Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons

Karl Y Bilimoria, MD, MS, FACS, Yaoming Liu, PhD, Jennifer L Paruch, MD, Lynn Zhou, PhD, Thomas E Kmiecik, PhD, Clifford Y Ko, MD, MS, MSHS, FACS, Mark E Cohen, PhD

BACKGROUND: Accurately estimating surgical risks is critical for shared decision making and informed consent. The Centers for Medicare and Medicaid Services may soon put forth a measure requiring surgeons to provide patients with patient-specific, empirically derived estimates of postoperative complications. Our objectives were to develop a universal surgical risk estimation tool, to compare performance of the universal vs previous procedure-specific surgical risk calculators, and to allow surgeons to empirically adjust the estimates of risk.

STUDY DESIGN: Using standardized clinical data from 393 ACS NSQIP hospitals, a web-based tool was developed to allow surgeons to easily enter 21 preoperative factors (demographics, comorbidities, procedure). Regression models were developed to predict 8 outcomes based on the preoperative risk factors. The universal model was compared with procedure-specific models. To incorporate surgeon input, a subjective surgeon adjustment score, allowing risk estimates to vary within the estimate's confidence interval, was introduced and tested with 80 surgeons using 10 case scenarios.

RESULTS: Based on 1,414,006 patients encompassing 1,557 unique CPT codes, a universal surgical risk calculator model was developed that had excellent performance for mortality (c-statistic = 0.944; Brier score = 0.011 [where scores approaching 0 are better]), morbidity (c-statistic = 0.816, Brier score = 0.069), and 6 additional complications (c-statistics > 0.8). Predictions were similarly robust for the universal calculator vs procedure-specific calculators (eg, colorectal). Surgeons demonstrated considerable agreement on the case scenario scoring (80% to 100% agreement), suggesting reliable score assignment between surgeons.

CONCLUSIONS: The ACS NSQIP surgical risk calculator is a decision-support tool based on reliable multi-institutional clinical data, which can be used to estimate the risks of most operations. The ACS NSQIP surgical risk calculator will allow clinicians and patients to make decisions using empirically derived, patient-specific postoperative risks. (J Am Coll Surg 2013;217: 833–842. © 2013 by the American College of Surgeons)

CME questions for this article available at http://jacscme.facs.org

Disclosure Information: Authors have nothing to disclose. Timothy J Eberlein, Editor-in-Chief, has nothing to disclose.

Funding: Supported in part by the Agency for Healthcare Research and Quality.

Presented at the 2013 Annual ACS NSQIP Conference, San Diego, CA, July 2013.

Received May 7, 2013; Revised July 8, 2013; Accepted July 10, 2013.

From the Division of Research and Optimal Patient Care, American College of Surgeons (Bilimoria, Liu, Paruch, Zhou, Ko, Cohen) and the Surgical Outcomes and Quality Improvement Center, Department of Surgery, Feinberg School of Medicine, Northwestern University (Bilimoria, Kmiecik), Chicago, IL; and the Department of Surgery, University of California, Los Angeles (UCLA) and VA Greater Los Angeles Healthcare System, Los Angeles, CA (Ko).

Correspondence address: Karl Y Bilimoria, MD, MS, FACS, Division of Research and Optimal Patient Care, American College of Surgeons, 633 N St Clair St, 22nd Floor, Chicago, IL 60611. email: kbilimoria@facs.org

Understanding the risks of surgery is clearly important for both patients and surgeons in the shared decision making process.[1-4] Informed consent requires that patients have a thorough understanding of the potential risks of surgery. Moreover, clinicians and patients also need information regarding surgical risks in order to make decisions on the type of operation or whether surgery should be performed at all. Importantly, the Centers for Medicare and Medicaid Services (CMS) may soon incentivize surgeons through the Physician Quality Reporting System to discuss empirically derived, patient-specific risks with the patient before every elective operation performed in the US.[5]

However, predicting postoperative risks and identifying patients at a higher risk of adverse events have traditionally been based on individual surgeon experience and

© 2013 by the American College of Surgeons
Published by Elsevier Inc.

ISSN 1072-7515/13/$36.00
http://dx.doi.org/10.1016/j.jamcollsurg.2013.07.385

augmented by published rates in the literature, either from single institution studies or clinical trials. Unfortunately, these estimates are typically not specific to an individual patient's risk factors.

The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) collects high-quality, standardized clinical data on preoperative risk factors and postoperative complications from more than 500 hospitals in the US.[6-8] These data are used to provide hospitals with risk-adjusted 30-day outcomes comparisons, and we have previously leveraged these data to develop a risk prediction tool.[9] However, our initial procedure-specific surgical risk calculators were limited in that only a few operations and outcomes were included.

There is an increasing need for and interest in being able to empirically estimate customized, patient-specific risks for virtually all surgical operations in a user-friendly format. The intended use would be to counsel patients and facilitate decision making for elective surgery in an office-based setting or to discuss risks for more emergent or urgent surgery in the inpatient setting. Our objective was to leverage the high-quality clinical data collected by ACS NSQIP in order to develop an effective surgical risk estimation tool, the ACS NSQIP surgical risk calculator (http://riskcalculator.facs.org); to compare the performance of the new universal surgical risk calculator for multiple specialties with our previous procedure-specific risk calculators; and to develop an approach for clinicians to reasonably and empirically adjust risk estimates based on their clinical judgment and experience.

METHODS

Data source and patients

Data were obtained from ACS NSQIP. As described extensively elsewhere,[6,8,10,11] ACS NSQIP collects reliable and validated data on patient demographics, laboratories, comorbidities, and 30-day postoperative outcomes for patients undergoing a broad range of operations across all surgical subspecialties, with the exception of transplant and trauma. Data are collected by trained and audited surgical clinical reviewers (SCR) at each individual hospital using data definitions that are standardized across all hospitals.[11] Thirty-day outcomes are ascertained from the medical record or patients are contacted after discharge. Outcomes are ascertained irrespective of whether the patient was an inpatient, outpatient, or admitted to another facility.[12]

From hospitals participating in ACS NSQIP, patients were identified who underwent operations from January 1, 2009 to June 30, 2012, spanning all surgical subspecialties. The performance of the universal surgical risk calculator, which encompasses many specialties and hundreds of operations, was compared with the performance of our previous procedure-specific risk calculators. This was done for individual multiple operations (eg, colectomy, laparoscopic colectomy, pancreatectomy), but because the results were comparable, we focused on the comparison with the colectomy risk calculator, as it is currently the most commonly used risk calculator. From the overall dataset, 88,334 cases were identified as colon operations based on primary Current Procedural Terminology (CPT) codes (44140, 44141, 44143, 44144, 44145, 44146, 44147, 44150, 44151, 44160, 44204, 44205, 44206, 44207, 44208, or 44210).

Preoperative risk factors

Preoperative risk factors to be used in calculating patient-specific risks of surgical events were selected a priori based on predictive value, routine availability to the surgeon before the operation, and clinical face validity. The variables used in the universal and colon-specific models were similar except for how adjustment for the procedure is performed (Table 1). Missing data were handled with imputation using the Buck's method per the standard ACS NSQIP modeling approach.[8] In previous procedure-specific risk calculators, the operations were grouped into surgery subtypes based on these CPT codes (6 groups for colectomy) and into surgical indication categories based on International Classification of Disease (ICD-9) codes (8 groups for colectomy). For the universal surgical risk calculator model, a CPT-specific linear risk (different for each outcome) replaced CPT procedure categories in the procedure-specific model, and the universal model did not include an indication variable. The individual CPT-specific linear risks were logit transformed predicted probabilities, from preliminary models in which CPT (2,805 different CPTs), as a random effect in a hierarchical model, was used to predict each outcome.

Risk predication models

Random intercept, fixed slope hierarchical models (using SAS GLIMMIX), which account for clustering of cases within hospitals and impose an empirical Bayes-type shrinkage adjustment, were used.[8] Only fixed (patient-level) effects were used for risk prediction, although this methodology would permit inclusion of hospital-specific effects in later versions. Models for 8 surgical outcomes were evaluated including mortality, morbidity (any of the following intraoperative or postoperative events: surgical site infection, wound disruption, pneumonia,

Table 1. ACS NSQIP Variables Used in the Prior Colon-Specific and the New Universal Surgical Risk Calculators

Variable	Categories	Colon-specific	Universal
Age group, y	<65, 65–74, 75-84, ≥85	✓	✓
Sex	Male, female	✓	✓
Functional status	Independent, partially dependent, totally dependent	✓	✓
Emergency case	Yes, no	✓	✓
ASA class	1 or 2, 3, 4, or 5	✓	✓
Steroid use for chronic condition	Yes, no	✓	✓
Ascites within 30 d preoperatively	Yes, no	✓	✓
System sepsis within 48 h preoperatively	None, SIRS, sepsis, septic shock	✓	✓
Ventilator dependent	Yes, no	✓	✓
Disseminated cancer	Yes, no	✓	✓
Diabetes	No, oral, insulin	✓	✓
Hypertension requiring medication	Yes, no	✓	✓
Previous cardiac event	Yes, no	✓	✓
Congestive heart failure in 30 d preoperatively	Yes, no	✓	✓
Dyspnea	Yes, no	✓	✓
Current smoker within 1 y	Yes, no	✓	✓
History of COPD	Yes, no	✓	✓
Dialysis	Yes, no	✓	✓
Acute renal failure	Yes, no	✓	✓
BMI class	Underweight, normal, overweight, obese 1, obese 2, obese 3	✓	✓
Colon surgery group (colectomy)	Partial lap with anastomosis, partial lap with ostomy, partial open with anastomosis, partial open with ostomy, total lap with ostomy, total open with ostomy	✓	
Indication for colon surgery	Diverticulitis, enteritis/colitis, hemorrhage, neoplasm, obstruction/perforation, vascular insufficiency, volvulus, other	✓	
CPT-specific linear risk	2,805 values		✓

ASA, American Society of Anesthesiologists; BMI, body mass index; SIRS, systemic inflammatory response syndrome.

unplanned intubation, pulmonary embolism, on ventilator > 48 hours, progressive renal insufficiency, acute renal failure, urinary tract infection, stroke/cerebral vascular accident, cardiac arrest, myocardial infarction, deep venous thrombosis, systemic sepsis), pneumonia, cardiac event (cardiac arrest or myocardial infarction), surgical site infection, urinary tract infection, deep venous thrombosis, and renal failure (progressive renal insufficiency or acute renal failure). The same approach was used for the universal and procedure-specific surgical risk calculators.

Comparison of universal vs procedure-specific models

Performance of the universal vs the procedure-specific surgical risk calculator model was evaluating using 3 metrics: the c-statistic; the Hosmer-Lemeshow (HL) statistic, though in revised form for graphical representation; and the Brier score. First, the c-statistic is a measure of discrimination also referred to as area under the receiver operating characteristic (ROC) curve (the sensitivity vs 1-specificity plot). For our purposes, the c-statistic is not the ideal index of performance because it is based on rank, focuses on category comparisons, and does not directly evaluate the accuracy of prediction.[13]

Second, the Hosmer-Lemeshow (HL) statistic is a measure of calibration that reflects bias in predicting risk across the range of risk.[14] If there is a tendency to over- or underestimate risk for different risk groups, the chi-square statistic will become larger. One problem with the Hosmer-Lemeshow test is that because it is (asymptotically) chi-square distributed, smaller deviations from perfect calibration will be statistically significant as the sample size gets larger. For this reason, we prefer to rely on a graphic representation rather than statistical significance. For ease of interpretation, we constructed graphs based on sequential groups having equal numbers of observed events rather than equal numbers of patients (low risk groups will, therefore, tend to have more patients than high risk groups).

Third, the Brier score is defined as the average squared difference between patients' predicted probabilities and observed outcomes (1 or 0 depending on event or nonevent).[15,16] Because the Brier score is computed from differences between actual events and predicted probabilities, it is usually more informative than the rank-based c-statistic. As a model's predicted scores approach 0 and 1 for nonevents and events, respectively, the Brier score will approach 0.0 (perfect prediction). Another useful Brier score benchmark is its value when the observed overall event rate is assigned to each patient. The value for estimates coming out of this "null model" allows one to evaluate the added predictive contribution from individual patient-level risk. The Brier score reflects discrimination and calibration simultaneously. For our purposes of evaluating accuracy of risk prediction and comparing the universal to the procedure-specific surgical risk calculators, the Brier score may be the most appropriate measure of model performance.

Results were also compared when 2 years of data were used to develop models that were validated on the third year of data. No important differences were observed in these results, which is consistent with expectations for stable results due to the very large sample size used for both model creation and validation steps.

Surgeon risk adjustment

Because the postoperative complication risks estimated by the surgical risk calculator may not capture every potential comorbidity, we sought to create an ad hoc opportunity for surgeons to reasonably modify the estimated risks. Given that there is a degree of uncertainty (confidence interval) around the estimated risks, we allowed clinicians to increase the risk of surgery within the confidence interval for each specific CPT. The default estimated risk from the model is designated as a surgeon adjustment score (SAS) of 1. The surgeon can then increase the risks to an SAS of 2 (+1 standard deviation of predicted risks for that CPT) or to an SAS of 3 (+2 standard deviations). If the estimated risk for an individual patient is greater than 2 standard deviations above the mean risk for that CPT (risk associated with a SAS of 3), then the estimated risk from the model would remain unmodified because the patient's predicted risk is already higher than the maximum SAS modification.

Next, agreement among surgeons on the SAS was assessed using 10 clinical scenarios. The 10 scenarios varied in complexity and included additional comorbidities or complicating factors not included in the surgical risk calculator. Scenarios were piloted with 10 surgeons and then applied to a convenience sample of 80 surgeons evaluated at the 2012 American College of Surgeons Clinical Congress. Surgeon agreement on the predicted risks of the cases and SAS assignments were assessed using a 5-point scale. Agreement was defined as the proportion of scores that were the mode or within 1 point of the mode score. All analyses were performed using SAS, version 9.1.3. This study was reviewed by the Northwestern University Institutional Review Board.

RESULTS

Universal ACS NSQIP surgical risk calculator

From 393 ACS NSQIP hospitals, 1,414,006 patients were identified and used in developing the universal surgical risk calculator (Appendix 1, online only). Subspecialties included general surgery, gynecology, neurosurgery, orthopaedics, otolaryngology, plastic surgery, cardiothoracic surgery, urology, and vascular surgery (Table 1). Of all the cases, 99% were within CPT codes that had at least 25 cases reported and 94% were within CPT codes that had at least 200 cases reported. Although the model was developed using all cases, the surgical risk calculator reports risk estimates only if more than 25 cases could be used as the basis of determining the risk of that CPT code (Appendix 2, online only). This resulted in 1,557 unique CPT codes being available in the ACS surgical risk calculator. The universal surgical risk calculator included 21 predictor variables and reported on eight 30-day postoperative outcomes.

Event rates ranged from 0.6% for renal failure to 9.0% for overall morbidity (Table 2). C-statistics for the universal model were reasonable and ranged from 0.806 for UTI to 0.944 for mortality. The Brier scores were generally small, reflecting good prediction. The relationship between observed and predicted rates for 633 CPT codes with 200 or more cases for 3 representative outcomes is shown in Figure 1. Points center on the diagonal, suggesting that the universal model provides estimates consistent with observed rates, though greater dispersion from the diagonal would be seen if we included all CPT codes, where smaller counts would likely result in less reliable estimates. When interquartile ranges (IQR) for each CPT are added on the right column of the figure, it is possible to see the assumed influence of differences in patient risk within CPT codes.

Universal vs procedure-specific models

For the 88,334 identified colorectal patients, predictions from the colon-specific model were compared with those from the universal model applied to the same colon surgery patients (Table 3). For both the c-statistic and the Brier score, the colon-specific model was slightly better than the universal model. Calibration for both

Table 2. Complication Rates and Model Statistics for the Universal Surgical Risk Calculator Model (n = 1,414,006)

Outcomes	Events, n (%)	c-statistic*	Brier score*	Brier score (null model)*
Mortality	18,909 (1.3)	0.944	0.011	0.0132
Morbidity	126,921 (9.0)	0.816	0.069	0.0817
Pneumonia	17,183 (1.2)	0.870	0.011	0.0120
Cardiac	10,676 (0.8)	0.895	0.007	0.0075
SSI	50,611 (3.6)	0.817	0.032	0.0346
UTI	20,777 (1.5)	0.806	0.014	0.0145
VTE	12,671 (0.9)	0.819	0.009	0.0089
Renal failure	8,996 (0.6)	0.903	0.006	0.0063

*The c-statistic is a measure of discrimination, that ranges from 0.5 (chance) to 1.0 (perfect), which reflects the extent to which cases are properly classified as having or not having an event. The Brier score describes the averaged squared difference between patients' predicted probability and the actual outcome (0 for a nonevent and 1 for an event). If all patients without an event are assigned a predicted probability of 0, and all patients with an event are assigned a predicted probability of 1, the Brier score will be 0, indicating perfect prediction. For the null model Brier score, the overall event rate (say, 0.05 for a 5% mortality rate) is assigned to each patient. This indexes predictive value when using this information but no patient-specific factors.
SSI, surgical site infection; UTI, urinary tract infection; VTE, deep venous thrombosis.

the procedure-specific and universal models was similar and acceptable (Appendix 3, online only).

To further evaluate discriminatory performance, we looked at mean predicted probabilities separately for colon-surgery patients who did and did not experience an event. Predictions for colon-specific and universal models were very similar (Appendix 4, online only). There is a small tendency for the universal model to yield slightly larger predicted probabilities than the colon-specific model. This results in somewhat greater error (based on Brier score) for the universal model compared with the colon-specific model when there is not an event, but less error when there is an event. Compared with the colon-specific model, the universal model slightly overestimates risk for patients who do not experience an event, but appropriately assigns higher risk to patients who do experience an event. The universal model was more accurate in 23 of the 48 (47.9%) colon models presented. Among models with differences between colon-specific and universal model predicted rates of at least 0.01, the universal model was more accurate in 13 of 17 models (76.5%). We expect that similar performance attributes would apply when our models are applied to the dataset where we do not know the actual outcome.

Surgeon risk estimate adjustment

Based on the uncertainty of the risk estimates, the Surgeon Adjustment Score (SAS) allows clinicians to increase the estimated risks. For example, for CPT 44140 (hemicolectomy), the mean predicted mortality was 1.2% (SAS = 1), the mean + 1 standard deviation was 9.6% (SAS = 2), and the mean + 2 standard deviations was 17.9% (SAS = 3) (Table 4). Surgeons demonstrated considerable agreement on the case scenario SAS scoring, ranging from 80% to 100% agreement.

DISCUSSION

Accurate estimates of postoperative complication risks are undoubtedly important to patients, caregivers, and clinicians. However, there is no risk estimation tool currently available that covers nearly all operations across multiple subspecialties. Using the standardized validated, high-quality clinical data from ACS NSQIP, we developed the universal surgical risk calculator, which had good discrimination and calibration and also performed with only slight differences from previous procedure-specific risk calculators. Moreover, a Surgeon Adjustment Score was included to allow clinicians to reasonably modify estimated risks based on their impression of the patient. The ACS NSQIP surgical risk calculator offers benefits to patients, their families, and health care providers.

Universal surgical risk calculator

Existing risk calculators, including those previously developed by our team, have several limitations. First, they are often for a single procedure, indication, or complication (eg, colectomy, pancreatitis, surgical site infection).[9,17-20] Second, currently available risk estimation tools are often from a single or small number of institutions that specialize in the procedure in question. They may also be from a multi-institutional clinical trial, but these are highly selected patients typically at specialized centers. Third, some risk estimation tools are based on administrative data, which are not as accurate as clinical data, and are often limited to only complications that occur in the inpatient setting.[12,21] Fourth, many do not use statistically acceptable modeling approaches. To address these limitations, we created a surgical risk estimation tool, the ACS NSQIP surgical risk calculator, which allows surgeons to easily enter 21 preoperative risk factors

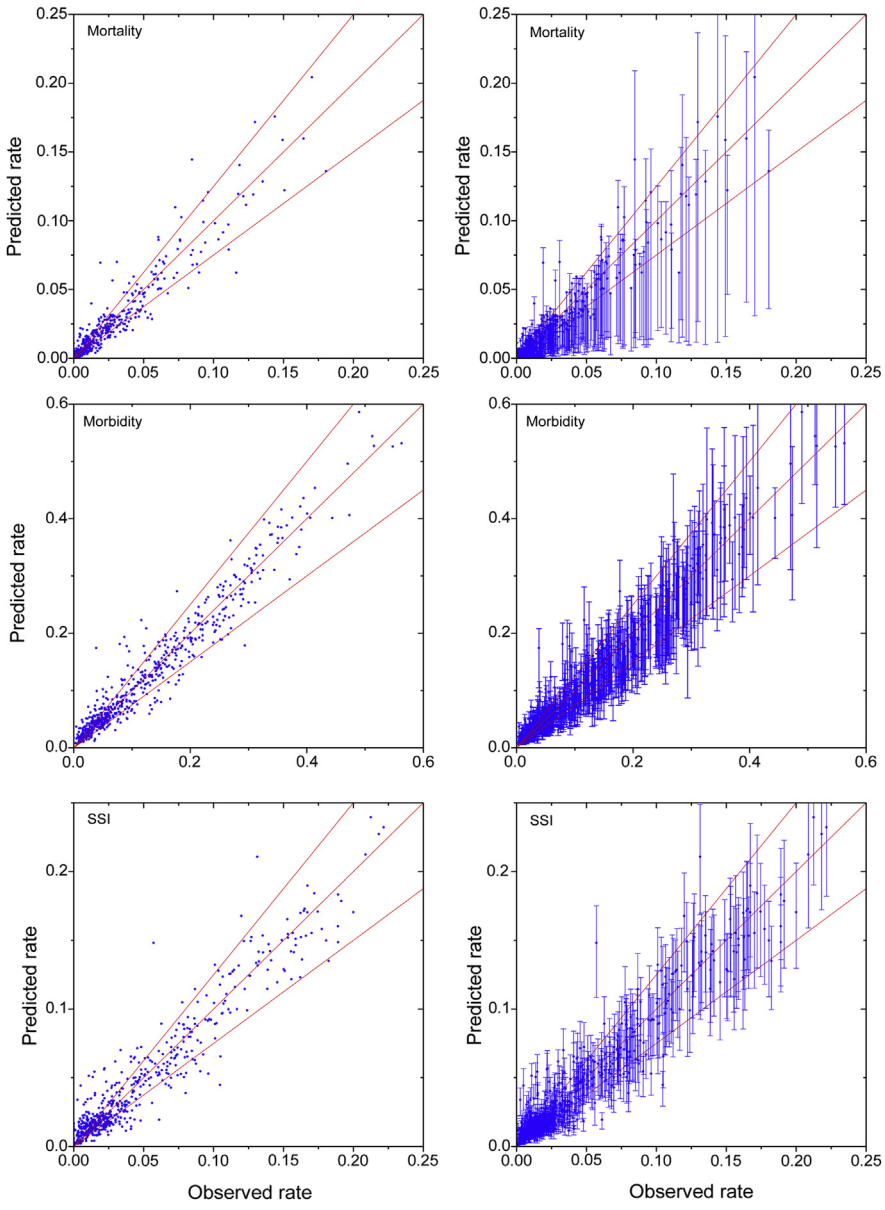


Figure 1. Plots of mean observed rates and universal-model-predicted rates for Current Procedural Terminology (CPT) codes. *Only CPTs with at least 200 cases are included. The limit lines on either side of the diagonal representing perfect agreement (observed = predicted) are set at ±25% (from the diagonal). Scatter plots are in the left column; the right column includes the interquartile range (IQR) around the predicted values for each CPT group. The IQR spread represents differences in patient-predicted risk within each CPT code.

for a specific patient (Fig. 2A) and to then receive a report of the customized, patient-specific risks of surgery (Fig. 2B). The surgical risk calculator can be used for more than 1,500 CPTs across all surgical subspecialties. Discrimination and calibration were found to be reasonable in comparison to other currently available risk calculators.

Procedure-specific vs universal risk calculator

An important aspect of creating the universal surgical risk calculator was to ensure that it performed equally well as procedure-specific risk calculators developed by our team and by other groups. Intuitively, one may believe that a calculator designed for a specific operation and indication (eg, pancreaticoduodenectomy for cancer) would

Table 3. Comparison of Colon-Specific and Universal Surgical Risk Calculator Models for Colon Surgery Outcomes and Indications

			c-statistic		Brier score		
Variable	n	Events, n (%)	Colon-specific model	Universal model	Colon-specific model	Universal model	Null model
Mortality	88,334	3,508 (4.0)	0.9193	0.9142	0.0291	0.0299	0.0381
Morbidity	88,334	21,523 (24.4)	0.7265	0.7150	0.1598	0.1625	0.1843
Pneumonia	88,107	2,881 (3.3)	0.8195	0.8087	0.0296	0.0300	0.0316
Cardiac	88,334	1,613 (1.8)	0.8441	0.8345	0.0171	0.0172	0.0179
SSI	87,531	10,408 (11.9)	0.6712	0.6494	0.1004	0.1015	0.1048
UTI	88,053	2,815 (3.2)	0.7270	0.7067	0.0302	0.0304	0.0309
VTE	88,053	2,010 (2.3)	0.7384	0.7203	0.0218	0.0218	0.0222
Renal failure	88,185	1,516 (1.7)	0.8202	0.8087	0.0162	0.0164	0.0169
Mortality							
Diverticulitis	16,329	214 (1.3)	0.9428	0.9402	0.0103	0.0107	0.0129
Enteritis colitis	5,056	99 (2.0)	0.9511	0.9517	0.0149	0.0162	0.0192
Hemorrhage	1,263	117 (9.3)	0.8150	0.8115	0.0732	0.0749	0.0841
Neoplasm	41,784	858 (2.1)	0.8613	0.8566	0.0182	0.0184	0.0201
Obstruction/perforation	5,391	615 (11.4)	0.8611	0.8539	0.0781	0.0810	0.1011
Other	14,655	992 (6.8)	0.9201	0.9162	0.0451	0.0463	0.0631
Vascular insufficiency	2,012	502 (25.0)	0.7937	0.7797	0.1510	0.1579	0.1873
Volvulus	1,844	111 (6.0)	0.8728	0.8691	0.0473	0.0476	0.0566
Morbidity							
Diverticulitis	16,329	3,117 (19.1)	0.6992	0.6883	0.1405	0.1434	0.1544
Enteritis colitis	5,056	1,232 (24.4)	0.6894	0.6769	0.1658	0.1693	0.1843
Hemorrhage	1,263	512 (40.5)	0.7031	0.6898	0.2119	0.2194	0.2410
Neoplasm	41,784	8,362 (20.0)	0.6899	0.6771	0.1479	0.1495	0.1601
Obstruction/perforation	5,391	2,327 (43.2)	0.7119	0.7029	0.2128	0.2188	0.2453
Other	14,655	4,420 (30.2)	0.7354	0.7262	0.1786	0.1819	0.2106
Vascular insufficiency	2,012	1,049 (52.1)	0.6970	0.6865	0.2201	0.2255	0.2495
Volvulus	1844	504 (27.3)	0.7014	0.6884	0.1774	0.1812	0.1986

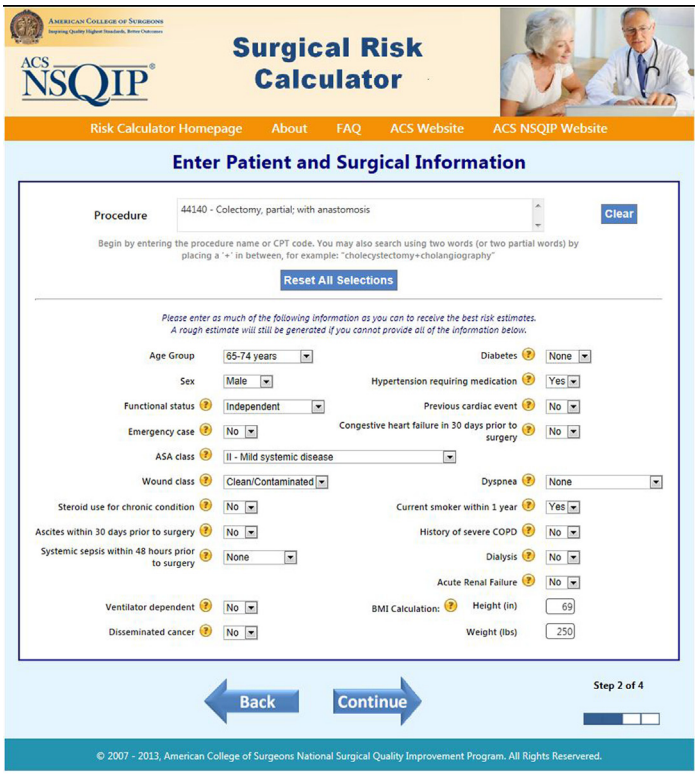
SSI, surgical site infection; UTI, urinary tract infection; VTE, deep venous thrombosis.

offer better predictions than a universal calculator. We found that the universal surgical risk calculator generally performed similarly to our previous procedure-specific surgical risk calculators. The differences, if any, were quite small. The main disadvantage of the procedure-specific risk calculators is that we would have to create more than 800 models (8 complications for approximately 100 procedure groups) to encompass the most commonly used CPT codes, as opposed to simply creating 8 universal models. So, any slight decrement in prediction is offset by the opportunity to have a universal prediction tool.

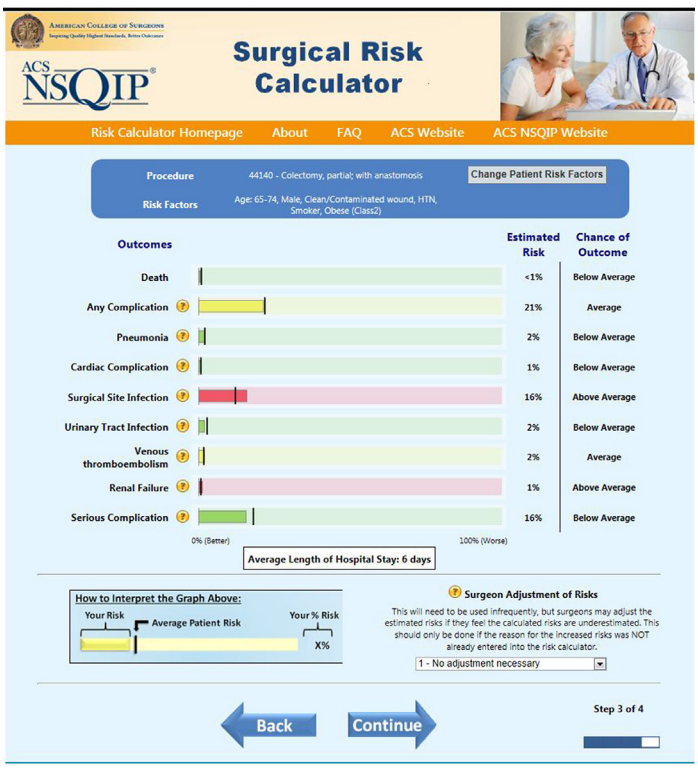
Table 4. Examples of Predicted Probability of the Event and +1 or +2 Standard Deviations for CPT 44140 (Hemicolectomy) for the Surgeon Adjustment Score

Event	Minimum, %	SAS = 1, mean, %	SAS = 2, mean+1.0σ, %	SAS = 3, mean+2σ, %	Maximum, %
Morbidity	1.9	22.1	34.3	46.5	91.7
Mortality	0.1	1.2	9.6	17.9	95.7
SSI	1.5	12.4	17.6	22.8	51.0
Cardiac	0.1	1.1	4.1	7.1	41.2
VTE	0.5	2.0	3.4	4.8	17.2
Pneumonia	0.2	2.3	6.0	9.6	48.5
UTI	0.3	3.1	5.7	8.2	34.1

The estimated risk is designated as an SAS of 1. The surgeon can then increase the risks to an SAS of 2 (+1 standard deviation) or to an SAS of 3 (+2 standard deviations). If the estimated risk was greater than the risk associated with an SAS of 2 or 3, then the estimated risk remains unchanged.
SAS, Surgeon Adjustment Score; SSI, surgical site infection; UTI, urinary tract infection; VTE, deep venous thrombosis.



A



B

Figure 2. Screenshots of the ACS NSQIP Surgical Risk Calculator (http://riskcalculator.facs.org). (A) Risk factor entry screen. (B) Report screen (includes recently added models for serious morbidity and length of stay).

Surgeon Adjustment Score

The surgical risk calculator is based on 21 preoperative risk factors, but there could be many other factors that increase a patient's risk of postoperative complications. Thus, the surgeon's evaluation may be that the risks are higher than those predicted by the surgical risk calculator.[22] The Surgeon Adjustment Score (SAS) allows clinicians to increase the risk of surgery within the confidence interval for the predicted risk. This will allow surgeons to better counsel patients using both the modeled estimate and that estimate adjusted by the clinician's evaluation and impression. Surgeon agreement on the clinical scenarios corresponding to the Surgeon Adjustment Scores was quite high, suggesting that surgeons will use the SAS similarly, but additional evaluation will be needed. Although surgeons may believe that this adjustment is necessary, it might be the case that comorbidities not entered into the surgical risk calculator are already reflected, indirectly, in the other predictors. Under this scenario, the inclusion of extra modeling adjustments may not actually offer any improvement in the prediction.

Limitations

First, the data for the surgical risk calculator come from 393 hospitals that participate in ACS NSQIP—approximately 10% of hospitals in the US. However, ACS NSQIP hospitals perform approximately 30% of all operations in the US. Second, only clinical preoperative variables collected by ACS NSQIP could be used in estimating postoperative risks. Although additional variables may seem clinically important, the vast majority of risk adjustment can be done with 10 or fewer variables.[23,24] Third, because the SAS is a modification that is performed to augment the estimated risks outside of the modeling process, there is no quantitative evidence that these adjusted risks are more accurate. Nevertheless, the SAS offers the surgeon an option to better counsel patients. Fourth, the risks estimated by the surgical risk calculator are from nearly 400 hospitals and thousands of surgeons. Certainly, there will be variation in outcomes by hospital and by surgeon. Overall, however, adjustment for the surgeon and/or hospital is a relatively small component because patient comorbidities are more important in predicting postoperative risks.[25] Fifth, our current universal surgical risk calculator does not account for the indication for the procedure. However, the universal and procedure-specific calculators perform similarly despite the exclusion of indication as other variables may account for some of the risk by indication (eg, CPT, emergency case). Because surgical indication is certainly important for the clinical face validity of the surgical risk calculator, we have already begun work to include this in the next iteration. Finally, predictions from the surgical risk calculator are only estimates and are best used under the guidance of a clinician. Additional work will need to focus on how to best present the information to patients and to assess whether patients find the information understandable and useful.

CONCLUSIONS

Risk assessments are already used as quality indicators in other fields. The Centers for Medicare and Medicaid Services are considering requiring clinicians to discuss empirically derived, customized risk assessments with patients before any elective operation. The ACS NSQIP surgical risk calculator offers surgeons the ability to quickly and easily estimate important, patient-specific postoperative risks and present the information in a patient-friendly format. Discussions of these risks may better inform patient and caregiver expectations, help surgeons and patients decide which operation to perform, and even offer insights about whether the operative risk is prohibitive. The surgical risk calculator offers an opportunity to improve shared decision making and informed consent, and therefore, improve patient care.

Author Contributions

Study conception and design: Bilimoria, Ko, Cohen
Acquisition of data: Bilimoria, Liu, Paruch, Zhou, Kmiecik, Ko, Cohen
Analysis and interpretation of data: Bilimoria, Liu, Paruch, Zhou, Kmiecik, Ko, Cohen
Drafting of manuscript: Bilimoria, Cohen
Critical revision: Bilimoria, Liu, Paruch, Zhou, Kmiecik, Ko, Cohen

REFERENCES

1. Walker JA. What is the effect of preoperative information on patient satisfaction? Br J Nurs 2007;16:27–32.
2. Leclercq WK, Keulers BJ, Scheltinga MR, et al. A review of surgical informed consent: past, present, and future. A quest to help patients make better decisions. World J Surg 2010; 34:1406–1415.
3. Knops AM, Legemate DA, Goossens A, et al. Decision aids for patients facing a surgical treatment decision: a systematic review and meta-analysis. Ann Surg 2013;257:860–866.
4. Schenker Y, Fernandez A, Sudore R, et al. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: a systematic review. Med Decis Making 2011;31:151–173.
5. Measure Application Partnership Pre-Rule Making Report 2013. Available at: http://www.qualityforum.org/Publications/2013/02/MAP_Pre-Rulemaking_Report_-_February_2013.aspx. Accessed May 2, 2013.

6. 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.
7. Ingraham AM, Richards KE, Hall BL, et al. Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. Adv Surg 2010;44:251–267.
8. 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 April 26 [Epub ahead of print].
9. 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.
10. Hall BL, Richards K, Ingraham A, et al. New approaches to the National Surgical Quality Improvement Program: the American College of Surgeons experience. Am J Surg 2009; 198:S56–S62.
11. Shiloach M, Frencher SK Jr, Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. J Am Coll Surg 2010;210:6–16.
12. Bilimoria KY, Cohen ME, Ingraham AM, et al. Effect of post-discharge morbidity and mortality on comparisons of hospital surgical quality. Ann Surg 2010;252:183–190.
13. Merkow RP, Hall BL, Cohen ME, et al. Relevance of the c-statistic when evaluating risk-adjustment models in surgery. J Am Coll Surg 2012;214:822–830.
14. 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.
15. Brier GW. Verification of forcasts expressed in terms of probability. Monthly Weather Review 1950;78:1–3.
16. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–138.
17. Finks JF, Kole KL, Yenumula PR, et al. Predicting risk for serious complications with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative. Ann Surg 2011; 254:633–640.
18. Gupta PK, Franck C, Miller WJ, et al. Development and validation of a bariatric surgery morbidity risk calculator using the prospective, multicenter NSQIP dataset. J Am Coll Surg 2011; 212:301–309.
19. Gupta PK, Gupta H, Sundaram A, et al. Development and validation of a risk calculator for prediction of cardiac risk after surgery. Circulation 2011;124:381–387.
20. Parikh P, Shiloach M, Cohen ME, et al. Pancreatectomy risk calculator: an ACS-NSQIP resource. HPB (Oxford) 2010;12: 488–497.
21. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. Ann Surg 2012;256:973–981.
22. Woodfield JC, Pettigrew RA, Plank LD, et al. Accuracy of the surgeons' clinical prediction of perioperative complications using a visual analog scale. World J Surg 2007;31:1912–1920.
23. Schilling PL, Hallstrom BR, Birkmeyer JD, et al. Prioritizing perioperative quality improvement in orthopaedic surgery. J Bone Joint Surg Am 2010;92:1884–1889.
24. 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.
25. Ko CY, Chang JT, Chaudhry S, et al. Are high-volume surgeons and hospitals the most important predictors of in-hospital outcome for colon cancer resection? Surgery 2002; 132:268–273.

Appendix 1. Patient and Hospital Characteristics for the Universal ACS NSQIP Surgical Risk Calculator

Variable	Category	n	%
Patients		1,414,006	
Sex	Female	809,990	57.3
	Male	604,016	42.7
Functional status	Independent	1,344,929	95.1
	Partially dependent	52,500	3.7
	Totally dependent	16,577	1.2
Emergency case	No	1,253,610	88.7
	Yes	160,396	11.3
ASA class	1/2-No/mild disturb	777,115	55
	3-Severe disturb	541,404	38.3
	4/5-Life threat/Moribund	95,487	6.8
Steroid use for chronic condition	No	1,370,710	96.9
	Yes	43,296	3.1
Ascites within 30 days prior to surgery	No	1,405,309	99.4
	Yes	8,697	0.6
Systemic sepsis within 48 hours before surgery	None	1,316,645	93.1
	SIRS	55,090	3.9
	Sepsis	33,725	2.4
	Septic shock	8,546	0.6
Ventilator dependent	No	1,403,887	99.3
	Yes	10,119	0.7
Disseminated cancer	No	1,385,833	98
	Yes	28,173	2
Diabetes	Insulin	82,581	5.8
	No	1,198,826	84.8
	Oral	132,599	9.4
Hypertension requiring medication	No	755,035	53.4
	Yes	658,971	46.6
Previous cardiac event	No	1,309,750	92.6
	Yes	104,256	7.4
Congestive heart failure in 30 days before surgery	No	1,401,593	99.1
	Yes	12,413	0.9
Dyspnea	At rest	15,571	1.1
	Moderate exertion	110,720	7.8
	No	1,287,715	91.1
Current smoker	No	1,141,684	80.7
	Yes	272,322	19.3
History of severe COPD	No	1,345,737	95.2
	Yes	68,269	4.8
Dialysis	No	1,391,177	98.4
	Yes	22,829	1.6
Acute renal failure	No	1,406,903	99.5
	Yes	7,103	0.5
BMI class	Class 1obese	283,975	20.1
	Class 2 obese	142,527	10.1
	Class 3 obese	147,044	10.4
	Normal	367,871	26
	Overweight	444,283	31.4
	Underweight	28,306	2
Specialty	Cardiac surgery	11,170	0.8
	General surgery	840,071	59.4
	Gynecology	74,737	5.3
	Neurosurgery	44,603	3.2
	Orthopaedics	170,280	12
	Otolaryngology	32,489	2.3
	Vascular	137,678	9.7
	Plastics	29,284	2.1
	Thoracic	14,939	1.1
	Urology	58,749	4.2
Hospitals			
Teaching affiliation*	Major	147	39.0
	Minor	112	29.7
	Nonteaching	118	31.3
Number of total medical/surgical beds*	<100	19	5.5
	100–299	39	11.2
	300–499	139	40.1
	500+	150	43.2

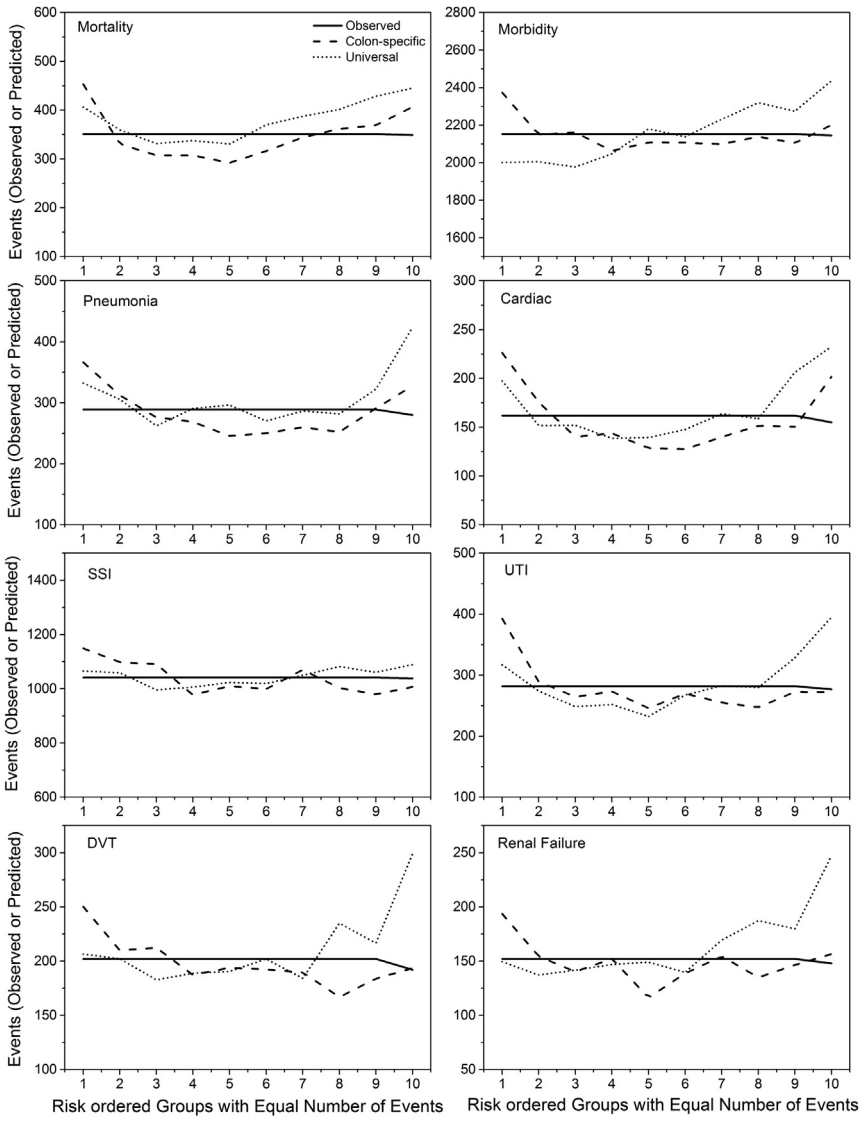
*n = 377 hospitals, but information on the characteristics for some hospitals is missing.

ASA, American Society of Anesthesiologists; BMI, body mass index; SIRS, systemic inflammatory response syndrome.

Appendix 2. Distribution of Current Procedural Terminology Codes and Cases

Number of Cases within each CPT ≥	Number of CPTs	Cases, n (%)
1	2,805	1,414,006 (100.0)
25	1,557	1,403,137 (99.2)
50	1,227	1,391,627 (98.4)
75	1,022	1,379,094 (97.5)
100	894	1,368,155 (96.8)
200	633	1,331,357 (94.2)
500	373	1,247,717 (88.2)
1,000	224	1,143,523 (80.9)

CPT, Current Procedural Terminology.



Appendix 3. Plots of Hosmer-Lemeshow Type Calibration for Procedure-Specific and Universal Models Applied to the Eight Studied Outcomes. DVT, deep venous thrombosis; SSI, surgical site infection; UTI, urinary tract infection.

Appendix 4. Predicted Rates Without and With Events for Colon-Specific vs Universal Models

	Patients without an event (observed rate = 0%)			Patients with an event (observed rate = 100%)		
	Mean predicted rates, %			Mean predicted rates, %		
Event	Colon-specific model	Universal model	Difference*	Colon-specific model	Universal model	Difference*
Outcomes						
Mortality	2.97	3.19	0.22	27.59	28.98	−1.39
Morbidity	21.15	20.38	−0.77	34.28	33.57	0.71
Pneumonia	3.03	3.25	0.23	9.47	9.92	−0.44
Cardiac	1.70	1.83	0.13	6.89	7.39	−0.50
SSI	11.40	11.30	−0.10	15.28	14.44	0.84
UTI	3.09	3.28	0.19	5.28	5.29	−0.01
VTE	2.20	2.35	0.16	4.12	4.37	−0.25
Renal failure	1.62	1.74	0.12	5.66	6.52	−0.86
Mortality for:						
Diverticulitis	1.08	1.83	0.75	18.12	28.34	−10.22
Enteritis colitis	1.48	2.39	0.91	25.36	39.23	−13.87
Hemorrhage	7.82	8.06	0.25	22.88	23.85	−0.97
Neoplasm	1.84	1.88	0.03	11.02	12.04	−1.02
Obstruction/perforation	8.57	8.62	0.05	33.12	32.75	0.37
Other	4.75	4.90	0.15	34.24	36.42	−2.17
Vascular insufficiency	18.96	16.95	−2.00	42.93	39.19	3.74
Volvulus	5.00	5.60	0.61	21.37	24.00	−2.64
Morbidity for:						
Diverticulitis	17.33	17.52	0.18	26.41	27.82	−1.41
Enteritis colitis	21.77	20.80	−0.97	32.36	32.08	0.28
Hemorrhage	35.75	30.16	−5.59	47.66	42.46	5.20
Neoplasm	18.55	18.10	−0.45	25.73	25.18	0.55
Obstruction/perforation	37.37	32.01	−5.36	50.88	46.02	4.86
Other	25.70	24.20	−1.51	40.47	40.18	0.29
Vascular insufficiency	45.52	43.89	−1.63	58.33	57.58	0.75
Volvulus	24.42	26.28	1.85	35.00	37.67	−2.67

*Positive values indicate that the colon-specific model is more accurate, negative values indicate that the universal model is more accurate. To achieve this structure, the directionality of subtractions (colon-specific minus universal vs universal minus colon-specific) was adjusted depending on which model yielded predictions closest to the true outcome. For this dataset, in contrast to programmatic use of the calculator in the medical setting, we know actual outcomes for each patient.

SSI, surgical site infection; UTI, urinary tract infection; VTE, venous thromboembolism.