JAMA Otolaryngology–Head & Neck Surgery | Original Investigation

# Ability of the National Surgical Quality Improvement Program Risk Calculator to Predict Complications Following Total Laryngectomy

Alexander L. Schneider, MD; Christopher R. Deig, BA; Kumar G. Prasad, MD; Benton G. Nelson, MD; Avinash V. Mantravadi, MD; Joseph S. Brigance, MD, MS; Mark P. Langer, MD; Mark W. McDonald, MD; Peter A. Johnstone, MD; Michael G. Moore, MD

**IMPORTANCE** The accuracy of the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) risk calculator has been assessed in multiple surgical subspecialties; however, there have been no publications doing the same in the head and neck surgery literature.

**OBJECTIVE** To evaluate the accuracy of the calculator's predictions in a single institution's total laryngectomy (TL) population.

**DESIGN, SETTING, AND PARTICIPANTS** Total laryngectomies performed between 2013 and 2014 at a tertiary referral academic center were evaluated using the risk calculator. Predicted 30-day outcomes were compared with observed outcomes for return to operating room, surgical site infection, postoperative pneumonia, length of stay, and venous thromboembolism.

**MAIN OUTCOMES AND MEASURES** Comparison of the NSQIP risk calculator's predicted postoperative complication rates and length of stay to what occurred in this patient cohort using percent error, Brier scores, area under the receiver operating characteristic curve, and Pearson correlation analysis.

**RESULTS** Of 49 patients undergoing TL, the mean (SD) age at operation was 59 (9.3) years, with 67% male. The risk calculator had limited efficacy predicting perioperative complications in this group of patients undergoing TL with or without free tissue reconstruction or preoperative chemoradiation or radiation therapy with a few exceptions. The calculator overestimated the occurrence of pneumonia by 165%, but underestimated surgical site infection by 7%, return to operating room by 24%, and length of stay by 13%. The calculator had good sensitivity and specificity of predicting surgical site infection for patients undergoing TL with free flap reconstruction (area under the curve, 0.83). For all other subgroups, however, the calculator had poor sensitivity and specificity for predicting complications.

**CONCLUSIONS AND RELEVANCE** The risk calculator has limited utility for predicting perioperative complications in patients undergoing TL. This is likely due to the complexity of the treatment of patients with head and neck cancer and factors not taken into account when calculating a patient's risk.

Invited Commentary page 979

*JAMA Otolaryngol Head Neck Surg*. 2016;142(10):972-979. doi:10.1001/jamaoto.2016.1809
Published online July 28, 2016.

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author**: Alexander L. Schneider, MD, Northwestern University Feinberg School of Medicine, Department of Otolaryngology–Head and Neck Surgery, 675 N St Clair St, Galter Pavilion 15th Floor, Room 200, Chicago, IL 60611 (alexander.schneider@northwestern.edu).

A thorough and accurate understanding of the risk of perioperative complications is crucial to both the informed consent process and the surgeon's choice of management. The Institute of Medicine in 2013 identified shared decision making, particularly with regard to knowledge of treatment benefit and harm, as an important marker of high-quality cancer care.[1(pp91-152)] In addition, the Centers for Medicare & Medicaid Services may soon implement a measure that would financially reward surgeons for discussing patient-specific complication risks with surgical candidates before elective surgical procedures.[2] Today's clinicians are quickly coming to terms with the fact that reimbursements, predominantly for care of Medicare and Medicaid beneficiaries, will in large part depend on quality or value, which are adversely affected by surgical complications.[3]

The American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP) Surgical Risk Calculator was developed in 2013.[4] The calculator is based on a database that contains patient data including many preoperative characteristics and postoperative complications (within 30 days of a procedure) from 393 hospitals and more than 1.4 million operations from 2009 through 2012.[5] Surgeons enter a *Current Procedural Terminology* (CPT) code and their patient's preoperative variables, and the calculator produces an empirically derived predicted complication rate based on an individual patient's characteristics. Thus, using the calculator to accurately predict a patient's outcomes could prove to be medically, ethically, and financially beneficial.

However, multiple variables that head and neck surgeons regularly contend with are not included in the calculator, such as the preoperative exposure to chemotherapy and radiation therapy (RT), as well as changes in speech and swallowing physiology at the time of surgery. Early studies using the NSQIP database demonstrated that preoperative RT is indeed predictive of an increased complication rate after laryngectomy.[6] There is also an intrinsic complexity of the population of patients with head and neck cancer that is difficult to account for using the calculator including the need for flap reconstruction, the status of primary malignant neoplasm, nutritional status, swallowing physiology, and the effect of tobacco and alcohol use on tissue quality.

Our goal was to evaluate the accuracy of the risk-adjusted complication rate predicted by the NSQIP calculator in patients undergoing laryngectomy who received surgery alone or chemoradiation therapy (CRT) or RT followed by surgery. We also evaluate the accuracy of the calculator in specific subsets of laryngectomies, such as patients who underwent free tissue transfer to repair the laryngectomy defect or those with specific preoperative risk factors. We hypothesized that the NSQIP calculator would poorly predict perioperative outcomes in this patient population due to the innate complexity of head and neck procedures, as well as patient population–specific factors unaccounted for by the calculator.

### Key Points

**Question** Does the National Surgical Quality Improvement Program (NSQIP) calculator reliably predict complications in patients undergoing laryngectomy?

**Finding** This study found that the calculator had limited utility in predicting postoperative pneumonia, surgical site infection, return to operating room, and venous thromboembolism, with few exceptions.

**Meaning** There are a number of factors in the laryngectomy population that make the NSQIP calculator less reliable in this group of patients.

## Methods

### Study Design

Approval was obtained from the Indiana University School of Medicine Institutional Review Board. Laryngectomies from the years 2013 and 2014 were evaluated that had been coded with the following CPT codes: 31360 (total laryngectomy without neck dissection), 31365 (total laryngectomy with radical neck dissection), 31390 (pharyngolaryngectomy with radical neck dissection, without reconstruction), and 31395 (pharyngolaryngectomy with radical neck dissection, with reconstruction). Laryngectomies and pharyngolaryngectomies were selected in an attempt to select a relatively standardized procedure and control for the impact that changes in airway and swallowing physiology would have on perioperative outcomes in our patient population. We reviewed the medical records of each eligible patient and documented age, sex, and the preoperative risk factors that are specified by the ACS NSQIP Risk Calculator: height, weight, American Society of Anesthesiologists classification, wound class contamination, functional status, history of long-term steroid use, history of diabetes, history of hypertension requiring medication, history of congestive heart failure in the past 30 days, history of severe chronic obstructive pulmonary disease, history of dialysis treatment, and history of tobacco use in the year prior to the index surgery. We recorded the anatomic location of the primary lesion requiring resection and/or reconstruction, all major resection CPT codes for each patient, the type of reconstruction, and the presence or absence of preoperative CRT or RT. We also documented the following postoperative morbidities tracked by the NSQIP according to the definitions set forth by the NSQIP calculator: pneumonia, surgical site infection (SSI), venous thromboembolism (VTE), length of stay (LOS), and 30-day return to operating room (ROR). Postoperative progress and clinic notes were evaluated for the presence of the aforementioned morbidities for up to 30 days after the index surgical procedure. Patients were excluded if these preoperative variables were not recorded or if there was not data for the first 30 days of the perioperative period.

Once all of the required preoperative variables were compiled, risk was calculated for all patients by entering either 31360, 31365, or 31390 into the risk calculator even if they subsequently received free flap reconstruction. The risk calculator allows the surgeon to modify the patient's preoperative risk based on the surgeon's assessment of their overall health status and any other comorbid conditions not included in the NSQIP calculator. All of our patients were identified as 1: "no adjustment necessary."

Table. Demographic and Medical Information for All Patients Undergoing Laryngectomy

| Characteristic | Value (N = 49) |
|---|---|
| Age, mean (SD), y | 59 (9.3) |
| Sex, No. (%) | |
| Female | 16 (33) |
| Male | 33 (67) |
| Height, mean (SD), m | 1.72 (0.08) |
| Weight, mean (SD), kg | 69.7 (17.2) |
| BMI, mean (SD) | 23.7 (5.0) |
| Received upfront total laryngectomy, No. (%) | 17 (35) |
| Received preoperative radiation therapy, No. (%) | 15 (31) |
| Received preoperative chemotherapy only, No. (%) | 0 |
| Received chemoradiation therapy, No. (%) | 17 (35) |
| Wound class contaminated, No. (%) | 49 (100) |
| Independent functional status, No. (%) | 49 (100) |
| American Society of Anesthesiologists class, No. (%) | |
| III | 46 (94) |
| IV | 3 (6) |
| Hypertension requiring medication, No. (%) | |
| Yes | 24 (49) |
| No | 25 (51) |
| Smoker within 1 y, No. (%) | |
| Yes | 23 (47) |
| No | 26 (53) |
| Type 2 diabetes, No. (%) | |
| Yes | 11 (22) |
| No | 38 (78) |
| Previous cardiac event, No. (%) | |
| Yes | 10 (20) |
| No | 39 (80) |
| Severe chronic obstructive pulmonary disease history, No. (%) | |
| Yes | 15 (31) |
| No | 34 (69) |
| Symptom of dyspnea, No. (%) | |
| Yes | 11 (22) |
| No | 38 (78) |
| Steroid use for chronic condition, No. (%) | |
| Yes | 4 (8) |
| No | 45 (92) |
| Chronic heart failure 30 d prior to surgery, No. (%) | |
| Yes | 1 (2) |
| No | 48 (98) |
| Requirement of dialysis, No. (%) | |
| Yes | 1 (2) |
| No | 48 (98) |
| Use of mechanical ventilation, No. (%) | 0 |
| Disseminated cancer, No. (%) | 0 |
| Acute renal failure, No. (%) | 0 |
| Ascites within past 30 d, No. (%) | 0 |
| Sepsis within past 48 h, No. (%) | 0 |

Abbreviation: BMI, body mass index, calculated as weight in kilograms divided by height in meters squared.

By convention, at our institution, laryngectomies and pharyngolaryngectomies that require reconstruction are performed using separate ablative and reconstructive teams. As a result, all codes for these portions of the procedure are entered separately. In an effort to try to include as much of the perioperative risk as possible in the risk calculator, we did a subgroup analysis for all patients undergoing laryngectomy or pharyngolaryngectomy with flap patch or tubed reconstruction, substituting CPT 31395 for the previously used code (31360, 31365, or 31390).

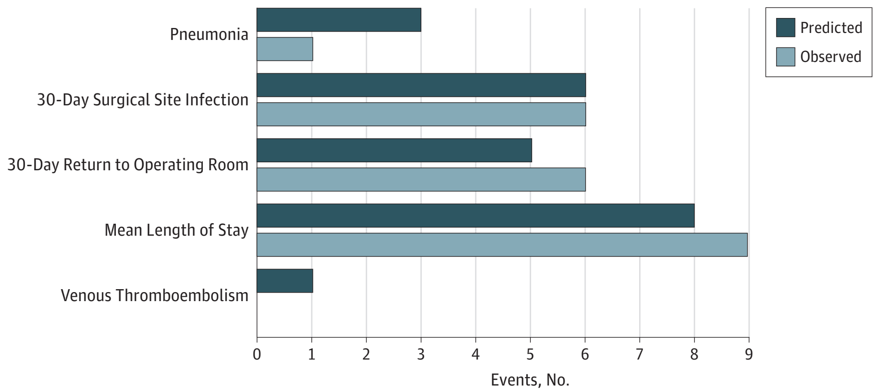
## Statistical Analysis

The predicted outcomes data were compared with observed outcomes data for the following variables: pneumonia, SSI, VTE, LOS, and ROR. Microsoft Excel and SPSS 22 were used for all statistical analysis. Brier scores were calculated for an estimation of the NSQIP calculator's predictive value of binary outcomes. The Brier score is a statistical model used to compare predicted with observed outcomes. Scores range from 0 (best) to 1 (worst). The score is calculated by assigning a value of 0 to each nonevent and 1 to each event. A score of 0.01 was used as a threshold for good performance, as illustrated in prior studies, which corresponds to approximately a 90% forecast accuracy.[4,7,8] Receiver operating characteristic (ROC) curves and the area under the curve (AUC) were used as an ancillary measure of forecast modeling because they provide information about the true- and false-positive rate of the NSQIP calculator. Receiver operating characteristic curves assess the ability of the forecast model to discriminate between events and nonevents. An AUC greater than 0.80 was considered a test with good sensitivity and specificity. The null hypothesis states that the AUC = 0.5, indicating that the test has no predictive value. This is demonstrated as a diagonal line on an ROC curve.[9] The Pearson product-moment correlation coefficient (Pearson *r*) was used to provide insight into the calculator's prediction of patient LOS. A Pearson *r* value of 1 indicates a perfect positive correlation, a score of −1 indicates a perfect negative correlation, and a score of 0 indicates no correlation. Percent error was used as a means to directly compare predicted rates of complications with observed rates of complications.

# Results

There were a total of 49 patients undergoing total laryngectomy (TL) included in this study. The mean (SD) age at operation was 59 (9.3) years, with 33 (67%) male and 16 (33%) female patients, and the mean (SD) body mass index (calculated as weight in kilograms divided by height in meters squared) was 23.7 (5.0) (**Table**). Seventeen patients underwent upfront TL (35%), while 15 (31%) had received preoperative RT and 17 (35%) had received preoperative CRT. Twenty-four (49%) patients in this cohort had a history of hypertension requiring treatment, 23 (47%) were smokers within the past year, 15 (31%) had a history of severe chronic obstructive pulmonary disease, 11 (22%) had dyspnea on exertion prior to surgery, 11 (22%) had type 2 diabetes, and 10 (20%) had a history of a prior cardiac event. The Table provides other demographic information in this cohort.

Figure 1. Predicted vs Observed Number of Complications




The predicted number of complications was derived from the National Surgical Quality Improvement Program estimation of risk. This graph provides a general comparison of total predicted vs observed complications without directly analyzing the predictive accuracy of the calculator; N = 49.

Twenty-seven patients were coded as 31360, 16 were coded as 31365, and 6 were coded as 31390. Twenty-eight patients underwent immediate free tissue transfer reconstruction whereas the remaining 21 received either primary closure (n = 8) or a pedicled pectoralis myocutaneous flap closure (n = 13). One of these patients receiving a pectoralis flap also received a free tissue transfer. Of the 28 patients receiving free tissue transfer, 15 (54%) received a radial forearm free flap for closure, 10 (36%) received an anterior lateral thigh free flap for closure, 2 (7%) received a supraclavicular free flap, and 1 (4%) received a fibula free flap (patient had an additional mandibular defect).

Thirty-two patients had a patch or tubed reconstruction of their laryngectomy or pharyngolaryngectomy defect. For these patients, an additional subgroup analysis was performed using 31395 as the CPT code to see whether this would make the calculator more predictive. Six of these patients had undergone a total laryngopharyngectomy. Unfortunately, because of the small number of patients in this cohort, as well as the fact that there was no consistent recording of the amount of residual pharynx after resection, it was not possible to perform a separate subgroup analysis based on the defect size.

The total numbers of complications observed were as follows: pneumonia, 1 (2%); SSI, 6 (12%, all superficial and/or associated with a fistula); VTE, 0; and ROR, 6 (12%: 4 for washout of fistula, 1 for repair of a chyle leak, and 1 for fibula free flap removal [not part of the laryngectomy closure]). **Figure 1** shows the comparison of the number of complications predicted by the NSQIP calculator vs what was observed.

Overall, the NSQIP risk calculator was a poor predictor of perioperative complications in our cohort of laryngectomies with or without a free flap reconstruction or preoperative RT or CRT. The percent error of complications predicted by the NSQIP calculator vs what was observed was broken down by type. Total predicted surgical site infections had the closest estimation whereas ROR and mean LOS were underestimated by the calculator. Percent error between the calculator's prediction and observed complication rate for those subgroups was as follows: pneumonia, 165% overestimation; SSI, 7% underestimation; ROR, 24% underestimation; and mean LOS, 13% underestimation.

As mentioned in the Statistical Analysis section, Brier scores were calculated to provide an estimate of the risk calculator's predictive value of binary outcomes, with a score of 0.01 used as a threshold for good performance corresponding to approximately a 90% forecast accuracy.[4,7,8] Brier scores for the group of all patients undergoing TL (N = 49) were greater than 0.01 with regard to forecasting postoperative pneumonia, SSI, and ROR, indicating a performance worse than our theoretical threshold of a test with good efficacy (**Figure 2**). Furthermore, the majority of TL subgroups analyzed had Brier scores greater than 0.01 as well.

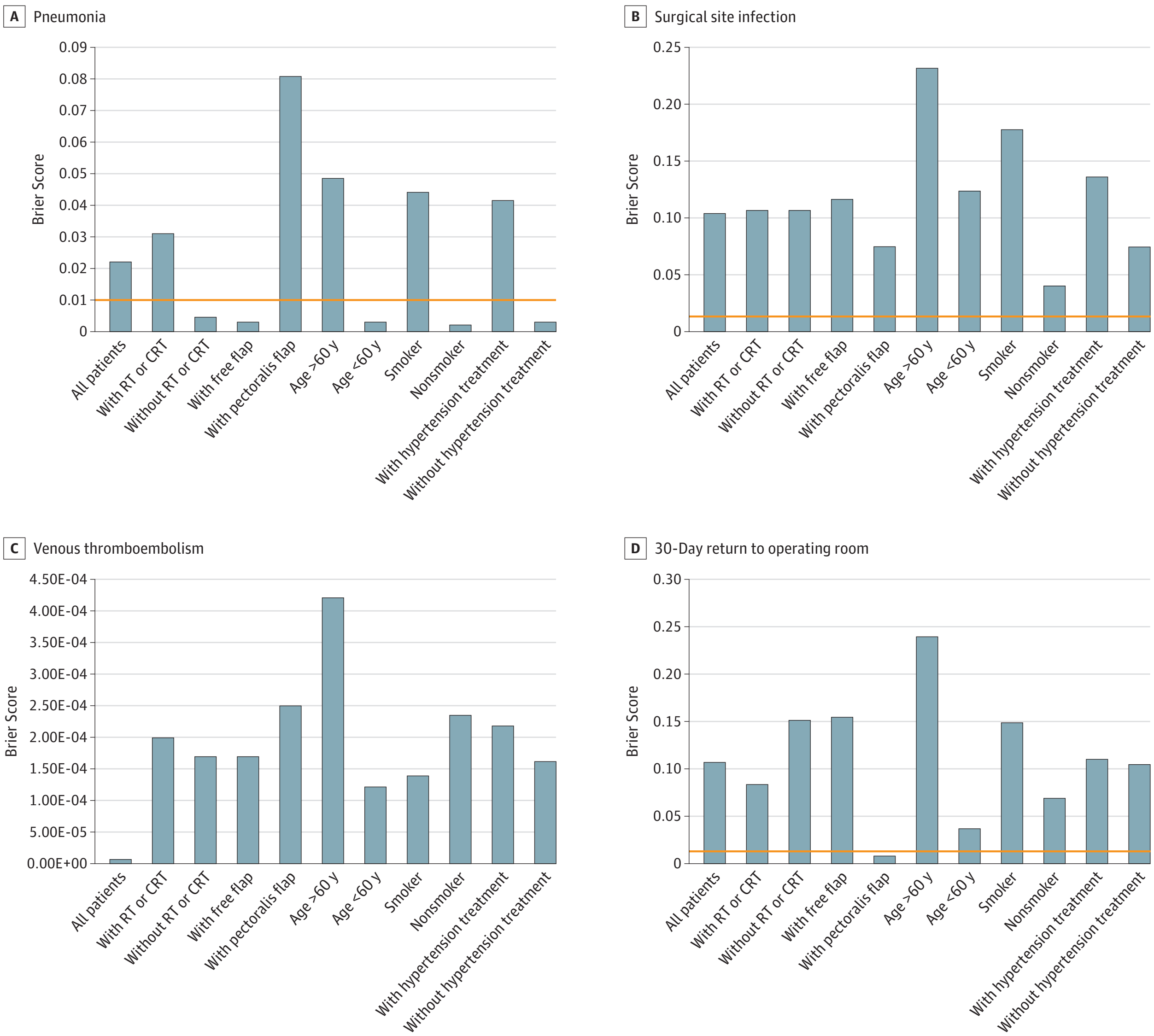
A few exceptions were noted on subgroup analysis with a Brier score of less than 0.01, including 30-day ROR for patients undergoing laryngectomy who underwent pectoralis flap reconstruction (Brier score = 0.008). Brier scores of less than 0.01 for postoperative pneumonia were observed in patients undergoing laryngectomy who did not receive preoperative RT or CRT (0.005), were younger than 60 years (0.002), had never received treatment for hypertension (0.003), and who were nonsmokers (0.002); however, there were no observed postoperative cases of pneumonia in these subgroups, which precludes full interpretation of results.

Likewise, Brier scores less than 0.01 were calculated for all subgroups when predicting VTE. However, there were no observed VTEs in our patient population; thus, interpretation is at high risk for type I error (false positive).

In the subgroup analysis using the CPT code 31395, there was not a significant improvement in the calculator's efficacy at predicting perioperative complications. No pneumonias or VTEs were actually observed in this subgroup, which precluded interpretation. All ROC curves were less than 0.8 with pneumonia, SSI, ROR, and VTE. Using the calculator, the risk of SSI was underestimated by 72% (Brier score, 0.108), ROR was overestimated by 12% (Brier score, 0.086), and LOS was underestimated by 29%.

Receiver operating characteristic curve analysis revealed that nearly all TL subgroups had a calculated AUC of less than 0.8 for each complication, suggesting that the NSQIP calculator has poor sensitivity and specificity for predicting a potential complication in this cohort. However, the calculator had relatively good sensitivity and specificity of predicting SSI for patients undergoing TL with free flap reconstruction as shown

Figure 2. Brier Scores of All Perioperative Complications Investigated in Patients Undergoing Total Laryngectomy




A value of less than 0.01 indicates approximately a 90% forecast accuracy. The solid horizontal line illustrates this theoretical threshold for an acceptable predictive model. There were no observed venous thromboembolisms during this study; therefore, interpretation of predicted venous thromboembolisms is limited. CRT indicates chemoradiation therapy, and RT, radiation therapy.

in **Figure 3** (AUC = 0.83). The calculator also had an AUC = 0.88 when predicting SSI in nonsmoking patients undergoing TL, indicating good sensitivity and specificity for predicting SSI in nonsmoking patients undergoing TL.

Pearson product-moment correlation coefficients were calculated for each subgroup comparing estimated LOS with observed LOS as shown in **Figure 4**. The Pearson *r* value for the all TL group of patients was −0.072, indicating there was no correlation between estimated and observed LOS when all patients undergoing TL were examined as a whole. On subgroup analysis, the *r* value for patients undergoing TL who did not receive preoperative RT or CRT was 0.28, again indicating poor correlation between estimated and observed LOS for this subgroup. The *r* values for patients who were younger than 60 years (−0.006), had never received treatment for hypertension (0.015), and were nonsmokers (−0.19) similarly indicated poor correlation between estimated and observed LOS. A similar lack of correlation was observed in the subgroup analysis of patients using the CPT code of 31395 (LOS was underestimated by 29%, *r* = 0.10). On the whole, there were no positive or negative correlations when analyzing any subgroup for LOS in patients undergoing TL.

## Discussion

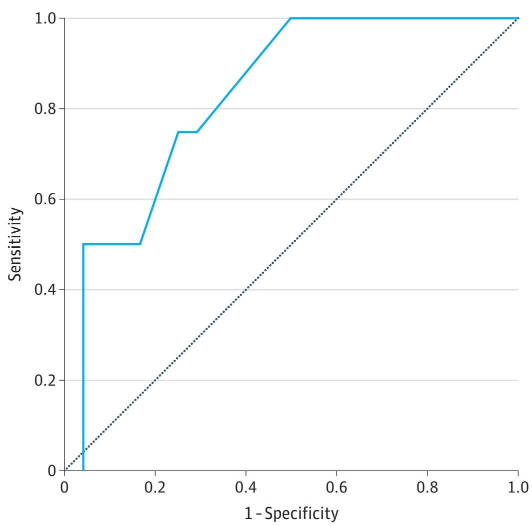
Since its inception and publication, the NSQIP risk calculator has been evaluated across several different surgical subspecialties including general surgery, gynecologic oncology, and orthopedic surgery, with each study having drawn different conclusions.[7,10,11] To our knowledge, this is the only otolaryngologic evaluation of the calculator's accuracy in predicting the risk of perioperative complications in a specific major head and neck surgery procedure. The results of this study indicate that the ACS NSQIP risk calculator had limited efficacy at predicting most perioperative complications in individual head and neck oncology patients undergoing TL. The calculator was inaccurate with regard to predictions made for patients who were preoperatively treated with RT or CRT. Also, it was generally inaccurate in patients who received surgery alone, making it difficult to draw a definitive conclusion about the effects of RT and/or CRT on the calculator's predictions.

Of note, the Brier score for 30-day ROR for patients undergoing laryngectomy who underwent pectoralis flap reconstruction was 0.008; however, the ROC curve had an AUC less than 0.8. These findings indicate that although the NSQIP calculator may have some predictive value, it has poor sensitivity and specificity for predicting an individual patient's risk of 30-day ROR.

It is also worth noting that Brier scores for predicting VTE were less than 0.01 in all subgroups, although Brier scores are at risk of committing type I error when few positive events (complications) are recorded. In our study, there were no recorded VTEs in the perioperative period; therefore, it is difficult to ascertain whether the NSQIP calculator has useful predictive value for predicting VTE from this cohort. Figure 3 depicts the ROC curve for SSI in patients undergoing laryngectomy who received free flap reconstruction, which reveals a predictive model in this subgroup (AUC = 0.83). This finding is concordant with a 7% error comparing total predicted SSI vs total SSI observed. Because of this group's Brier score greater than 0.01 with a favorable ROC curve analysis and relatively low percent error, these conflicting results may suggest that the NSQIP calculator has some poorly characterized role in predicting SSI in patients undergoing laryngectomy who receive free flap reconstruction.

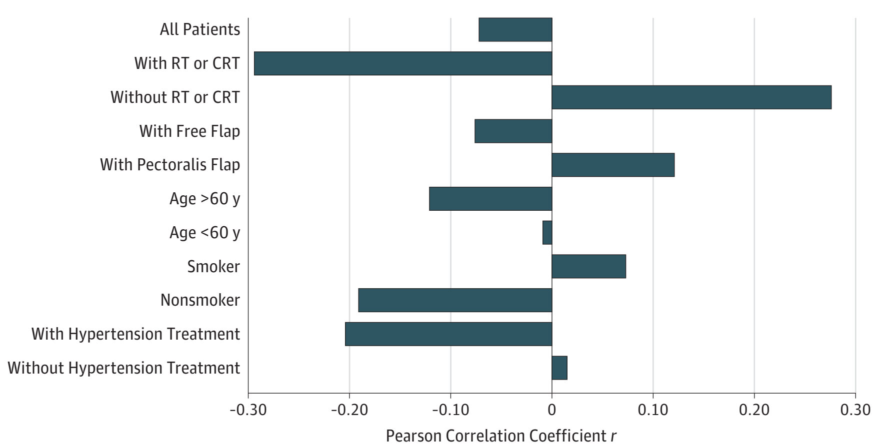
A review of the NSQIP data set found that RT was predictive of an increased risk of complication after laryngectomy.[6] A more recent review of the NSQIP database published in 2014 did not find that preoperative RT or CRT was indicative of increased risk of wound complications; however, the authors made a valid observation that is pertinent to our study: the NSQIP database underrepresents the utilization of laryngectomy nationwide.[12] Mlodinow et al[12] were able to include 713 laryngectomies from the NSQIP data set spanning 2006 through 2012. Using the Nationwide Inpatient Sample and *International Statistical Classification of Diseases* codes, a 2012 publication found that 3414 laryngectomies were performed

Figure 3. Receiver Operating Characteristic Curve for Surgical Site Infection (SSI) in Patients Undergoing Total Laryngectomy With Free Flap Reconstruction




This graph helps to represent the sensitivity and specificity of the National Surgical Quality Improvement Program calculator for this subgroup. There is a significant difference between the receiver operating characteristic curve for predicted SSI and the null hypothesis (diagonal line), indicating that the NSQIP calculator is better than chance alone at predicting SSI in this subgroup; area under the curve = 0.83.

Figure 4. Pearson *r* Values Comparing Predicted to Observed Length of Stay in All Total Laryngectomy Subgroups




There are no correlations in any subgroup between predicted and observed length of stay. CRT indicates chemoradiation therapy, and RT, radiation therapy.

nationwide in 2008 alone.[13] The number of otolaryngologic cases included in the calculator from the years 2009 through 2012 is 32 489, representing only 2.3% of the more than 1.4 million cases used to compile the calculator, and the number of laryngectomies included in the calculator is unclear.

There are also several characteristics of the head and neck oncology population that are not accounted for in the calculator. This may have affected the outcome of our study especially in light of the fact that the calculator overestimated perioperative pneumonia and underestimated total SSI, mean LOS, and ROR. In addition to cofounding factors such as RT and CRT, additional changes in airway and swallowing physiology may play a role. For example, a patient undergoing laryngectomy would be considered to be at lower risk for perioperative pneumonia than many other patients undergoing large head and neck procedures due to the fact that the pharynx and esophagus are separated from the airway during the procedure. Moreover, LOS predicted showed poor correlation across all subgroup analyses, which could be partially explained by the complexity of head and neck oncology patients, many of whom require subacute rehabilitation placement and are in tenuous psychosocial situations.

Currently, there is no good tool that uses preoperative RT or CRT as a variable to help predict perioperative complications such as SSI or ROR. Moreover, the full extent of resection (eg, amount of pharynx resected) and the complexity of the reconstruction are often not accurately captured simply by using CPT codes. In addition, perioperative variables such as operative time and the use of salivary bypass tubes are not included. These issues, and others mentioned herein, may be the subject of future investigation. Ideally, a specialty-specific calculator would be available to include these and other variables relevant to patients with head and neck cancer to more precisely predict outcomes and better counsel patients.

There are several limitations to our study. First, all information was recorded retrospectively, which carries a risk of introducing recall bias. Our study is a relatively small series, and all of our patients were treated at a tertiary referral academic center. Therefore, these results may not be generalizable to the community at large and the power of the study may be inadequate to observe more subtle trends suggested by the calculator. Furthermore, few of the patients included in our study were receiving steroids, dialysis, or mechanical ventilation or had disseminated cancer, which were factors included in the NSQIP calculator. This decreased the perioperative risk of complications according to the NSQIP calculator; however, those patients are generally not operative candidates for TLs.

Given the limited utility of the NSQIP calculator in our data set for predicting most complications after TL, it should be validated against a larger data set of patients undergoing TL before being used as a quality measure. We believe that this study is an important step for otolaryngologists, and head and neck surgeons in particular, toward active involvement in defining the quality measures by which many of our reimbursements will soon be judged.[14]

## Conclusions

The ACS NSQIP risk calculator appears to be of limited utility in our pilot cohort for predicting the individualized risk following TL, irrespective of preoperative receipt of RT and CRT, as well as the type of closure. This study represents the first otolaryngologic study to evaluate a risk calculator's predictions in a specific major head and neck procedure and is important in the pursuit of helping to define quality standards in the field of otolaryngology.

### ARTICLE INFORMATION

**Accepted for Publication:** May 25, 2016.

**Published Online:** July 28, 2016. doi:10.1001/jamaoto.2016.1809.

**Author Affiliations:** Department of Otolaryngology–Head and Neck Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Schneider); Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana (Deig, Prasad, Nelson, Mantravadi, Brigance, Moore); Department of Radiation Oncology, Indiana University School of Medicine, Indianapolis, Indiana (Langer); Department of Radiation Oncology, Winship Cancer Institute of Emory University, Atlanta, Georgia (McDonald); Department of Radiation Oncology, Moffitt Cancer Center, Tampa, Florida (Johnstone).

**Author Contributions:** Drs Schneider and Prasad had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
*Study concept and design:* Schneider, Prasad, Mantravadi, Johnstone, Moore.
*Acquisition, analysis, or interpretation of data:* Schneider, Deig, Prasad, Nelson, Brigance, Langer, McDonald, Moore.
*Drafting of the manuscript:* Schneider, Deig, Prasad, Nelson.
*Critical revision of the manuscript for important intellectual content:* All authors.
*Statistical analysis:* Schneider, Deig, Prasad, Nelson, Langer.
*Administrative, technical, or material support:* Deig, Mantravadi, Moore.
*Study supervision:* Prasad, Mantravadi, Langer, Johnstone, Moore.

**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Moore has received honoraria from AO North America. No other disclosures are reported.

### REFERENCES

1. Levit LA, Balogh EP, Nass SJ, Ganz PA. *Delivering High-Quality Cancer Care: Charting a New Course for a System in Crisis*. Washington, DC: Institute of Medicine of the National Academies; 2013.

2. Measure Applications Partnership. MAP Pre-Rulemaking Report: 2013 Recommendations on Measures Under Consideration by HHS. Washington, DC: National Quality Forum; 2013.

3. Cutler DM. Payment reform is about to become a reality. *JAMA*. 2015;313(16):1606-1607.

4. 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(5):833-842.e3.

5. American College of Surgeons National Surgical Quality Improvement Program. ACS NSQIP Surgical Risk Calculator. 2014. http://www.riskcalculator.facs.org/RiskCalculator/. Accessed June 28, 2016.

6. Schwartz SR, Yueh B, Maynard C, Daley J, Henderson W, Khuri SF. Predictors of wound complications after laryngectomy: a study of over 2000 patients. *Otolaryngol Head Neck Surg*. 2004;131(1):61-68.

7. Szender JB, Frederick PJ, Eng KH, Akers SN, Lele SB, Odunsi K. Evaluation of the National Surgical Quality Improvement Program universal surgical risk calculator for a gynecologic oncology service. *Int J Gynecol Cancer*. 2015;25(3):512-520.

8. Cohen ME, Bilimoria KY, Ko CY, Hall BL. 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(6):1009-1016.

9. Tazhibi M, Bashardoost N, Ahmadi M. Kernel smoothing for ROC curve and estimation for thyroid

stimulating hormone. *Int J Public Health Res.* 2011 (special issue):239-242.

10. 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.

11. Cologne KG, Keller DS, Liwanag L, Devaraj B, Senagore AJ. Use of the American College of Surgeons NSQIP surgical risk calculator for laparoscopic colectomy: how good is it and how can we improve it? *J Am Coll Surg*. 2015;220(3):281-286.

12. Mlodinow A, Khavanin N, Shires C, Samant S, Halen JV, Kim J. Predictors of adverse events after total laryngectomy: an analysis of the 2005-2012 NSQIP datasets. *Ann Otolaryngol Rhinol.* 2014;1(2): 1009-1014.

13. Maddox PT, Davies L. Trends in total laryngectomy in the era of organ preservation: a population-based study. *Otolaryngol Head Neck Surg*. 2012;147(1):85-90.

14. Gourin CG, Couch ME. Defining quality in the era of health care reform. *JAMA Otolaryngol Head Neck Surg*. 2014;140(11):997-998.

Invited Commentary

# Testing the Ability of the NSQIP Risk Calculator to Predict Laryngectomy Complications

Vic Velanovich, MD

Related article page 972

**The American College of Surgeons** National Surgical Quality Improvement Program (ACS-NSQIP) started in the Veterans Affairs (VA) health care system when the US Congress mandated that the VA improve the surgical care provided at its facilities. To those charged with this task, it was evident that not all VA hospitals provided care to the same mix of patient acuity. To be able to fairly compare hospitals, they developed and validated statistical methods to risk-adjust outcomes, based on preoperative risk factors. By 1991, this program became operational in the VA system.[1] As data were gathered on the performance of individual hospitals, this information was then confidentially reported to the individual VA hospitals in hopes that they would use it to improve quality. In 2001, the ACS received a grant to assess whether this program could be used in the non-VA facilities with a test of voluntary pilot hospitals. This proved to be feasible, and in 2004, the program was expanded to include any private and public (non-VA) hospital.[2] As of June 2016, 767 hospitals now participate in ACS-NSQIP.[3]

Underlying this expansion is the need for reliable data so that individual hospitals can understand their present performance and improve their quality. NSQIP's raison d'être is hospital quality improvement; it has been retrofitted as a database for research. The number of publications reporting results using NSQIP data has exploded. As NSQIP has tried to expand its usefulness, the database was used to develop a preoperative risk assessment tool.[3] And this is the rub: whenever data are coopted for purposes for which they were not originally intended (think post hoc subset analysis of clinical trials with negative results), there may be issues of interpretability. There have been studies previously showing that the NSQIP database may not adequately predict postoperative complications in complex surgical procedures. Specifically, in complex gastrointestinal oncologic operations[4] and pancreatectomy,[5] among others, the NSQIP risk database has been shown to be poorly predictive of postoperative morbidity. It was believed that in highly technical operations, in which complications are more related to surgical technique rather than patient comorbidities, a risk-adjusted tool based on preoperative risk factors may not adequately predict technique-related complications. It therefore is not surprising that the NSQIP risk calculator, a tool for which the NSQIP database was not originally intended, has a spotty track record. For example, a recent publication pertaining to microvascular head and neck reconstruction shows poor prediction performance.[6] Now Schneider et al[7] have added total laryngectomy to the list of complex procedures for which the NSQIP risk calculator may not be as accurate in predicting postoperative adverse events as we would like.

What Schneider and colleagues[7] have shown is that the NSQIP risk calculator in a cohort of 49 patients from a single, academic, tertiary medical center overestimated some risks, underestimated others, and was pretty close to a few. Their appropriate use of the Brier score to compare the predicted outcomes with the actual outcomes proves these discrepancies. They backed up this statistical model by analyzing the data with receiver operating characteristic curves showing relatively poor predictive value of the risk calculation for actual outcome. Their conclusions are clear.

Yet, this begs the question as to why this is. The origins of postoperative complications are manifold. They include patient factors (the presence of comorbidities and their severity), disease factors (the severity of the disease necessitating the operation), surgeon factors (how well the surgeon performs the operation), process factors (eg, administration of preoperative antibiotics), ancillary care factors (eg, anesthesia and nursing care), and postdischarge care factors (eg, home health care, rehabilitation). These factors may be clearly linked to an adverse event (the tracheostomy tube fell out in the intensive care unit, leading to acute respiratory distress); or, probably more commonly, they interact in opaque, nonlinear ways. Although the NSQIP risk calculator can enumerate the presence of comorbidities, it does not take into account how hard the operation was, how well the surgeon did it, whether best practices were followed, and how well patients were cared for after they left the hospital. It has always been the intent of NSQIP, as a quality improvement tool, to allow individual hospitals to investigate and modify these factors.

So, is the poor predictive value an intrinsic quality of the NSQIP database or a result of the patient mix and quality of care provided at the Indiana University Health University Hospital? I am sure that the surgeons are excellent and their