Reliability of Surgical Risk Calculator Performance Assessments in Single-Institution Data

Chelsea P. Fischer, MD, MS[1,2], **Mark E. Cohen, PhD**[2], **Ryan P. Merkow, MD, MS, FACS**[3]

[1]Department of Surgery, Loyola University Medical Center, Maywood, IL

[2]Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL

[3]Department of Surgery, Northwestern University Medical Center, Chicago, IL

Surgical risk calculators (SRCs) are tools commonly used by clinicians - initially developed out of the need for accurate patient risk prediction before surgery, both to guide clinician decision-making and aid in patient counseling. SRCs allow clinicians to input patient factors, such as demographics and comorbidities, and generate outcome predications based on the patient's individualized risk profile. The first large-scale SRC, developed by the American College of Surgeons (ACS) from the National Surgical Quality Improvement Program (NSQIP) data registry, includes risk profiles for over 1,500 different procedures.[1]

SRC use in clinical and research settings has expanded over time, and its accuracy evaluated for multiple surgical procedures. Vos et. al. examined the performance of the ACS NSQIP SRC in patients undergoing total gastrectomy for gastric cancer. The authors examined the predictive accuracy of the SRC for 12 adverse outcomes by validating the estimates against an institutionally-collected database. Performance of the SRC was shown to be inconsistent in patients undergoing total gastrectomy for gastric cancer, with underprediction of complications overall. The SRC was noted to perform well for death, renal failure, cardiac complication, and discharge to rehab/nursing home.

It is important to determine the conditions under which the SRC has good predictive accuracy and where it might not be insufficiently reliably to guide providers and patients. However, there are important limitations in this study that need to be acknowledged when translating these results to real-world practice. This study was performed in a single institution, and included a small number of patients undergoing a single, complex procedure. There is inherent unreliability of event rate estimates from small sample sizes, which makes it difficult to validate the SRC.[2] Only the authors' outcomes of any complication, SSI, and length of stay (LOS) outcomes have a sufficient number of cases with events (>100) to provide an adequate test.

The SRC is designed to make predictions for patients treated at the average NSQIP hospital and performance will decline when applied to hospitals which diverge substantially from

Corresponding Author: Ryan P. Merkow, MD, MS, FACS, Surgical Outcomes & Quality Improvement Center, 633 North Saint Clair Street, 20th Floor, Chicago, IL 60611, ryan.merkow@nm.org.

the average.[3] This may be particularly relevant in this study, as all cases were performed at a highly specialized hospital treating only cancer patients. Therefore, it may inappropriate to validate an SRC in a setting that is certainly not representative of almost all hospitals in the US. Finally, limiting an external validation of the SRC to a specific procedure reduces observed discrimination, as patients undergoing the same procedure are relatively homogeneous with similar risk profiles.[3,4]

These limitations are not to imply external validation of the SRC are not a worthwhile endeavor. Investigation of procedure-specific variables may improve the performance of calculators and additionally add clinical relevance for clinicians. Ideally, external validation should be based on large, multi-institution data sets as rate estimates from single institutions with low volume tend to be unstable. Importantly, SRCs are not designed to generate perfect predictions, nor should they be a replacement for institution experience, thoughtful patient selection, and patient-centered decision making.

Surgical risk calculators are valuable tools to incorporate into clinical practice and guide patient-centered decision making. Practicing clinicians who utilize such tools should understand both their limitations and the scenarios for clinically meaningful use.

References:

1. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. Journal of the American College of Surgeons. 2013;217(5):833–842. e833. [PubMed: 24055383]

2. Liu Y, Cohen ME, Hall BL, Ko CY, Bilimoria KY. Evaluation and enhancement of calibration in the American College of Surgeons NSQIP Surgical Risk Calculator. Journal of the American College of Surgeons. 2016;223(2):231–239. [PubMed: 27212006]

3. Cohen ME, Liu Y, Ko CY, Hall BL. An examination of American College of Surgeons NSQIP surgical risk calculator accuracy. Journal of the American College of Surgeons. 2017;224(5):787–795. e781. [PubMed: 28389191]

4. Merkow RP, Hall BL, Cohen ME, et al. Relevance of the c-statistic when evaluating risk-adjustment models in surgery. Journal of the American College of Surgeons. 2012;214(5):822–830. [PubMed: 22440055]