

Surgical Innovation

Crowdsourcing to Assess Surgical Skill

Thomas S. Lendvay, MD; Lee White, PhD; Timothy Kowalewski, PhD

What Is the Innovation?

Surgical skills impact patient outcomes.¹ Our profession needs methods that accurately and objectively evaluate surgical skills.² These methods must provide timely and meaningful feedback, minimize review time burden, and scale for widespread use. This evaluation hurdle may be overcome by leveraging crowdsourcing to help triage outliers and focus improvement efforts. Crowdsourcing is the process of using large groups of decentralized, independent people providing aggregated feedback.³ Crowdsourcing to assess surgical skill is a method by which surgical technique can be assessed by crowds of reviewers, some of whom are nonmedically trained. Crowdsourcing has enjoyed broad success in health care—discovering protein-folding patterns, assisting disabled patients, locating automatic defibrillators within cities, and annotating electronic medical records.³

What Are the Key Advantages Over Existing Approaches?

Surgeons in the United States annually perform more than 51 million operations, and no existing assessment method can scale to such magnitude. Some groups have performed assessments of their colleagues as part of quality-improvement collaboratives.¹ These initiatives add time burden to normal practice and may not be scalable outside of small voluntary groups or research endeavors.

The current state of objective surgical assessment involves either observing surgical performance in real-time by experts/coaches or by capturing video and executing postperformance appraisal using structured assessment tools. Although these tools are considered the gold standard, they are seldom used because they require extensive time commitments from the reviewers to achieve inter-rater reliability and time to watch the performances. These also in-

roduce appraisal bias because reviewers may be from within the same institution or within the same practicing community.

Crowdsourcing can leverage the readily available pool of millions of anonymous online crowdworkers to mitigate reviewer bias.³ Feedback with crowdsourcing is also timely as the responses are registered within minutes of posting surveys and completed within hours.⁴⁻⁷

How Will This Impact Clinical Care?

As governments and health care organizations demand more accountability, we will need objective performance metrics that directly correlate with patient outcomes. Until now, we have lacked scalable proficiency assessment solutions that can provide surgeons with meaningful feedback, which may improve future practice. The trend is emerging for evaluation of practicing clinicians' skills and linking these data to patient outcomes.¹

Is There Evidence Supporting the Benefits of the Innovation?

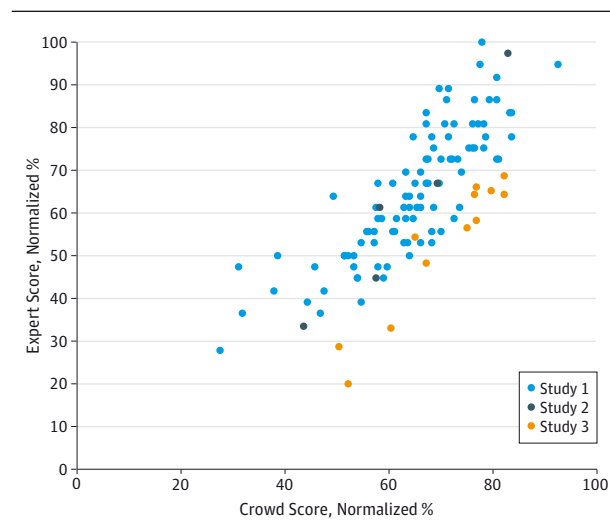
An increasing body of evidence suggests that crowdsourced evaluations strongly agree with surgical expert evaluations when all evaluators use identical assessment tools for appraising the same deidentified videos of laparoscopic and robotic skills.⁴⁻⁷ The Figure combines data from 3 different studies into a single plot, covering 115 unique surgical task performances evaluated by crowds and experts.⁴⁻⁶

To test this method in animate surgery, Holst et al⁶ observed similar correlations with surgeons doing live robotic porcine bladder closures. Previously, the Michigan Urologic Surgery Improvement Collaborative established that crowdsourcing skills assessment of robotic prostatectomies performed by 12 clinicians in the urologic community could adequately correlate with a panel of expert peer reviewers.⁷ Both crowds and surgeon reviewers identified the same bottom 5 surgeons ranked in order of their technical skill; this finding is relevant to a collaborative of surgeons looking to identify surgeons who may benefit from coaching.

What Are the Barriers to Implementing This Innovation More Broadly?

Crowdsourcing to assess surgical skill faces barriers to widespread adoption. First, surgeons may doubt that nonmedically trained groups can accurately grade their skills. To overcome this barrier, evidence from multiple respected organizations testing this method will be required. In a culture of rapid feedback, waiting weeks to receive peer-reviewed appraisals may not satisfy a tenet of feedback, which is that it be immediate or near-immediate. Surgeons will be able to receive feedback from crowds within hours to a day, thus benefitting the next patient on whom the surgeon is operating.

Figure. Plot Showing Composite Results of Crowd-to-Expert Surgical Skills Assessment Scores From 3 Studies



A second barrier is capturing the performances for review. Proliferating technologies address this: new operating rooms have integrated video capture capabilities and wearable technology (ie, Google Glass or video cameras in head-mounted lights). Furthermore, increasingly used minimally invasive surgical approaches make video capture routine. While striking the balance between patient privacy and skill transparency is crucial, as health care costs become more shared by individuals and employers, objective data on surgeons may be demanded.

The third barrier is reconciling the role technical skill has on outcomes relative to other influencers such as surgical selection and judgment, adherence to best practices, systems errors, and patient education. All are necessary but not individually sufficient.

The fourth and most challenging barrier will be changing the culture among surgeons to be receptive to feedback. In residency, we are familiar with frequent assessments but in practice, our technical skills are rarely evaluated. If assessment is used primarily as a punitive process or mandated by payers, adoption will be slow or non-existent. We submit that when such tools are driven by surgical initiatives and not payer directives, physicians welcome information on how to improve. There are avenues for continuing medical education, which address the choice of clinicians to enhance their practice, and surgical skills feedback could become an additional continuing medical education opportunity.

In What Time Frame Will This Innovation Likely Be Applied Routinely?

Crowdsourcing to assess surgical skill is currently being applied in practice, and as researchers continue validation, more evidence will emerge of its efficacy in improving clinical outcomes. Key opinion leaders within surgical societies espousing its value will help drive adoption. Other means of adoption will be through implementation in residency training programs. As the results are realized, graduate medical education bodies and surgical boards may choose to adopt similar assessment methods for proficiency benchmarking, advancement, and maintenance of certification.

Conclusions

We believe surgeons are inherently driven to improve their technical skills to advance patient outcomes. Instead of waiting for directives from the medical administrative apparatus, we have an opportunity to change the paradigm of technical skills assessment so that all surgeons may benefit from rapid, accurate, objective, and meaningful feedback. Listening to surgeons about this value and the feedback delivery mechanisms will be required to ensure that surgeons do not feel threatened by punitive action or a lack of discretionary control that could impact the surgeons' practices.

ARTICLE INFORMATION

Author Affiliations: Department of Urology, Seattle Children's Hospital, University of Washington, Seattle (Lendvay); Stanford University, Stanford, California (White); Department of Mechanical Engineering, University of Minnesota, Minneapolis (Kowalewski).

Corresponding Author: Thomas S. Lendvay, MD, Department of Urology, University of Washington, SP1266, 1st Floor, Surgery Pavilion, Box 356158, 1959 NE Pacific St, Seattle, WA 98195 (thomas.lendvay@seattlechildrens.org).

Section Editor: Justin B. Dimick, MD, MPH.

Published Online: September 30, 2015.
doi:10.1001/jamasurg.2015.2405.

Conflict of Interest Disclosures: All research directed by the 3 authors presented in this article was completed prior to the completion of the commercialization of C-SATS Inc, a University of Washington startup company cofounded by the authors of this article with equity interests. The Michigan Urologic Surgery Improvement

Collaborative Study was directed by the members of the Michigan Urologic Surgery Improvement Collaborative organization.

Submissions: Authors should contact Justin B. Dimick, MD, MPH, at jdimick@med.umich.edu if they wish to submit Surgical Innovation papers.

REFERENCES

1. 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(15):1434-1442.
2. Hampton T. Efforts seek to develop systematic ways to objectively assess surgeons' skills. *JAMA*. 2015;313(8):782-784.
3. Ranard BL, Ha YP, Meisel ZF, et al. Crowdsourcing: harnessing the masses to advance health and medicine, a systematic review. *J Gen Intern Med*. 2014;29(1):187-203.
4. White LW, Kowalewski TM, Dockter RL, Comstock B, Hannaford B, Lendvay TS.

Crowd-sourced assessment of technical skill: a valid method for discriminating basic robotic surgery skills [published online August 24, 2015]. *J Endourol*. doi:10.1089/end.2015.0191.

5. Holst D, Kowalewski TM, White LW, et al. Crowd-sourced assessment of technical skills: an adjunct to urology resident surgical simulation training. *J Endourol*. 2015;29(5):604-609.

6. Holst D, Kowalewski T, White LW, et al. Crowd-Sourced Assessment of Technical Skills (C-SATS): differentiating animate surgical skill through the wisdom of crowds [published online April 13, 2015]. *J Endourol*. doi:10.1089/end.2015.0104.

7. Peabody J, Miller D, Lane B, et al. Wisdom of the crowds: use of crowdsourcing to assess surgical skill of robot-assisted radical prostatectomy in a statewide surgical collaborative. *Eur Urol Supplements*. 2015; 193(4):e655-e656. doi:10.1016/j.juro.2015.02.1826.