

Surgical Sabermetrics: Applying Athletics Data Science to Enhance Operative Performance

Steven Yule, PhD,*,[†] Allison Janda, MD,[‡] and Donald S. Likosky, PhD[§]

Surgeons lack high fidelity, quantitative performance feedback regarding their individual, and team-level strengths and weaknesses. Existing gold-standard surgical registries are the equivalent of box scores for athletic events (baseball, soccer, national pro fast-pitch league), presenting essential data for assessing performance; within surgery, these registries track patient, process, and short-term outcome measures. However, these data explain only a small fraction of variability in hospital performance¹ and also lack important factors that impact quality and outcomes including the surgeon's technical skills, approach to the operation, and team-level nontechnical skills (eg, situation awareness, decision-making, communication, teamwork, leadership).² Additionally, the intraoperative environment has regular rotation of learners (akin to an athletic team's "Development League") who rotate on and off service such as resident physicians, as well as other team members including surgical assistants, anesthesiologists, nurses, and perfusionists who may be assigned to a given operating room and whose participation can fluctuate between and within cases. The integration of this dynamic team with standing players complicates team performance assessments. As the medical field has learned a great deal about patient safety from aviation, we could also learn from the data science approach taken by athletic leagues to assess the performance of our own players (surgical team members) and the broader operative team.

Twenty-first century baseball may leverage the benefits of data science more at present than surgery. The empirical analysis of baseball, specifically the quantification of in-game play to evaluate individual and team-based performance, is coined "sabermetrics." This emerging multidisciplinary science augments traditional in-game tally sheets with data science to uncover new insights beyond traditional box score metrics (eg, home runs, strikes). Sabermetrics enables individual athletes and teams to 1) use real-world, in-game data to evaluate individual

and team-level contributions to wins and losses and 2) leverage these data to make informed decisions about offensive and defensive strategies. The substrate for these sabermetric innovations is inexpensive, high-quality video combined with computer analytics. At an individual level, simulation workshops with multidimensional high-speed video cameras have been used to evaluate pitchers' arm angle and throwing mechanics,³ an investment that led to dramatic improvements in pitching statistics (eg, strikeouts, earned run average). High-definition video allows data analyses that are fine-grained enough to guide targeted modification of a pitcher's finger position on a baseball to alter its rotational speed, to identify and quantify fatigue and analyze batter and pitcher performance in high-stakes situations (eg, full count, or bases loaded). Analogies with operating room team performance are clear; the ability to contextualize surgical performance (eg, a surgeon's economy of motion, instrument handling, patient outcomes) and understand the role of intensity, pressure, and situational factors on individuals and teams could be invaluable. Documented examples of minor adjustments in positioning and orientation are limited in scale by human observers.⁴ Video-based surgical sabermetrics has the transformative potential to democratize quantified operative performance feedback (eg, technical and nontechnical skills box scores).

At an organizational level, Major League Baseball has invested in sabermetrics in part to reduce errors and improve the quality of umpire judgments. By tracking a baseball as it is pitched, Doppler technology determines its trajectory through the strike zone, and the message "ball," "strike," or "did not track" is relayed in real time through an earpiece to the umpire, who then makes the call. Accuracy is enhanced because the digital audiovisual system performs at a higher level of reliability and is less susceptible to bias than its human counterpart. Underlining the importance of sabermetrics in sports, data science and video skills are becoming essential skills in the coach and general manager job description.⁵ Just as Major League Baseball players and fans have an emotional and financial investment in the outcome of an inning, game, or season, surgeons, anesthesiologists, nurses, operative teams, and patients have an emotional and physical stake in the outcome of the surgery. This high-stakes environment also necessitates an unbiased assessment of events and performance.

The rise of "surgical sabermetrics" to augment existing surgical registries therefore offers an unprecedented opportunity for advancing surgical outcomes by pairing clinical data with video-based assessments of technical and nontechnical skills. Advanced analytics of digitally recorded surgical training and operative procedures can enhance insight, support professional development, and optimize clinical, safety, and financial outcomes. In Figure 1, we highlight 1) the perspective of established sabermetrics within baseball, as exemplar of how another industry has adopted data science to enhance performance and 2) potential analogous benefits of sabermetrics applied to the surgical specialty. In particular, augmenting traditional surgical metrics with sabermetrics may explain additional variation in patient outcomes and have a wide reaching impact in support of patient safety, quality, and education including personalized training, performance feedback, team optimization, and real-time clinical guidance. These innovations are founded on the

From the * Department of Clinical Surgery, University of Edinburgh, Edinburgh, Scotland; [†]Department of Surgery, Brigham & Women's Hospital/Harvard Medical School, Boston, MA; [‡]Department of Anesthesiology, Michigan Medicine, Ann Arbor, MI; and [§]Department of Cardiac Surgery, Michigan Medicine, Ann Arbor, MI.

Disclosure: The authors declare that they have nothing to disclose. This work was supported by a grant from the National Heart, Lung, and Blood Institute (R01HL146619). Dr. Janda is funded by a US National Institutes of Health T32 Grant - National Institute of General Medical Sciences (T32GM103730).

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Reprints: Steven Yule, PhD, FRCSEd, Department of Clinical Surgery, College of Medicine and Veterinary Medicine, University of Edinburgh, Scotland, United Kingdom. E-mail: steven.yule@ed.ac.uk

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Annals of Surgery (2021) 2:e054

Received: 28 January 2021; Accepted 13 February 2021

Published online 29 March 2021

DOI: 10.1097/AS9.0000000000000054

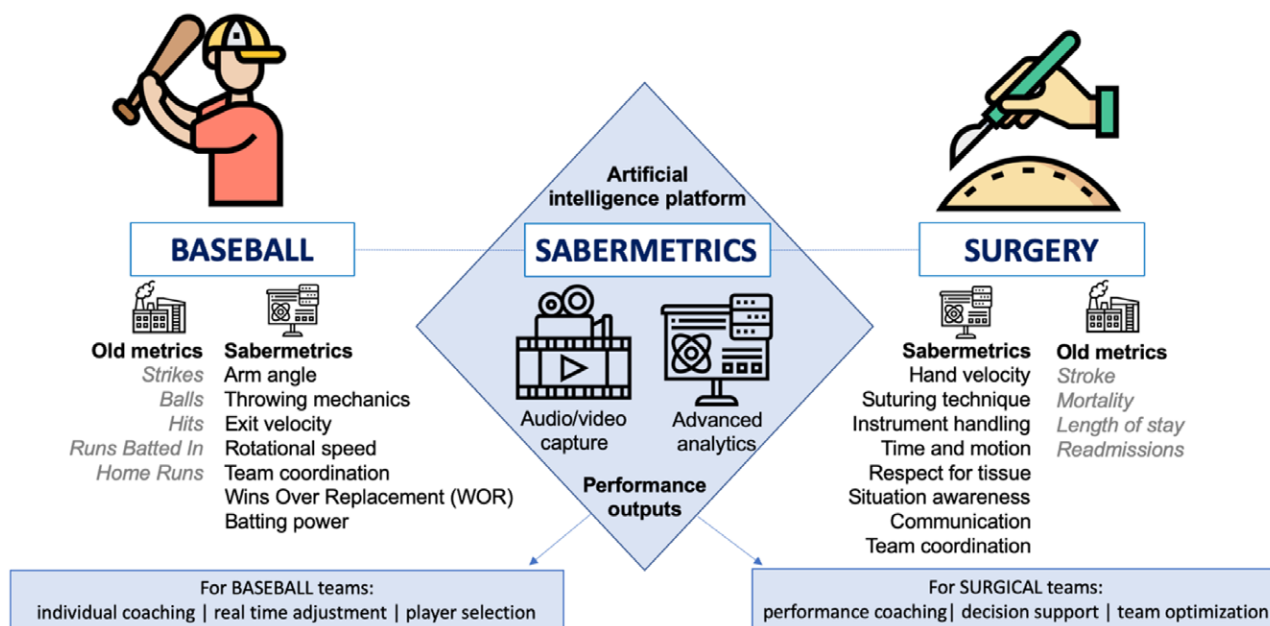


FIGURE 1. Surgical sabermetrics applies baseball-inspired data science to provide new insights into operating room performance that enhances patient safety and saves lives.

rigorous analysis of surgical videos as an assessment tool for learning, an approach that is arguably underutilized within this specialty and despite advantages also comes with a unique set of challenges.⁶

Artificial intelligence-enabled video with resulting analysis and targeted action has the potential for transformative improvements in surgery,⁷ including reduced surgical error, higher quality patient care, and enhanced professional development for clinicians. Within laparoscopic bariatric surgery, peer assessments from video of an individual surgeon's technical skills (eg, evaluating the movement of suturing technique or instrument handling) have provided meaningful performance data that are predictive of outcomes.⁸ There is unprecedented potential for artificial intelligence to enhance this process; by leveraging computer vision and machine learning to automatically detect operative phases, clinical decision support systems are being developed to transform surgical safety and quality in real time.⁹

However, individual skill alone does not ensure consistently high performance. A substantial proportion of success and failure in surgery may be attributed to team dynamics, communication, situation awareness, adaptability, and decision-making. In response, sabermetric-type capabilities for analysis of team-based performance are gradually seeing application within surgery. Crowd-sourcing video platforms leveraging taxonomies for assessing technical and nontechnical skills¹⁰ have emerged for surgeons, anesthesiologists, scrub practitioners, and entire operating room teams. Embedded artificial intelligence within these platforms to identify previously unobservable practice patterns signals that a surgical sabermetrics revolution is imminent.

As evidence, researchers at the University of Toronto have developed and successfully pilot-tested a comprehensive, prospective multipoint data recorder called the operating room black box, that utilizes several inputs (eg, audiovisual, physiologic, environmental, and device-related data streams) continuously captured during surgery. Critical analyses and expert reflection on surgical performance are linked to the operative video. During a 1-year implementation, this technology identified and characterized intraoperative errors, events, and distractions among 132 consecutive patients undergoing elective laparoscopic surgery.¹¹ A median of 20 errors (interquartile range, 14–36) and 8 events (interquartile range, 4–12) and 138

auditory distractions were identified per procedure. Cognitive distractions were identified in 64% of procedures. Surgical sabermetrics like the operating room black box are filling a knowledge gap left by gold-standard surgical data registries and demonstrate a data science approach to quantifying the human, technical, and system factors that combine to define successful surgery. These factors have always been thought to influence performance, but until now have not been systematically measured in the operating room, or linked with clinical outcomes using data science.

The ubiquity of video, paired with advancements in technology and data sciences, provides the surgical community with myriad opportunities to transform healthcare delivery and outcomes. We envision data from sabermetrics could be used to provide much needed foundation for coaching providers on technical¹² and nontechnical skills.¹³ Nonetheless, wide-scale adoption of surgical sabermetrics by our profession will require addressing a number of technical and cultural challenges, including how to obtain multicenter, granular audiovisual, electronic health record and patient outcomes data in surgery, the need for embedded tools with validity evidence, acceptability for sabermetrics among team members, and delineating individual versus team perspectives. We propose four challenges to the surgical community that if addressed can pave the way for sabermetric innovations to enhance operative performance:

- 1) Implementing technology for digitally capturing and reviewing surgeons' technical skills and team nontechnical skills at scale and gathering sufficiently large samples to facilitate training of artificial intelligence models.
- 2) Identifying structured frameworks to analyze and learn from past performance and gathering validity evidence to incorporate findings from video analyses to augment surgical decision-making and enhance nontechnical skills.
- 3) Defining metrics for artificial intelligence-enabled assessment of technical and nontechnical skills with context-specific validity evidence to reduce bias.
- 4) Encouraging a culture supporting the transparent use of digital recordings for quality improvement by engaging thought leaders and multidisciplinary clinical collaborators who are custodians of sabermetric data capture, storage, use, governance.

In summary, industries such as aerospace and sports routinely quantify the building blocks of individual and team performance, developing sophisticated analytic frameworks for advancing quality. Within athletics, millions of dollars are spent annually on selection, assessment, and coaching, enabling teams to fine tune aspects of behavior and skill to enhance performance. Organizations are able to select the team with the highest chance of success for the local conditions, time of day, and opponent and make real-time adjustments to avoid pitfalls, rescue errors, and take advantage of opportunities. Implementing sabermetrics can turn average players into superstars and transform losing teams into champions, with dramatic changes in technique seemingly impossible a decade ago. Disruptive change often requires outside perspective and influence, rather than incremental improvement from within. By addressing the four challenges highlighted in this perspective article, digital capture of audio and video, paired with advancements in technology and data sciences, has the potential to revolutionize clinical performance and enhance outcomes for clinicians, teams, hospitals, and patients everywhere.

ACKNOWLEDGMENT

The authors thank Francis Pagani, MD, PhD, for critical appraisal and guidance.

REFERENCES

1. Brescia AA, Rankin JS, Cyr DD, et al; Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative. Determinants of variation in pneumonia rates after coronary artery bypass grafting. *Ann Thorac Surg*. 2018;105:513–520.
2. Neily J, Mills PD, Young-Xu Y, et al. Association between implementation of a medical team training program and surgical mortality. *JAMA*. 2010;304:1693–1700.
3. Foley M. Adam ottavino's harlem hide-out. *The New York Times*. February 10, 2019. Available at: <https://www.nytimes.com/2019/02/10/sports/baseball/adam-ottavino-yankees.html>. Accessed March 9, 2021.
4. Gawande A. Personal best. Top athletes and singers have coaches. Should you?. *The New Yorker*. October 3, 2011.
5. Lemire J. Throw BP, know SQL: The modern baseball coach's job description. 2018. Available at: <https://www.sportstechie.com/mlb-data-analytics-sabermetrics-sam-fuld-jeff-luhnow/>. Accessed January 19, 2021.
6. Pugh CM, Hashimoto DA, Korndorffer JR Jr. The what? How? And who? Of video based assessment. *Am J Surg*. 2021;221:13–18.
7. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg*. 2019;270:414–421.
8. Birkmeyer JD, Finks JF, O'Reilly A, et al; Michigan Bariatric Surgery Collaborative. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369:1434–1442.
9. Hashimoto DA, Rosman G, Rus D, et al. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70–76.
10. Yule S, Gupta A, Gazarian D, et al. Construct and criterion validity testing of the Non-Technical Skills for Surgeons (NOTSS) behaviour assessment tool using videos of simulated operations. *Br J Surg*. 2018;105:719–727.
11. Jung JJ, Jüni P, Lebovic G, et al. First-year analysis of the operating room black box study. *Ann Surg*. 2020;271:122–127.
12. Greenberg CC, Ghousseini HN, Pavuluri Quamme SR, et al. Surgical coaching for individual performance improvement. *Ann Surg*. 2015;261:32–34.
13. Pradarelli JC, Yule S, Panda N, et al. Surgeons' coaching techniques in the surgical coaching for operative performance enhancement (SCOPE) program [published online ahead of print July 24, 2020]. *Ann Surg*. doi: 10.1097/SLA.0000000000004323.