Predicting complication risk in spine surgery: a prospective analysis of a novel risk assessment tool

Anand Veeravagu, MD,1 Amy Li, BA,1 Christian Swinney, BA,1 Lu Tian, ScD,2 Adrienne Moraff, MD,1 Tej D. Azad, BA,1 Ivan Cheng, MD,3 Todd Alamin, MD,3 Serena S. Hu, MD,3 Robert L. Anderson, MS,4 Lawrence Shuer, MD,1 Atman Desai, MD,1 Jon Park, MD,1 Richard A. Olshen, PhD,2 and John K. Ratliff, MD1

Departments of 1Neurosurgery, 2Biomedical Data Science, and 3Orthopedic Surgery, Stanford University School of Medicine; and 4LogicThread Inc., Stanford, California

OBJECTIVE The ability to assess the risk of adverse events based on known patient factors and comorbidities would provide more effective preoperative risk stratification. Present risk assessment in spine surgery is limited. An adverse event prediction tool was developed to predict the risk of complications after spine surgery and tested on a prospective patient cohort.

METHODS The spinal Risk Assessment Tool (RAT), a novel instrument for the assessment of risk for patients undergoing spine surgery that was developed based on an administrative claims database, was prospectively applied to 246 patients undergoing 257 spinal procedures over a 3-month period. Prospectively collected data were used to compare the RAT to the Charlson Comorbidity Index (CCI) and the American College of Surgeons National Surgery Quality Improvement Program (ACS NSQIP) Surgical Risk Calculator. Study end point was occurrence and type of complication after spine surgery.

RESULTS The authors identified 69 patients (73 procedures) who experienced a complication over the prospective study period. Cardiac complications were most common (10.2%). Receiver operating characteristic (ROC) curves were calculated to compare complication outcomes using the different assessment tools. Area under the curve (AUC) analysis showed comparable predictive accuracy between the RAT and the ACS NSQIP calculator (0.670 [95% CI 0.60–0.74] in RAT, 0.669 [95% CI 0.60–0.74] in NSQIP). The CCI was not accurate in predicting complication occurrence (0.55 [95% CI 0.48–0.62]). The RAT produced mean probabilities of 34.6% for patients who had a complication and 24% for patients who did not (p = 0.0003). The generated predicted values were stratified into low, medium, and high rates. For the RAT, the predicted complication rate was 10.1% in the low-risk group (observed rate 12.8%), 21.9% in the medium-risk group (observed 31.8%), and 49.7% in the high-risk group (observed 41.2%). The ACS NSQIP calculator consistently produced complication predictions that underestimated complication occurrence: 3.4% in the low-risk group (observed 12.6%), 5.9% in the medium-risk group (observed 34.5%), and 12.5% in the high-risk group (observed 38.8%). The RAT was more accurate than the ACS NSQIP calculator (p = 0.0018).

CONCLUSIONS While the RAT and ACS NSQIP calculator were both able to identify patients more likely to experience complications following spine surgery, both have substantial room for improvement. Risk stratification is feasible in spine surgery procedures; currently used measures have low accuracy.

https://thejns.org/doi/abs/10.3171/2016.12.SPINE16969

KEY WORDS spine surgery; complications; risk assessment
THERE has been a significant rise in the number of spine surgeries performed in the United States over the last 2 decades.\(^5\)\(^\text{21,41}\) Perioperative adverse events and complications negatively impact clinical and patient-reported outcomes as well as increasing costs.\(^15\)\(^\text{43}\) Understanding and reducing the incidence of complications is an important clinical goal.

Adverse event occurrence is a common theme in quality-improvement efforts. Many national quality-improvement activities champion efforts to decrease complication occurrence in surgical procedures. Adverse event occurrence may be interpreted as an indicator of lower quality of care.

The incidence of adverse events provides a foundation for numerous quality-reporting metrics. Accurate risk adjustment is necessary to compare performance between different clinicians. Several studies investigated risk factors affecting the occurrence of postoperative complications in spine surgery patients. Patient factors influence complication occurrence, with sicker patients having a greater risk of perioperative complications.\(^23\)\(^\text{27,28,36,45}\) Poor modeling of complication occurrence and poor risk adjustment impairs comparative-effectiveness research and makes reliable assessment of expected occurrence of complications impossible.

Attempts have been made at risk modeling. Two commonly used tools for the assessment of complication risk in surgical patients are the Charlson Comorbidity Index (CCI) and the American College of Surgeons National Surgery Quality Improvement Program (ACS NSQIP) Surgical Risk Calculator. The CCI categorizes each patient’s morbidities and assigns an appropriate weight according to the associated risk of mortality. A sum of all the weights allows for a single cumulative CCI score for each patient.\(^8\) A number of studies have found the CCI score to be a competent predictor of survival and outcomes following various treatments such as radical prostatectomy, elective noncardiac surgery, dialysis, and peritoneal dialysis.\(^7\)\(^\text{16,18,19,24}\) The CCI was not specifically designed to assess patients undergoing spine surgery, and the usefulness of the CCI score to accurately predict complication occurrence after spine surgery remains uncertain.\(^2,22,44\)

The ACS NSQIP Surgical Risk Calculator was developed from a multi-institutional database of surgical procedures across all surgical specialties and draws on a model utilizing 21 preoperative factors to predict postoperative outcomes in individual patients.\(^3\) The ACS NSQIP Surgical Risk Calculator has been shown to perform as well as procedure-specific risk calculators for colon surgery patients.\(^3\) No study has evaluated the accuracy of this universal prediction tool in spine surgery patients. The ACS NSQIP calculator is a foundational element noted in a National Quality Forum (NQF)—sanctioned Physician Quality Reporting System element (Measure 358: Patient-Centered Surgical Risk Assessment and Communication); this measure is endorsed by NQF as applicable to patients undergoing a variety of surgical procedures, including many spine surgery procedures.\(^6\)

A previously described Risk Assessment Tool (RAT) was developed for patients undergoing spine surgery. The tool was developed using statistical modeling of longitudinal prospective data from an administrative claims database.\(^34\) This model predicts adverse events after spine surgery. The RAT incorporates procedure, patient, and diagnosis elements to develop a prediction rule for individual patients. Better modeling of complication occurrence will improve comparative-effectiveness research, provide superior patient counseling, and foster shared decision making.

Methods

Developing the Spinal Risk Assessment Tool

The senior author’s research group had previously developed an algorithm from a statistical model using a longitudinal claims database (MarketScan) of 279,145 spine surgery patients in the United States from 2006 to 2010.\(^34\) The algorithm was trained to predict adverse events from the first 179,145 records, with the remaining 100,000 records used for validation. We constructed a prediction rule from the model using logistic regression with main effects and logistic regression with additive main effects, in addition to 2 and 3 factor interactions. Receiver operating characteristic (ROC) curve analysis showed an area under the ROC curve (AUC) of 0.7 for the final model in predicting adverse events after spine surgery.

Patient Selection and Study Design

We initially tested the RAT on a retrospective cohort of 200 adults who underwent spine surgery performed by a single surgeon (J.K.R.) at Stanford University from 2011 to 2013. This initial study was a simple longitudinal assessment of a single practice. No inclusion or exclusionary criteria were applied. Complication occurrence was recorded through manual chart abstraction.

After a retrospective review revealed promising results of the RAT, a prospective assessment was performed. To avoid the bias intrinsic to retrospective reviews and to increase sample diversity, we conducted this 3-month prospective evaluation from February to April 2015, including all adult patients who underwent spine surgery procedures in either the departments of neurosurgery or orthopedic surgery. A trained, independent auditor prospectively followed a total of 246 patients undergoing 257 procedures and captured postoperative complication occurrences. One patient was lost to follow-up and was excluded only from analyses comparing observed outcomes and model predictions. There were no exclusionary criteria applied to the prospectively tracked patients. All patients were followed at centralized outpatient facilities, improving capture of postoperative adverse event occurrence.

For all measures, we used a broad definition of postoperative adverse events.\(^25\)\(^35\) We captured all adverse events occurring in the postoperative period, regardless of a presumed direct or indirect relationship to the surgery being performed. The study period was limited to the 30 days after a given procedure. The auditor followed patients prospectively during their hospitalization and recorded postoperative adverse events occurring after discharge based upon chart review and direct contact with treating clinical teams. While the analysis focused on 4 common types of perioperative complications (pulmonary, cardiac, wound,
and thrombotic), a wide variety of complications were captured in our assessment (urinary tract infection, postoperative radiculopathy, dysphagia, delirium, and others).

We applied the algorithm in the RAT to the prospective cohort to test for accuracy in predicting complication risks following spine surgery. The prospective calculation of RAT did not affect preoperative surgical optimization; all patients followed similar standard preoperative protocols. Where necessary, we conducted a retrospective review of the prospectively collected data to calculate the CCI score and the ACS NSQIP surgical risk for each patient.

Assessments for the RAT

We applied the RAT to both the retrospective and prospective cohorts. Preoperative and intraoperative characteristics (9 factors total) for each patient were entered into the RAT’s interface. Required preoperative variables include age, sex, preoperative diagnosis or indication for surgery (degenerative, trauma, tumor, or infection/miscellaneous), and comorbidities. The medical comorbidities available to be selected include pulmonary dysfunction, neurological dysfunction, hypercholesterolemia, smoking (active and former), hypertension, cardiac dysfunction, diabetes mellitus, systemic malignancy, gastroesophageal dysfunction, substance abuse, and psychiatric disorder. Intraoperative characteristics include the surgical approach (anterior cervical, anterior thoracolumbar, posterior cervical, or posterior thoracolumbar), the use of bone morphogenetic protein (BMP), surgical fusion, number of operative spinal levels, and the placement of instrumentation. Based on preoperative factors and surgery performed, the RAT generates a series of risk scores for adverse event occurrence after the planned spine surgery. The risk scores include a general case probability for cardiovascular, thrombotic, pulmonary, wound and overall complications.

Assessment of the CCI and the ACS NSQIP Surgical Risk Calculator

The CCI and the ACS NSQIP Surgical Risk Calculator were applied to the prospective cohort only. Table 1 compares the measured factors between the RAT, CCI, and ACS NSQIP calculator. The CCI generates a summed score based on the number and type of patient morbidities. A higher CCI score indicates a higher risk for potential postoperative complications. The ACS NSQIP Surgical Risk Calculator estimates the risk of complications after surgery and is available as an interactive online tool (http://riskcalculator.facs.org/). The calculator was applied to each patient after entering the procedure name or CPT code as well as 21 preoperative factors in its online interface. The ACS NSQIP surgical risk output for each patient includes the percent risk of any or a serious complication, pneumonia, cardiac complication, surgical site
infection, urinary tract infection, venous thromboembolism, renal failure, return to operating room, death, and discharge to postacute care. A serious complication was defined by the calculator as cardiac arrest, myocardial infarction, pneumonia, progressive renal insufficiency, acute renal failure, pulmonary embolism, deep vein thrombosis, return to the operating room, deep incisional surgical site infection, organ space surgical site infection, systemic sepsis, unplanned intubation, urinary tract infection, or wound disruption. The calculator further defined the category of “any complication” as including all of the “serious complications” previously listed in addition to superficial incisional surgical site infection, ventilator usage for more than 48 hours, and stroke.

The ACS NSQIP calculator records complications similar to those captured in our assessment, but there are specific subcategories of complications in the ACS NSQIP calculator not measured in the RAT. Likewise, some rare, severe complications of spine surgery procedures, such as postoperative neurological deficit, are not captured in the ACS NSQIP assessment (Table 2). Our focus is on the overall complication risk measured in the RAT and ACS NSQIP calculator to provide an adequate comparison between the 2 models.

### Statistical Analysis

We compared the performance of 3 prediction models (RAT, CCI, and ACS NSQIP Surgical Risk Calculator) for predicting complication risk. First, an ROC curve was estimated for each model based on observed occurrence of complications, and 95% confidence intervals for the AUCs were constructed. Comparisons of the AUCs for the RAT, CCI, and ACS NSQIP Surgical Risk Calculator were performed using DeLong’s test. Next, we estimated the average distance between the predicted risk and the observed binary outcome indicating the occurrence of complication to measure prediction accuracy. The average prediction accuracy was compared among 3 predictions with paired t-tests. Lastly, we used a nonparametric spline estimator to estimate complication risk as a function of each of the risk predictions. If a risk prediction is accurate for the true risk, we would expect the nonparametrically estimated risk function to be similar to the identity function. If the estimated risk function is consistently above (or below) the identity function, then the corresponding prediction model systematically underestimates (or overestimates) the actual risk. We performed all data analysis using R (http://www.R-project.org).

### Results

#### Patient Population and Definition of a Complication

Patient demographics and comparisons of both the retrospective cohort (n = 200) and prospective cohort (n = 246) are summarized in Table 3. We used a previously validated, broad definition of perioperative complications, including medical complications and adverse events not directly related to surgery. We used a 30-day window for adverse event occurrence.

#### Retrospective Assessment

In the retrospective cohort (n = 200), the mean age of the patients (± SD) was 50.8 ± 14.0 years and 57.5% were male. The mean body mass index (BMI) was 28.6 ± 6.5. Most patients (92.5%) presented with at least 1 comorbidity, the most common being hypertension (39.0%), psychiatric disorder (35%), and smoking (34.0%). The primary diagnosis or indication for a spinal procedure was degenerative disease in 68.0% of patients, trauma in 11.0% of patients, tumor in 9.5% of patients, and infection/miscellaneous in 11.5% of patients. Five of the 200 patients in the retrospective cohort had 2 procedures as part of a staged operation on different dates for a total of 205 procedures. Approximately half (50.2%) of all procedures were performed in the posterior thoracolumbar spine. Other approaches were performed on the anterior cervical (22.4%), anterior thoracolumbar (2.9%), and posterior cervical spine (24.4%). Of the total group of procedures, 60.0% were performed on more than 1 spinal level, 3.4% used BMP, 67.3% included fusion, and 68.3% involved instrumentation placement. Seventy-two procedures (35.1%) led to at least 1 complication postoperatively. Of the 4 common complication types assessed (pulmonary, cardiac, thrombosis, and wound-related), the most common was cardiac (9.8%), followed by wound issues (4.9%), pulmonary (4.4%), and thrombosis (1.0%).

#### Prospective Assessment

In the prospective cohort (n = 246), the mean age of the patients was 61.6 ± 15.3 years and 57.3% were male. The mean BMI was 27.4 ± 5.5. Five patients did not have their BMI values recorded. A majority of patients (74.4%) presented with at least 1 comorbidity, the most common being hypertension (45.9%), hypercholesterolemia/hyperlipidemia (28.9%), and cardiac dysfunction (25.2%). The primary diagnosis or indication for spine surgery was

### Table 2. Complication types in outcomes measured by the RAT and the ACS NSQIP Surgical Risk Calculator

<table>
<thead>
<tr>
<th>Complication Type</th>
<th>RAT</th>
<th>ACS NSQIP Surgical Risk Calculator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac dysrhythmia</td>
<td>Cardiac complication (cardiac arrest or MI)</td>
<td></td>
</tr>
<tr>
<td>CHF</td>
<td>Death</td>
<td></td>
</tr>
<tr>
<td>Delirium</td>
<td>Discharge to nursing or rehab facility</td>
<td></td>
</tr>
<tr>
<td>DVT</td>
<td>Pneumonia</td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>Pulmonary embolism</td>
<td></td>
</tr>
<tr>
<td>Neurological</td>
<td>Readmission</td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>Renal failure (progressive renal insufficiency or acute renal failure)</td>
<td></td>
</tr>
<tr>
<td>Pulmonary</td>
<td>Return to OR</td>
<td></td>
</tr>
<tr>
<td>Pulmonary embolus</td>
<td>SSI (deep, superficial, or organ space)</td>
<td></td>
</tr>
<tr>
<td>Renal failure</td>
<td>Stroke</td>
<td></td>
</tr>
<tr>
<td>UTI</td>
<td>Systemic sepsis</td>
<td></td>
</tr>
<tr>
<td>Wound hematoma</td>
<td>Unplanned intubation</td>
<td></td>
</tr>
<tr>
<td>Wound infection</td>
<td>Ventilator &gt; 48 hrs</td>
<td></td>
</tr>
<tr>
<td>Other wound</td>
<td>Venous thromboembolism</td>
<td></td>
</tr>
<tr>
<td>complication</td>
<td>Unplanned intubation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wound disruption</td>
<td></td>
</tr>
</tbody>
</table>

DVT = deep venous thrombosis; OR = operating room; rehab = rehabilitation; SSI = surgical site infection; UTI = urinary tract infection.
spinal degenerative disease (87.9%), followed by trauma (6.5%), tumor (4.5%), and infection/miscellaneous (2.0%). Eleven of the 246 patients in the prospective cohort had 2 procedures on separate days as part of a staged operation, yielding a total of 257 procedures. The procedures largely involved the posterior thoracolumbar spine (63.8%), with 13.2% involving the anterior cervical spine, 9.3% involving the anterior thoracolumbar spine, and 13.6% involving the posterior cervical spine. Of the total group of procedures, 48.6% were performed on greater than 1 spinal level, 0.4% involved the use of BMP, 55.6% involved fusion, and 59.1% involved placement of instrumentation.

Of the 246 patients assessed prospectively, 1 patient was lost to follow-up and could not be assessed for postoperative outcomes following surgery. Of the remaining 245 patients with a total of 256 procedures, 69 patients (73 procedures, 28.5%) had at least 1 postoperative complication. We assessed 4 complications specifically (pulmonary, cardiac, thrombosis, and wound-related) and found the following rates of complications in the cohort: 10.2% cardiac, 9.8% wound-related, 4.3% pulmonary, and 0.8% thrombosis. Other complications (13.3%) were not classified into the 4 most common groups.

### Predicting Adverse Events With RAT

The RAT generates a series of risk scores for adverse events after spine surgery based on 9 preoperative factors: age, sex, diagnosis, comorbidity burden, surgical approach, BMP use, fusion status, number of spinal levels operated on, and instrumentation use. The output of the tool provides the probability of complication occurrence.

We initially tested the RAT on the retrospective cohort to determine its ability to distinguish between patients who would and who would not experience a postoperative complication. The mean predicted probability of experiencing a complication was significantly higher in patients with observed postoperative complications (n = 72, mean 0.449) compared with the no-complication group (n = 133, mean 0.371, p = 0.0436, t-test).

We then applied the RAT to a prospective cohort. Table 4 provides a summary of the predicted probabilities of occurrence of overall complications as well as specific complications. The mean probability of complication occurrence calculated by the application was 0.273 ± 0.196 in the entire cohort of 257 procedures in 246 patients (range 0.049−0.964). The remaining prospective cohort of 256 procedures (excluding the single procedure undergone by the patient lost to follow-up) was stratified into 3 equal groups based on the calculated probability for complication occurrence in RAT: a low probability score (n = 86, range 0.049−0.157), a medium probability score (n = 85, range 0.158–0.289), and a high probability score (n = 86, range 0.291–0.964). Out of those with low probability scores, 12.8% were observed to have postoperative complications. Out of those with medium probability scores, 31.8% ultimately had postoperative complications. The group with high probability scores had the highest percentage of observed postoperative complications at 41.2% (Fig. 1). The predicted complication rates from the RAT were similar to the observed complication rates in each low, medium, and high score group: 10.1% in the low score.

#### TABLE 3. Patient characteristics in retrospective and prospective cohorts

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Retrospective Cohort</th>
<th>Prospective Cohort</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>200</td>
<td>246</td>
<td></td>
</tr>
<tr>
<td>Mean age (yrs)</td>
<td>50.8 ± 14.0</td>
<td>61.6 ± 15.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean BMI</td>
<td>28.6 ± 6.5</td>
<td>27.4 ± 5.5</td>
<td>0.04</td>
</tr>
<tr>
<td>Male sex (%)</td>
<td>57.5</td>
<td>57.3</td>
<td>0.75</td>
</tr>
<tr>
<td>Comorbidity burden (%)</td>
<td>92.5</td>
<td>74.4</td>
<td>0.74</td>
</tr>
<tr>
<td>Pulmonary dysfunction (%)</td>
<td>31.0</td>
<td>11.4</td>
<td>0.98</td>
</tr>
<tr>
<td>Neurological dysfunction (%)</td>
<td>31.0</td>
<td>4.9</td>
<td>0.99</td>
</tr>
<tr>
<td>Hypercholesterolemia (%)</td>
<td>32.5</td>
<td>28.9</td>
<td>0.83</td>
</tr>
<tr>
<td>Smoking (active and former)</td>
<td>34.0</td>
<td>6.1</td>
<td>0.19</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>39.0</td>
<td>45.9</td>
<td>0.72</td>
</tr>
<tr>
<td>Cardiac dysfunction (%)</td>
<td>20.5</td>
<td>25.2</td>
<td>0.41</td>
</tr>
<tr>
<td>DM</td>
<td>17.0</td>
<td>15.4</td>
<td>0.46</td>
</tr>
<tr>
<td>Systemic malignancy (%)</td>
<td>11.5</td>
<td>11.4</td>
<td>0.32</td>
</tr>
<tr>
<td>Gastroesophageal dysfunction (%)</td>
<td>26.0</td>
<td>15.9</td>
<td>0.21</td>
</tr>
<tr>
<td>Substance abuse (%)</td>
<td>10.5</td>
<td>2.0</td>
<td>0.73</td>
</tr>
<tr>
<td>Psychiatric disorder (%)</td>
<td>35.0</td>
<td>20.7</td>
<td>0.75</td>
</tr>
<tr>
<td>Other</td>
<td>39.0</td>
<td>11.2</td>
<td>0.22</td>
</tr>
<tr>
<td>Preop diagnosis (%)</td>
<td></td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Degenerative disease (%)</td>
<td>68.0</td>
<td>87.9</td>
<td></td>
</tr>
<tr>
<td>Tumor</td>
<td>9.5</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>11.0</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>Infection/misc.</td>
<td>11.5</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>No. of procedures</td>
<td>205</td>
<td>257</td>
<td></td>
</tr>
<tr>
<td>Location of surgery (%)</td>
<td></td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Anterior cervical (%)</td>
<td>22.4</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>Posterior cervical (%)</td>
<td>24.4</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>Anterior thoracolumbar (%)</td>
<td>2.9</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>Posterior thoracolumbar (%)</td>
<td>50.2</td>
<td>63.8</td>
<td></td>
</tr>
<tr>
<td>Procedure factors (%)</td>
<td>78.0</td>
<td>68.9</td>
<td>0.20</td>
</tr>
<tr>
<td>&gt;1 level surgery</td>
<td>60.0</td>
<td>48.6</td>
<td>0.80</td>
</tr>
<tr>
<td>Use of BMP</td>
<td>3.4</td>
<td>0.4</td>
<td>0.01</td>
</tr>
<tr>
<td>Fusion status</td>
<td>67.3</td>
<td>55.6</td>
<td>0.64</td>
</tr>
<tr>
<td>Instrumentation status</td>
<td>68.3</td>
<td>59.1</td>
<td>0.97</td>
</tr>
<tr>
<td>Complications (%)</td>
<td>35.1</td>
<td>28.5</td>
<td>0.14</td>
</tr>
<tr>
<td>Pulmonary (%)</td>
<td>4.4</td>
<td>4.3</td>
<td>0.72</td>
</tr>
<tr>
<td>Cardiac (%)</td>
<td>9.8</td>
<td>10.2</td>
<td>0.79</td>
</tr>
<tr>
<td>Thrombosis (%)</td>
<td>1.0</td>
<td>0.8</td>
<td>0.001</td>
</tr>
<tr>
<td>Wound</td>
<td>4.9</td>
<td>9.8</td>
<td>0.70</td>
</tr>
<tr>
<td>Other</td>
<td>26.8</td>
<td>13.3</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Mean values are given with standard deviations. Percentages for location of surgery, procedure factors, and complications were calculated on the basis of the number of procedures in the given cohort. For all other variable categories, the percentages were calculated on the basis of the number of patients. One patient in the prospective cohort (who underwent 1 operation) was lost to follow-up; for this reason, rates of complications for the prospective cohort were calculated on the basis of 256 procedures rather than 257.

* Based on 2-sample t-test for continuous variables and Pearson's chi-square test for categorical variables.
A. Veeravagu et al.

GROUP 1
Pulmonary 4.3 25.8 (10.1–73.3) 0.6 (0.1–4.0)
Cardiac 10.2 41.8 (17.2–78.4) 0.3 (0–2.1)
Thrombosis 0.8 11.2 (3.4–28.3) 0.8 (0.2–4.4)
Wound/surgical site 9.8 22.6 (10.3–46.5) 1.5 (0.3–8.1)
UTI — 1.2 (0.2–6.5)
Renal failure — 0.2 (0–1.5)
Return to OR — 3.1 (0.8–13.7)
Death — 0.4 (0–7.0)
Discharge to post-acute care — 11.7 (0.9–77.4)

GROUP 2
Any complication risk (%) 28.5 27.3 (4.9–96.4) 7.2 (1.3–26.9)

GROUP 3

Complication rates were analyzed by patient BMI on a per-procedure basis in the prospective cohort (256 procedures), using standard BMI categories: less than 18.5 (n = 5, observed complications in 20.0%), 18.5–24.9 (n = 90, observed complications in 25.6%), 25.0–29.9 (n = 81, observed complications in 25.9%), and 30.0 or greater (n = 75, observed complications in 34.7%). Univariate regression showed no correlation between BMI and observed complication occurrence (p = 0.143).

Prospective patients who developed complications had, on average, a significantly higher preoperative predicted probability for experiencing an adverse event after spine surgery. The mean predicted probability of a complication was 0.346 for patients who experienced an actual complication (n = 73) compared with 0.242 for patients who had no postoperative complications (n = 183, p = 0.0003).

The RAT further calculated the probabilities of specific complication occurrence in the pulmonary, cardiac, thrombosis, and wound categories for each patient. When stratifying patients in the prospective cohort between those who experienced pulmonary complications postoperatively and those who did not, the mean calculated probability for pulmonary complications with the RAT was significantly different between the groups. Those who actually experienced pulmonary complications (n = 11) had a higher predicted probability of pulmonary complications (mean 0.432) compared with the no-pulmonary complication group (n = 245, mean 0.249, p = 0.0004, t-test). The RAT did not distinguish between patients who had a cardiac complication or wound complication and those who did not (n = 26, mean predicted probability 0.462 with a complication and n = 230, mean predicted probability 0.411 for patients without a complication, p = 0.153, for cardiac complication; n = 25, mean predicted probability 0.246 for patients with a complication and n = 231, mean predicted probability 0.223 for patients without a complication, p = 0.095, for wound complications). Only 2 patients had a thrombosis complication in the entire cohort; this sample size was insufficient to perform a statistical analysis.

CHARTER COMORBIDITY INDEX

We calculated the CCI score for the prospective cohort only. The overall mean CCI score was 0.853 (range 0–9). A low score (0–3) was calculated for 245 procedures, a medium score (4–5) for 6 procedures, and a high score (6+) for 6 procedures. The low numbers of patients in the CCI medium and high score groups may reflect the inability of the CCI score to reflect comorbidities that predict complication occurrence in spine surgery procedures.

One patient from the low score group was lost to follow-up and was subsequently excluded. Of those in the remaining low score group, 28.7% were observed to have postoperative complications. In the medium score group, 33.3% resulted in actual postoperative complications. In the highest score group, postoperative complications were observed after 16.7% of total procedures (Fig. 1). The CCI score did not distinguish between those who did and did not experience a complication. The mean CCI score for procedures after which patients experienced at least 1 complication was 1.0 compared with 0.8 for procedures after which patients had no complications (p = 0.332, t-test).

ACS NSQIP SURGICAL RISK CALCULATOR

In the prospective cohort, the ACS NSQIP Surgical Risk Calculator yielded individual percent risk scores for developing an adverse event after surgery in the following categories: serious complication (mean 5.4% ± 3.4%, range 1.3%–22.0%), any complication (mean 7.2% ± 4.7%, range 1.3%–26.9%), pneumonia (mean 0.6% ± 0.7%, range 0.1%–4.0%), cardiac complication (mean 0.3% ± 0.3%, range 0%–2.1%), surgical site infection (mean 1.5% ± 1.1%, range 0.3%–8.1%), urinary tract infection (mean 1.2% ± 1.0%, range 0.2%–6.5%), venous thromboembolism (mean 0.8% ± 0.7%, range 0.2%–4.4%), renal failure (0.2% ± 0.2%, range 0%–1.5%), return to operating room (mean 3.1% ± 1.9%, range 0.8%–13.7%), death (mean 0.4% ± 0.8%, range 0%–7.0%), and discharge to postacute care (mean 11.7% ± 14.4%, range 0.9%–77.4%) (Table 4).

The prospective cohort was stratified into 3 groups based on the estimated percent risk of experiencing any complications after surgery, as determined by the ACS NSQIP Surgical Risk Calculator preoperatively. The remaining cohort (n = 236 procedures) was classified as low (n = 87, range 1.3%–4.7%), medium (n = 84, range 4.8%–7.6%), and high (n = 85, range 7.7%–26.9%) risk score groups. Of those in the low “any complications” risk score group, 12.6% were observed to develop complications. In the medium risk score group, 34.5% were observed to have postoperative complications. Of those categorized in the highest risk score group, 38.8% ultimately developed at least 1 postoperative complication (Fig. 1). The predict-
ed complication rates from the ACS NSQIP Surgical Risk Calculator consistently underestimated the observed complication rates in each low, medium, and high score group. The calculated predicted complication rate was 3.4% in the low score group, 5.9% in the medium score group, and 12.5% in the high score group.

The ACS NSQIP Surgical Risk Calculator further generated chance of outcome statistics in the form of “above average,” “average,” and “below average” risks of any complication occurrence for each patient. Of the procedures leading to an “above average” risk of complication occurrence (n = 229), 27.5% (n = 63) were observed to have at least 1 postoperative complication. Those classified with an “average” risk (n = 21) had an observed complication rate of 42.9% (n = 9). Those classified with a “below average” risk (n = 6) had an observed complication rate of 16.7% (n = 1).

The ACS NSQIP Surgical Risk Calculator significantly distinguished between those who developed complications and those who did not. The mean calculated risk of a serious complication for procedures associated with a postoperative complication was 6.22% compared with 5.05% in those without any postoperative complications (p = 0.012, t-test). The mean calculated risk of having any complications at all for those who actually experienced a postoperative complication (8.55%) was also significantly higher than that in the no-complication group (6.72%, p = 0.004, t-test).

Comparing Prediction Models

In ROC analyses, RAT and the ACS NSQIP Surgical Risk Calculator similarly predicted adverse event occurrence after spine surgery with moderate efficacy. In the prospective cohort, the AUC for predicting complication occurrence was 0.670 (95% CI 0.599–0.741) in RAT and 0.669 (95% CI 0.601–0.737) in the ACS NSQIP Surgical Risk Calculator. The CCI was a poor predictor of complication occurrence in the spine patient population, with an AUC of 0.555 (95% CI 0.485–0.625). A summary of the ROC analyses is shown in Fig. 2.

The AUCs in the RAT and ACS NSQIP Surgical Risk Calculator were statistically indistinguishable from each other (p = 0.975). However, the AUC in the CCI score was significantly lower (worse) than that in the RAT and the ACS NSQIP Surgical Risk Calculator (p = 0.003, p = 0.003). Based on the ROC curves, the predictive accuracies of RAT and ACS NSQIP scores are comparably fair. While ROC curve analysis suggests a relatively high probability of concordance of predicted risk and actual risk of complication in both the RAT and ACS NSQIP models, it does not examine absolute value accuracy. We
measured the absolute value distance between the predicted probability risk and the observed binary complication occurrence in each model. The average prediction error scores for RAT and ACS NSQIP were 0.309 and 0.359, respectively. The difference between the 2 scores was $-0.051$ (95% CI $-0.082$ to $-0.019$, $p = 0.0018$). Therefore, the average prediction error of the scores generated by RAT was significantly lower than that generated by the ACS NSQIP Surgical Risk Calculator.

We further examined the relationship between the outcome of actual complication occurrence and predicted scores. A nonparametrically estimated risk as a function of the risk prediction (RAT or ACS NSQIP model) was graphically reported in Fig. 3. We found that the predicted risk scores from RAT were substantially closer to the observed complication risks, while the ACS NSQIP risk scores systematically underestimated the true complication risk for most patients. For the entire prospective population, the true complication rate was 28.5%. The average predicted complication rate from our model in RAT was 27.2% compared with the 7.2% rate predicted by the ACS NSQIP Surgical Risk Calculator.

Discussion

Risk Adjustment

Assessing the rate of adverse event occurrence is a theme of many quality-improvement efforts; decreasing the risk of complication occurrence in surgical procedures is a worthy goal. Adverse event occurrence may be interpreted as an indicator of lower quality of care; some public reporting efforts use complication occurrence as a proxy for the quality of care provided by a physician (https://projects.propublica.org/surgeons/).

Population-based assessments that do not incorporate specific patient and procedural factors may not adequately risk-adjust and may not appreciate the risk of complications in high-risk patients. Measures that fail to capture accurate relative risk profiles and fail to offer risk adjustment may not offer proper assessment of clinician performance. These measures may falsely impugn clinicians who care for high-risk patients.

Quality improvement efforts should incorporate risk adjustment to insure that patient and procedure factors are incorporated into assessing the performance of clinicians.

Risk Assessment Tool

Our group developed the Risk Assessment Tool (RAT) from an administrative claims database of over 270,000 spine patients to predict adverse events in patients undergoing spine surgery. This study evaluated the accuracy of the RAT in predicting adverse events after individual spine surgeries. We applied the algorithm prospectively to a patient cohort whose adverse event rate was 28.5% and found that the probability score generated by the RAT significantly distinguished between those that did and did not experience at least 1 postoperative complication. The model was effective at predicting general complication occurrence with an AUC of 0.670.

A streamlined approach by the RAT is feasible without losing predictive power because of our emphasis on specific preoperative and intraoperative risk factors applicable to spine surgery patients. Patient variables such as age, sex, and the number and type of medical comorbidities have been found to influence the risk of complications that include deep wound infection, cardiopulmonary complication, thromboembolic disease, pulmonary embolism, cerebrovascular accident, and neurological deficit. The types of medical comorbidities that the RAT incorporates are diverse, ranging from cardiac dysfunction to psychiatric disorders. Psychiatric disorders are well established in the spine population, considering the high
Predicting complication occurrence

Although a subset of these patients were incorporated in its development, the ACS NSQIP Surgical Risk Calculator was not designed specifically for spine surgery patients, although a subset of these patients were incorporated in its development.36 To our knowledge, no study has yet examined the accuracy of the universal ACS NSQIP Surgical Risk Calculator when applied only to spine surgery patients. Our study offers evidence that this risk assessment tool can be applied with fair predictive accuracy to spine surgery populations. The ACS NSQIP Surgical Risk Calculator, similar to the RAT, significantly differentiated between patients with and without adverse events after spine surgery. Its predictive accuracy was similar to the RAT with an AUC of 0.669. Despite their similar probabilities of concordance, the average prediction error of the scores generated by the RAT was significantly lower than for the ACS NSQIP Surgical Risk Calculator, suggesting a higher absolute value accuracy of the RAT for predicting adverse events in patients undergoing spine surgery.

The ACS NSQIP Surgical Risk Calculator, while proven to be an effective predictive tool from ROC analysis, consistently underestimates true risk in the spine surgery population. The underestimation of complication risk in the ACS NSQIP instrument was evident from another analysis performed outside of the nonparametric spline estimations. After stratifying the percent risk scores into 3 groups (low, medium, and high scores), the predicted complication rates for each group were compared with the observed complication rates. We found that the ACS NSQIP Surgical Risk Calculator consistently underestimated the risk of complications in our prospective cohort in each low, medium, and high score group. The predicted complication rates from the RAT were similar to the observed complication rates in the low, medium, and high probability score groups.

Comparison of the RAT and ACS NSQIP Calculator

The accuracy of the RAT’s prediction model is similar to that of the current universal adverse events predictor offered by the ACS NSQIP Surgical Risk Calculator in ROC analysis. The ACS NSQIP calculator, however, underestimated complication occurrence in our prospective patient population. The ACS NSQIP Surgical Risk Calculator was not designed specifically for spine surgery patients, although a subset of these patients were incorporated in its development.36 To our knowledge, no study has yet examined the accuracy of the universal ACS NSQIP Surgical Risk Calculator when applied only to spine surgery patients. Our study offers evidence that this risk assessment tool can be applied with fair predictive accuracy to spine surgery populations. The ACS NSQIP Surgical Risk Calculator, similar to the RAT, significantly differentiated between patients with and without adverse events after spine surgery. Its predictive accuracy was similar to the RAT with an AUC of 0.669. Despite their similar probabilities of concordance, the average prediction error of the scores generated by the RAT was significantly lower than for the ACS NSQIP Surgical Risk Calculator, suggesting a higher absolute value accuracy of the RAT for predicting adverse events in patients undergoing spine surgery.

The ACS NSQIP Surgical Risk Calculator, while proven to be an effective predictive tool from ROC analysis, consistently underestimates true risk in the spine surgery population. The underestimation of complication risk in the ACS NSQIP instrument was evident from another analysis performed outside of the nonparametric spline estimations. After stratifying the percent risk scores into 3 groups (low, medium, and high scores), the predicted complication rates for each group were compared with the observed complication rates. We found that the ACS NSQIP Surgical Risk Calculator consistently underestimated the risk of complications in our prospective cohort in each low, medium, and high score group. The predicted complication rates from the RAT were similar to the observed complication rates in the low, medium, and high probability score groups.

Limitations of the ACS NSQIP Calculator

The limitations of the ACS NSQIP Surgical Risk Calculator for spine surgery patients include consistently underestimating the risk of complications and generating chance-of-outcome statistics that may not be applicable. Such limitations may result from the universal calculator’s inability to capture neurological complications or other complications specific to spine surgery. While these complications are exceedingly rare, they are relevant to patient education and should be captured in prospective risk modeling.

Other aspects of the ACS NSQIP Surgical Risk Calculator may contribute to the difficulty of its use. The ACS NSQIP Surgical Risk Calculator (http://riskcalculator.facs.org/) requires an input of a single CPT code for the type of procedure performed. No option exists to enter the multiple CPT codes that are often required for accurate description of some operations. In this study, we only selected the first CPT code provided for each patient, though other aspects of the operation (e.g., instrumentation) were excluded. The ACS Surgical Risk Calculator requires the input of 21 preoperative factors into its online interface in addition to the CPT code. The RAT requires only 9 preoperative and intraoperative factors.

Charlson Comorbidity Index

The CCI is another well-studied risk assessment tool that successfully predicted outcomes after a variety of procedures.7,8,16,18,19,28 While it was not designed specifically for the spine surgery population, previous studies have analyzed its utility for predicting outcomes in spine surgery patients. One study found the CCI score to be a significant predictor of 30-day complications following surgery for metastatic spine disease.2 Another study of its use on spine surgery patients found the CCI score to be a useful comorbidity index but not a complete predictor of postoperative complications due to its inadequacy for capturing spine-specific comorbidities or assigning appropriate weights to conditions that may or may not affect outcomes after spine surgery.44 In this study, we found the CCI score to be a poor predictor of complication occurrence after spine surgery with an AUC of 0.555. In our assessment, the vast majority of our patients had low scores on the CCI. This may reflect the inability of the CCI to capture comorbidities that predict complication occurrence. A significant difference between the predictive model of the CCI and that of both RAT and ACS NSQIP Surgical Risk Calculator demonstrates the inadequacy of the CCI compared with the other models in this study.

Other Approaches to Risk Modeling in Spine Surgery

Others have explored risk modeling in spine surgery. Lee et al.20 developed a predictive algorithm (available to the public through SpineSage.com) based on known risk factors. This model was found to be effective, with an AUC of 0.76 for any medical complications and 0.81 for major medical complications. The authors used this model to generate an absolute risk percentage, which was then used to counsel patients regarding potential risks of surgery.20 This represents an advantage of their model, as the ability to provide an absolute percent likelihood of adverse events after spine surgery may be easier for patients to comprehend. While the generation of an absolute risk percentage is similar to the RAT, their model is limited by its basis on a much smaller patient size of 1476 drawn from only 2 academic institutions. In comparison, RAT is more robust, developed from records of 279,145 spine surgery patients nationwide.

Another new classification system, known as Spine Adverse Events Severity (SAVES), was also developed and tested at a single institution to determine its reliability as a tool for providers to identify complications after spine surgery.13 Further studies have validated its use in captur-
ing more adverse events as well as more people with adverse events after spine surgery.\textsuperscript{20,24} The SAVES system is primarily a grading tool for adverse events after they occur and, while useful for possible future refinements of adverse event predictor models, is not meant to be a predictive tool like the RAT for spine surgery patients.

Limitations

There are significant limitations in the RAT measure. The RAT measure was developed from an administrative database on which the predictive modeling was based. The development of the statistical model was based on a retrospective data analysis, and the results depend on the choice of cohort. The large administrative claims database was noted to have an underrepresentation of elderly patients, who represent a large portion of spine surgery patients. One should use caution in generalizing the current results to a different population. The present effort is based upon single-institution prospective assessment; broadening the assessment to other institutions will be required.

Dissemination of the RAT

We believe that wider use of the RAT may allow it to be further refined by other investigators. Similarly, the predictive algorithms used in the RAT calculations are too involved to be routinely used in clinical practice. To provide wider use of the tool, an iOS application incorporating the predictive algorithm has been developed and may be downloaded, free of charge, by interested parties (http://itunes.apple.com/app/ratool/id1087663216).

Conclusions

This prospective analysis of a predictive tool for assessing the risk of surgical complications illustrates that relative risk adjustment in musculoskeletal procedures is feasible. The predictive tool incorporates patient (demographic, comorbidities), procedure (type of surgery, number of levels, whether fusion and/or instrumentation was performed), and diagnostic (surgical indication) elements into a predictive model.

The investigation validates the algorithms used by the RAT to predict postoperative complications in spine surgery patients. Although the RAT and the ACS NSQIP Surgical Risk Calculator have substantial room for improvement, they both performed significantly better than the CCI. Moreover, the results of this study show that the predictive utility of the RAT is superior to that of the ACS NSQIP Surgical Risk Calculator for this patient cohort. The ACS NSQIP Surgical Risk Calculator consistently underestimated complication occurrence in our patient cohort. Patient counseling based upon the operative risk assessment provided by the ACS NSQIP instrument will provide an inaccurately low assessment of operative risks.

These findings may indicate that models particular to a given surgical field may prove more valuable than general, multispecialty tools.

Acknowledgments

Funding for this work was provided by the Stanford Medical Scholars Program, the Stanford Society of Physician Scholars, the Walsh Foundation, and the Orthopaedic Research Education Foundation.

References

17. Faciszewski T, Winter KB, Lonstein JE, Denis F, Johnson L:


Disclosures
Dr. Cheng reports consultant relationships with NuVasive, Globus Medical, and Stryker Spine.

Author Contributions

Correspondence
John Ratliff, Department of Neurosurgery, Stanford University School of Medicine, 300 Pasteur Dr., Stanford, CA 94305-5327. email: jratliff@stanford.edu.

J Neurosurg Spine Volume 27 • July 2017 91