	31 (0.9)		13958 (39.4)	35 (0.3)	
Stage 3 (GFR 30–59)	893 (12.2)	20 (2.2)		16678 (47.1)	87 (0.5)	
Stage 4 (GFR 15–29)	49 (0.7)	2 (4.1)		815 (2.3)	16 (2.0)	
Stage 5 (GFR <15 or dialysis)	36 (0.5)	0 (0.0)		334 (0.9)	4 (1.2)	
Missing	331 (4.5)	0 (0.0)		852 (2.4)	1 (0.1)	
Platelets (× 1000/μL)			0.160			<0.001
>120	6879 (93.9)	77 (1.1)		31408 (88.6)	112 (0.4)	
≤120	172 (2.4)	2 (1.2)		3599 (10.2)	41 (1.1)	
Missing	274 (3.7)	0 (0.0)		438 (1.2)	1 (0.2)	
Total bilirubin (mg/dL)			0.776			0.603
≤2.0	5290 (72.2)	58 (1.1)		34445 (97.2)	150 (0.4)	
>2.0	50 (0.7)	1 (2.0)		271 (0.8)	2 (0.7)	
Missing	1985 (27.1)	20 (1.0)		729 (2.0)	2 (0.3)	
AST (U/L)			0.463			0.188
≤100	5193 (70.9)	59 (1.1)		34811 (98.2)	151 (0.4)	
>100	39 (0.5)	1 (2.7)		151 (0.4)	2 (1.3)	
Missing	2093 (28.6)	19 (0.9)		483 (1.4)	1 (0.2)	
WBC (× 1000/μL)			0.004			0.003
≥3.5 to ≤9.0	5890 (80.4)	55 (0.9)		30481 (86.0)	120 (0.4)	
<3.5 or >9.0	1167 (15.9)	23 (2.0)		4235 (11.9)	32 (0.8)	
Missing	268 (3.7)	1 (0.4)		729 (2.1)	2 (0.3)	
		Median (IQR)				**Median (IQR)**
Length of stay						
Total		6 (4–8)				16 (12–25)
Survived		6 (4–8)				16 (12–25)
Died		6 (4–10)				10 (6–20)

Data are expressed as mean ± standard deviation or frequency (%).

T test/Wilcoxon Mann–Whitney test applied for continuous variables and chi-square/Fisher exact test applied for categorical variables.

ASA = American Society of Anesthesiologists, AST = aspartate aminotransferase, BMI = body mass index, COPD = chronic obstructive pulmonary disease, GFR = glomerular filtration rate, IQR = interquartile range, NCD = National Clinical Database, NSQIP = National Surgical Quality Improvement Program, PT-INR = prothrombin time–international normalized ratio, SIRS = systemic inflammatory response syndrome, WBC = white blood cell.

The NCD and ACS-NSQIP data have been used in prediction tools to facilitate risk stratification before surgery in a procedure-targeted manner.[16–24] More specifically, the ACS-NSQIP risk calculator works by utilizing information regarding the patient's risk factors related to the planned surgical procedure. The calculator then provides a predicted risk of complications after surgery.[25] However, the ability of risk models created using nationwide databases to predict the surgical risk for patients undergoing the same procedure in other countries has yet to be evaluated. In this study, we used the NCD and NSQIP databases to develop independent 30-day mortality risk models, and identified significant variables

TABLE 3. Univariate Analysis for 30-Day Mortality of Pancreaticoduodenectomy

	US/NSQIP (N = 5182) (Died = 133, Died [%] = 2.57%)			Japan/NCD (N = 15,527) (Died = 210, Died [%] = 1.35%)		
	Total (N [%])	Died (n [%])	*P*	Total (N [%])	Died (n [%])	*P*
Age			<0.001			<0.001
<60	1466 (28.3)	17 (1.2)		2186 (14.0)	14 (0.6)	
60–70	1732 (33.4)	44 (2.5)		5181 (33.4)	62 (1.2)	
70–80	1456 (28.1)	50 (3.4)		6394 (41.2)	99 (1.6)	
80–90	508 (9.8)	19 (3.7)		1755 (11.3)	35 (2.0)	
≥90	20 (0.4)	3 (15.0)		11 (0.1)	0 (0.0)	
Sex			0.029			0.004
Male	2749 (53.1)	83 (3.0)		9604 (61.8)	150 (1.6)	
Female	2433 (46.5)	50 (2.1)		5923 (38.2)	60 (1.0)	
Diabetes			0.842			0.425
Insulin	613 (11.8)	15 (2.5)		1637 (10.5)	18 (1.1)	
Noninsulin	4569 (88.2)	118 (2.6)		13890 (89.5)	192 (1.4)	
COPD			<0.001			0.028
No	4935 (95.2)	117 (2.4)		15115 (97.3)	199 (1.3)	
Yes	247 (4.8)	16 (6.5)		412 (2.7)	11 (2.7)	
Hypertension			<0.001			0.028
No	2314 (44.6)	34 (1.5)		10009 (64.5)	199 (2.0)	
Yes	2868 (55.4)	99 (3.5)		5518 (35.5)	11 (0.2)	
Congestive heart failure			1.000			0.058
No	5174 (99.8)	133 (2.6)		15462 (99.6)	207 (1.3)	
Yes	8 (0.2)	0 (0.0)		65 (0.4)	3 (4.6)	
Bleeding disorder			0.415			0.001
No	5040 (97.3)	128 (2.5)		15044 (96.9)	194 (1.3)	
Yes	142 (2.7)	5 (3.5)		483 (3.1)	16 (3.3)	
Emergency status			0.577			0.002
No	5149 (99.4)	132 (2.6)		15429 (99.4)	204 (1.3)	
Yes	33 (0.6)	1 (3.0)		98 (0.6)	6 (6.1)	
Functional status			0.304			<0.001
Independent	5118 (98.8)	130 (2.5)		15007 (96.7)	189 (1.3)	
Partially dependent	58 (1.1)	3 (5.2)		472 (3.0)	18 (3.8)	
Totally dependent	6 (0.1)	0 (0.0)		48 (0.3)	3 (6.3)	
ASA class			<0.001			<0.001
1-No disturb	37 (0.7)	0 (0.0)		4696 (30.2)	35 (0.8)	
2-Mild disturb	1341 (25.9)	15 (1.1)		9368 (60.3)	136 (1.5)	
3-Severe disturb	3502 (67.5)	100 (2.9)		1421 (9.2)	35 (2.5)	
4-Life threat	300 (5.8)	18 (6.0)		28 (0.2)	3 (10.7)	
5-Moribund	2 (0.1)	0 (0.0)		14 (0.1)	1 (7.1)	
Dyspnea			<0.001			<0.001
None	4794 (92.5)	110 (2.3)		15349 (98.8)	202 (1.3)	
Moderate	377 (7.3)	20 (5.3)		165 (1.1)	8 (4.9)	
At rest	11 (0.2)	3 (27.3)		13 (0.1)	0 (0.0)	
Sepsis			0.297			0.043
None	5091 (98.2)	129 (2.5)		15452 (99.5)	208 (1.4)	
SIRS/sepsis/septic shock	91 (1.8)	4 (4.4)		75 (0.5)	2 (2.7)	
Weight loss			0.960			0.105
No	4333 (83.6)	111 (2.6)		14422 (92.9)	189 (1.3)	
Yes	849 (16.4)	22 (2.6)		1105 (7.1)	21 (1.9)	
Disseminated cancer			0.014			0.294
No	5021 (96.9)	124 (2.5)		15455 (99.5)	208 (1.4)	
Yes	161 (3.1)	9 (5.6)		72 (0.5)	2 (2.8)	
BMI			0.016			<0.001
Underweight	140 (2.7)	5 (3.6)		2293 (14.8)	21 (0.9)	
Normal	1864 (36.0)	39 (2.1)		10938 (70.4)	137 (1.3)	
Overweight	1842 (35.5)	39 (2.1)		2068 (13.3)	47 (2.3)	
Obese 1	850 (16.4)	29 (3.4)		187 (1.2)	3 (1.6)	
Obese 2/3	486 (9.4)	21 (4.3)		41 (0.3)	3 (7.3)	

	US/NSQIP (N = 5182) (Died = 133, Died [%] = 2.57%)			Japan/NCD (N = 15,527) (Died = 210, Died [%] = 1.35%)		
	Total (N [%])	**Died (n [%])**	***P***	**Total (N [%])**	**Died (n [%])**	***P***
Steroid			0.237			0.323
No	5062 (97.7)	128 (2.5)		15350 (98.9)	208 (1.4)	
Yes	120 (2.3)	5 (4.2)		177 (1.1)	2 (1.1)	
Surgical complexity			0.009			0.020
PD alone	4444 (85.8)	104 (2.3)		12975 (83.6)	161 (1.2)	
With adjacent organ	240 (4.6)	13 (5.4)		623 (4.0)	14 (2.3)	
Vascular ± other organ	498 (9.6)	16 (3.2)		1929 (12.4)	35 (1.8)	
Albumin			0.003			0.012
≥3.5	3303 (63.7)	69 (2.1)		10963 (70.6)	129 (1.2)	
2.8–3.5	1030 (19.9)	34 (3.3)		3391 (21.8)	55 (1.6)	
≤2.8	414 (8.0)	20 (4.8)		706 (4.6)	16 (2.3)	
Missing	435 (8.4)	10 (2.3)		467 (3.0)	10 (2.1)	
INR			0.004			<0.001
≤1.25	4277 (82.5)	99 (2.3)		14086 (90.7)	175 (1.2)	
>1.25	272 (5.3)	15 (5.5)		627 (4.0)	21 (3.4)	
Missing	633 (12.2)	19 (3.0)		814 (5.3)	14 (1.7)	
Chronic kidney stage			0.002			<0.001
Stage 1 (GFR ≥90)	2139 (41.3)	39 (1.8)		1627 (10.5)	13 (0.8)	
Stage 2 (GFR 60–89)	2255 (43.5)	62 (2.8)		6857 (44.1)	69 (1.0)	
Stage 3 (GFR 30–59)	641 (12.4)	24 (3.7)		6382 (41.1)	110 (1.7)	
Stage 4 (GFR 15–29)	30 (0.6)	2 (6.7)		281 (1.8)	6 (2.1)	
Stage 5 (GFR <15 or dialysis)	25 (0.5)	3 (12.0)		149 (1.0)	7 (4.7)	
Missing	92 (1.8)	3 (3.3)		231 (1.5)	5 (2.2)	
Platelets (× 1000/μL)			0.045			<0.001
>120	4982 (96.1)	123 (2.5)		9602 (61.8)	73 (0.8)	
≤120	104 (2.0)	6 (5.8)		5795 (37.3)	134 (2.3)	
Missing	96 (1.9)	4 (4.2)		130 (0.9)	3 (2.3)	
Total bilirubin (mg/dL)			0.576			0.156
≤2.0	3482 (67.2)	86 (2.5)		11578 (74.5)	147 (1.3)	
>2.0	1211 (23.4)	31 (2.6)		3800 (24.5)	59 (1.6)	
Missing	489 (9.4)	16 (3.3)		149 (1.0)	4 (2.7)	
AST (U/L)			0.260			0.612
≤100	3993 (77.1)	97 (2.4)		13510 (87.0)	183 (1.4)	
>100	757 (14.6)	26 (3.4)		1887 (12.2)	24 (1.3)	
Missing	432 (8.3)	10 (2.3)		130 (0.8)	3 (2.3)	
WBC (× 1000/μL)			0.169			0.303
≥3.5 to ≤9.0	3966 (76.5)	96 (2.4)		13434 (86.5)	177 (1.3)	
<3.5 or >9.0	1122 (21.7)	32 (2.9)		1944 (12.5)	29 (1.5)	
Missing	94 (1.8)	5 (5.3)		149 (1.0)	4 (2.7)	
		Median (IQR)				**Median (IQR)**
Length of stay						
Total		9 (7–14)				31 (22–43)
Survived		9 (7–14)				31 (22–43)
Died		11 (6–17)				15 (7–24)

Data are expressed as mean ± standard deviation or frequency (%).

T test/Wilcoxon Mann–Whitney test applied for continuous variables and chi-square/Fisher exact test applied for categorical variables.

ASA = American Society of Anesthesiologists, AST = aspartate aminotransferase, BMI = body mass index, COPD = chronic obstructive pulmonary disease, GFR = glomerular filtration rate, INR = international normalized ratio, IQR = interquartile range, NCD = National Clinical Database, NSQIP = National Surgical Quality Improvement Program, PD = pancreaticoduodenectomy, SIRS = systemic inflammatory response syndrome, WBC = white blood cell.

from both datasets to create NCD/ACS-NSQIP risk models using a common set of variables. For the purpose of estimating risk, the 2 models based on the 2 country's own dataset were able to adequately predict mortality with a good c-index and similar ORs observed for each variable (Table 4).

We found that discrimination decreased when we ran each risk model using the other country's data. When we focused on a measure of calibration (the Hosmer–Lemeshow plot), we found that both NCD and ACS-NSQIP models accurately predicted the number of deaths in their respective datasets. However, calibration diminished when data from the other country were

TABLE 4. Risk Models of Preoperative Factors for 30-Day Mortality Rates After RH, LAR, and PD

RH Variables	US/NSQIP			Japan/NCD		
		95% CI			95% CI	
	Odds Ratio	Lower CI	Upper CI	Odds Ratio	Lower CI	Upper CI
Age 60–70	3.641	1.984	6.682	1.068	0.563	2.147
Age 70–80	5.427	3.028	9.726	1.426	0.796	2.761
Age ≥80	10.473	5.890	18.623	2.23	1.249	4.309
Sex (male)	1.615	1.267	2.059	1.332	1.006	1.766
Bleeding disorder (yes)	1.826	1.249	2.668	1.721	1.134	2.534
Emergency status (yes)	1.653	1.146	2.384	1.873	1.304	2.644
Functional status (partially dependent)	1.923	1.317	2.810	2.767	1.994	3.815
Functional status (totally dependent)	3.429	1.644	7.153	4.241	2.83	6.263
Dyspnea at rest	2.697	1.414	5.144	2.519	1.22	4.844
Dyspnea moderate exertion	1.404	1.026	1.923	1.485	0.909	2.327
Sepsis (yes)	2.113	1.424	3.135	2.899	1.829	4.524
Weight loss (yes)	1.678	1.149	2.451	1.244	0.854	1.774
Disseminated cancer (yes)	3.045	2.233	4.151	3.385	2.452	4.616
Albumin (g/dL) 2.8–3.5	2.315	1.674	3.200	2.022	1.432	2.856
Albumin (g/dL) ≤2.8	3.991	2.777	5.736	3.447	2.385	4.989
Albumin (g/dL) missing	1.162	0.808	1.670	1.571	0.767	2.943
CKD missing	0.617	0.225	1.691	4.115	1.69	8.835
CKD stage 3	1.242	0.940	1.642	1.352	0.994	1.845
CKD stage 4/5	2.699	1.730	4.211	2.123	1.383	3.21
WBC (×1000/μL) <3.5 or >9.0	1.572	1.210	2.043	1.885	1.423	2.487
WBC (×1000/μL) missing	1.490	0.628	3.534	0.475	0.133	1.422
C-statistic	0.840			0.857		
Hosmer–Lemeshow chi-square	5.866 ($P = 0.662$)			11.243 ($P = 0.188$)		

LAR Variables	US/NSQIP			Japan/NCD		
		95% CI			95% CI	
	Odds Ratio	Lower CI	Upper CI	Odds Ratio	Lower CI	Upper CI
Age 60–70	2.833	1.299	6.180	2.725	1.372	6.033
Age 70–80	4.976	2.332	10.619	4.584	2.372	9.984
Age ≥80	7.220	3.265	15.967	7.988	4.009	17.757
Sex (male)	1.880	1.140	3.100	1.742	1.212	2.558
Bleeding disorder (yes)	2.250	1.028	4.925	1.987	1.091	3.366
Emergency status (yes)	2.433	0.924	6.403	1.768	0.608	4.227
Functional status (partially dependent)	1.931	0.810	4.603	1.91	1.111	3.132
Functional status (totally dependent)	5.118	0.608	43.110	3.798	1.734	7.534
Dyspnea at rest	5.606	1.816	17.307	2.21	0.346	7.743
Dyspnea moderate exertion	1.370	0.688	2.728	1.1	0.446	2.313
Sepsis (yes)	2.020	0.765	5.333	3.79	1.132	10.471
Weight loss (yes)	2.585	1.315	5.082	1.798	1.003	3.036
Disseminated cancer (yes)	2.736	1.520	4.925	3.087	1.887	4.835
BMI class (obese)	1.118	0.662	1.888	3.723	1.728	7.069
BMI class (underweight)	0.721	0.162	3.199	1.16	0.74	1.761
Steroid use (yes)	0.934	0.273	3.190	3.102	0.936	7.584
Albumin (g/dL) 2.8–3.5	2.190	1.127	4.254	2.091	1.406	3.066
Albumin (g/dL) ≤2.8	5.425	2.547	11.557	1.887	0.952	3.464
Albumin (g/dl) missing	1.533	0.854	2.752	0.775	0.272	1.729
C-statistic	0.822			0.782		
Hosmer–Lemeshow chi-square	10.946 ($P = 0.205$)			5.236 ($P = 0.631$)		

PD Variables	US/NSQIP			Japan/NCD		
		95% CI			95% CI	
	Odds Ratio	Lower CI	Upper CI	Odds Ratio	Lower CI	Upper CI
Age 60–70	1.933	1.085	3.442	1.706	0.977	3.193
Age 70–80	2.591	1.447	4.638	1.989	1.157	3.688
Age ≥80	3.280	1.656	6.498	2.462	1.317	4.844
Sex (male)	1.519	1.059	2.181	1.325	0.949	1.866
Bleeding disorder (yes)	1.013	0.401	2.556	1.829	1.033	3.025
Emergency status (yes)	0.651	0.084	5.056	4.389	1.666	9.566
Functional status (dependent)	1.378	0.415	4.572	2.394	1.44	3.789
Disseminated cancer (yes)	2.161	1.044	4.471	1.854	0.3	6.113
Hypertension (yes)	1.738	1.145	2.639	1.393	1.048	1.85
Albumin (g/dL) 2.8–3.5	1.520	0.994	2.324	1.233	0.885	1.696
Albumin (g/dL) ≤2.8	1.985	1.174	3.357	1.515	0.846	2.54
Albumin (g/dL) missing	1.030	0.486	2.182	1.643	0.747	3.209
CKD missing	1.665	0.448	6.192	2.148	0.69	5.471
CKD stage 3	1.183	0.736	1.903	1.421	1.031	1.969
CKD stage 4/5	3.020	1.133	8.044	2.054	1.057	3.704
Surgical complexity (adjacent organ)	2.281	1.234	4.214	1.796	0.986	3.018
Surgical complexity (vascular)	1.524	0.881	2.633	1.701	1.154	2.44
COPD (yes)	2.382	1.371	4.137	1.541	0.775	2.758
BMI class (obese)	1.905	1.306	2.779	1.53	0.535	3.434
BMI class (underweight)	1.823	0.712	4.668	0.659	0.404	1.023
C-statistic	0.719			0.782		
Hosmer–Lemeshow chi-square	8.619 ($P = 0.375$)			9.908 ($P = 0.272$)		

BMI = body mass index, CI = confidence interval, CKD = chronic kidney disease, COPD = chronic obstructive pulmonary disease, LAR = low anterior resection, NCD = National Clinical Database, NSQIP = National Surgical Quality Improvement Program, PD = pancreaticoduodenectomy, RH = right hemicolectomy, US = United States, WBC = white blood cell.

TABLE 5. Observed and Expected Mortality after RH, LAR, and PD

Mortality	US/NSQIP Model	Japan/NCD Model	Difference in Mortality Rate (%)
RH			
US/NSQIP mortality	(Observed)	(Expected)	
%	1.88	0.60	−68
C-statistic	0.840	0.789	
Hosmer–Lemeshow chi-square	5.866 ($P = 0.662$)	171.01 ($P < 0.001$)	
Japan/NCD mortality	(Expected)	(Observed)	
%	3.83	0.76	404
C-statistic	0.828	0.857	
Hosmer–Lemeshow chi-square	955.233($P < 0.001$)	11.243 ($P = 0.188$)	
LAR			
US/NSQIP mortality	(Observed)	(Expected)	
%	1.08	0.60	−44
C-statistic	0.822	0.786	
Hosmer–Lemeshow chi-square	10.946 ($P = 0.205$)	49.54 ($P < 0.001$)	
Japan/NCD mortality	(Expected)	(Observed)	
%	1.08	0.43	151
C-statistic	0.778	0.782	
Hosmer–Lemeshow chi-square	145.375 ($P < 0.001$)	5.236 ($P = 0.631$)	
PD			
US/NSQIP mortality	(Observed)	(Expected)	
%	2.57	2.41	−6
C-statistic	0.719	0.674	
Hosmer–Lemeshow chi-square	8.619 ($P = 0.375$)	8.817 ($P = 0.358$)	
Japan/NCD mortality	(Expected)	(Observed)	
%	4.23	1.35	213
C-statistic	0.540	0.782	
Hosmer–Lemeshow chi-square	366.217 ($P < 0.001$)	9.908 ($P = 0.272$)	

LAR = low anterior resection, NCD = National Clinical Database, NSQIP = National Surgical Quality Improvement Program, PD = pancreaticoduodenectomy, RH = right hemicolectomy, US = United States.

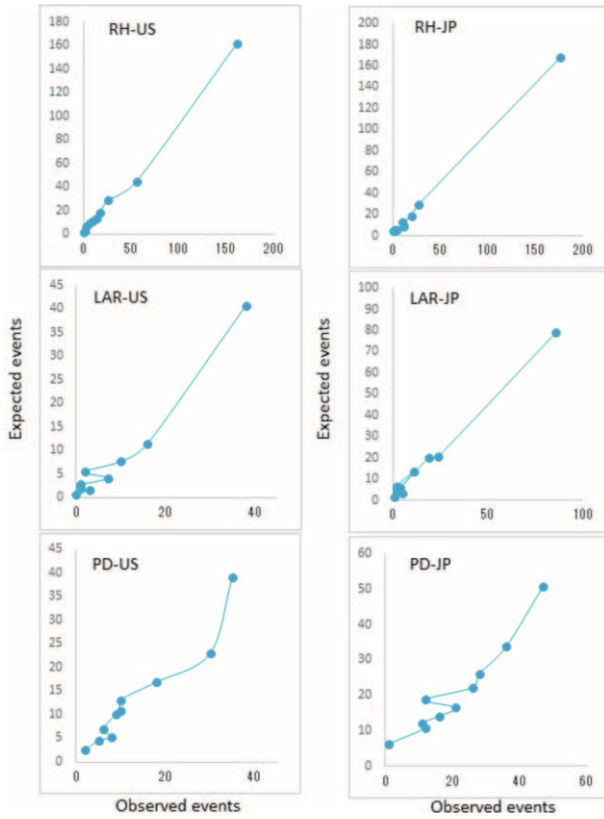

FIGURE 1. Calibration for 30-day mortality models for RH, LAR, and PD based on the US data using the US/ACS-NSQIP model (US) and the Japanese data using the Japan/NCD model (JP). ACS-NSQIP = American College of Surgeons National Surgical Quality Improvement Program, JP = Japan, LAR = low anterior resection, PD = pancreaticoduodenectomy, RH = right hemicolectomy, US = United States.

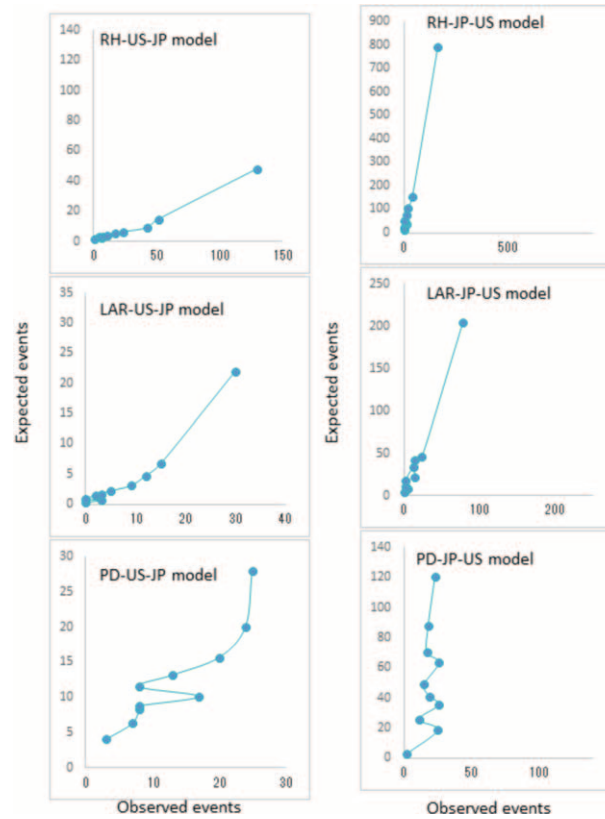

FIGURE 2. Calibration for 30-day mortality models for RH, LAR, and PD based on the US data using the Japan/NCD model (US-JP model) and the Japanese data using the US/ACS-NSQIP model (JP-US model). ACS-NSQIP = American College of Surgeons National Surgical Quality Improvement Program, JP = Japan, LAR = low anterior resection, PD = pancreaticoduodenectomy, RH = right hemicolectomy, US = United States.

used. These results indicate that risk models based on local data accurately predict mortality rate; however, risk models based on data from other countries are unable to accurately predict mortality rate. When evaluating the performance of a prediction model in adherence to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guideline,[26] investigators should pay attention to the discrepancy, involving the use of participant data collected by another country for external validation.

We considered reasons for this discrepancy. Differences in the prevalence of risk factors are unlikely to have a significant impact on model performance since we conducted a risk-adjusted analysis. However, there may be several risk factors not included in our model. These are likely to be ethnicity; operative information (operation time, amount of bleeding, and transfusion amount); and incidence and management of postoperative complications. Cytokine response has been shown to differ between races[27,28]; it is reasonably assumed that this difference may lead to different outcomes. Because the incidence of severe morbidity affects mortality, successful prophylactic management as a team may reduce the incidence of morbidity and decrease mortality rates.[10,29] Relatively longer hospital stays after surgery due to the insurance system in Japan[30] may protect patients with high morbidity after surgery, but this assumption needs to be fully assessed in future comparative studies.

This study should be interpreted with the appreciation of several limitations. We were unable to combine data from the 2 datasets due to the prohibition by NCD for security reasons. The backgrounds of the databases may be different. Although the NCD/Japan contains nearly 95% of surgical cases from all hospitals in Japan, ACS-NSQIP contains samples from selected hospitals in US only. This may be a source of bias if there was a difference in surgical practice or hospital procedural volume. Other differences in patient factors, including social, economic, and racial differences, have not been considered. Secondly, 30-day mortality was the only outcome studied. The 30-day mortality likely underestimates treatment-associated mortality by not including mortality occurring 30 days after operations. Thirdly, the impact of perioperative and postoperative complications that potentially affects surgical mortality are unknown due to a lack of data regarding these variables.

In conclusion, we found significantly different mortality rates, comorbidity prevalences, and procedural practices between Japan and the US. Risk-prediction models that can be reasonably used for both patient groups should be developed while recognizing that some risk predictors may be population-specific. This study demonstrates the feasibility and utility of

international collaborative research between Japan and the US, but risk models based on local data remain essential for quality assessment and improvement.

ACKNOWLEDGMENTS

We wish to thank all of the data managers and hospitals that participated in the NCD project and the ACS-NSQIP for their great efforts in data entry. In addition, we wish to express appreciation to all the people and academies that cooperated in this project.

REFERENCES

1. Hall BL, Hamilton BH, Richards K, et al. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg.* 2009;250:363–376.
2. Miyata H, Gotoh M, Hashimoto H, et al. Challenges and prospects of a clinical database linked to the board certification system. *Surg Today.* 2014;44:1991–1999.
3. Anazawa T, Miyata H, Gotoh M. Cancer registries in Japan: National Clinical Database and site-specific cancer registries. *Int J Clin Oncol.* 2015;20:5–10.
4. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg.* 2013;217:336–346.
5. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. *Ann Surg.* 2008;248:329–336.
6. User Guide for the 2013 Participant Use Data File, 2014. Available at: http://site.acsnsqip.org/participant-use-data-file/. Accessed January 1, 2015.
7. Dimick JB, Osborne NH, Hall BL, et al. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg.* 2010;210:503–508.
8. Anderson RP, Jin R, Grunkemeier GL. Understanding logistic regression analysis in clinical reports: an introduction. *Ann Thorac Surg.* 2003;75:753–757.
9. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med.* 2007;35:2052–2056.
10. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet.* 2012;380:1059–1065.
11. Munasinghe A, Markar SR, Mamidanna R, et al. Is it time to centralize high-risk cancer care in the United States? Comparison of outcomes of esophagectomy between England and the United States. *Ann Surg.* 2015;262:79–85.
12. Osborne NH, Nicholas LH, Ryan AM, et al. Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for Medicare beneficiaries. *JAMA.* 2015;313:496–504.
13. Etzioni DA, Wasif N, Dueck AC, et al. Association of hospital participation in a surgical outcomes monitoring program with inpatient complications and mortality. *JAMA.* 2015;313:505–511.
14. Lawson EH, Wang X, Cohen ME, et al. Morbidity and mortality after colorectal procedures: comparison of data from the American College of Surgeons case log system and the ACS NSQIP. *J Am Coll Surg.* 2011;212:1077–1085.
15. Konno H, Wakabayashi G, Udagawa H, et al. Annual Report of National Clinical Database in Gastroenterological Surgery 2011–2012. *Jpn J Gastroenterol Surg.* 2013;46:952–963.
16. Kenjo A, Miyata H, Gotoh M, et al. Risk stratification of 7,732 hepatectomy cases in 2011 from the National Clinical Database for Japan. *J Am Coll Surg.* 2014;218:412–422.
17. Kimura W, Miyata H, Gotoh M, et al. A pancreaticoduodenectomy risk model derived from 8575 cases from a national single-race population (Japanese) using a web-based data entry system: the 30-day and in-hospital mortality rates for pancreaticoduodenectomy. *Ann Surg.* 2014;259:773–780.
18. Kobayashi H, Miyata H, Gotoh M, et al. Risk model for right hemicolectomy based on 19,070 Japanese patients in the National Clinical Database. *J Gastroenterol.* 2014;49:1047–1055.
19. Matsubara N, Miyata H, Gotoh M, et al. Mortality after common rectal surgery in Japan: a study on low anterior resection from a newly established nationwide large-scale clinical database. *Dis Colon Rectum.* 2014;57:1075–1081.
20. Nakagoe T, Miyata H, Gotoh M, et al. Surgical risk model for acute diffuse peritonitis based on a Japanese nationwide database: an initial report of 30-day and operative mortality. *Surg Today.* 2014In press.
21. Takeuchi H, Miyata H, Gotoh M, et al. A risk model for esophagectomy using data of 5354 patients included in a Japanese nationwide web-based database. *Ann Surg.* 2014;260:259–266.
22. Watanabe M, Miyata H, Gotoh M, et al. Total gastrectomy risk model: data from 20,011 Japanese patients in a nationwide internet-based database. *Ann Surg.* 2014;260:1034–1039.
23. Cohen ME, Bilimoria KY, Ko CY, et al. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg.* 2009;208:1009–1016.
24. Greenblatt DY, Kelly KJ, Rajamanickam V, et al. Preoperative factors predict perioperative morbidity and mortality after pancreaticoduodenectomy. *Ann Surg Oncol.* 2011;18:2126–2135.
25. 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;217:833–842.
26. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg.* 2015;102:148–158.
27. Hassan MI, Aschner Y, Manning CH, et al. Racial differences in selected cytokine allelic and genotypic frequencies among healthy, pregnant women in North Carolina. *Cytokine.* 2003;21:10–16.
28. Ness RB, Haggerty CL, Harger G, et al. Differential distribution of allelic variants in cytokine genes among African Americans and White Americans. *Am J Epidemiol.* 2004;160:1033–1038.
29. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med.* 2009;360:491–499.
30. Ikegami N, Yoo BK, Hashimoto H, et al. Japanese universal health coverage: evolution, achievements, and challenges. *Lancet.* 2011;378:1106–1115.