



Published in final edited form as:

*J Am Coll Surg*. 2020 December ; 231(6): 650–656. doi:10.1016/j.jamcollsurg.2020.09.023.

## Performance of the American College of Surgeons NSQIP Surgical Risk Calculator for Total Gastrectomy

Elvira L Vos, MD, PhD<sup>a</sup>, Ashley E Russo, MD<sup>a</sup>, Alexandra Hohmann<sup>a</sup>, Sam S Yoon, MD, FACS<sup>a</sup>, Daniel G Coit, MD, FACS<sup>a</sup>, Clifford Ko, MD, FACS<sup>b</sup>, Vivian E Strong, MD, FACS<sup>a</sup>

<sup>a</sup>Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY

<sup>b</sup>Department of Surgery, University of California Los Angeles Medical Center, Los Angeles, CA; Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL

### Abstract

**Background:** To encourage implementation of the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) Risk Calculator for total gastrectomy for gastric cancer, its predictive performance for this specific procedure should be validated. We assessed its discriminatory accuracy and goodness of fit for predicting 12 adverse outcomes.

**Study Design:** Data on all patients with gastric cancer that underwent total gastrectomy with curative intent at Memorial Sloan Kettering Cancer Center between 2002-2017 were collected. Preoperative risk factors from the electronic medical record were manually inserted into the ACS-NSQIP Risk Calculator. Predictions for adverse outcomes were compared with observed outcomes by Brier scores, c-statistics, and Hosmer-Lemeshow p value.

**Results:** In a total of 452 patients, the predicted rate of all complications (29%) was lower than the observed rate (45%). Brier scores varied between 0.017 for death and 0.272 for any complication. C-statistics were moderate (0.7–0.8) for death and renal failure, good (0.8–0.9) for cardiac complication, and excellent (0.9) for discharge to nursing or rehabilitation facility. Hosmer-Lemeshow p value found poor goodness of fit for pneumonia only.

**Conclusions:** For adverse outcomes after total gastrectomy with curative intent in gastric cancer patients, performance of the ACS-NSQIP Risk Calculator is variable. Its predictive performance was best for cardiac complications, renal failure, death, and discharge to nursing or rehabilitation facility.

### Precis

We assessed the predictive performance of the ACS-NSQIP Risk Calculator for total gastrectomy for gastric cancer. We found that its performance was variable, and best for cardiac complication, renal failure, death, and discharge to nursing or rehabilitation facility.

---

Corresponding author: Vivian Strong, MD, Department of Surgery, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY, 10065, Phone: 212-639-5056, strongv@mskcc.org.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure Information: Nothing to disclose.

## Introduction

More than 26,000 individuals were newly diagnosed with gastric cancer in the United States in 2018 (1). Surgery remains the mainstay of treatment, but carries high risk of morbidity and mortality; reported rates range from 29 to 50% and 2 to 6%, respectively (2-4).

To accurately inform patients of the risks of gastrectomy, surgical risk calculators may be used. The most widely used calculator is the universal surgical risk calculator developed by the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) based on data from nearly 400 hospitals and 1.4 million patients undergoing 1,500 different surgical procedures (5). It estimates procedure-specific and case-mix adjusted risk for adverse outcomes using preoperative patient characteristics.

Although large internal evaluations showed excellent performance of the ACS-NSQIP Risk Calculator (6), its accuracy for total gastrectomy has not yet been validated. The primary aim of this study was to assess the predictive performance of the ACS-NSQIP Risk Calculator for adverse events after total gastrectomy for gastric cancer in terms of discrimination and goodness of fit. The secondary aim was to identify the adverse outcomes that the calculator accurately predicts and for which it is thus clinically applicable.

## Methods

### Patients and Data Collection

For all patients who underwent total gastrectomy with Roux-en-Y reconstruction for gastric cancer between 2002 and 2017 at Memorial Sloan Kettering Cancer Center (MSKCC), data regarding the 20 preoperative risk factors included in the ACS-NSQIP calculator (listed in Table 1) were collected from medical records. Data were manually entered into the online ACS-NSQIP Risk Calculator in June 2018, and its predictions for the following 12 adverse outcomes were recorded: any complication, pneumonia, cardiac complication, surgical site infection, urinary tract infection, venous thromboembolism, renal failure, readmission, return to operating room (OR), death, discharge to nursing or rehabilitation facility, length of hospital stay. The sum of the calculated risks for all patients was the predicted number of events.

The actual observed adverse outcomes for each individual patient were collected from the institutional Surgical Secondary Events (SSE) database. This prospectively maintained database tracks deviations from the expected postoperative course that occur within the first 30 days after surgery. Complications are defined by body system and graded on a 1 to 5 scale of severity according to the widely used modified Clavien-Dindo classification (7). A previous audit of the accuracy of the SSE database found that 91% of complications were correctly entered, and nearly all that were missed were grade 1-2 (7).

### Statistical Analysis

The mean predicted and observed number of events were calculated for each of the 12 adverse outcomes studied. The predictive performance of the risk calculator was assessed by

Brier score, c-statistic, and Hosmer-Lemeshow. Overall performance is most often measured by the Brier score, the average squared difference between the predicted probability and the observed probability of an event, reflecting both discrimination and goodness of fit. A lower Brier score indicates a smaller difference between the predicted and observed event rates and thus a better fit; a score of 0 reflects perfect predictive accuracy (8). The c-statistic represents the model's ability to discriminate between patients who will or will not experience an event. A higher c-statistic means better discriminatory performance: 0.9 is generally considered excellent, 0.8–0.9 is good, 0.7–0.8 is moderate, 0.6–0.7 is fair and 0.5 means that the model is no better than random chance. The Hosmer-Lemeshow statistic evaluates differences in the probability of observed and predicted events across deciles of increasing predicted risk. A Hosmer-Lemeshow p value < 0.05 leads to rejection of the null hypothesis that the model is well-calibrated (8). Hosmer-Lemeshow calibration plots were generated by graphing the observed versus predicted risk for each decile.

The mean predicted and observed lengths of stay (the only continuous outcome) were compared by paired Student's t-test and the correlation was assessed by Pearson correlation coefficient. To assess the risk calculator's performance for this variable using the Brier score and c-statistic, data were categorized as shorter or longer than the mean. P values < 0.05 were considered statistically significant. Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY.

## Results

A total of 452 patients underwent total gastrectomy between January 2002 and June 2017 at MSKCC. Data on the preoperative risk factors included in the risk calculator are displayed in Table 1. The majority (61%) of patients had an American Society of Anesthesiologists (ASA) physical status of 3, and 35% had an ASA status of 2. More than one-third of patients (37%) had hypertension requiring medication, 13% were diabetic, 14% were current smokers, and the mean BMI was  $27 \pm 5.1$ .

The predicted rate of events from the ACS-NSQIP Risk Calculator and the observed rate and number of events from the institutional SSE database for each adverse outcome are displayed in Table 2. The observed rate considerably exceeded the predicted rate for any complication (45% vs. 29%) and for surgical site infection (25% vs. 16%), and was considerably lower than the expected rate for return to OR (5.8% vs. 8.8%) and for discharge to nursing or rehabilitation facility (2.0% vs. 11%).

The predictive performance of the ACS-NSQIP Risk Calculator of each adverse outcome as assessed by Brier score, c-statistic, and Hosmer-Lemeshow p value is displayed in Table 3. Adverse outcomes with relatively low Brier scores, indicating better overall performance of the model, included renal failure (0.015), death (0.017), and cardiac complications (0.019). Adverse outcomes with the highest c-statistics, indicating better discrimination between patients who will or will not experience an event, included discharge to nursing or rehabilitation facility (0.90) and cardiac complications (0.83). Only pneumonia had a Hosmer-Lemeshow p value < 0.05, indicating that the model does not fit well for predicting this adverse outcome for patients undergoing total gastrectomy. As Hosmer-Lemeshow p

values can only be calculated for categorical variables, the calculator's accuracy in predicting length of hospital stay was assessed by paired Student's t-test, which revealed a significant difference between the mean predicted and observed lengths of hospital stay;  $p = 0.005$ . Figure 1 compares the distributions of observed vs. predicted lengths of hospital stay.

Hosmer-Lemeshow calibration curves are shown in Figure 2 and indicate relatively good correlation between the predicted and observed rates of most adverse outcomes. The most noticeable exception is pneumonia, for which the regression line has a negative slope, indicating higher predicted probability but lower observed risk, consistent with the low  $p$  value and  $c$ -statistic for this adverse outcome.

## Discussion

The performance of the ACS-NSQIP Risk Calculator for total gastrectomy at our institution was found to be variable among outcomes. The adverse outcomes for which the calculator was well-fitted (Hosmer-Lemeshow  $p$ -value of  $> 0.05$ ) and showed at least moderate discrimination ( $c$ -statistic  $> 0.7$ ) were the same as those for which it had the best overall performance (lowest Brier scores), namely cardiac complications, renal failure, death, and discharge to nursing or rehabilitation facility.

Our study complements a previous assessment of the accuracy of the ACS-NSQIP Risk Calculator for gastric cancer surgery by the U.S. Gastric Cancer Collaborative (9). The population in that study differed from ours, in that the majority underwent distal rather than total gastrectomy, and 11% of patients had distant metastatic disease. Brier scores in that study were similar to those we found for any complication, pneumonia, cardiac complication, renal failure, and return to OR, and the 3 outcomes with the lowest Brier scores similarly included cardiac complication and renal failure. On the other hand, we found a twofold lower Brier score for death, which may result from the lower observed death rate in our population (1.8% vs. 3.9%), as a lower incidence limits the maximum Brier score (8). While we found discriminatory power assessed by the  $c$ -statistic to be moderate or better ( $> 0.7$ ) for 4 outcomes, the  $c$ -statistics found in the Gastric Cancer Collaborative study were all below 0.7.

The predicted rate of any complication (29%) was lower than the observed rate of complications (45%). This discrepancy is similar to that previously found in assessments of the ACS-NSQIP Risk Calculator in colorectal cancer, head and neck cancer, and lung cancer surgery (10-12), which found that it underestimates the overall complication rate, though this was not seen in breast cancer surgery (13). An explanation for the difference could be a higher report rate of complications in our hospital compared with the average of the hundreds of hospitals in the NSQIP database, similar to the higher report rate for surgical site infections at our hospital compared with that in NSQIP (6.5% vs. 4%) (14). As illustrated in the latter study, differences in event rates can also result from differences in definitions, suggesting that standardizing definitions would improve consistency among databases (14).

Ours is not the first study to find that the ACS-NSQIP Risk Calculator shows variable predictive strength for adverse outcomes of specific surgeries; the same has been found for many other procedures (11, 15-26). The ACS examined the quality of the design of these external validations, as well as others with more concordant conclusions, and found that only 19 of 21 had sufficient sample size for evaluation of performance metrics, which requires a minimum of 100 events, and ideally 200 or more (27). In addition, they found that studying a homogeneous study population that underwent one type of procedure at a single institution limits discriminatory power and generalizability, because there is less information to discriminate events from non-events. The addition of surgery-specific predictors has been shown to improve the ACS-NSQIP Risk Calculator's performance (15, 16). Despite its limitations for specific procedures, the calculator remains a useful tool for general risk predictions and benchmarking, as its discriminatory power and accuracy has been shown to be excellent in validation studies using data from millions of patients including hundreds of hospitals and types of surgeries (6).

In addition to the 12 adverse outcomes studied here, the ACS-NSQIP Risk Calculator also estimates the risk of serious complications. This was not addressed in the current study because of the discrepancy in definitions between the NSQIP and our institute's SSE database. The SSE database grades complications on a 1–5 scale according to the widely used modified Clavien-Dindo classification, which classifies events of grades 3–5 as major events. However, the NSQIP definition of serious complications includes some grade 1–2 events and excludes some grade 3 events. For example, small intra-abdominal abscesses that do not require intervention and urinary tract infections that require only oral antibiotics are defined as serious complications by NSQIP. Further, serious complications that are relatively common in the study population (e.g., dysphagia requiring anastomotic stricture dilatation and dyspnea requiring pleural effusion drainage) are not defined as serious by NSQIP. Finally, NSQIP may not adequately capture anastomotic leaks, as Rickles et al. showed that 75% of these were not entered in the database after colectomies (28). The estimated serious complication rate by the ACS-NSQIP Risk Calculator in the current study population was 26% and the actual incidence of grade 3–5 events as recorded in the SSE database was 19%. The current updated ACS-NSQIP Risk Calculator also includes the “surgeon adjustment of risk” function, allowing adjustment of the risk for all complications based on risk factors that are not accounted for in the current set of preoperative predictors by 1 or 2 standard deviations.

The current study included 202 events with any type of complication, making it one of the largest studies assessing the performance of the ACS-NSQIP Risk Calculator. Another strength of this study was the inclusion of only patients undergoing total gastrectomy for gastric cancer with curative intent, increasing the reliability and applicability of our findings to this population. Our source data are also highly reliable, as previous studies have found that the SSE database is highly accurate and concordant with NSQIP, at least for surgical site infections (7, 14). A disadvantage was the lack of assessment of the reliability of the calculator in predicting serious complications.

## Conclusions

For total gastrectomy with curative intent in gastric cancer patients, the predictive accuracy of the ACS-NSQIP Risk Calculator varies among outcomes. Its predictive performance was best for cardiac complication, renal failure, death, and discharge to nursing or rehabilitation facility.

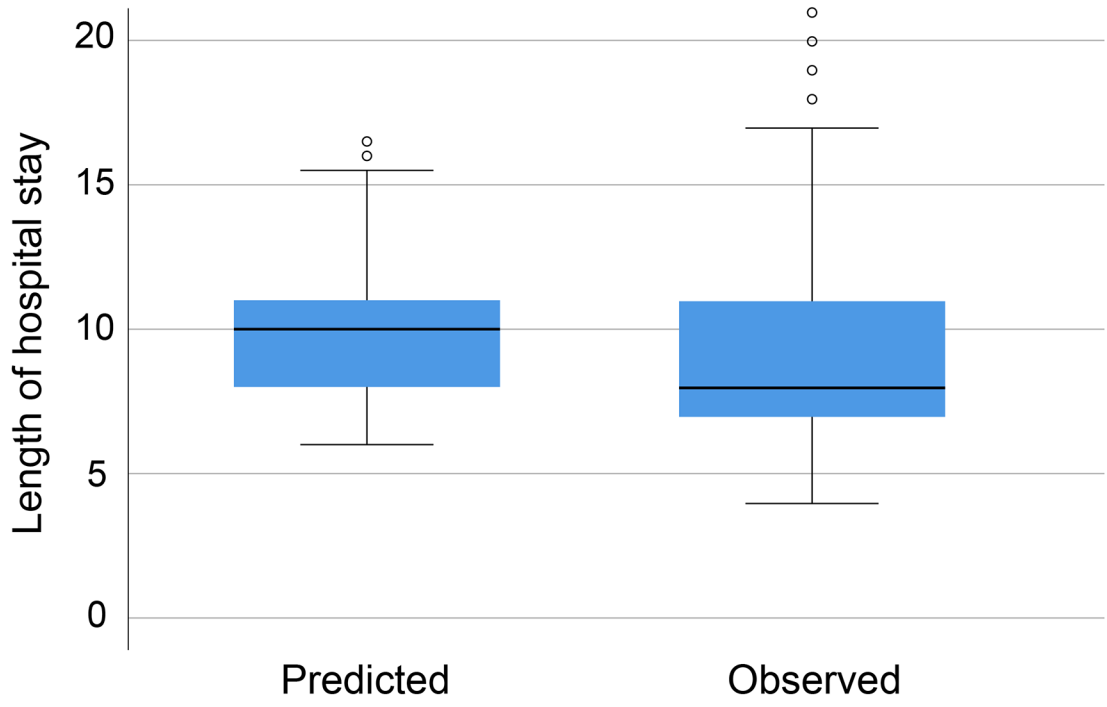
## Acknowledgment

The authors acknowledge Jessica Moore, MS, for editorial assistance.

## References

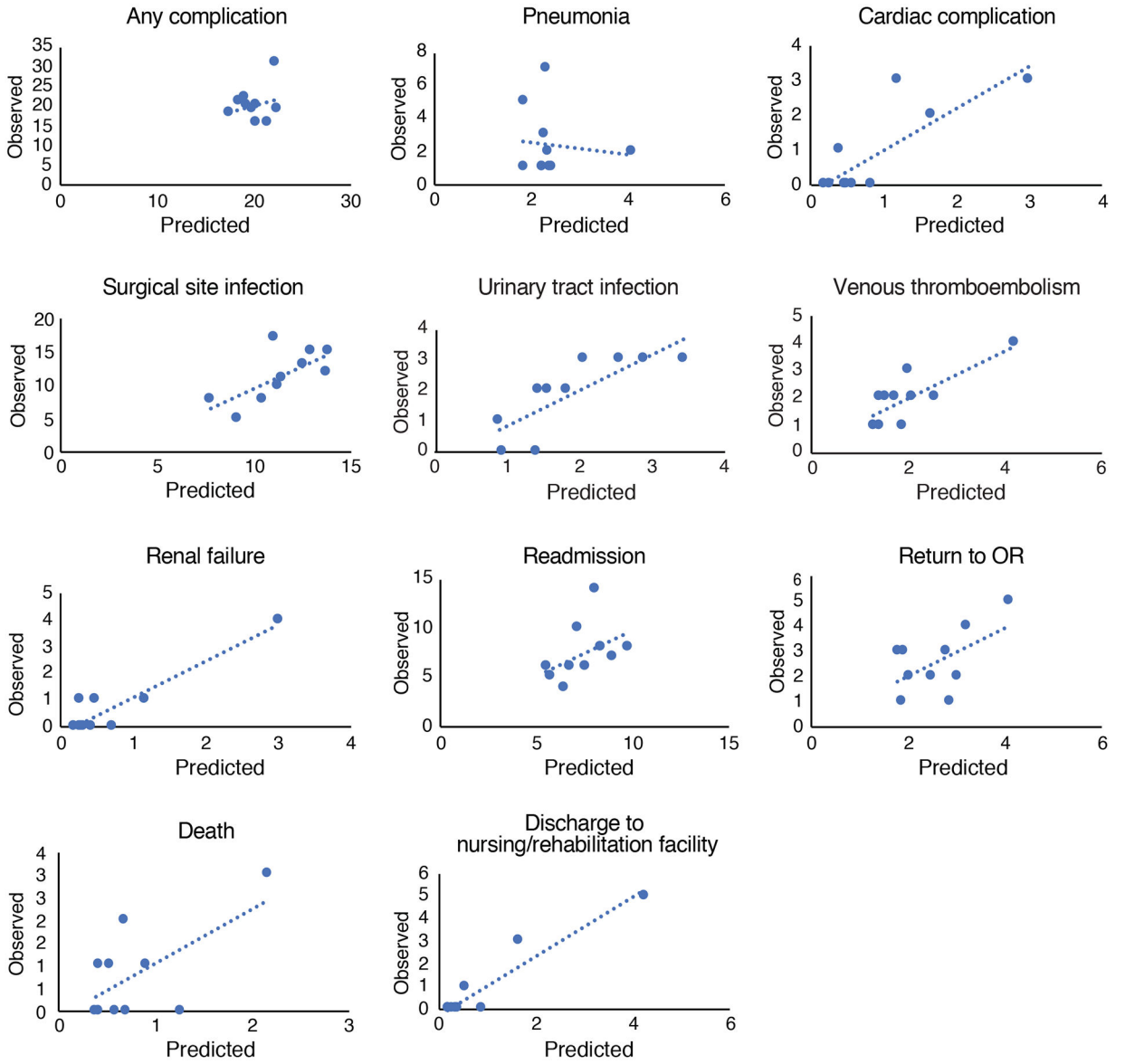
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30. [PubMed: 29313949]
2. Al-Batran SE, Homann N, Pauligk C, et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *Lancet*. 2019;393(10184):1948–57. [PubMed: 30982686]
3. Cunningham D, Allum WH, Stenning SP, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med*. 2006;355(1):11–20. [PubMed: 16822992]
4. Papenfuss WA, Kukar M, Oxenberg J, et al. Morbidity and mortality associated with gastrectomy for gastric cancer. *Ann Surg Oncol*. 2014;21(9):3008–14. [PubMed: 24700300]
5. 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;117(5):833–42 e1–3. [PubMed: 24055383]
6. 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(2):231–9. [PubMed: 27212006]
7. Strong VE, Selby LV, Sovel M, et al. Development and assessment of Memorial Sloan Kettering Cancer Center's Surgical Secondary Events grading system. *Ann Surg Oncol*. 2015;22(4):1061–7. [PubMed: 25319579]
8. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38. [PubMed: 20010215]
9. Beal EW, Saunders ND, Kearney JF, et al. Accuracy of the ACS NSQIP Online Risk Calculator Depends on How You Look at It: Results from the United States Gastric Cancer Collaborative. *Am Surg*. 2018;84(3):358–64. [PubMed: 29559049]
10. Adegboyega TO, Borgert AJ, Lambert PJ, Jarman BT. Applying the National Surgical Quality Improvement Program risk calculator to patients undergoing colorectal surgery: theory vs reality. *Am J Surg*. 2017;123(1):30–5. [PubMed: 27424043]
11. Samson P, Robinson CG, Bradley J, et al. The National Surgical Quality Improvement Program risk calculator does not adequately stratify risk for patients with clinical stage I non-small cell lung cancer. *J Thorac Cardiovasc Surg*. 2016;151(3):697–705 e1. [PubMed: 26410001]
12. Vosler PS, Orsini M, Enepekides DJ, Higgins KM. Predicting complications of major head and neck oncological surgery: an evaluation of the ACS NSQIP surgical risk calculator. *J Otolaryngol Head Neck Surg*. 2018;147(1):21. [PubMed: 29566750]
13. Lyle B, Landercasper J, Johnson JM, et al. Is the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator applicable for breast cancer patients undergoing breast-conserving surgery? *Am J Surg*. 2016;211(4):820–3. [PubMed: 26489987]

14. Selby LV, Sjoberg DD, Cassella D, et al. Comparing surgical infections in National Surgical Quality Improvement Project and an Institutional Database. *J Surg Res.* 2015;196(2):416–20. [PubMed: 25840487]
15. McMillan MT, Allegrini V, Asbun HJ, et al. Incorporation of Procedure-specific Risk Into the ACS-NSQIP Surgical Risk Calculator Improves the Prediction of Morbidity and Mortality After Pancreatoduodenectomy. *Ann Surg.* 2017;265(5):978–86. [PubMed: 27232260]
16. Liu JB, Sosa JA, Grogan RH, et al. Variation of Thyroidectomy-Specific Outcomes Among Hospitals and Their Association With Risk Adjustment and Hospital Performance. *JAMA Surg.* 2018;153(1):e174593. [PubMed: 29188293]
17. 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(5):570–5. [PubMed: 27627780]
18. Schneider AL, Deig CR, Prasad KG, et al. Ability of the National Surgical Quality Improvement Program Risk Calculator to Predict Complications Following Total Laryngectomy. *JAMA Otolaryngol Head Neck Surg.* 2016;142(10):972–9. [PubMed: 27467967]
19. Prasad KG, Nelson BG, Deig CR, et al. ACS NSQIP Risk Calculator: An Accurate Predictor of Complications in Major Head and Neck Surgery? *Otolaryngol Head Neck Surg.* 2016;155(5):740–2. [PubMed: 27329422]
20. Arce K, Moore EJ, Lohse CM, et al. The American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator Does Not Accurately Predict Risk of 30-Day Complications Among Patients Undergoing Microvascular Head and Neck Reconstruction. *J Oral Maxillofac Surg.* 2016;74(9):1850–8. [PubMed: 27019416]
21. Rivard C, Nahum R, Slagle E, et al. Evaluation of the performance of the ACS NSQIP surgical risk calculator in gynecologic oncology patients undergoing laparotomy. *Gynecol Oncol.* 2016;141(2):281–6. [PubMed: 26899020]
22. Keller DS, Cologne KG, Senagore AJ, Haas EM. Does one score fit all? Measuring risk in ulcerative colitis. *Am J Surg.* 2016;212(3):433–9. [PubMed: 26837227]
23. Edelstein AI, Kwasny MJ, Suleiman LI, et al. Can the American College of Surgeons Risk Calculator Predict 30-Day Complications After Knee and Hip Arthroplasty? *J Arthroplasty.* 2015;30(9 Suppl):5–10. [PubMed: 26165953]
24. 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(1):48–54. [PubMed: 27789387]
25. Johnson C, Campwala I, Gupta S. Examining the validity of the ACS-NSQIP Risk Calculator in plastic surgery: lack of input specificity, outcome variability and imprecise risk calculations. *J Investig Med.* 2017;65(3):722–5.
26. Massoumi RL, Trevino CM, Webb TP. Postoperative Complications of Laparoscopic Cholecystectomy for Acute Cholecystitis: A Comparison to the ACS-NSQIP Risk Calculator and the Tokyo Guidelines. *World J Surg.* 2017;41(4):935–9. [PubMed: 27834012]
27. 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(5):787–95 e1. [PubMed: 28389191]
28. Rickles AS, Iannuzzi JC, Kelly KN, et al. Anastomotic leak or organ space surgical site infection: What are we missing in our quality improvement programs? *Surgery.* 2013;154(4):680–7. [PubMed: 24074406]



**Figure 1.** Distributions of observed vs predicted length of hospital stay.





**Figure 2.**  
Hosmer-Lemeshow calibration curves.

**Table 1.**

Preoperative Risk Factors (n=452).

Variable	Data
Age, y, mean (SD)	62 (14)
Sex, m, n (%)	292 (65)
Functional status, n (%)	
Independent	446 (99)
Partially dependent	6 (1.3)
Totally dependent	0
Emergency case, n	0
ASA physical status, n (%)	
1	10 (2.2)
2	157 (35)
3	274 (61)
4	11 (2.4)
Steroid use for chronic condition, n (%)	3 (0.7)
Ascites within 30 days prior to surgery, n	0
Systemic sepsis within 48 h prior to surgery, n	0
Ventilator dependence, n	0
Cancer dissemination, n (%)	5 (1.1)
Diabetes, n (%)	58 (13)
Hypertension requiring medication, n (%)	165 (37)
Congestive heart failure in 30 days prior to surgery, n	0
Dyspnea with moderate exertion, n (%)	3 (0.7)
Current smoker within 1 year, n (%)	61 (14)
History of severe COPD, n (%)	3 (0.7)
Dialysis, n	0
Acute renal failure, n	0
BMI, kg/m <sup>2</sup> , mean (SD)	27 (5.1)

ASA, American Society of Anesthesiologists

**Table 2.**

Mean Predicted and Observed Rates of Adverse Outcomes

Variable	Predicted	Observed	Observed no. of events
Any complication, %	29	45	202
Pneumonia, %	7.0	5.3	24
Cardiac complication, %	0.8	2.0	9
Surgical site infection, %	16	25	114
Urinary tract infection, %	3.2	4.2	19
Venous thromboembolism, %	2.1	4.4	20
Renal failure, %	1.8	1.5	7
Readmission, %	12	16	74
Return to operating room, %	8.8	5.8	26
Death, %	1.8	1.8	8
Discharge to nursing or rehabilitation facility, %	11	2.0	9
Length of stay, d	9.8	11	n/a
Longer length of stay ( mean of 12 days), %	22	24	107

n/a, not applicable

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

Performance of American College of Surgeons NSQIP Risk Calculator for 12 Adverse Outcomes.

<b>Variable</b>	<b>Brier score</b>	<b>C-statistic</b>	<b>Hosmer-Lemeshow p value</b>
Any complication	0.272	0.53	0.083
Pneumonia	0.052	0.49	0.008
Cardiac complication	0.019	0.83	0.568
Surgical site infection	0.195	0.60	0.373
Urinary tract infection	0.040	0.67	0.906
Venous thromboembolism	0.043	0.57	0.987
Renal failure	0.015	0.79	0.727
Readmission	0.140	0.44	0.313
Return to OR	0.055	0.58	0.841
Death	0.017	0.71	0.470
Discharge to nursing or rehabilitation facility	0.031	0.90	0.832
Length of stay longer than mean of 12 days	0.235	0.56	n/a

n/a, not applicable

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript