Severity Weighting of Postoperative Adverse Events in Orthopedic Surgery


Authors:
Daniel D. Bohl, MD, MPH Junyoung Ahn, BS Adam M. Lukasiewicz, MSc Andre M. Samuel, BBA Matthew L. Webb, AB Bryce A. Basques, MD Nicholas S. Golinvaux, MD Kern Singh, MD Jonathan N. Grauer, MD

Author Affiliation | Disclosures

Authors’ Disclosure Statement: The authors report no actual or potential conflict of interest in relation to this article.

Take-Home Points

- Studies of AEs after orthopedic surgery commonly use composite AE outcomes.
- These types of outcomes treat AEs with different clinical significance similarly.
- This study created a single severity-weighted outcome that can be used to characterize the overall severity of a given patient’s postoperative course.
- Future studies may benefit from using this new severity-weighted outcome score.

Recently there has been an increase in the use of national databases for orthopedic surgery research. Studies commonly compare rates of postoperative adverse events (AEs) across different demographic, comorbidity, and procedural characteristics. Their conclusions often highlight different modifiable and/or nonmodifiable risk factors associated with the occurrence of postoperative events.

The several dozen AEs that have been investigated range from very severe (eg, death, myocardial infarction, coma) to less severe (eg, urinary tract infection [UTI], anemia requiring blood transfusion). A common approach for these studies is to consider many AEs together in the same analysis, asking a question such as, “What are risk factors for the occurrence of ‘adverse events’ after spine surgery?” Such studies test for associations with the occurrence of “any adverse event,” the occurrence of any “serious adverse event,” or similar composite outcomes. How common this type of study has become is indicated by the fact that in 2013 and 2014, at least 12 such studies were published in Clinical Orthopaedics and Related Research and the Journal of Bone and Joint Surgery, and many more in other orthopedic journals. However, there is a problem in using this type of composite outcome to perform such analyses: AEs with highly varying degrees of severity have identical impacts on the outcome variable, changing it from negative (“no adverse event”) to positive (“at least one adverse event”). As a result, the system may treat a very severe AE such as death and a very minor AE such as UTI similarly. Even in studies that use the slightly more specific composite outcome of “serious adverse events,” death and a nonlethal thromboembolic event would be treated similarly. Failure to differentiate these AEs in terms of their clinical significance detracts from the clinical applicability of conclusions drawn from studies using these types of composite AE outcomes.

In one of many examples that can be considered, a retrospective cohort study compared general and spinal
anesthesia used in total knee arthroplasty.\textsuperscript{10} The rate of any AEs was higher with general anesthesia than with spinal anesthesia (12.34\% vs 10.72\%; \( P = .003 \)). However, the only 2 specific AEs that had statistically significant differences were anemia requiring blood transfusion (6.07\% vs 5.02\%; \( P = .009 \)) and superficial surgical-site infection (SSI; 0.92\% vs 0.68\%; \( P < .001 \)). These 2 AEs are of relatively low severity; nevertheless, because these AEs are common, their differences constituted the majority of the difference in the rate of any AEs. In contrast, differences in the more severe AEs, such as death (0.11\% vs 0.22\%; \( P > .05 \)), septic shock (0.14\% vs 0.12\%; \( P > .05 \)), and myocardial infarction (0.20\% vs 0.20\%; \( P > .05 \)), were small and not statistically significant. Had more weight been given to these more severe events, the outcome of the study likely would have been “no difference.”

To address this shortcoming in orthopedic research methodology, we created a severity-weighted outcome score that can be used to determine the overall “severity” of any given patient’s postoperative course. We also tested this novel outcome score for correlation with procedure type and patient characteristics using orthopedic patients from the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP). Our intention is for database investigators to be able to use this outcome score in place of the composite outcomes that are dominating this type of research.

**Methods**

**Generation of Severity Weights**

Our method is described generally as utility weighting, assigning value weights reflective of overall impact to differing outcome states.\textsuperscript{24} Parallel methods have been used to generate the disability weights used to determine disability-adjusted life years for the Global Burden of Disease project\textsuperscript{25} and many other areas of health, economic, and policy research.

All orthopedic faculty members at 2 geographically disparate, large US academic institutions were invited to participate in a severity-weighting exercise. Each surgeon who agreed to participate performed the exercise independently.

Each participant was given a stack of 23 index cards, each listing the name and description of an AE monitored by
ACS-NSQIP (Table 1). In addition, in the upper right corner of each card was a box in which the participant could write a number. Each stack of cards was provided in a distinct randomized order. Written instructions for participants were exactly as follows:

- **STEP 1:** Please reorder the AE cards by your perception of “severity” for a patient experiencing that event after an orthopedic procedure.
- **STEP 2:** Once your cards are in order, please determine how many postoperative occurrences of each event you would “trade” for 1 patient experiencing postoperative death. Place this number of occurrences in the box in the upper right corner of each card.
- **NOTES:** As you consider each AE:
  - Please consider an “average” occurrence of that AE, but note that in no case does the AE result in perioperative death.
  - Please consider only the “severity” for the patient. (Do not consider the extent to which the event may be related to surgical error.)
  - Please consider that the numbers you assign are relative to each other. Hence, if you would trade 20 of “event A” for 1 death, and if you would trade 40 of “event B” for 1 death, the implication is that you would trade 20 of “event A” for 40 of “event B.”
  - You may may adjust the order of your cards at any point.

Participants’ responses were recorded. For each number provided by each participant, the inverse (reciprocal) was taken and multiplied by 100%. This new number was taken to be the percentage severity of death that the given participant considered the given AE to embody. For example, as a hypothetical on one end of the spectrum, if a participant reported 1 (he/she would trade 1 AE X for 1 death), then the severity would be \(\frac{1}{1} \times 100\% = 100\%\) of death, a very severe AE. Conversely, if a participant reported a very large number like 100,000 (he/she would trade 100,000 AEs X for 1 death), then the severity would be \(\frac{1}{100,000} \times 100\% = 0.001\%\) of death, a very minor AE. More commonly, a participant will report a number like 25, which would translate to 4% of death (1/25 × 100% = 4%). For each AE, weights were then averaged across participants to derive a mean severity weight to be used to generate a novel composite outcome score.

**Definition of Novel Composite Outcome Score**

The novel composite outcome score would be expressed as a percentage to be interpreted as percentage severity of death, which we termed *severity-weighted outcome relative to death* (SWORD). For each patient, SWORD was defined as no AE (0%) or postoperative death (100%), with other AEs assigned mean severity weights based on faculty members’ survey responses. A patient with multiple AEs would be assigned the weight for the more severe AE. This method was chosen over summing the AE weights because in many cases the AEs were thought to overlap; hence, summing would be inappropriate. For example, generally a deep SSI would result in a return to the operating room, and one would not want to double-count this AE. Similarly, it would not make sense for a patient who died of a complication to have a SWORD of >100%, which would be the summing result.

**Application to ACS-NSQIP Patients**

ACS-NSQIP is a surgical registry that prospectively identifies patients undergoing major surgery at any of >500 institutions nationwide. Patients are characterized at baseline and are followed for AEs over the first 30 postoperative days.
Patients undergoing any of 8 common orthopedic procedures were identified in the 2012 ACS-NSQIP database using International Classification of Diseases, Ninth Revision (ICD-9) codes and Current Procedural Terminology (CPT) codes (Table 2). Any patient with missing data was excluded from this population before analysis.

First, mean SWORD was calculated and reported for patients undergoing each of the 8 procedures. Analysis of variance (ANOVA) was used to test for associations of mean SWORD with type of procedure both before and after multivariate adjustment for demographics (sex; age in years, <40, 40-49, 50-59, 60-69, 70-79, 80-89, ≥90) and comorbidities (diabetes, hypertension, chronic obstructive pulmonary disease, exertional dyspnea, end-stage renal disease, congestive heart failure).

Second, patients undergoing the procedure with the highest mean SWORD (hip fracture surgery) were examined in depth. Among only these patients, multivariate ANOVA was used to test for associations of mean SWORD with the same demographics and comorbidities.

All statistical tests were 2-tailed. Significance was set at α = 0.05 (P < .05).

All 23 institution A faculty members (100%) and 24 (89%) of the 27 institution B faculty members completed the exercise.

Total number of participants was 47, and the overall response rate was 94%. Participant characteristics are listed in Table 3.

In the ACS-NSQIP database, 85,109 patients were identified on the basis of the initial inclusion criteria.
After patients with missing data were excluded, 85,031 remained for analysis. Patient characteristics are listed in Table 4.

Results

Figure 1 shows mean severity weights and standard errors generated from faculty responses. Mean (standard error) severity weight for UTI was 0.23% (0.08%); blood transfusion, 0.28% (0.09%); pneumonia, 0.55% (0.15%); hospital readmission, 0.59% (0.23%); wound dehiscence, 0.64% (0.17%); deep vein thrombosis, 0.64% (0.19%); superficial SSI, 0.68% (0.23%); return to operating room, 0.91% (0.29%); progressive renal insufficiency, 0.93% (0.27%); graft/prosthesis/flap failure, 1.20% (0.34%); unplanned intubation, 1.38% (0.53%); deep SSI, 1.45% (0.38%); failure to wean from ventilator, 1.45% (0.48%); organ/space SSI, 1.76% (0.46%); sepsis without shock, 1.77% (0.42%); peripheral nerve injury, 1.83% (0.47%); pulmonary embolism, 2.99% (0.76%); acute renal failure, 3.95% (0.85%); myocardial infarction, 4.16% (0.98%); septic shock, 7.17% (1.36%); stroke, 8.73% (1.74%); cardiac arrest requiring cardiopulmonary resuscitation, 9.97% (2.46%); and coma, 15.14% (3.04%).

Among ACS-NSQIP patients, mean SWORD ranged from 0.2% (elective anterior cervical decompression and fusion) to 6.0% (hip fracture surgery) (Figure 2).
Mean SWORD was associated with procedure type both before ($P < .001$) and after ($P < .001$) controlling for demographic and comorbidity differences between populations. Among ACS-NSQIP patients having hip fracture surgery, mean SWORD was independently associated with older age, male sex, and 4 of 6 tested comorbidities ($Ps < .05$) (Figure 3).

**Discussion**

The use of national databases in studies has become increasingly common in orthopedic surgery.\textsuperscript{1-4}

However, many of these studies use composite outcomes such as “any adverse events” and “serious adverse events” to generate primary results.\textsuperscript{5-23} Such methods implicitly consider the severity of markedly different AEs (death, UTI) to be the same. Our study provides orthopedics researchers with a tool that can be used to overcome this methodologic deficit.

The academic orthopedic surgeons who participated in our severity-weighting exercise thought the various AEs have markedly different severities. The least severe AE (UTI) was considered 0.23% as severe as postoperative death, with other events spanning the range up to 15.14% as severe as death. This wide range of severities demonstrates the problem with composite outcomes that implicitly consider all AEs similarly severe. Use of these...
markedly disparate weights in the development of SWORD enables this outcome to be more clinically applicable than outcomes such as “any adverse events.”

SWORD was highly associated with procedure type both before and after adjustment for demographics and comorbidities. Among patients undergoing the highest SWORD procedure (hip fracture surgery), SWORD was also associated with age, sex, and 4 of 6 tested comorbidities. Together, our findings show how SWORD is intended to be used in studies: to identify demographic, comorbidity, and procedural risk factors for an adverse postoperative course. We propose that researchers use our weighted outcome as their primary outcome—it is more meaningful than the simpler composite outcomes commonly used.

Outside orthopedic surgery, a small series of studies has addressed severity weighting of postoperative AEs. However, their approach was very different, as they were not designed to generate weights that could be transferred to future studies; rather, they simply compared severities of postoperative courses for patients within each individual study. In each study, a review of each original patient record was required, as the severity of each patient’s postoperative course was characterized according to the degree of any postoperative intervention—from no intervention to minor interventions such as placement of an intravenous catheter and major interventions such as endoscopic, radiologic, and surgical procedures. Only after the degree of intervention was defined could an outcome score be assigned to a given patient. However, databases do not depict the degree of intervention with nearly enough detail for this type of approach; they typically identify only occurrence or nonoccurrence of each event. Our work, which arose independently from this body of literature, enables an entirely different type of analysis. SWORD, which is not based on degree of intervention but on perceived severity of an “average” event, enables direct application of severity weights to large databases that store simple information on occurrence and nonoccurrence of specific AEs.

This study had several limitations. Most significantly, the generated severity weights were based on the surgeons’ subjective perceptions of severity, not on definitive assessments of the impacts of specific AEs on actual patients. We did not query the specialists who treat the complications or who present data on the costs and disabilities that may arise from these AEs. In addition, to develop our severity weighting scale, we queried faculty at only 2 institutions. A survey of surgeons throughout the United States would be more representative and would minimize selection bias. This is a potential research area. Another limitation is that scoring was subjective, based on surgeons’ perceptions of patients—in contrast to the Global Burden of Disease project, in which severity was based more objectively on epidemiologic data from >150 countries.

Orthopedic database research itself has often-noted limitations, including inability to sufficiently control for confounders, potential inaccuracies in data coding, limited follow-up, and lack of orthopedic-specific outcomes. However, this research also has much to offer, has increased tremendously over the past several years, and is expected to continue to expand. Many of the limitations of database studies cannot be entirely reversed. In providing a system for weighting postoperative AEs, our study fills a methodologic void. Future studies in orthopedics may benefit from using the severity-weighted outcome score presented here. Other fields with growth in database research may consider using similar methods to create severity-weighting systems of their own.

Am J Orthop. 2017;46(4):E235-E243. Copyright Frontline Medical Communications Inc. 2017. All rights reserved.
Key Info

Figures/Tables

References

References


21. Lovecchio F, Beal M, Kwasn M, Manning D. Do patients with insulin-dependent and noninsulin-


Multimedia

Product Guide

Product Guide

- STRATAFIX™ Symmetric PDS™ Plus Knotless Tissue Control Device
- STRATAFIX™ Spiral Knotless Tissue Control Device
- BioComposite SwiveLock Anchor
- BioComposite SwiveLock C, with White/Black TigerTape™ Loop

Citation


Daniel D. Bohl, MD, MPH