Analysis and Review of Automated Risk Calculators Used to Predict Postoperative Complications After Orthopedic Surgery

Robert K. Merrill[1] · John M. Ibrahim[1] · Anthony S. Machi[1] · James S. Raphael[1]

Abstract
Purpose of Review To discuss the automated risk calculators that have been developed and evaluated in orthopedic surgery.
Recent Findings Identifying predictors of adverse outcomes following orthopedic surgery is vital in the decision-making process for surgeons and patients. Recently, automated risk calculators have been developed to quantify patient-specific preoperative risk associated with certain orthopedic procedures. Automated risk calculators may provide the orthopedic surgeon with a valuable tool for clinical decision-making, informed consent, and the shared decision-making process with the patient. Understanding how an automated risk calculator was developed is arguably as important as the performance of the calculator. Additionally, conveying and interpreting the results of these risk calculators with the patient and its influence on surgical decision-making are paramount.
Summary The most abundant research on automated risk calculators has been conducted in the spine, total hip and knee arthroplasty, and trauma literature. Currently, many risk calculators show promise, but much research is still needed to improve them. We recommend they be used only as adjuncts to clinical decision-making. Understanding how a calculator was developed, and accurate communication of results to the patient, is paramount.

Keywords Automated risk calculator · Orthopedic surgery · Postoperative complications · Predictive modeling

Introduction

An integral part of the informed consent and shared decision-making process includes ensuring that the patient has a full understanding of the potential risks, benefits, and alternatives to treatment [1, 2]. Additionally, an adequate understanding of risk helps surgeons to choose the most appropriate treatment for his or her patient. Several risk factors have been identified for specific orthopedic procedures, which further helps surgeons risk stratify patients on an individual basis. Identifying modifiable risk factors, such as uncontrolled diabetes or smoking, can allow the patient and surgeon to make

preoperative interventions to reduce the risk of postoperative complications [3]. A thorough understanding of risk by both the surgeon and patient facilitates the shared decision-making process and optimizes the treatment plan.

From a quality improvement perspective, risk estimation may help reduce adverse events and The Centers for Medicare and Medicaid Services (CMS) may soon incentivize surgeons to incorporate risk assessment tools into the informed consent process for elective surgery [4]. Growing interest and emphasis on the importance of risk prediction in the orthopedic community have led to the development of several automated risk calculators [5–17]. Automated risk calculators allow the user to input specific patient characteristics into a calculator for a specific procedure to produce an automated risk prediction of adverse outcomes. While not a surgical risk calculator, the fracture risk assessment tool (FRAX) is an example of an automated risk calculator to predict the 10-year probability of hip and major osteoporotic fractures, and aids in the decision-making process to begin pharmacologic treatment for osteoporosis [18]. For surgery, accurate risk calculation is an

---

This article is part of the Topical Collection on *The Use of Technology in Orthopaedic Surgery—Intraoperative and Post-Operative Management*

✉ Robert K. Merrill
  merrillr@einstein.edu

[1] Department of Orthopedic Surgery, Albert Einstein Medical Center, 5501 Old York Road, Philadelphia, PA 19141, USA

invaluable tool for the orthopedic surgeon during the perioperative decision-making process.

The largest body of research for automated risk calculators in orthopedics revolves around the fields of spine, total joint arthroplasty, and trauma. The objective of this review is to provide a summary of the most extensively evaluated automated risk calculators used to predict postoperative complications after orthopedic surgery.

Risk Calculator Evaluation

Prior to presenting the literature on automated risk calculators, we will briefly discuss the metrics used to evaluate the performance of automated risk calculators to familiarize the reader with the commonly used terms and tests presented in the literature. A review by Mansmann et al. provides an excellent summary on the methods behind development and interpretation of automated risk calculators [19•]. This comprehensive explanation is beyond the scope of this review, but is a good reference to review a background on risk calculators.

Discrimination is the ability of a predictive model to separate patients who experienced the outcome from those who did not. Discrimination is measured by the area under the receiver operating characteristic curve (AUC), and is synonymous with the concordance statistic (c-statistic). The value of the AUC or C-statistic ranges from 0 to 1 and is the probability that a randomly selected individual in which the predicted outcome occurred had a higher predicted probability than a randomly selected individual who did not experience the outcome. Therefore, a model with an AUC of 0.5 is no better than random chance [20]. In addition, values of 0.6–0.7 indicate poor prediction, 0.7–0.8 indicate fair prediction, 0.8–0.9 indicate good prediction, and 0.9–1 indicate excellent prediction. Of note, terms like "fair" and "good" are not necessarily standardized in the orthopedic literature, and the performance of clinical prediction models is often assessed by the authors evaluating them.

Calibration is the second major metric used to evaluate predictive models, and indicates how similar the predicted risk is to the true, observed risk. An accurate model is well calibrated and the predicted risk percentages will be close to the observed risk percentages [20]. The Hosmer-Lemeshow statistical test assesses calibration and indicates good calibration if the p value is > 0.05. The Brier score is another measure of calibration with a perfect calibration having a Brier score of 0 [21].

To summarize, an automated risk calculator with a high AUC, high Hosmer and Lemeshow p value, and low Brier's score suggests a model with good discrimination and calibration.

Spine

The field of spine surgery has a variety of automated risk calculators that have been developed and validated for complications after various types of procedures and diagnoses (Table 1).

Risk Assessment Tool

Ratliff et al. used data from the MarketScan database on 279,145 patients to develop a spine surgery risk assessment tool (RAT), available publicly on iOS devices [17]. Development of the RAT involved any patient undergoing spine surgery. The RAT uses surgical factors and patient characteristics to predict 30-day postoperative complications. A benefit of the RAT is that it was developed from spine surgery patients, making it more relevant to the field of spine surgery. The RAT had an AUC score of 0.70 for predicting complications after all procedures, and AUC scores ranging from 0.66 to 0.73, depending on the approach and region of the spine, suggesting relatively good clinical prediction. The authors note that the MarketScan database contains almost no patients of Medicare-age, which is one of the most rapidly growing age groups in spine surgery [17]. Veeravagu et al. expanded on the previous study by Ratliff and prospectively compared the predictive ability of the RAT and the National Surgical Quality Improvement Program Surgical Risk Calculator (NSQIP-SRC) for determining complications in patients undergoing spine surgery. The authors found the RAT and NSQIP-SRC to have respective AUC values of 0.67 and 0.70, with no statistically significant difference between the two. The authors also found, however, that the RAT had significantly lower prediction error compared with the NSQIP-SRC, and that the NSQIP-SRC significantly underestimated the risk of developing complications [24]. Therefore, the RAT and NSQIP-SRC had similar discrimination, but the RAT had better calibration.

NSQIP-SRC

The NSQIP-SRC has been evaluated by a number of independent studies for spine surgery. A major drawback of the NSQIP-SRC is that it was developed using a heterogenous patient population. In an attempt to evaluate the NSQIP-SRC on a more homogenized group, Sebastian et al. used single-level lumbar spine fusion cases from the 2015 NSQIP database, and compared the development of actual complications with complications predicted by the NSQIP-SRC [22]. The authors found that in general for 30-day complication risk, the NSQIP-SRC was suboptimal with a c-statistic of 0.61 for any complication. The NSQIP-SRC performed best for VTE, although it was still limited (c-statistic 0.66). The authors suggest the NSQIP-SRC may be limited because it

Table 1 Summary of automated risk calculators assessed in the spine literature

Calculator	Link	Procedure	Main outcome	Evaluation metrics	Author comments
Risk Assessment Tool (RAT) [17]	https://apps.apple.com/us/app/risk-assessment-tool-for-spine-surgery-procedures/id1087663216	Spine surgery (unspecified)	30 day post-operative complications	AUC: 0.70	Database does not contain Medicare-aged patients Lower prediction error compared with NSQIP-SRC
NSQIP-SRC [22, 23]	http://riskcalculator.facs.org/RiskCalculator/	1-level lumbar fusion	30 day post-operative complications VTE	AUC: 0.61 AUC: 0.66	Developed with wide variety of patients Only accepts 1 CPT code Inaccurate predictor of complications
SpineSage [15]	http://depts.washington.edu/spinersk/	Spine surgery (unspecified)	30 day medical complications 30 day major medical complications	AUC: 0.76 AUC: 0.81	Accurate predictor of complications Advantages: specific patient comorbidities and surgical variables Disadvantages: developed with small cohort (1476 patients)
Global Spine Tumour Study Group (GSTSG) Risk Calculator [16]	https://spinemet.com/	Spine surgery for symptomatic spinal metastasis	Death at 2 years after surgery	AUC: 0.68	Developed from a single cohort at one center

was developed with a wide variety of patients and that the development of risk calculators may require specialized populations. Another reason the authors suggest the calculator may have performed poorly is because it only accepts 1 CPT code which may underestimate the complexity of spine surgery patients [22]. For elderly (> 60 years) patients undergoing laminectomy without fusion at an institution in China, Wang et al. evaluated the NSQIP-SRC and concluded that it was not an accurate predictor of complications following surgery based on AUC scores and Brier scores (0.68 and 0.32, respectively) [23].

SpineSage

Another publicly available online automated risk calculator is SpineSage™, developed by Lee et al. using 1476 patients from the Spine End Result Registry (SERR) [15]. On internal validation, the model was found to be accurate with AUC values for 30-day medical and major medical complications of 0.76 and 0.81, respectively. This calculator was further assessed by Kasparek et al. on 273 patients and found similar AUC values of 0.85 for medical complications and 0.71 for major medical complications. The authors concluded that the calculator was accurate at determining complications following various spine procedures [14, 15]. The benefits of this calculator are that it includes several surgery-specific variables and patient comorbidities. The drawback is that it was developed using only 1476 patients which may not be a large enough cohort for the development of a predictive model.

Global Spine Tumour Study Group Risk Calculator

Choi et al. developed an automated, publicly available online risk calculator to predict mortality of patients with spinal metastases at various time points based on specific patient and tumor characteristics at the time of treatment [16]. The model was developed based on 1264 patients from the Global Spine Tumour Study Group (GSTSG). The calculator performed better than commonly used prognostic scoring systems described by Tomita et al. and Tokuhashi et al. for predicting death at 2 years after surgery [25, 26]. The authors highlight important methodologic considerations for the development of the calculator, and suggest it may be a valuable adjunct to clinician decision-making when discussing surgical options for patients with spinal metastases [16].

While several automated risk calculators have been developed to predict complications after spine surgery, further studies are required to confirm their accuracy and benefit. The consensus seems that the ACS-NSQIP-SRC underestimates the risk of postoperative complications and may be suboptimal. The RAT, SpineSage™, and GSTSG are more optimal tools to guide spine surgeons in understanding risk of postoperative complications after surgery.

Total Joints

The total joints literature has an abundance of risk calculators that have been internally and externally validated (Table 2). Some of these calculators use heterogenous populations like the NSQIP-SRC, and others use total joint arthroplasty–specific patients and complications.

NSQIP-SRC

A number of studies have assessed the accuracy of the NSQIP-SRC to predict 30-day complications after THA and TKA. Edelstein et al. assessed the predictive ability of the risk calculator for 1764 Medicare patients at their institution who underwent primary unilateral THA or TKA [28]. The total complication rate in their cohort was 11.54%, 90.2% of which were considered serious complications. For the entire cohort, any complication, cardiac complication, pneumonia, and discharge to a rehab or nursing facility had risk estimates that were associated with an observed event occurrence, though the authors indicated that none of these outcomes had a strong c-statistic (> 0.8). When stratified for THA patients, the authors found that cardiac complications, pneumonia, and discharge to rehab facility risk estimates were associated with an actual event occurrence. In TKA patients, risk estimates for any complication, cardiac complication, pneumonia, and discharge to rehab were associated with actual event occurrence. None of the c-statistics for either the THA or TKA groups was considered strong (> 0.8). While the authors conclude that the NSQIP-SRC cannot accurately predict 30-day complications after TKA or THA, they did set a high standard for c-statistic values > 0.8. Some of the outcomes predicted, such as discharge to a rehab facility in THA patients, had relatively strong c-statistics (0.74) compared with other risk calculators that have been studied highlighting the need for individualized assessments.

Wingert et al. assessed the accuracy of the NSQIP-SRC specifically for predicting PJI after THA or TKA [29]. The authors analyzed 1620 total joint replacements and found that the NSQIP-SRC had an AUC of 0.74 for predicting PJI within 30 days, and an AUC of 0.71 for predicting PJI within 90 days and therefore conclude that the calculator is only a fair predictor of detecting acute postoperative PJI. Aspects of the study that may have led to suboptimal results include combining THA and TKA patients into a single cohort. Additionally, the NSQIP-SRC was not specifically designed to detect SSI, not PJI, and it is possible the authors had patients with a simple SSI rather than PJI, which could explain some of the discrepancy between the predicted and actual risk observed.

Goltz et al. analyzed 496 TKAs and 413 THAs using the NSQIP-SRC to predict discharge to SNF/rehab, DVT, 90-day readmission, PJI, and return to OR [30]. The authors included the above complications if they occurred within 90 days of surgery. The authors found the calculator to be most suitable to predict discharge to SNF/rehab with an AUC of 0.72 (TKA 0.75, THA 0.68). The remainder of the tested outcomes had either low discrimination (AUC < 0.7) or were not statistically significant, leading the authors to conclude that the risk calculator was only suitable for predicting discharge to SNF/rehab. The authors also noted that the calculator had very similar predicted length of stay to actual length of stay, only overestimating the actual length of stay by 0.2 days, despite this difference being statistically significant.

American Total Joint Replacement Registry Risk Calculator

The American Total Joint Replacement Registry Risk Calculator (AJRR) was developed using a sample of 65,499 Medicare patients who underwent THA, and 137,546 Medicare patients who underwent TKA. It inputs 30 patient variables to return a prediction of 90-day mortality and prosthetic joint infection within 2 years. Harris et al. performed an external validation of the AJRR using Medicare eligible patient from the Veterans Affairs Surgical Quality Improvement Program (VASQIP) to predict 90-day mortality after either TKA or THA [10]. The authors observed poor discrimination and calibration of the risk calculator, with a c-statistic of 0.62. They concluded the calculator was not accurate at predicting 90-day mortality in Medicare eligible VA patients, but acknowledged that the calculator's performance in other populations remains unknown. The authors suggest the poor performance of the calculator may be due to the relatively low mortality rate that occurs after THA and TKA, making it difficult to develop and accurate model around this outcome. Also, the differences in the sample used to create the calculator and used to test the calculator cannot be overlooked when performing external validation [10].

Total Knee Replacement Surgery – Risk IQ Tool

HealthGrades Inc. (Denver, CO) has a publicly available risk calculator for complications after TKA. The calculator differs somewhat from other calculators, in that it is intended for patient use and variables are input by the patients themselves. It also does not provide any risk percentage, but rather provides a qualitative assessment of risk. Romine et al. retrospectively reviewed 2284 primary TKAs performed at their institution and assessed the accuracy of the HealthGrades calculator to predict postoperative complications during the first 14 days postoperatively [27]. They included common postoperative complications such as PE, sepsis, wound complications, and many others. The authors observed a 3.6% complication rate in their patient cohort, which was significantly lower than that predicted by the calculator (12.4%), with an AUC value of 0.61. The authors highlight several reasons for this overestimation of risk, namely that the calculator was

Table 2 Summary of automated risk calculators used in the total joints literature

Calculator	Link	Procedure	Main outcome	Evaluation metrics	Author comments
Total-Knee Arthroplasty Revision Probability Calculator [13]	https://jordanbstarr.shinyapps.io/TKARevCalc/	Primary TKA	Revision TKA	Mean absolute error (1 year): 0.1% / Mean absolute error (5 years): 3.6%	Generalizability: developed from VA database / Has not been externally validated
PJI Calculator [11]	https://apps.apple.com/us/app/pji-calculator/id1435025720	Surgical Treatment for PJI	Treatment success as defined by Delphi criteria*	AUC: 0.69	Several modifiable risk factors were identified
PJI Risk Calculator	https://icmphilly.com/ortho-applications/prosthetic-joint-infection-pji-risk-calculator/	Primary or revision TKA or THA	PJI / Antibiotic-resistant PJI / S. aureus PJI	AUC**: 0.83, 0.84 / AUC**: 0.86, 0.83 / AUC**: 0.86, 0.73	Contains detailed surgical variables / Strong AUC value / Generalizability: externally validated at a high-volume institution, may not be applicable at low-volume institutions
Total Knee Replacement Surgery - Risk IQ tool [27]	https://www.healthgrades.com/risk-iq/total-knee-replacement-surgery	Primary TKA	Complications up to 14 days postoperatively	AUC: 0.61	Lower observed complication rate than predicted complication rate / Generalizability: developed from Medicare patients
American Joint Replacement Registry (AJRR) Total Joint Replacement Risk Calculator [10]	https://teamwork.aaos.org/ajrr/SitePages/Risk%20Calculator.aspx	Primary TKA or THA	Mortality within 90 days / PJI within 2 years	AUC: 0.62	AJRR calculator is inaccurate in predicting outcomes of Veterans / AJRR calculator has poor discrimination and calibration / AJRR calculator likely lacks a statistical model / AJRR should undergo external validation
NSQIP-SRC [28–30]	http://riskcalculator.facs.org/RiskCalculator/	TKA or THA	Specific complications within 30 days / PJI within 30 days / PJI within 90 days / WIthin 90 days: discharge to SNF/rehab, DVT, readmission, PJI, return to OR	AUC for each complication: < 0.80 / AUC: 0.74 / AUC: 0.71 / AUC, SNF/rehab: 0.72 / AUC, all other complications: <0.70	Risk estimates were associated with actual event occurrence / NSQIP-SRC is designed to predict SSI, not PJI / Calculator may overestimate risk of PJI / Fair predictor of discharge to SNF/rehab / Limited use in predicting other complications / Accurate predictor of hospital length of stay

*Delphi criteria for surgical PJI treatment success: (1) infection eradication characterized by a healed wound without pain or infection recurrence caused by the same organism strain, (2) no subsequent surgical intervention for infection after reimplantation surgery, and (3) no occurrence of PJI related mortality

**AUC values include interval AUC first, followed by AUC value from external validation

developed using Medicare data, while the patients at their institution comprise a more diverse mix of patients. Medicare patients may have more inherent risk in developing a complication, leading to overestimation. Other reasons for overestimation by the calculator included possibly missing complications due to coding inaccuracy, and not capturing complications for patients who were readmitted to an outside institution which would have artificially lowered the observed complication rate in the author's cohort [27]. It is also important to note that the 14-day postoperative window would only capture the most acute complications following a TKA.

PJI Risk Calculator

Tan et al. examined 27,717 patients who underwent either THA or TKA and developed an automated calculator to predict periprosthetic joint infection [31]. Forty-two risk factors were initially analyzed, and 25 of them were found not to be significant. The remaining 17 factors were incorporated into the calculator, with the most significant being previous open surgical procedure, drug abuse, revision procedure, and HIV/ AIDs. The authors reported good AUC values on external validation of the calculator, with AUC values for any PJI, antibiotic-resistant PJI, and *S. aureus* PJI of 0.84, 0.83, and 0.73, respectively. One of the major benefits of this calculator is that it contains detailed surgical variables such as primary vs. revision procedure and number of prior surgeries, and weighs some of the strongest risk factors, such as drug abuse. While it requires further external validation, it has potential to be a valuable tool to stratify high-risk patients for developing PJI after total joint arthroplasty.

Kheir et al. developed an automated risk calculator to predict failure of surgical treatment of prosthetic joint infection after total knee or hip arthroplasty [11]. A total of 1438 patients with PJI were used to develop the calculator, incorporating 63 risk factors related to patient characteristics, microbiology data, and surgical variables. The final analysis yielded 10 significant risk factors that were used in the final calculator. The AUC for the calculator was 0.69. Several of the risk factors included in the calculator are modifiable, suggesting that patients may be optimized prior to surgical treatment of PJI, and the calculator may be a beneficial tool to communicate risk of failure to patients, though it has relatively poor discrimination.

Total-Knee Arthroplasty Revision Probability Calculator

In 2018, Starr et al. developed a risk calculator to predict risk of revision TKA [13]. The risk calculator was developed using 32,297 patients from the Veterans Affairs (VA) informatics and computing infrastructure who underwent a TKA and subsequent revision TKA. The variables included in the model were age, gender, BMI, DM, CKD, and chronic opioid use.

The authors included chronic opioid use within the calculator because it has been recently associated with early revision TKA [32, 33]. The mean absolute error of the calculator at 1 year was 0.1%, and 3.6% at 5 years. The calculator may not be generalizable because of its development from a VA database. Additionally, there are no other studies that have evaluated this risk calculator. Therefore, it should be used cautiously until further literature can support its accuracy.

Many of the authors that evaluated the NSQIP-SRC concluded that it is a poor predictor of postoperative complications after TKA or THA. Additionally, AAJR also was found to be inaccurate for predicting postoperative complications. Some of the calculators used to predict revision or success of PJI treatment performed the best, which may be because of the total joint specific populations and variables used to develop them.

Trauma and Fractures

Risk calculators for postoperative outcomes after fracture surgery have been less extensively studied. Hip fractures are of particular interest due to the increasing incidence in the elderly population and the clinical challenge they pose due to high morbidity related to a number of elderly comorbid conditions. Additionally, hip fractures pose a high economic burden, and the ability to predict and stratify high-risk patients may improve outcomes and reduce costs. The only other injury with an available automated risk calculator is distal radius fracture (Table 3).

NSQIP-SRC

Wang et al. used the NSQIP-SRC to analyze 410 elderly (age ≥ 60) patients with a femoral neck fracture who underwent hemiarthroplasty [34]. The authors set a cutoff of c-statistic > 0.83 to indicate the calculator being a good predictive tool. Only the c-statistic for mortality was above this cutoff (0.93), with the remainder of complications having a c-statistic below that cutoff value of 0.83. Incidence of reoperation had the lowest Brier's score, which was the only complication below the threshold the authors set of <0 .01. In conclusion, the authors suggest that the NSQIP-SRC is valuable for predicting mortality and reoperation after hip fracture surgery in elderly patients, but lacks the accuracy needed to predict some of the more common complications that occur. Lack of accuracy may be attributed to several elderly related comorbid conditions that are not included within the risk calculator. Further refinement of the calculator is needed, specifically with respect to surgical patients, before it becomes a commonly used tool for hip fracture patients [34].

Pugely et al. provide links to automated risk calculators for predicting 30-day mortality, any complication, major complication, and minor complication in patients undergoing surgery

Table 3 Summary of automated risk calculators used in the trauma literature

Calculator	Link	Procedure	Main outcome	Evaluation metrics	Author comments
NSQIP-SRC	http://riskcalculator.facs. org/RiskCalculator/	Femoral head replacement for femoral neck fractures in Chinese patients age ≥ 60	30-day mortality: C-statistic 30-day reoperation incidence: Brier's score	C-statistic: 0.93 Brier's score: < 0.01	Valuable for predicting mortality and reoperation after hip fracture surgery in elderly patients Not predictive of other common complications after hip surgery Does not account for variability in comorbidities in elderly populations Population was specifically elderly Chinese individuals
		ORIF, closed fixation, THA, HA for hip fractures	30-day morbidity and mortality in hip fracture surgery	C-statistic, mortality: 0.77 C-statistic, any complication: 0.67 C-statistic, major complication: 0.71 C-statistic, minor complication: 0.67	Predictive of both 30-day morbidity and mortality Stronger predictor of mortality than other complications Be aware that 1-year mortality for hip fracture is significantly greater than 30 days Delay in treatment did not affect mortality independently
Surgical Outcome Risk Tool (SORT)	www.sortsurgery.com	All orthopedic surgeries Surgical treatment of hip fracture	30-day mortality after all orthopedic surgery 30-day morbidity in hip fracture surgery	C-statistic: 0.93 C-statistic: 0.70	Developed as a generalized surgical risk calculator Fast and easy to use (6 variables) Variables are relatively subjective (Surgical urgency/severity) Poor calibration to Anesthesia Sprint Audit of Practice database of hip fracture cases Needs further examination to be valuable for predicting hip fracture surgery outcomes
Nottingham Hip Fracture Score (NHFS)	http://www. riskprediction.org. uk/index-nhfs.php	Surgical fixation of femoral neck fracture	30-day and 1-year morbidity in hip fracture surgery	Kaplan-Meier curve: $p < 0.001$ C-statistic: 0.71	NHFS is an accurate predictor of 30-day and 1-year mortality after hip fracture surgery From a single hospital in the UK Extensively validated, but not necessarily generalizable
Edinburgh Wrist Calculator (EWC)	www.trauma.co. uk/wristcalc	Closed reduction and casting of distal radius fracture	Malalignment within 6 weeks of reduction and casting	Accuracy: 0.77 Sensitivity: 0.95 Negative predictive value: 0.97	EWC is a better predictor of outcome when compared with expert opinion and majority rule
			Malalignment within 2 weeks of reduction and casting	AUC: 0.47 Specificity: 0.95 Sensitivity: 0.02 Positive predictive value: 0.33 Negative predictive value: 0.38	EWC is a poor predictor of fracture displacement EWC needs further validation

for hip fractures [8]. Patient data from the NSQIP database were used to evaluate the risk calculators. Various patient demographics and comorbid conditions are included as risk factors in the calculator. Predicting risk of mortality had the highest accuracy with a *c*-statistic of 0.7. All remaining outcomes had *c*-statistics less than 0.7, leading the authors to suggest that the strongest risk calculator was for predicting mortality. For mortality, the greatest risk factors included age > 80, male gender, decreased functional status, ASA class 3 or 4, and a history of cancer. The authors highlight that delay in time to surgery was not an independent risk factor for increased mortality, suggesting that the sickest patients are often those delayed, and it is their morbidity rather than delay that contributes to increased mortality rates [8].

Surgical Outcome Risk Tool, Nottingham Hip Fracture Score

Another automated risk calculator to predict 30 day postoperative mortality after non-cardiac surgery is the Surgical Outcome Risk Tool (SORT; sortsurgery.com), developed by Protopapa et al. using the National Confidential Enquiry into Perioperative Death (NCEPOD) data from 16,788 patents [7]. The SORT uses 6 variables: ASA-PS, urgency of surgery, high-risk specialty, surgical severity, cancer, and age ≥ 65 years. Marufu et al. used patients from the Anesthesia Sprint Audit of Practice to validate SORT for predicting 30-day mortality after hip fracture surgery and compared it with the Nottingham Hip Fracture Score (NHFS) [35]. Both the NHFS and SORT had good discrimination with *c*-statistics of 0.71 and 0.70, respectively, but the SORT showed poor calibration. Part of the weakness of the SORT is that many of the variables remain somewhat subjective, such as surgical severity and urgency. Nonetheless, it is an efficient tool to predict mortality, though it requires more validation for use with orthopedic procedures.

Edinburgh Wrist Calculator

Distal radius fractures are another injury that has gained the attention of an automated risk calculator. Certain distal radius fracture patterns may predispose to loss of alignment and malunion. Predicting loss of reduction may help patients and clinicians decide between operative and nonoperative management. The Edinburgh wrist calculator (EWC) was derived based on the evaluation and outcomes of 4000 distal radius fractures [36]. Luokkala et al. evaluated the EWC against expert opinion and majority rule for 71 distal radius fractures initially treated with closed reduction and casting [6]. The authors evaluated patients at 6 weeks post reduction to determine if reduction was maintained or lost. They found that the EWC had the highest accuracy (0.77%), highest sensitivity (0.95%), and negative predictive value (0.97%). Walenkamp et al. also evaluated the EWC for loss of distal radius fracture

alignment [37]. In contrast to the previous study that evaluated fractures at 6 weeks, Walenkamp et al. evaluated loss of reduction within 2 weeks after casting. The authors determined the sensitivity, specificity, and AUC of the EWC for predicting loss of reduction compared with established prediction criteria [36]. Out of 515 patients, the EWC had a poor AUC (0.47), low sensitivity (1.6%), high specificity (95%), low PPV (33%), and low NPV (38%). These results conflict with those of the prior study, suggesting the need for further validation of the EWC before its routine use in clinical practice. It may prove to be an easy, valuable tool for upper extremity and trauma surgeons treating distal radius fractures, but requires additional experimentation.

Tumor

The only study available in the literature on tumor reconstruction is a study published by Slump et al. that evaluates the NSQIP-SRC for flap reconstruction after soft tissue sarcoma resection [38]. The authors evaluated 265 patients who underwent either pedicle or free flap reconstruction after soft tissue sarcoma resection on the trunk or extremities. The authors found that the NSQIP-SRC underestimated the risk of any serious complication, with statistically significant differences between predicted and observed. The AUC value for any complication was 0.62, and the Brier's score was 0.24, suggesting the model is not a good predictor of complications after flap reconstruction for soft tissue sarcoma resection. The authors highlight that there are several factors relevant to sarcoma resection such as site and size of tumor, adjuvant therapies, and multiple procedures that are not accounted for in the NSQIP-SRC which may lead to poor discrimination. The authors suggest the NSQIP-SRC should not be used for this type of procedure and a more disease specific calculator may provide better results [38].

Discussion

Several automated risk calculators have been developed in the field of orthopedics. The most abundant research has been conducted in the total joint arthroplasty and spine literature. With the development of automated risk calculators, it is important for surgeons to understand how and when to employ them.

Despite advances in developing and evaluating automated risk calculators, they must be used judiciously. Risk calculators should serve as a tool to help clinical decision-making, promote individualized medicine, and aid in the shared decision-making process [19•]. Many of the cited studies in this review have highlighted the shortcomings of automated risk calculators, namely that the populations used to develop the risk calculators may be vastly different from the

population on which they are used [5]. When using risk calculators, it is important for the physician to understand how the calculator was developed, and discrepancies between the population of the physician's practice and that used to develop the risk calculator. Literature suggests that in general, a physician's understanding of statistical models is lacking [39, 40]. Wegworth et al. conducted a randomized study of primary care physicians to assess their understanding of cancer screening statistics by randomizing primary care physicians into two groups presented with different scenarios. One group was presented with a hypothetical screening test that increased early detection and improved 5-year survival rates in cancer patients, and the other group was presented with a hypothetical screening test that reduced mortality and increased incidence. The latter scenario, termed relevant evidence, clearly shows a benefit to cancer patients by reduced mortality, whereas the former, termed irrelevant evidence, simply shows that improved 5 year survival may be a function of earlier detection. When asked if the physicians would recommend the test, 69% presented with irrelevant evidence stated their recommendation, but only 23% presented with relevant evidence recommended the screening test [40•]. Additionally, physicians were not able to distinguish between which scenario was relevant versus irrelevant. While cancer screening tests differ from automated risk calculators, this highlights physicians' lack of understanding of basic statistical modeling. Physicians should understand the tools they are using in order to affect the most benefit and avoid giving incorrect treatment recommendations. Similar studies should be performed regarding physicians use and understanding of automated risk calculators. Many of the studies we have cited report poor discrimination and calibration of the investigated risk calculators. It is almost universal that a downside of the studied risk calculator is that it was developed with a patient population different from the population on which it was evaluated.

Another important aspect of risk calculator use is accurate communication of information. If used as a tool for decision-making and informed consent, it is important that the patient be aware of the risk calculator, the results of the calculator, and if it has affected any decision-making to undergo a surgical procedure. Even more important is how risk information is communicated to the patient in order to ensure their understanding. Brian Zikmund-Fisher published an article on the taxonomy of risk communication [39•]. Zikmund-Fisher emphasizes that risk communication methods should vary based on why a patient desires risk information. For example, patients who only seek to know if they are "high" or "low" risk for a procedure may not benefit from specific risk numbers or percentages, and providing that type of information may in fact be counterproductive. Contrarily, a patient who seeks to reduce risk by a specific margin may benefit from a more complex, quantitative presentation of risk. Physicians also often assume that providing detailed percentages of risk means

they are fully informing their patients, when in fact they may not be meeting the needs of the patient by communicating the risk in a way that most benefits the patient [41]. Different formats for presenting risk effects each patient differently. This concept of risk format was assessed in a prospective study by Bodemer et al. on 1234 patients presented with hypothetical scenarios about a treatment that either increased or decreased baseline risk [42•]. One scenario presented baseline risk in frequency format, and the other presented it in percentage format. The results demonstrated that patients had a much better understanding of relative risk reduction or increase when baseline risk was presented in frequency format. Therefore, the dialogue between clinician and patient about the patient's goals regarding risk information and accurately communicating that information may be as important as the actual information generated by the risk calculator. If an orthopedic surgeon uses an automated risk calculator, he or she should strongly consider divulging the results to the patient and also be prepared to accurately communicate that information.

An understanding of risk calculator development and communication of risk is important, but the actual accuracy of the risk calculator must be sound. As the literature discussed in this review demonstrates, many of the risk calculators used in orthopedic surgery provide only mediocre results (Tables 1, 2, and 3). Several shortcomings still exist with most risk calculators currently used. Despite this, a number of the risk calculators discussed show promise as adjuncts to the informed consent and decision-making process. Many of the calculators with better discrimination and calibration were developed for specific procedures or outcomes, further emphasizing the importance of the patient population used to develop a risk calculator. Broader databases with more generalized patient populations, like the NSQP-SRC, seldom show promising results for more specific procedures. Choosing to use a risk calculator for clinical decision-making mandates a calculator with good discrimination and calibration. Aside from the aforementioned notion of communication, the calculator must be sound. At this time, none of the calculators demonstrate outstanding results, but several are promising and could serve as valuable tools for patient evaluation and surgical decision-making.

Conclusions

Several publicly available automated risk calculators exist in the field of orthopedics, with the greatest number present in spine and total joints. These calculators provide a useful tool to guide surgeons and patients during the informed consent and shared decision-making process. While a number of risk calculators show promise regarding discrimination and calibration, none perform well enough to be recommended as a must-use tool for surgical decision-making. At this time, we

recommend these calculators be used as adjuncts to surgeon and patient judgement. An important aspect of their use resides in a thorough understanding by the surgeon of how they were developed, and tactful communication of the results to the patient.

Compliance with Ethical Standards

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance

1. Godolphin W. The role of risk communication in shared decision making. BMJ. 2003;327(7417):692–3.
2. Leclercq WKG, Keulers BJ, Scheltinga MRM, Spauwen PHM, van der Wilt G-J. A review of surgical informed consent: past, present, and future. A quest to help patients make better decisions. World J Surg. 2010;34(7):1406–15. https://doi.org/10.1007/s00268-010-0542-0.
3. Wong J, Lam DP, Abrishami A, Chan MTV, Chung F. Short-term preoperative smoking cessation and postoperative complications: a systematic review and meta-analysis. Can J Anaesth J Can Anesth. 2012;59(3):268–79. https://doi.org/10.1007/s12630-011-9652-x.
4. Hackbarth G, Reischauer R, Mutti A. Collective accountability for medical care — toward bundled Medicare payments. N Engl J Med. 2008;359(1):3–5. https://doi.org/10.1056/NEJMp0803749.
5. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aide and informed consent tool for patients and surgeons. J Am Coll Surg. 2013;217(5):833–842.e3. https://doi.org/10.1016/j.jamcollsurg.2013.07.385.
6. Luokkala T, Flinkkilä T, Paloneva J, Karjalainen TV. Comparison of expert opinion, majority rule, and a clinical prediction rule to estimate distal radius malalignment. J Orthop Trauma. 2018;32(3):e97–e101. https://doi.org/10.1097/BOT.0000000000001022.
7. Protopapa KL, Simpson JC, Smith NCE, Moonesinghe SR. Development and validation of the Surgical Outcome Risk Tool (SORT). Br J Surg. 2014;101(13):1774–83. https://doi.org/10.1002/bjs.9638.
8. Pugely AJ, Martin CT, Gao Y, Klocke NF, Callaghan JJ, Marsh JL. A risk calculator for short-term morbidity and mortality after hip fracture surgery. J Orthop Trauma. 2014;28(2):63–9. https://doi.org/10.1097/BOT.0b013e3182a22744.
9. Rosinsky PJ, Go CC, Shapira J, Maldonado DR, Lall AC, Domb BG. Validation of a risk calculator for conversion of hip arthroscopy to total hip arthroplasty in a consecutive series of 1400 patients. J Arthroplast. 2019;34(8):1700–6. https://doi.org/10.1016/j.arth.2019.04.013.
10. Harris AHS, Kuo AC, Bozic KJ, Lau E, Bowe T, Gupta S, et al. American Joint Replacement Registry risk calculator does not predict 90-day mortality in veterans undergoing total joint replacement. Clin Orthop. 2018;476(9):1869–75. https://doi.org/10.1097/CORR.0000000000000377.
11. Kheir MM, Tan TL, George J, Higuera CA, Maltenfort MG, Parvizi J. Development and evaluation of a prognostic calculator for the surgical treatment of periprosthetic joint infection. J Arthroplast. 2018;33(9):2986–2992.e1. https://doi.org/10.1016/j.arth.2018.04.034.
12. Paxton EW, Inacio MCS, Khatod M, Yue E, Funahashi T, Barber T. Risk calculators predict failures of knee and hip arthroplasties: findings from a large health maintenance organization. Clin Orthop. 2015;473(12):3965–73. https://doi.org/10.1007/s11999-015-4506-4.
13. Starr J, Rozet I, Ben-Ari A. A risk calculator using preoperative opioids for prediction of total knee revision arthroplasty. Clin J Pain. 2018;34(4):328–31. https://doi.org/10.1097/AJP.0000000000000544.
14. Kasparek MF, Boettner F, Rienmueller A, Weber M, Funovics PT, Krepler P, et al. Predicting medical complications in spine surgery: evaluation of a novel online risk calculator. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc. 2018;27(10):2449–56. https://doi.org/10.1007/s00586-018-5707-9.
15. Lee MJ, Cizik AM, Hamilton D, Chapman JR. Predicting medical complications after spine surgery: a validated model using a prospective surgical registry. Spine J Off J North Am Spine Soc. 2014;14(2):291–9. https://doi.org/10.1016/j.spinee.2013.10.043.
16. Choi D, Pavlou M, Omar R, et al. A novel risk calculator to predict outcome after surgery for symptomatic spinal metastases; use of a large prospective patient database to personalise surgical management. Eur J Cancer Oxf Engl 1990. 2019;107:28–36. https://doi.org/10.1016/j.ejca.2018.11.011.
17. Ratliff JK, Balise R, Veeravagu A, Cole TS, Cheng I, Olshen RA, et al. Predicting occurrence of spine surgery complications using "big data" modeling of an administrative claims database. J Bone Joint Surg Am. 2016;98(10):824–34. https://doi.org/10.2106/JBJS.15.00301.
18. Kanis JA, Harvey NC, Johansson H, Odén A, Leslie WD, McCloskey EV. FRAX and fracture prediction without bone mineral density. Climacteric. 2015;18(sup2):2–9. https://doi.org/10.3109/13697137.2015.1092342.
19.• Mansmann U, Rieger A, Strahwald B, Crispin A. Risk calculators-methods, development, implementation, and validation. Int J Colorectal Dis. 2016;31(6):1111–6. https://doi.org/10.1007/s00384-016-2589-3
20. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA. 2017;318(14):1377–84. https://doi.org/10.1001/jama.2017.12126.
21. Dankers FJWM, Traverso A, Wee L, van Kuijk SMJ. Prediction modeling methodology. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of clinical data science. Cham (CH): Springer; 2019. http://www.ncbi.nlm.nih.gov/books/NBK543534/. Accessed 27 Sept 2019.
22. Sebastian A, Goyal A, Alvi MA, Wahood W, Elminawy M, Habermann EB, et al. Assessing the performance of National Surgical Quality Improvement Program Surgical Risk Calculator in elective spine surgery: insights from patients undergoing single-level posterior lumbar fusion. World Neurosurg. 2019;126:e323–9. https://doi.org/10.1016/j.wneu.2019.02.049.
23. Wang X, Hu Y, Zhao B, Su Y. Predictive validity of the ACS-NSQIP surgical risk calculator in geriatric patients undergoing lumbar surgery. Medicine (Baltimore). 2017;96(43):e8416. https://doi.org/10.1097/MD.0000000000008416.

24. Veeravagu A, Li A, Swinney C, Tian L, Moraff A, Azad TD, et al. Predicting complication risk in spine surgery: a prospective analysis of a novel risk assessment tool. J Neurosurg Spine. 2017;27(1):81–91. https://doi.org/10.3171/2016.12.SPINE16969.

25. Tomita K, Kawahara N, Kobayashi T, Yoshida A, Murakami H, Akamaru T. Surgical strategy for spinal metastases. Spine. 2001;26(3):298–306. https://doi.org/10.1097/00007632-200102010-00016.

26. Tokuhashi Y, Matsuzaki H, Oda H, Oshima M, Ryu J. A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis. Spine. 2005;30(19):2186–91. https://doi.org/10.1097/01.brs.0000180401.06919.a5.

27. Romine LB, May RG, Taylor HD, Chimento GF. Accuracy and clinical utility of a peri-operative risk calculator for total knee arthroplasty. J Arthroplast. 2013;28(3):445–8. https://doi.org/10.1016/j.arth.2012.08.014.

28. Edelstein AI, Kwasny MJ, Suleiman LI, Khakhkhar RH, Moore MA, Beal MD, et al. Can the American College of Surgeons risk calculator predict 30-day complications after knee and hip arthroplasty? J Arthroplast. 2015;30(9 Suppl):5–10. https://doi.org/10.1016/j.arth.2015.01.057.

29. Wingert NC, Gotoff J, Parrilla E, Gotoff R, Hou L, Ghanem E. The ACS NSQIP risk calculator is a fair predictor of acute periprosthetic joint infection. Clin Orthop. 2016;474(7):1643–8. https://doi.org/10.1007/s11999-016-4717-3.

30. Goltz DE, Baumgartner BT, Politzer CS, DiLallo M, Bolognesi MP, Seyler TM. The American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator has a role in predicting discharge to post-acute care in total joint arthroplasty. J Arthroplast. 2018;33(1):25–9. https://doi.org/10.1016/j.arth.2017.08.008.

31. Tan TL, Maltenfort MG, Chen AF, Shahi AS, Higuera CA, Siqueira M, et al. Development and evaluation of a preoperative risk calculator for periprosthetic joint infection following total joint arthroplasty. J Bone Joint Surg Am. 2018;100(9):777–85. https://doi.org/10.2106/JBJS.16.01435.

32. Bedard NA, DeMik DE, Dowdle SB, Owens JM, Liu SS, Callaghan JJ. Preoperative opioid use and its association with early revision of total knee arthroplasty. J Arthroplast. 2018;33(11):3520–3. https://doi.org/10.1016/j.arth.2018.06.005.

33. Ben-Ari A, Chansky H, Rozet I. Preoperative opioid use is associated with early revision after total knee arthroplasty: a study of male

patients treated in the Veterans Affairs system. J Bone Joint Surg Am. 2017;99(1):1–9. https://doi.org/10.2106/JBJS.16.00167.

34. Wang X, Zhao BJ, Su Y. Can we predict postoperative complications in elderly Chinese patients with hip fractures using the surgical risk calculator? Clin Interv Aging. 2017;12:1515–20. https://doi.org/10.2147/CIA.S142748.

35. Marufu TC, White SM, Griffiths R, Moonesinghe SR, Moppett IK. Prediction of 30-day mortality after hip fracture surgery by the Nottingham hip fracture score and the surgical outcome risk tool. Anaesthesia. 2016;71(5):515–21. https://doi.org/10.1111/anae.13418.

36. Mackenney PJ, McQueen MM, Elton R. Prediction of instability in distal radial fractures. J Bone Joint Surg Am. 2006;88(9):1944–51. https://doi.org/10.2106/JBJS.D.02520.

37. Walenkamp MMJ, Mulders MAM, van Hilst J, Goslings JC, Schep NWL. Prediction of distal radius fracture redisplacement: a validation study. J Orthop Trauma. 2018;32(3):e92–6. https://doi.org/10.1097/BOT.0000000000001105.

38. Slump J, Ferguson PC, Wunder JS, Griffin A, Hoekstra HJ, Bagher S, et al. Can the ACS-NSQIP surgical risk calculator predict postoperative complications in patients undergoing flap reconstruction following soft tissue sarcoma resection? J Surg Oncol. 2016;114(5):570–5. https://doi.org/10.1002/jso.24357.

39.• Zikmund-Fisher BJ. The right tool is what they need, not what we have: a taxonomy of appropriate levels of precision in patient risk communication. Med Care Res Rev MCRR. 2013;70(1 Suppl):37S–49S. https://doi.org/10.1177/1077558712458541

40.• Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. Ann Intern Med. 2012;156(5):340–9. https://doi.org/10.7326/0003-4819-156-5-201203060-00005

41. Heath C, Heath D. Made to stick: why some ideas survive and others die. 1st ed. New York: Random House; 2007.

42.• Bodemer N, Meder B, Gigerenzer G. Communicating relative risk changes with baseline risk: presentation format and numeracy matter. Med Decis Mak Int J Soc Med Decis Mak. 2014;34(5):615–26. https://doi.org/10.1177/0272989X14526305