Evaluating Surgeon Scorecards

Shaina Sekhri¹; James Mossner¹; Rahul Iyengar¹; Andrew Mullard^{1,2}; Michael J. Englesbe^{1,2}; Joseph Papin IV¹

Introduction

There is an increasing demand for transparency in the field of health care, particularly as it pertains to the quality of care that hospitals and physicians provide patients. Transparency is defined as “making available information about the cost and quality of healthcare services, so that patients can become informed consumers.”¹ Transparency increases trust and improves dynamics between patients and physicians by providing complete, objective, and high-quality data to all involved stakeholders.^{2,3} Increased transparency in health care will improve the health and wellness of patient populations.

Transparency can be improved through surgeon scorecards, which provide a framework to evaluate performance and make information available to patients.^{1,4,5} The technical skill of practicing surgeons can vary widely, and this greater skill is associated with fewer postoperative complications.⁶ However, fewer than 1% of surgical outcomes are being measured, leaving surgeons and hospitals “unaware of

how their patients fare collectively over time.”^{7,8} High-quality surgeon scorecards can help the medical community hold surgeons accountable and empower patients to make informed decisions about their care.

Current Surgeon Rating Efforts

One of the largest and most well-known organizations to rank physicians is ProPublica.⁹ This independent, nonprofit newsroom published a *Surgeon Scorecard* in July 2015 with adjusted complication rates for nearly 17 000 surgeons in 8 inpatient procedures. Their goal is to provide patients and the health care community with “reliable and actionable data points, at both the level of the surgeon and the hospital, in the form of a publicly available online searchable database.”⁸

ProPublica employs a rigorous approach in creating its scorecard. The *Surgeon Scorecard* utilizes administrative billing data from Medicare, which is reliable for certain reporting purposes.^{8,10}

Author Affiliations:

¹University of Michigan Medical School, Ann Arbor, MI 48109, USA;
²Department of Surgery, University of Michigan Health System, Ann Arbor, MI 48109, USA

Corresponding Author:

Shaina Sekhri (ssekhhri@med.umich.edu)

While this data restriction limits the case volumes for surgeons who see fewer Medicare patients, ProPublica verified, with state-level clinical data, that “low-volume” Medicare surgeons had lower overall case volumes; lower case volumes are correlated with higher complication rates for certain procedures.^{8,11}

The *Surgeon Scorecard* also employs a strategy used successfully by Dimick et al to identify a uniform patient cohort.¹² Analysis is restricted to 8 common, elective procedures generally performed on healthy patients. Complex cases including revision surgeries, emergency admissions, transfers from other care facilities, or uncommon principal diagnoses indicating a complication were excluded.⁸ To control for comorbidities, a Health Score is created using the Van Walraven technique to create an index of Elixhauser comorbidities for each patient.^{13,14} Adverse outcomes are identified as death at the index admission or 30-day readmission with a relevant principal diagnosis. From these values, an Adjusted Complication Rate is created for each surgeon and reported in the *Surgeon Scorecard* on ProPublica’s website.^{8,9}

Response to the *Surgeon Scorecard*

While the *Surgeon Scorecard* is promising, it has received widespread criticism from the medical community for its inability to convey a cohesive story.¹⁵ Core to the critiques against the *Surgeon Scorecard* is the derivation of the Adjusted Complication Rate, which does not reflect true complication rates. The Adjusted Complication Rate used by ProPublica only encompasses an adjusted 30-day readmission rate and a small number of deaths. This readmission figure excludes complications during the index hospitalization, complications that did not require readmission, and complications after the 30-day period.¹⁵ Data from the National Surgical Quality Improvement Program (NSQIP) suggest that for 7 of the 8 procedures reported in the *Surgeon Scorecard*, 88% of 30-day complications occurred during the index hospitalization;

these complications are not captured by the *Surgeon Scorecard*.^{16,17}

In designing their hierarchical models, the ProPublica researchers also set the hospital random effect to zero, with surgeons operating at a “hypothetical average hospital” with an average patient pool.¹⁷ In attempting to “level the playing field among surgeons,” this method undervalues the differences in care received by patients in low- and high-performing hospitals.¹⁵ A physician with above-average outcomes could very well be given a worse ranking due to his or her association with a hospital that has poor outcomes.¹⁸ In addition, good surgeons likely practice in good hospitals; the random hospital effect would adjust their outcomes toward the mean.¹⁵

Finally, ProPublica utilizes Medicare administrative claims data without conducting any validation against representative samples of clinical data. Using exclusively Medicare data leads to an inadequate sample size for reliable measures, particularly given the rarity of morbidity and mortality for the selected procedures.^{19,20} These data are also inconsistent, with significant observed variation in Part A and Part B Medicare claims data for surgical procedures.²¹ Reviewers from the RAND Corporation have identified specific misattributions in the *Surgeon Scorecard*, such as the listing of inapplicable surgeons, suggesting that the accuracy of performance data for surgeons is questionable.¹⁷ Most importantly, the results of the *Surgeon Scorecard* are not tested for reliability in predicting future performance of surgeons. This is the main factor that should drive patient decisions, and it highlights a critical flaw in ProPublica’s approach.

Future Directions

Surgeon scorecards have the capacity to be a valuable tool for guiding patient decisions; however, better data and measures are needed. The ProPublica authors discuss the potential for state-level clinical data to offer groundbreaking insights into patient safety that administrative claims data cannot.⁸ Furthermore, mortality and 30-day

readmissions are not sufficient measures with which to evaluate surgical performance. Perioperative mortality is rare, and using mortality for physician scorecards without case-mix adjustment will have unintended consequences.²² Within hospitals and practice groups, the most challenging operations with the highest risk of mortality are managed by select surgeons; heavily weighing mortality may unfairly punish these surgeons. Thirty-day readmission figures, particularly when exclusions are applied, also fail to accurately represent surgical complications that occur perioperatively.^{15,17}

In Michigan, we have developed a unique approach to creating clinically useful surgeon scorecards. Surgeons within the Michigan Surgical Quality Collaborative (MSQC) identified relevant measures to assess surgeon performance for a variety of procedures. For example, surgeons

selected postoperative morbidity, mortality, compliance with best practices, resource utilization, and anastomotic leak as being the most relevant criteria for a colectomy. Based on these criteria, we then utilized granular, state-level clinical data from the MSQC, in combination with process and utilization data, to assign composite scores to surgeons in the state of Michigan. While we are currently in the process of collecting data to robustly evaluate these scorecards, our preliminary findings suggest they may reliably predict future surgeon performance. They also identify specific domains of strength and weakness, providing actionable feedback to surgeons and hospitals. These scorecards will have the ability to improve health care transparency, ultimately empowering patients to make informed decisions while strengthening patient-provider relationships.

References

1. Mongan JJ, Ferris TG, Lee TH. Options for slowing the growth of health care costs. *N Engl J Med*. 2008;358(14):1509-1514.
2. Blendon RJ, Benson JM, Hero JO. Public trust in physicians—U.S. medicine in international perspective. *N Engl J Med*. 2014;371(17):1570-1572.
3. Levey NN. Medical professionalism and the future of public trust in physicians. *JAMA*. 2015;313(18):1827-1828.
4. Shea K, Shih A, Davis K. Health care opinion leaders' views on the transparency of health care quality and price information in the United States. November 2007. http://www.commonwealthfund.org/~media/files/surveys/2007/the-commonwealth-fund--modern-healthcare--health-care-opinion-leaders-survey--transparency-of-health/hcol_transparency_survey_data_brief.pdf.
5. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med*. 2008;148(2):111-123.
6. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369(15):1434-1442.
7. Lyu H, Cooper M, Patel K, Daniel M, Makary MA. Prevalence and data transparency of national clinical registries in the United States. *J Healthc Qual*. April 24, 2015. [Epub ahead of print.]
8. Pierce O, Allen M. Assessing surgeon-level risk of patient harm during elective surgery for public reporting (as of August 4, 2015). White paper. ProPublica, 2015. <https://static.propublica.org/projects/patient-safety/methodology/surgeon-level-risk-methodology.pdf>. Accessed September 11, 2015.
9. Wei S, Pierce O, Allen M. Surgeon scorecard. Online tool. ProPublica, 2015. <https://projects.ProPublica.org/surgeons/>. Accessed September 11, 2015.
10. Krumholz HM, Lin Z, Drye EE, et al. (2011). An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. 2011;4(2), 243-252.
11. Lau RL, Perruccio AV, Gandhi R., Mahomed NN. The role of surgeon volume on patient outcome in total knee arthroplasty: a systematic review of the literature. *BMC Musculoskelet Disord*. 2012;13(1):250.
12. Dimick JB, Staiger DO, Baser O, Birkmeyer JD. Composite measures for predicting surgical mortality in the hospital. *Health Affairs*. 2009;28(4):1189-1198.
13. van Walraven C, Austin, PC, Jennings A, Quan H, Forster, AJ. A modification of the Elixhauser

- comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47(6):626-633.
14. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
 15. Friedberg MW, Pronovost PJ, Shahian DM, et al. *A Methodological Critique of the ProPublica Surgeon Scorecard*. Santa Monica, CA: RAND Corporation; 2015. <http://www.rand.org/pubs/perspectives/PE170>.
 16. American College of Surgeons. *National Surgical Quality Improvement Program: Semiannual Report*. July 20, 2015.
 17. Friedberg MW, Bilimoria KY, Pronovost PJ, Shahian DM, Damberg CL, Zaslavsky AM. *Response to ProPublica's Rebuttal of Our Critique of the Surgeon Scorecard*. Santa Monica, CA: RAND Corporation; 2015. <http://www.rand.org/pubs/perspectives/PE170z1.html>.
 18. Hall BL, Huffman KM, Hamilton BH, et al. Profiling individual surgeon performance using information from a high-quality clinical registry: opportunities and limitations. *J Am Coll Surg*. 2015;221(5):901-913.
 19. Rosenbaum L. Scoring no goal—further adventures in transparency. *N Engl J Med*. 2015;373(15):1385-1388.
 20. Shih T, Cole AI, Al-attar PM, et al. Reliability of surgeon-specific reporting of complications after colectomy. *Ann Surg*. 2015;261(5):920-925.
 21. Dowd B, Kane R, Parashuram S, Swenson T, and Coulam RF. Alternative approaches to measuring physician resource use: final report. April 9, 2012. Centers for Medicare and Medicaid Services Web site. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/Research-Reports-Items/Alternative-Approaches-to-Measuring-Physician-Resource-Use.html>. Accessed December 1, 2015.
 22. Shahian DM, He X, Jacobs JP, Rankin JS, Peterson ED, Welke KF, et al. Issues in quality measurement: target population, risk adjustment, and ratings. *Ann Thorac Surg*. 2013;96(2):718-726.
 23. Bilimoria KY, Cohen ME, Ingraham AM, et al. Effect of postdischarge morbidity and mortality on comparisons of hospital surgical quality. *Ann Surg*. 2010;252(1):183-190.