Can the American College of Surgeons NSQIP Surgical Risk Calculator Accurately Predict Adverse Postoperative Outcomes in Emergency Abdominal Surgery? An Italian Multicenter Analysis

Giovanni Scotton, MD, Antonio La Greca, MD, FACS, Chiara Lirusso, MD, Diego Mariani, FEBSEMSURG, Mauro Zago, FACS, FEBSEMSURG, Massimo Chiarugi, FACS, Dario Tartaglia, MD, FACS, PhD, Nicolò de Manzini, MD, FRCS, Alan Biloslavo, MD, FACS, for the Emergency Surgery Collaborative

BACKGROUND: The American College of Surgeons NSQIP surgical risk calculator provides an estimation of 30-day postoperative adverse outcomes. It is useful in the identification of high-risk patients needing clinical optimization and supports the informed consent process. The purpose of this study is to validate its predictive value in the Italian emergency setting.

STUDY DESIGN: Six Italian institutions were included. Inclusion diagnoses were acute cholecystitis, appendicitis, gastrointestinal perforation or obstruction. Areas under the receiving operating characteristic curves, Brier score, Hosmer-Lemeshow index, and observed-to-expected event ratio were measured to assess both discrimination and calibration. Effect of the Surgeon Adjustment Score on calibration was then tested. A patient's personal risk ratio was obtained, and a cutoff was chosen to predict mortality with a high negative predicted value.

RESULTS: A total of 2,749 emergency procedures were considered for the analysis. The areas under the receiving operating characteristic curve were 0.932 for death (0.921 to 0.941, $p < 0.0001$; Brier 0.041) and 0.918 for discharge to nursing or rehabilitation facility (0.907 to 0.929, $p < 0.0001$; 0.070). Discrimination was also strong (area under the receiving operating characteristic curve >0.8) for renal failure, cardiac complication, pneumonia, venous thromboembolism, serious complication, and any complication. Brier score was informative (<0.25) for all the presented variables. The observed-to-expected event ratios were 1.0 for death and 0.8 for discharge to facility. For almost all other variables, there was a general risk underestimation, but the use of the Surgeon Adjustment Score permitted a better calibration of the model. A risk ratio >3.00 predicted the onset of death with sensitivity = 86%, specificity = 77%, and negative predicted value = 99%.

CONCLUSIONS: The American College of Surgeons NSQIP surgical risk calculator has proved to be a reliable predictor of adverse postoperative outcomes also in Italian emergency settings, with particular regard to mortality. We therefore recommend the use of the surgical risk calculator in the multidisciplinary care of patients undergoing emergency abdominal surgery. (J Am Coll Surg 2023;236:387–398. © 2022 by the American College of Surgeons. Published by Wolters Kluwer Health, Inc. All rights reserved.)

Disclosure Information: Nothing to disclose.

Members of the Emergency Surgery Collaborative who coauthored this article are listed in the Appendix.

Received June 9, 2022; Revised August 14, 2022; Accepted September 6, 2022.

From the Department of General Surgery, ASUITS, Cattinara Hospital, Trieste, Italy (Scotton, de Manzini, Biloslavo); Emergency and Trauma Surgery, Fondazione Policlinico Universitario A. Gemelli IRCCS Roma – Università Cattolica del Sacro Cuore, Italy (La Greca); Division of General Surgery, Department of Surgery, Santa Maria della Misericordia (ASUFC) University Hospital, Udine, Italy (Lirusso); Department of Emergency General Surgery, Legnano Hospital, ASST Ovest Milanese, Legnano, MI, Italy (Mariani); General and Emergency Surgery Department, A. Manzoni Hospital, Lecco, Italy (Zago); and General, Emergency and Trauma Surgery, Pisa University Hospital, Pisa, Italy (Chiarugi, Tartaglia).

Correspondence address: Giovanni Scotton, MD, Department of General Surgery, ASUGI Cattinara Hospital, Strada di Fiume 447, 34149 Trieste (TS) Italy. email: gioscotton@hotmail.it

Supplemental digital content is available for this article.

Abbreviations and Acronyms

ACS	=	American College of Surgeons
AUC	=	areas under the ROC curves
H-L	=	Hosmer-Lemeshow index
IQR	=	Interquartile range
O/E	=	Observed to expected event ratio
ROC	=	Receiving operating characteristic
RR	=	Risk ratio
SAS	=	Surgeon Adjustment Score
SRC	=	Surgical risk calculator
SSI	=	Surgical site infection
UTI	=	Urinary tract infection
VTE	=	Venous thromboembolism

Emergency surgery represents a unique scenario, totally different from elective surgery where procedures and strategies can be accurately planned and controlled.[1] In emergency cases, optimization efforts may be very hard due to lack of time, leading to increased stress for the patient, family, and surgeon.[2] In such a setting, the surgeon needs to face a great variability of situations, where similar diagnoses can present with different clinical states and can quickly transit to a decompensated state. Although emergency procedures significantly contribute to the total volume of general surgery procedures, surgical quality improvement efforts have often focused on elective general surgery outcomes. In addition to the greater intrinsic risk of emergency procedures, hospital performance is not always highly consistent across emergency and elective general surgery patients.[3] Nevertheless, through past years perioperative mortality and morbidity have significantly improved, thanks to the advances in anesthetic care, with perioperative complications remaining largely limited to the group of high-risk patients.[4] Because the opportunity for a true preoperative optimization is limited, an early recognition of high-risk patients is crucial to facilitate surgical decision-making and improve perioperative management. Therefore, the assessment of surgical risk represents the first step in planning every possible action that can contribute to mitigating it.[5] It can be essential in the choice of an interventional approach (definitive, damage control, palliative) and to guide postoperative care (intensive care unit, individualized and targeted therapies). Risk stratification represents an essential element in the preoperative informed consent process, providing the patient with ample information about possible outcomes, alternatives in treatment and their risks.[6]

The perioperative risk stratification depends on multiple interacting factors related to patient-specific characteristics, the specific surgical procedure, and anesthetic care. The surgeon's risk perception, based both on the surgeon's personal experience and emotional, social, and cultural factors, may significantly differ from the real risk.[5] Thus, the personal perception of risk can have a negative impact on the decision-making process.

During the past year, the creation of a great number of risk stratification tools have permitted an increasing objectification of perioperative risk, also allowing a fair comparison between centers.[7] Perioperative risk factors and procedure type impact postoperative morbidity and mortality differently in emergency surgery vs nonemergency surgery. Therefore, evidence suggested the need to differentiate predictive models in emergency and elective surgeries.[8]

The American College of Surgeons NSQIP Surgical Risk Calculator (ASC NSQIP SRC) arouses particular interest in the surgical community, because it permits accurate procedure-specific outcomes predictions within 30 days after elective and emergency surgery.[9] The ASC NSQIP SRC has been tested in different surgical specialties with variable results, especially considering surgical outcomes in elective surgical procedures.[10-16] Because insufficient data had been published on the application of the Surgical Risk Calculator (SRC) in emergency general surgery, in 2020 we conducted a pilot study of external validation of the calculator on a cohort of 317 emergency procedures from a single surgical center.[17] We found that the SRC was accurate in predicting adverse postoperative outcomes such as serious complications, death, and discharge to a nursing or rehabilitation facility. We introduced the concept of risk ratio (RR) cutoff, a practical method to forecast the onset of a specific complication in a single patient. The study had the big limitation that data were collected from a single surgical center with a small sample size.[17] We therefore decided to extend our cohort to include patient data collected at 6 Italian high-volume centers.

METHODS

Study design and data collection

The study was a multicenter retrospective analysis of data collected in 6 high-volume Italian general surgery departments. None of the institutions was a member of the NSQIP at the time of data collection. Each center obtained the ethical committee's approval in accordance with Italian law, and the study was conducted in accordance with Good Clinical Practice. Inclusion and exclusion criteria were previously reported and are synthetized in Table 1.[17]

Data regarding demographics, surgical procedures, and complications were collected de novo by retrospectively reviewing the prospectively maintained institutional databases. Particular attention was paid to adhering to

Table 1. Inclusion and Exclusion Criteria

Inclusion	Exclusion
Age 18 y or older	Aged <18 y
Admitted from January 2018 to the end of December 2019	Multiple simultaneous surgical procedures (>1 CPT code)
Admission diagnosis: acute cholecystitis, appendicitis, nontraumatic gastrointestinal perforation, and bowel obstruction*	Bowel obstruction or perforation as a complication of abdominal wall hernia or occurring as a postoperative complication
Emergency surgery is the first choice or a salvage strategy if nonoperative management failed	Missing preoperative or postoperative data

*Both neoplastic and nonneoplastic causes of obstruction and perforation were included.

the NSQIP data definitions reported on the official website. There was no use of preexisting NSQIP data sets. Missing data were recovered by revision of the electronic medical records by 3 members of each research team. Data on postoperative complications were captured from medical documentation recorded within 30 days after surgery (discharge letters, laboratory analyses, follow-up surgical examinations, notes, specialty consultations, and imaging). A unique code was assigned to each patient meeting the inclusion criteria, and records were maintained in a designated database, encrypted, and password protected.

Individual patient characteristics were manually entered into the ASC NSQIP SRC (https://riskcalculator.facs.org/RiskCalculator/PatientInfo.jsp). For the question "Are there other potential appropriate treatment options?" "none" was systematically selected, and risks were calculated for each of the 3 levels of the "Surgeon Adjustment of Risk" (1 – No adjustment necessary; 2 – Risk somewhat higher than estimated; and 3 – Risk significantly higher). The Surgeon Adjustment Score (SAS) was created to provide an opportunity for surgeons to adjust the risks in cases where the surgeon believes that the actual risk is much higher than the predicted risk based on risk factors not provided by the SRC. An SAS of 2 or 3 will move the predicted risk to the first and the second standard deviation, respectively, from the mean for that CPT code. If the patient's baseline risk is estimated by the SRC to be >2 SDs above the mean for the CPT code, the displayed risks will remain unchanged.

Last, the estimated patient personal risks ("Your risk") for each postoperative adverse event were recorded as a percentage.

Statistical analysis

Demographics and patients' characteristics were summarized as mean ± SD for continuous variables and percentage value for categorical variables. The length of postoperative hospital stay was expressed as median ± interquartile range (IQR). The performance of the calculator was studied by assessing discrimination and calibration. A brief explanation of these statistical methods was previously reported.[17] Discrimination was characterized using the area under (AUC) the receiver operating characteristic (ROC) curve or C statistic and was considered poor with an AUC of 0.6 to 0.69, adequate with an AUC of 0.7 to 0.79, strong with an AUC of 0.8 to 0.89, and excellent with an AUC of 0.9 to 1.0. A sample size of 300 patients (α 0.05, power 95%) was calculated to show that an AUC of 0.7 differs significantly from the null hypothesis value 0.5. Calibration was assessed with a graphical representation of the Hosmer-Lemeshow (H-L) test, the Brier score, and the observed-to-expected ratio (O/E). The H-L test assesses whether or not the observed event rates match expected event rates in subgroups of the model population. The null hypothesis is that the observed and expected proportions are the same ($p > 0.05$). The alternative hypothesis is that the observed and expected proportions are not the same ($p \le 0.05$). The Brier score reflects both discrimination and calibration of a model, and its values range between 0 and 1, with 0 indicating a perfect model of prediction. A risk prediction model with a Brier score of 0.25 or higher is considered as noninformative.[18]

C statistic, H-L test, Brier score, and O/E ratio were calculated for each of the 3 levels of the SAS (1 – No adjustment necessary; 2 – Risk somewhat higher than estimate; and 3 – Risk significantly higher) to demonstrate potential improvement in risk prediction accuracy.

We finally conducted a subanalysis related to variable mortality. The estimated patient personal risk was divided by the "Average Risk" provided by the SRC to calculate a personal RR. These RRs were used to generate an additional ROC curve. By referring to this curve, a range of all possible RR cutoff points was chosen to keep both sensitivity and specificity values at least >60%. We arbitrarily assumed that a sensitivity and specificity simultaneous value of a minimum >60% was indispensable to consider the RR cutoff range as informative. To calculate sensitivity and specificity, we then chose an intermediate value of this range to obtain a high negative predicted value to permit the exclusion of mortality in the single patient.

For all the presented statistical analysis, p values of <0.05 were considered statistically significant. Statistical analyses were performed using IBM SPSS Statistics version 21 (IBM Corporation, Armonk, NY) and MedCalc Statistical Software version 17.2 (MedCalc Software bvba, Ostend, Belgium).

RESULTS

From January 2018 to December 2019, a total of 2,749 patients satisfied the inclusion criteria and were included in the study. Preoperative diagnoses were acute cholecystitis, appendicitis, bowel obstruction, or gastrointestinal perforation. The contribution of each center is summarized in Table 2. Each center contributed with the following number of surgeons: 8 in Trieste, 7 in Udine, 7 in Legnano, 7 in Lecco, 5 in Pisa, and 3 in Rome. Some of them are specifically trained in emergency surgery, others in colorectal, hepato-pancreato-biliary, or upper gastrointestinal surgery. The most performed procedure was appendectomy (875 patients), followed by cholecystectomy (768 patients), interventions due to bowel obstruction (628 patients), and gastrointestinal perforation (478 patients). Specific performed procedures, with specific CPT code, are summarized in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A168. Patients' demographics and preoperative characteristics are summarized in **Supplemental Digital Content 2,** http://links.lww.com/JACS/A168. Observed complications are reported in Table 3.

Assessment of discrimination and calibration with surgeon adjustment score = 1 (no adjustment)

The assessment of discrimination and calibration was initially performed without the use of the surgeon adjustment score (SAS = 1) and the results are summarized in Table 4. Discriminative performance was excellent for death with an AUC of 0.932 (95% CI 0.921 to 0.941, $p < 0.0001$) and discharge to nursing or rehabilitation facility with an AUC of 0.918 (95% CI 0.907 to 0.929, $p < 0.0001$). Renal failure (AUC = 0.876), cardiac complication (AUC = 0.873), pneumonia (AUC = 0.842), venous thromboembolism (VTE; AUC = 0.842), serious complication (AUC = 0.833), and any complication (AUC = 0.816) showed strong discriminative performance. Return to operating room (AUC = 0.774), surgical site infection (SSI; AUC = 0.759), and urinary tract infection (UTI; AUC = 0.724; $p < 0.0001$) had adequate discriminative performance (Fig. 1). Discrimination was inadequate for ileus, anastomotic leak, readmission, and sepsis.

The Brier score reflects both discrimination and calibration. It was informative (<0.25) for all the considered outcomes, and particularly good (<0.05) for death (0.041), renal failure (0.028), cardiac complication (0.039), pneumonia (0.049), VTE (0.009), UTI (0.019), and sepsis (0.049).

For the analysis of the H-L test calibration plots were preferred, permitting a visual analysis in each of the different deciles of risk of the model population (see Fig. 2). The blue line approached and followed the black line of perfect prediction model for death, discharge to facility, serious complication, any complication, pneumonia, cardiac complication, return to operating room, renal failure, UTI, SSI, and sepsis, but with a general trend of underestimation of the expected events. For VTE, ileus, anastomotic leak, and readmission, the model showed bad calibration, even with an inversion of the curve.

The number of observed events matched the expected events for death (131 vs 133 cases, O/E = 1.0) and discharge to facility (291 vs 375 cases, O/E = 0.8), but, for the other outcomes, the lack in calibration was confirmed by the O/E ratios, as shown in Table 3. The SRC overestimated the number of hospital readmissions, VTE, and ileus, with a general underestimation of the remaining outcome events, particularly marked for cardiac

Table 2. Diagnosis and Center Contributions

Diagnosis	Trieste	Udine	Roma	Legnano	Lecco	Pisa	Total
Acute cholecystitis	174	150	202	135	52	55	768
Appendicitis	184	200	226	174	60	31	875
Bowel obstruction	149	185	160	107	27	—	628
Gastroduodenal	—	2	2	2	—	—	6
Small-bowel obstruction	86	130	107	62	18	—	403
Colorectal (neoplastic)	59	33	31	34	4	—	161
Colorectal (divert/volvulus)	4	20	20	9	5	—	58
Gastrointestinal perforation	97	136	113	75	11	46	478
Gastroduodenal	37	26	25	8	3	15	114
Colorectal (diverticular)	30	67	45	28	5	31	206
Colorectal (neoplastic)	17	23	22	11	—	—	73
Small bowel	13	20	21	28	3	—	85
Total	604	671	701	491	150	132	2749

Table 3. Observed vs Expected Complication

Complication	Observed n (%)	Expected n (%)	Observed/expected ratio
Serious complication	462 (16.8)	341 (12.4)	1.4
Any complication	650 (23.6)	424 (15.4)	1.5
Pneumonia	153 (5.6)	80 (2.9)	1.9
Cardiac complication	119 (4.3)	41 (1.5)	2.9
Surgical site infection	300 (10.9)	116 (4.2)	2.6
Urinary tract infection	51 (1.9)	36 (1.3)	1.4
Venous thromboembolism.	24 (0.9)	43 (1.6)	0.6
Renal failure (n = 2,647 patients)	84 (3.2)	26 (1.0)	3.2
Readmission	135 (4.9)	220 (8.0)	0.6
Return to operating room	168 (6.1)	89 (3.2)	1.9
Death	131 (4.8)	133 (4.8)	1.0
Discharge to facility (n = 2,624 patients)	291 (11.1)	375 (14.3)	0.8
Sepsis (n = 2,412 patients)	126 (5.2)	60 (2.5)	2.1
Ileus (n = 328 patients)	47 (14.3)	109 (33.2)	0.4
Anastomotic leak (n = 331 patients)	27 (8.2)	11 (3.3)	2.5

Table 4. Discrimination and Calibration with SAS = 1

Complication	Area under curve (95% CI)	p Value	Brier score
Death	0.932 (0.921–0.941)	<0.0001	0.041
Discharge to facility	0.918 (0.907–0.929)	<0.0001	0.070
Renal failure	0.876 (0.862–0.889)	<0.0001	0.028
Cardiac complication	0.873 (0.860–0.886)	<0.0001	0.039
Pneumonia	0.842 (0.828–0.856)	<0.0001	0.049
Venous thromboembolism	0.842 (0.827–0.856)	<0.0001	0.009
Serious complication	0.833 (0.818–0.847)	<0.0001	0.109
Any complication	0.816 (0.801–0.831)	<0.0001	0.138
Return to operating room	0.774 (0.757–0.790)	<0.0001	0.055
Surgical site infection	0.759 (0.742–0.776)	<0.0001	0.092
Urinary tract infection	0.724 (0.706–0.741)	<0.0001	0.019
Ileus (n = 328 patients)	0.551 (0.491–0.606)	0.342	0.179
Anastomotic leak (n = 331 patients)	0.543 (0.483–0.602)	0.564	0.060
Readmission	0.530 (0.511–0.550)	0.2018	0.054
Sepsis	0.569 (0.548–0.589)	0.0582	0.049

complication (O/E = 2.9), SSI (O/E = 2.6), and renal failure (O/E = 3.2).

Assessment of calibration with surgeon adjustment score = 2 and surgeon adjustment score = 3

The variables with an underestimation of the risk were then considered for recalculation after application of the surgeon adjustment score (SAS = 2 and SAS = 3). The results are summarized in Table 5. Asterisks represent the level of the SAS with the best O/E ratio. The effect of recalibration was graphically represented with calibration plots. Examples are illustrated in Figure 3. For most of the outcome variables, the best performance was obtained with an SAS = 3.

Interprocedural variability

The procedures with a minimum number of 140 patients were selected for separate analysis. Open and laparoscopic appendectomy (860 patients with CPT 44950, 44960, 44970), cholecystectomy (769 patients with CPT 47562, 47600), enterolysis (192 patients with CPT 44005), resection of small intestine with single anastomosis (149

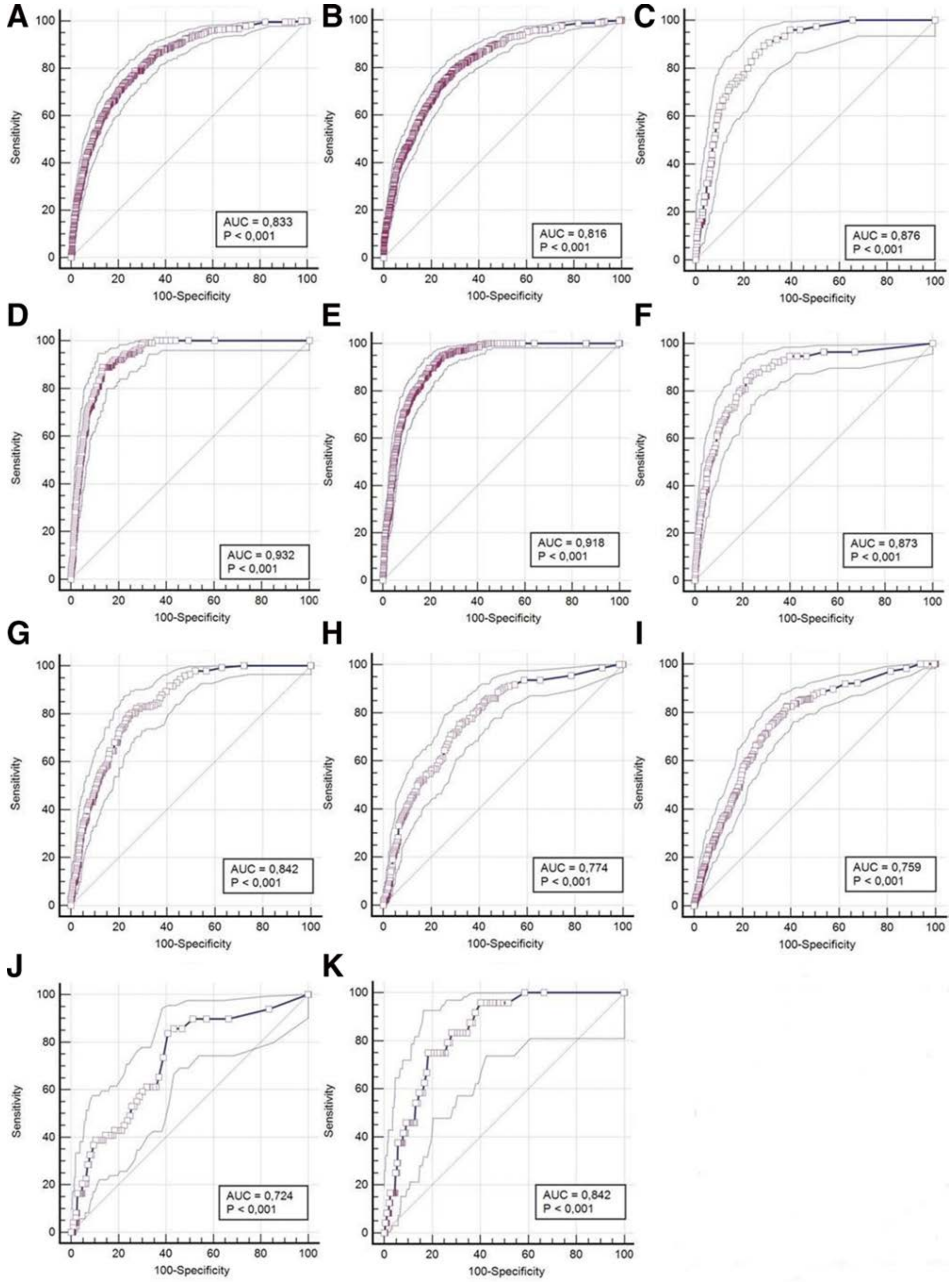

Figure 1. Receiver operating characteristic curves. (A) Serious complication. (B) Any complication. (C) Renal failure. (D) Death. (E) Discharge to facility. (F) Cardiac complication. (G) Pneumonia. (H) Return to operating room (OR). (I) Surgical site infection. (J) Urinary tract infection. (K) Venous thromboembolism. AUC, area under the curve.

patients with CPT 44120), and Hartmann procedure (142 patients with CPT 44143) showed different performances in terms of discrimination and calibration. Particularly good AUCs and Brier scores were calculated for appendectomy and cholecystectomy; these procedures had the lowest complication rate. Conversely, the SRC showed a general decline of performance with more complex procedures, while maintaining adequate discrimination and calibration values (see Table 6).

Differences between centers

To analyze the impact of the center on the performance of the SRC, the ROC curves and the O/E ratio were

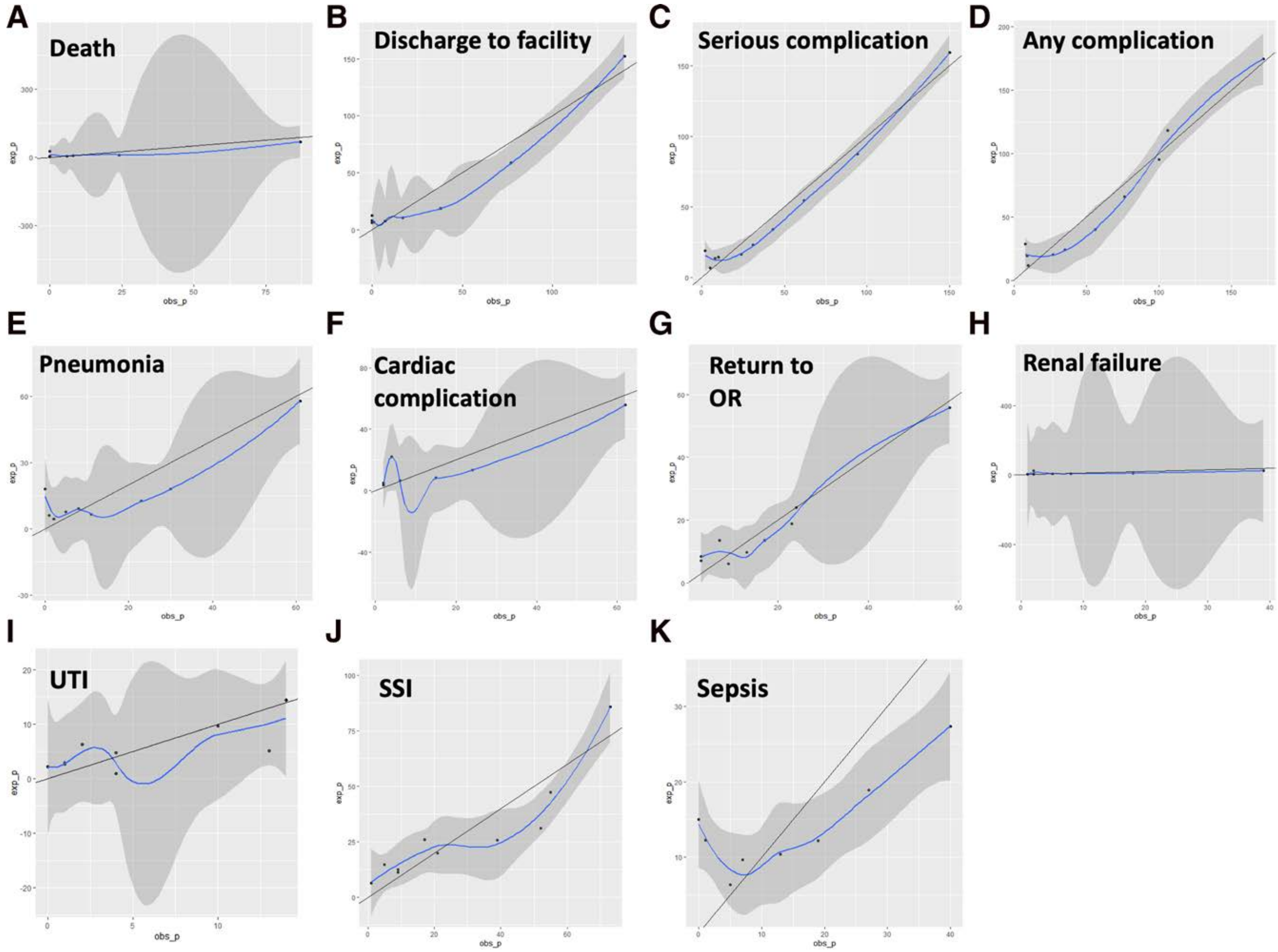

Figure 2. Hosmer-Lemeshow index test calibration plots. (A) Death. (B) Discharge to facility. (C) Serious complication. (D) Any complication. (E) Pneumonia. (F) Cardiac complication. (G) Return to operating room (OR). (H) Renal failure. (I) Urinary tract infection (UTI). (J) Surgical site infection (SSI). (K) Sepsis.

separately calculated. The discriminative power of the SRC was homogeneous between the centers. The calibration (E/O) showed a general underestimation of risk, but with greater variability between centers. **Supplemental Digital Content 3** (http://links.lww.com/JACS/A168) summarizes the differences for some outcomes of interest. The 2 centers with the lowest number of patients were excluded from this analysis due to the low number of events (Lecco and Pisa).

Length of stay

The median observed length of stay was 5 (IQR 3 to 9) days, compared with a predicted length of stay of 2 (IQR 1 to 7.5) days (Mood's Median test, p = 0.000). The use of the SAS did not improve the prediction of stay: 2.5 (IQR 1.5 to 11.5) vs 5 (IQR 3 to 9) days with an SAS of 2 (p = 0.002) and 3 (IQR 2 to 14.5) vs 5 (IQR 3 to 9) days with an SAS of 3 (p = 0.015).

Mortality risk ratio cutoff

The calculated personal RRs for death were used to generate an additional ROC curve (Fig. 4) with an AUC of 0.857 (95% CI 0.843 to 0.870, p < 0.0001). A cutoff value range in the RR between 1.00 and 5.00 was chosen to maximize sensitivity and specificity (the reported range refers to cutoff values whose related sensitivity and specificity are simultaneously >0.6) and an RR cutoff of 3.00 (that is the intermediate value of the previous range) was considered. An RR >3.00 predicted the onset of death with a sensitivity of 86%, a specificity of 77%, and an negative predicted value of 99% (Fig. 4). In a patient who presented a RR <3, mortality was consequently excluded with a confidence of 99%.

DISCUSSION

Surgical risk calculators stand as useful tools that can facilitate surgical decision-making and support dialogue

Table 5. Effect of the Surgeon Adjustment Score

	Observed	SAS = 2		SAS = 3	
Complication	**n (%)**	**Expected**	**O/E**	**Expected**	**O/E**
Serious	462 (16.8)	435*	1.06*	521	0.89
Any	650 (23.6)	539	1.20	649*	1.00*
Pneumonia	153 (5.6)	120	1.28	155*	0.99*
Cardiac	119 (4.3)	76	1.57	105*	1.13*
Surgical site infection	300 (10.9)	145	2.07	194*	1.55*
Urinary tract infection	51 (1.9)	43	1.19	50*	1.02*
Renal failure	84 (3.2)	54	1.56	79*	1.06*
Return to operating room	168 (6.1)	111	1.51	130*	1.29*
Sepsis	126 (5.2)	84	1.5	101*	1.25*
Anastomotic leak (n = 331 patients)	27 (8.2)	13	2.08	16*	1.69*

*Level of the SAS with the best O/E ratio.
O/E, observed/expected; SAS, Surgeon Adjustment Score.

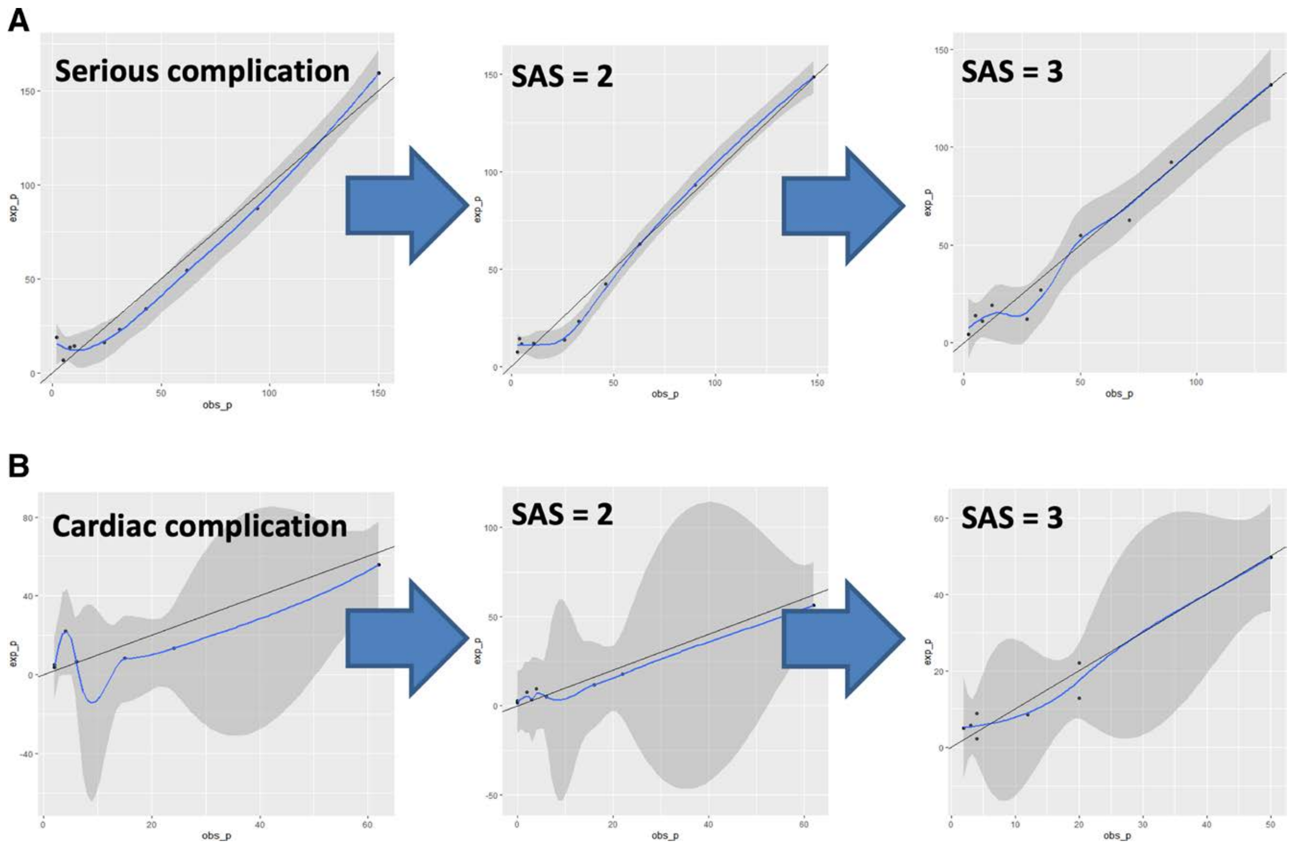

Figure 3. Hosmer-Lemeshow index test calibration plots after surgeon adjustment score (SAS) application. Recalibration improves the predictive accuracy of the surgical risk calculator with the progressive approach of the blue line to the black line of perfect prediction model. (A) Serious complication. (B) Cardiac complication.

with patients and their families during the informed consent process regarding alternative managements, risks, and adverse outcomes. They provide the statistical basis for the early identification of high-risk patients who will need an effective preoperative optimization, permitting the notification of medical teams such as intensive care and anesthetics in supporting the surgical team managing these patients, given their risk of sudden deterioration.

A recent review examined all the existing risk stratification tools that can be applied to emergency surgical care.[7]

Table 6. Interprocedural Variability in Discrimination and Calibration

	Appendectomy (n = 860)		Cholecystectomy (n = 769)		Enterolysis (n = 192)		Small-bowel resection (n = 149)		Hartmann (n = 142)	
Complication	**AUC**	**Brier**	**AUC**	**Brier**	**AUC**	**Brier**	**AUC**	**Brier**	**AUC**	**Brier**
Death	0.975	0.0018	0.986	0.0087	0.762	0.0490	0.758	0.1000	0.746	0.1609
Discharge to facility	0.988	0.0109	0.909	0.0519	0.776	0.1368	0.845	0.1413	0.724	0.2150
Serious complication	0.776	0.0449	0.809	0.0774	0.788	0.1593	0.663	0.2326	0.711	0.2504
Pneumonia	0.877	0.0089	0.833	0.0294	0.835	0.0998	0.729	0.1063	0.653	0.1544
Cardiac	0.752	0.0069	0.823	0.2156	0.731	0.0712	0.822	0.0788	0.749	0.1735
Renal failure	0.984	0.0066	0.887	0.0098	0.691	0.0466	0.651	0.0771	0.677	0.1423

AUC, area under the curve.

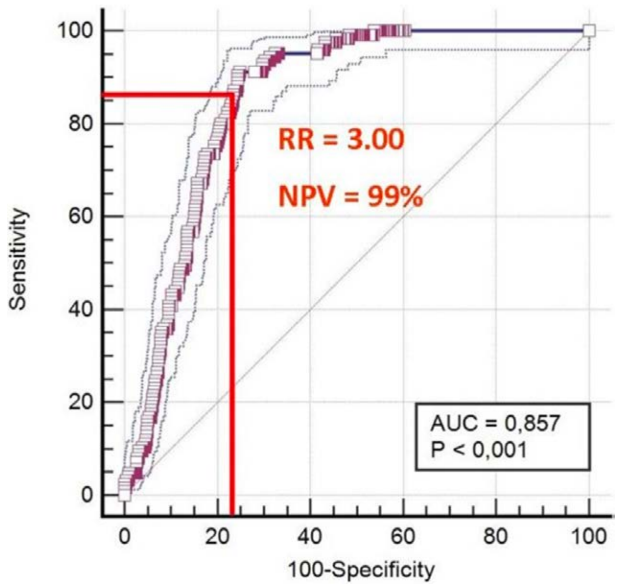

Figure 4. Mortality risk ratio cutoff. AUC, area under the curve; NPV, negative predictive value; RR, risk ratio.

The authors considered the ASC NSQIP SRC to be the best fit for the definition of the ideal scoring tool at this time, because it accurately quantifies both morbidity and mortality in the emergency setting, with readily obtainable objective data that can be used in the early phases of decision-making and care.

In the analyzed population, these authors found a slight underestimation of risk in the emergency group compared with the elective one. However, these differences were small and did not affect the mortality estimate, where the ratio of observed to predicted deaths was 1.03.[7]

One of the most comprehensive analyses of this issue was published in 2016 by Hyder and colleagues.[19] In this work, 56,942 patients undergoing urgent surgical procedures were compared with 136,311 patients undergoing elective surgeries. The analysis showed a slight underestimation of the risk in emergency patients compared with the elective ones, more accentuated in patients undergoing gastrointestinal procedures. The difference was attributed to the greater clinical heterogeneity of patients and the wider range of performed procedures. Nevertheless, the results did not discourage the use of the ASC NSQIP SRC in emergency high-risk patients.[19]

The calculator was developed initially from cases exclusively collected within US NSQIP hospitals, including both elective and emergency cases. It has not yet been validated for acute general surgery in a wide cohort of patients outside the US, such as in the Italian setting. We previously published a single-center external validation on a cohort of 317 emergency procedures, demonstrating accurate predictions of complication rates.[17] These results encouraged the planning of the current study to overcome the intrinsic limitations of the previous publication, such as the small sample size and its monocentric nature.

This study involves 6 high-volume general and emergency surgery departments in northern and central Italy, with a total number of 2,749 emergency procedures. The ASC NSQIP SRC demonstrated excellent discrimination for death (AUC = 0.932) and discharge to facility (AUC = 0.918), surpassing the results of our previous publication. With particular regard to death, its prediction represents the best indicator of the performance of the calculator, being intrinsically less affected by data collection inaccuracies or misinterpretation. Brier score (0.041) and calibration plot were excellent for death, with an O/E ratio of 1.0. Such predictions were obtained with an SAS = 1, demonstrating that no adjustment was necessary for this variable in the Italian setting. This result is of fundamental importance, because death represents the most feared complication and the most decisive in the choice between surgery and alternative treatments.

Similar results were recently obtained by an Australian meta-analysis that included 6 studies with a wide geographical distribution (from Asia, Europe, the US, New Zealand, and Australia) and involved emergency intra-abdominal procedures such as emergency laparotomy, appendectomy, and cholecystectomy.[20] A total of 1,835 patients were included in the study. Overall, the performance of SRC in its ability to predict mortality was accurate with an O/E ratio of 1.06 (95% CI 0.74 to 1.51). Nevertheless, there was statistically significant heterogeneity between the 6 studies, demonstrating a variable performance of the calculator.[20]

With respect to the other outcomes, our study showed strong discrimination (AUC >0.8) for renal failure, cardiac complication, pneumonia, VTE, serious complication and any complication, and adequate discrimination (AUC >0.7) for return to operating room, SSI, and UTI. With an SAS = 1, Brier score was informative (<0.25) for all the considered outcomes, but O/E ratio showed a general underestimation of the risk (see Table 3), particularly accentuated for cardiac complications, SSI, and renal failure. This lack in calibration was confirmed by the H-L plots. The use of the SAS permitted the best recalibration of the model for each single variable in our population, in particular, an SAS of 2 was necessary to obtain good predictions for serious complication, and an SAS of 3 was necessary to obtain good predictions for any complication, pneumonia, cardiac complication, SSI, UTI, renal failure, and return to the operating room.

Ileus, anastomotic leak, readmission, and sepsis showed bad discrimination and/or calibration, and the application of the SAS produced no improvement in the prediction rate.

The most frequent surgeries in our population were laparoscopic or open appendectomy (860 procedures, 31.3% of patients), laparoscopic or open cholecystectomy (769 procedures, 30.0% of patients), enterolysis (192 procedures, 7.0% of patients), small-bowel resection (149 procedures, 5.4% of patients), and Hartmann's procedure (142 procedures, 5.2% of patients). All remaining surgeries (637 procedures for gastrointestinal perforation or occlusion, for a total of 23.2% of patients) included 40 different abdominal surgical procedures (40 CPT codes), thus demonstrating a wide variability of procedures for a relatively small number of diagnoses. A subanalysis was consequently possible only for the groups with the highest sample size.

Calibration and discrimination showed some interprocedural variability. The best results were obtained for appendectomy and cholecystectomy. These 2 groups showed the highest sample size and the lowest number of observed complications. Both these conditions account for the very high AUC values (up to 0.975 for mortality in appendectomy and up to 0.986 for mortality in cholecystectomy) and very low Brier score values. Other procedures with a smaller sample size showed worse discrimination and calibration values, although they remained at adequate levels. In accordance with our findings, Golden and colleagues[21] published data on 1,693 acute care surgical procedures performed during a 5-year period. They found that the SRC had good discriminative performance in predicting both serious and any complication (AUC of 0.81 and 0.79, respectively) when considering the population as a single group. However, the accuracy was lower when analyzing different procedures individually. Therefore, subanalysis of smaller groups is not informative, because the number of events in each group (ie deaths) will consequently became insufficient to accurately evaluate calibration and discrimination. Cohen and colleagues[22] recommended that a minimum number of at least 100 events would be best to produce reliable estimates with small standard errors.

The determination of the RR transforms an absolute risk (expressed as percentage) to a relative risk (related to the average risk of a complication in the population undergoing the specific procedure). In this way, the only measurable parameter remains discrimination, overcoming the possible lack of calibration of the model. The selection of an RR cutoff permits the surgeon to answer the question, "Is it possible to exclude a postoperative complication?" instead of the question, "What probability has the patient of developing a complication?" This simple calculation provides a more practical method to exclude the onset of a specific outcome in a single patient. Its use can be useful in the discussion between surgeon and patient to reassure the patient that a particularly feared complication will not occur. For example, we found that an RR <3.00 excluded the onset of death with an negative predicted value of 99%.

Our study has limitations. First , although the sample size of 2,749 surgical procedures far exceeds that of 317 surgeries in our previous study, it remains small when compared with the >4.3 million surgeries performed in the 780 US hospitals originally involved in the design of the SRC. Two centers showed a smaller sample size that did not allow for an adequate center-specific analysis of the SRC performance. The aim of the multicenter study is to overcome the limitations of the single center. Discrimination was excellent and homogeneously distributed between centers. Nevertheless, calibration (expressed as O/E) showed more variability. These data could reflect small differences in center quality or in surgeons' skills and competences. As a consequence, the application of the SAS should be tailored to the specific results of the single center and ideally of the single surgeon. This could be the object of further investigation. Second, the average quality of Italian surgical centers could significantly differ from the NSQIP average. Therefore, what we consider as an underestimation of the risk may actually reflect a real higher rate of nonfatal postoperative complications. If that were the case, however, we would likewise expect an underestimation of the mortality rate. Therefore, other factors may influence these results, such as socioeconomic and cultural differences existing between NSQIP patients and international patients.[20,23] For example, length of stay was significantly longer than expected in our cohort, but the hospital readmission rate was conversely significantly lower. These data definitely reflect the NSQIP hospitals' efforts toward patients' early discharge, but at the cost of a higher risk of readmission. Nonclinical factors contribute to sustained international differences in length of stay, and differences in risk prediction cannot be considered as a failure of the SRC. These factors may include professional or cultural norms, differing payment schemes, and access to long-term care facilities. For example, shorter hospital stays reduce the burden of medical fees and increase the bed turnover rate, which in turn increases the profit margin of hospitals, while lowering the overall social costs. Finally, inadequate documentation and data collection may have negatively influenced the risk assessment.

CONCLUSIONS

The ASC NSQIP SRC is effective in the prediction of postoperative outcomes in patients undergoing emergency general procedures. The SRC performance was excellent in predicting mortality and discharge to nursing or rehabilitation facilities. It showed good discrimination in estimating renal failure, cardiac complications, pneumonia, VTE, serious complications and any complications, but with a general underestimation of the expected events. The application of the SAS permitted a satisfactory recalibration for most variables. The performance was conversely inadequate for postoperative ileus, anastomotic fistula, readmission, sepsis, and length of stay. The introduction of an RR cutoff offers the surgeon a practical method to predict the onset of a specific complication in a single patient. Based on the results of our study, we recommend the use of the ASC NSQIP SRC in general emergency surgery even outside the US.

APPENDIX

Members of the Emergency Surgery Collaborative: Riccardo Penazzi, MD; Davide Cosola, MD; Department of General Surgery, ASUITS, Cattinara Hospital, Trieste, Italy; Marta di Grezia, MD; Emergency and Trauma Surgery, Fondazione Policlinico Universitario A. Gemelli IRCCS Roma – Università Cattolica del Sacro Cuore, Italy; Sara Crestale, MD; Roberto Silvestro, Division of General Surgery, Department of Surgery, Santa Maria della Misericordia (ASUFC) University Hospital, Udine, Italy; MD; Giulia Lo Bianco, MD; Giulia Bonfanti, MD; General and Emergency Surgery Department, A. Manzoni Hospital, Lecco, Italy.

Author Contributions

Conceptualization: Scotton, Tartaglia, Biloslavo
Data curation: Scotton, La Greca, Lirusso, Mariani, Zago, Chiarugi
Formal analysis: Scotton
Methodology: Scotton, Biloslavo
Supervision: Chiarugi, de Manzini, Biloslavo
Writing – original draft: Scotton
Writing – review & editing: La Greca, Lirusso, Mariani, Zago, Chiarugi, Biloslavo

REFERENCES

1. Bohnen JD, Ramly EP, Sangji NF, et al. Perioperative risk factors impact outcomes in emergency versus nonemergency surgery differently: time to separate our national risk-adjustment models? J Trauma Acute Care Surg 2016;81:122–130.
2. Havens JM, Peetz AB, Do WS, et al. The excess morbidity and mortality of emergency general surgery. J Trauma Acute Care Surg 2015;78:306–311.
3. Ingraham AM, Cohen ME, Raval MV, et al. Comparison of hospital performance in emergency versus elective general surgery operations at 198 hospitals. J Am Coll Surg 2011;212:20–28.e1.
4. Bose S, Talmor D. Who is a high-risk surgical patient? Curr Opin Crit Care 2018;24:547–553.
5. Lacey J, Khan N, Oliver CM. Perioperative risk. Br J Hosp Med (Lond) 2017;78:616–621.
6. Cocanour CS. Informed consent-it's more than a signature on a piece of paper. Am J Surg 2017;214:993–997.
7. Havens JM, Columbus AB, Seshadri AJ, et al. Risk stratification tools in emergency general surgery. Trauma Surg Acute Care Open 2018;3:e000160.
8. Bohnen JD, Ramly EP, Sangji NF, et al. Perioperative risk factors impact outcomes in emergency versus nonemergency surgery differently: time to separate our national risk-adjustment models? J Trauma Acute Care Surg 2016;81:122–130.
9. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013;217:833–42.e1.
10. Sahara K, Paredes AZ, Merath K, et al. Evaluation of the ACS NSQIP surgical risk calculator in elderly patients undergoing hepatectomy for hepatocellular carcinoma. J Gastrointest Surg 2020;24:551–559.
11. Vaziri S, Wilson J, Abbatematteo J, et al. Predictive performance of the american college of surgeons universal risk calculator in neurosurgical patients. J Neurosurg 2018;128:942–947.
12. Slump J, Ferguson PC, Wunder JS, et al. Can the ACS-NSQIP surgical risk calculator predict post-operative complications in

patients undergoing flap reconstruction following soft tissue sarcoma resection? J Surg Oncol 2016;114:570–575.
13. Wingert NC, Gotoff J, Parrilla E, et al. The ACS NSQIP risk calculator is a fair predictor of acute periprosthetic joint infection. Clin Orthop Relat Res 2016;474:1643–1648.
14. Teoh D, Halloway RN, Heim J, et al. Evaluation of the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator in gynecologic oncology patients undergoing minimally invasive surgery. J Minim Invasive Gynecol 2017;24:48–54.
15. Lone Z, Hall S, Terakawa T, et al. Accuracy of American College of Surgeons National Surgical Quality Improvement Program universal surgical risk calculator in predicting complications following robot-assisted radical cystectomy at a national comprehensive cancer center. J Endourol 2019;33:383–388.
16. Keller DS, Ho JW, Mercadel AJ, et al. Are we taking a risk with risk assessment tools? evaluating the relationship between NSQIP and the ACS risk calculator in colorectal surgery. Am J Surg 2018;216:645–651.
17. Scotton G, Del Zotto G, Bernardi L, et al. Is the ACS-NSQIP risk calculator accurate in predicting adverse postoperative outcomes in the emergency setting? an Italian single-center preliminary study. World J Surg 2020;44:3710–3719.
18. 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.
19. Hyder JA, Reznor G, Wakeam E, et al. Risk prediction accuracy differs for emergency versus elective cases in the ACS-NSQIP. Ann Surg 2016;264:959–965.
20. Parkin CJ, Moritz P, Kirkland O, Glover A. What is the accuracy of the ACS-NSQIP surgical risk calculator in emergency abdominal surgery? A meta-analysis. J Surg Res 2021;268:300–307.
21. Golden DL, Ata A, Kusupati V, et al. Predicting postoperative complications after acute care surgery: how accurate is the ACS NSQIP surgical risk calculator? Am Surg 2019;85:335–341.
22. Cohen ME, Liu Y, Ko CY, Hall BL. An examination of American College of Surgeons NSQIP surgical risk calculator accuracy. J Am Coll Surg 2017;224:787–795.e1.
23. Liu Y, Cohen ME, Hall BL, et al. Evaluation and enhancement of calibration in the American College of Surgeons NSQIP surgical risk calculator. J Am Coll Surg 2016;223:231–239.