RESEARCH ARTICLE JSLS

Association between Surgical Technical Skills and Clinical Outcomes: A Systematic Literature Review and Meta-Analysis

Michael S. Woods, MD, MMM, Joshua N. Liberman, PhD, MBA, Pinyao Rui, MPH, Emily Wiggins, MPH, Joan White, APRN, MSN, Bruce Ramshaw, MD, Jonah J. Stulberg, MD, PhD, MPH

ABSTRACT

Background: A systematic literature review and meta-analysis was conducted to assess the association between intraoperative surgical skill and clinical outcomes.

Methods: Peer-reviewed, original research articles published through August 31, 2021 were identified from PubMed and Embase. From the 1,513 potential articles, seven met eligibility requirements, reporting on 151 surgeons and 17,932 procedures. All included retrospective assessment of operative videos. Associations between surgical skill and outcomes were assessed by pooling odds ratios (OR) using random-effects models with the inverse variance method. Eligible studies included pancreaticoduodenectomy, gastric bypass, laparoscopic gastrectomy, prostatectomy, colorectal, and hemicolectomy procedures.

Results: Meta-analytic pooling identified significant associations between the highest vs. lowest quartile of surgical skill and reoperation (OR: 0.44; 95% confidence interval [CI]: 0.23, 0.83), hemorrhage (OR: 0.66; 95% CI, 0.65, 0.68), obstruction (OR: 0.33; 95% CI, 0.30, 0.35), and any medical complication (OR: 0.23, 95% CI, 0.19, 0.27). Nonsignificant inverse associations were noted between skill and readmission, emergency department visit, mortality, leak, infection, venous thromboembolism, and cardiac and pulmonary complications.

Conclusions: Overall, surgeon technical skill appears to predict clinical outcomes. However, there are surprisingly few articles that evaluate this association. The authors recommend a thoughtful approach for the development of a comprehensive surgical quality infrastructure that could significantly reduce the challenges identified by this study.

Key Words: Surgical outcome, Surgical technical skill, Video-based assessment.

Caresyntax Corp., Mequon, WI. (Dr. Woods and Ms. White)

Health Analytics LLC., Columbia, MD. (Dr. Liberman, Mss. Rui and Wiggins)

CQInsights PBC, Knoxville, TN. (Dr. Ramshaw)

Department of Surgery, McGovern Medical School at the University of Texas Health Sciences Center of Houston, Houston, TX. (Dr. Stulberg)

Data availability: The datasets generated and analyzed during the current study are available from the corresponding author upon request.

Acknowledgements: The authors wish to acknowledge Perri Beach for her administrative support in the preparation of this manuscript.

Disclosure: Health Analytics received funding to conduct the research.

Conflict of interests: Dr. Ramshaw via CQ Insights is a paid consultant to Caresyntax.

Funding sources: Caresyntax sponsored this research project.

Informed consent: Dr. Joshua N. Liberman declares that since this article is a systematic literature review, no patient informed consent was appropriate or possible.

Address correspondence to: Dr. Joshua N. Liberman, PhD, MBA, Health Analytics, LLC, 9200 Rumsey Road, Suite 215, Columbia, MD 21045, Telephone: 925.301.3969, E-mail: j.liberman@healthanalytics.com.

DOI: 10.4293/JSLS.2022.00076

INTRODUCTION

The operating room is incredibly complex and potentially very dangerous. This complexity has led to substantial variation in patient outcomes, even among surgical centers of excellence.[1,2] Decades of research have led to the identification of risk factors for poor intraoperative and postoperative outcomes, including characteristics related to the patient, the institution, the procedure, the surgical team, and the surgeon. Commonly reported patient-related risk factors include age,[3–5] anatomic complexity,[6] comorbidities,[7–9] and frailty.[10,11] Institution-related factors typically focus on experience with a specific procedure.[12–15] Procedure-related factors include total operative or anesthesia time,[16,17] procedure complexity,[18–20] and type of procedure.[5,21] The most common measures of a surgeon's performance relate to experience and procedure volume.[14,22,23]

However, more direct assessments of a surgeon's technical performance have been developed and validated. These measures, such as the Objective Structured Assessment of Technical Skills (OSATS),[24] Global Evaluative Assessment of Laparoscopic Skills (GOALS),[25] and Global Evaluative Assessment of Robotic Skills (GEARS),[26] require direct evaluation of the surgeon's technical skills while performing a surgical procedure. In addition, there are procedure-specific assessments, such as the Colorectal Objective Structured Assessment of Technical Skills (COSATS)[27] in use and others in development such as one for hiatal hernia[28] and one for laparoscopic fundoplication.[29]

A common method for scoring surgical performance is to review surgeon-specific operative videos. Video-based assessment (VBA) in training can document consistent surgical technical skill improvement.[30–32] Further, given the substantial variability in technical skills among surgeons, recent calls-to-action recommend the widespread implementation of VBA for continual quality improvement among practicing surgeons.[32–34] While improving technical skills is valuable for surgical training and potentially valuable for continual learning, a technical skill score in the absence of understanding impact on patient outcome is of little meaning. Therefore, it is essential to understand whether better technical proficiency correlates with better patient outcomes.

Until recently, few research studies have explored the relationship between standardized measures of surgeon technical skills in relationship to patient outcomes.[35,36] These studies demonstrated that surgeons scoring in the upper quartile generally had better patient outcomes than those surgeons scoring in the lowest quartile. However, these studies document this association in specific surgical settings, for specific procedures, and among specific specialties.

To assess the current state of knowledge we conducted a systematic literature review and meta-analysis to document the relationship between surgeon technical skill and patient outcomes with a focus on: 1) the consistency and magnitude of any association between technical skill and patient outcomes and 2) gaps in the inclusion of surgical specialties, procedures, and outcomes.

METHODS

Search Strategy and Eligibility

The study was conducted in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[37] and Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines.[38]

Eligible manuscripts were English-language, original research studies that included both measurement of surgical technical skills and patient-specific intraoperative or postoperative clinical outcomes. Eligible articles were required to report on surgeon technical skills using observation of a surgeon's performance during real-world surgical procedures. Articles that met any of the following criteria were deemed ineligible: reported surgeon technical skill or performance based on outcomes (e.g. evaluation of surgical success by completion of procedure-specific tasks such as quality of suturing, or of imaging that documents correct placement of a device); nonprimary research (abstracts, conference proceedings, posters, commentaries, and editorials); review articles; or studies related to dental surgery, obstetrics, or ophthalmology.

A systematic literature search was conducted from database inception to August 31, 2021 using the PubMed and Embase library databases to identify peer-reviewed original research manuscripts that assessed the association between surgical skills and clinical outcomes. Search terms were derived from Medical Subject Heading (MeSH) and non-MeSH terms using Boolean operators applied to surgery (MeSH: "surgical procedures, operative"), level of skill (MeSH: "professional competence" and "clinical competence"), clinical outcomes (MeSH: "treatment outcome"), and surgical technical quality (see **Appendix Table 1**).

Two authors (JNL and EW) independently evaluated each manuscript title, abstracts for articles deemed potentially eligible following title review, and full manuscripts deemed potentially eligible following abstract review. To ensure complete capture of eligible articles, the reviewers screened the references and citations of each eligible manuscript. The two authors independently extracted the following data from each manuscript: study design; geographic location; surgeon and surgical patient sample size; surgeon specialty; procedure type; technical skills evaluation methodology; and outcomes. Each article was evaluated for the following: technical skill assessment method (direct observation, deidentified video submission); number and type of raters evaluating technical skill; and technical skill assessment scale used. The risk of bias in each manuscript was assessed using a modified version of the Newcastle-Ottawa Scale **(Appendix Tables 2 and 3)**.[39]

Analysis

Data abstracted from the full manuscript review was used to create an analytical file in Microsoft Excel. This file contained data on all outcomes reported in the eligible manuscripts, the sample size of surgeries by surgeon technical skill group (i.e. quartile 1 – quartile 4), the number of outcome events in each surgeon technical skill group, the average technical skill score by presence/absence of outcome, precalculated effect sizes and variances, etc. Eligible studies all reported technical skill scores by quartiles except for MacKenzie, et al.[40] which reported three categories based on cumulative sum chart and validated in prior publications. These three categories were assigned as Q1, Q2 – Q3, and Q4 to allow comparison and inclusion in the analysis.

The data were imported into R version 40.10.1 (R Project for Statistical Computing) and the R package 'meta' was used to synthesize results across studies to derive pooled effect sizes. Random effects models were used to estimate effect sizes and variances due to the assumption of between-study heterogeneity. Mantel-Haenszel odds ratios and Knapp-Hartung adjusted 95% confidence intervals (CI) of surgeon technical skill quartile associated with surgical complications were calculated by pooling study-specific odds ratios using random-effects models with invariance method to incorporate the heterogeneity of differences across studies. Between-study heterogeneity was measured using the Paule-Mandel method of calculating the heterogeneity variance τ^2.

A separate meta-analysis was conducted to include studies that either reported precalculated odds ratios or information to derive odds ratios. The association between surgeon technical skill and any postoperative complications was assessed by pooling odds ratios using random-effects models with the inverse variance method. Between-study heterogeneity was measured using the Paule-Mandel method of calculating the heterogeneity variance τ^2. Publication bias was evaluated using funnel plots and Egger's test. Statistical tests were two-sided and used a significance threshold of $P < .05$.

RESULTS

A total 1,513 articles were identified by the systematic search criteria and from that search, four[35,36,40,41] were identified as eligible (**Figure 1**). Following a complete reference and citation search among these articles, an additional three articles[42–44] were deemed eligible for inclusion (**Figure 1**). Noting that Hogg, et al.[44] reported the number of procedures but not surgeons, the seven eligible studies reported on 151 surgeons and a total of 17,932 procedures. Each study reported on a different procedure: laparoscopic sleeve gastrectomy, laparoscopic right hemicolectomy, laparoscopic gastric bypass, laparoscopic gastrectomy for gastric adenocarcinoma, laparoscopic colorectal surgery, robot-assisted radical prostatectomy, and robot-assisted pancreaticoduodenectomy (**Table 1**).

Among the four studies in a U.S. surgical population, two were from the Michigan Bariatric Surgery Collaborative, one from the Illinois Surgical Quality Improvement Collaborative, and one from the University of Pittsburgh Medical Center (**Table 1**). Two studies reported on a Canadian surgical population, one from the University of Toronto and one for the University of British Columbia, and one reported on a U.K. population from the National Training Programme for Laparoscopic Colorectal Surgery (**Table 1**).

Technical Skills Assessment and Presentation

The technical skill assessment was completed by review of deidentified operative videos, with five articles reporting on an assessment derived from a participating surgeon's self-selected representative surgical video and three articles reporting on surgery-specific videos (**Appendix Table 2**). The number of raters for each video varied from 1 to 10 (or more), with the raters' specialty and experience varying from surgeons familiar with the surgery to individuals familiar with the surgery.

The most frequently used skills assessment instrument was the OSATS (or modified OSATS) used as the primary tool in five of the studies, followed by the GEARS[43] and the competency assessment tool.[40] The COSATS[36] and the Generic Error Rating Tool (GERT) [42,43] were also included complementary assessments.

Four studies[35,36,40,41] reported outcomes as a function of technical skill, typically categorizing technical skill scores into quartiles (**Figure 2**). Three studies[42–44] reported technical skill scores among individuals with and without poor postoperative outcomes and odds ratios of the association between surgeon technical skill and poor postoperative outcomes (**Figure 3**).

Outcomes

Two studies[40,44] did not report the timeframe over which outcomes were measured. One article[43] reported on achieving postprostatectomy continence by three-months. The remaining five articles reported 30-day postoperative outcomes (**Table 1**). The most frequently reported

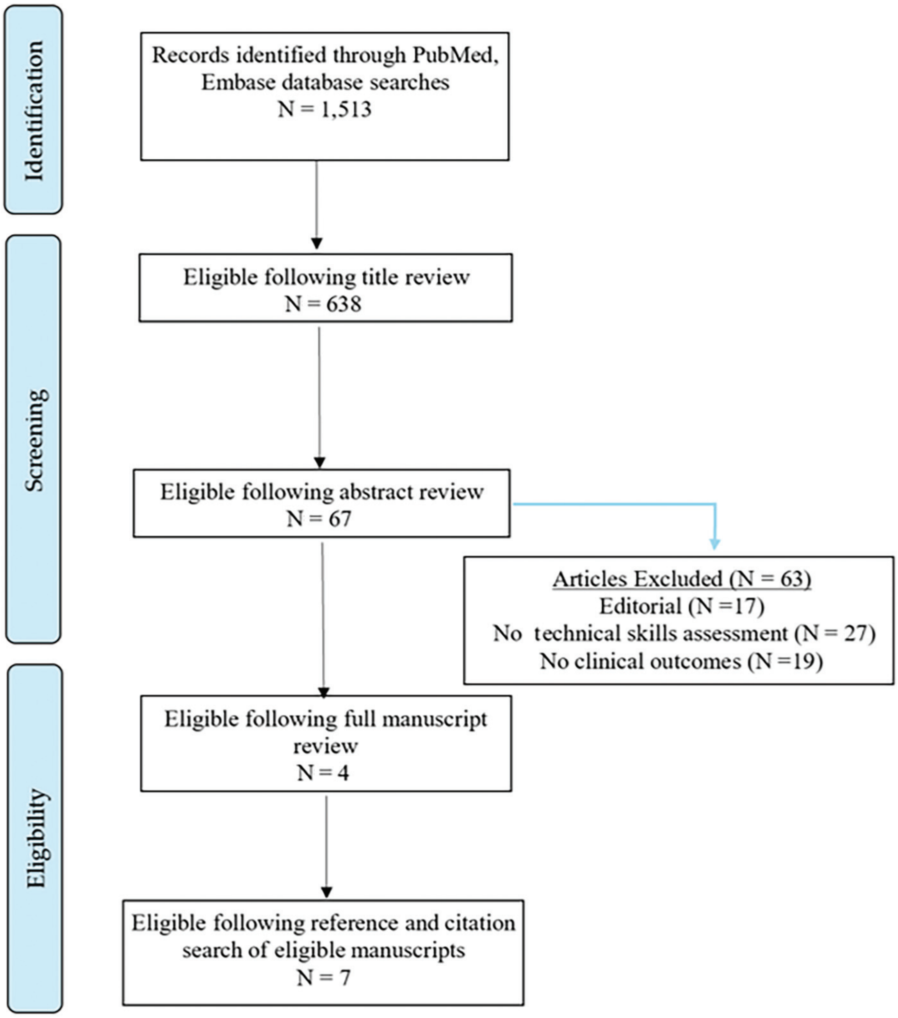

Figure 1. Study selection flowchart.

outcomes (reported in three studies) were mortality, reoperation, readmission, and infection, followed by the following outcomes reported in two studies: emergency department (ED) visit, any surgical complication, any medical complication, leak, obstruction, hemorrhage, venous thromboembolism (VTE), cardiac complication, and pulmonary complication. A composite measure for any postoperative complication was created to include three studies[42–44] that reported precalculated ORs of the association between surgeon technical skill and postoperative complications (**Figure 4**) and three studies[35,36,40] that reported the number of postoperative complications occurring in each quartile of surgeon technical skill in which an OR could be derived (**Figure 4**). The measure included the outcomes of any complication reported in four studies,[35,36,40,45] postoperative pancreatic fistula reported in Hogg 2016,[44] and incontinence reported in Goldenberg 2017,[43] for a total of six studies (**Figure 4**).

Surgical Skill and Outcomes

No studies reported that higher technical skill was associated with poorer patient outcomes. All reported significant associations between surgeon technical skill and at least one outcome. Four articles categorized surgeons by technical skill scores and reported outcome events by those categories.[35,36,40,41] Birkmeyer et al.[35] and Stulberg, et al.[36] reported on three categories of technical skill (bottom quartile, middle 50%, and top quartile) (**Figure 2**) while Varban, et al.,[41] reported on top quartile vs. bottom quartile only (**Figure 3**). Three

Table 1.
Review Development Literature for Standardized Instruments

Authors	Sampling Frame	Surgery	Sample Size (Surgeons)	Sample Size (Patients/ Surgeries)	Time Frame	Skill Assessment Instrument	Skill Data Source	Evaluators	Post-Operative Outcomes
Birkmeyer et al. (2013)	Michigan Bariatric Surgery Collaborative, Michigan, U.S.A.	Laparoscopic gastric bypass	20	10,343	8/28/2006 – 8/1/2012	Modified OSATS	Self-selected representative video	10+ participating surgeon-raters	Any postoperative complication, incl. surgical and medical. Surgical: SSI, wound infection requiring reoperation; abdominal abscess; a leak; anastomotic stricture; bowel obstruction; bleeding. Medical: pneumonia, respiratory failure, renal failure, VTE, AMI, cardiac arrest, death. Also, 30-day events of: mortality, unplanned reoperation, readmission, and ED visit.
Fecso, et al. (2019)	University of Toronto, Canada	Laparoscopic gastrectomy for gastric adenocarcinoma	3	61	1/1/2009 – 12/31/2015	OSATS, GERT	Patient-specific operative videos	Single rater	30-day rates of surgical complications classified by Clavien-Dindo categorized into two groups: no or minor complications vs. major complications (CD ≥ 3)
Goldenberg, et al. (2017)	University of British Columbia, Canada	Robot-assisted radical prostatectomy	1	46	NS	GEARS, GERT	Patient-specific operative videos	Single rater	Continence at three-month postoperatively
Hogg, et al. (2016)	University of Pittsburgh Medical Center, U.S.	Robot-assisted pancreaticoduodenectomy	NS	133	11/2011 – 07/2015	Modified OSATS	Patient-specific operative videos	Two hepatobiliary surgeons	Postoperative pancreatic fistula.
Mackenzie et al. (2015)	National Training Programme for Laparoscopic Colorectal Surgery, U.K.	Laparoscopic colorectal surgery	85	171	9/2009 – 2/2013	Competency assessment tool:1	Self-selected representative video	Two raters	Any surgical outcome: anastomotic leak, bleeding, abdominal collection, ileus, obstruction, and wound infection; Any medical complications: respiratory, cardiac, cerebrovascular.
Stulberg et al. (2020)	Illinois Surgical Quality Improvement Collaborative, Illinois, U.S.A.	Laparoscopic right hemicolectomy	17	1,120	9/23/2016 to 2/10/2018	Modified OSATS, COSATS	Self-selected representative video	10+ participant-reviewers and 2 colorectal surgeons.	Measured colorectal skills against colorectal surgical outcomes AND against noncolorectal surgical outcomes

Table 1. Continued

Authors	Sampling Frame	Surgery	Sample Size (Surgeons)	Sample Size (Patients/ Surgeries)	Time Frame	Skill Assessment Instrument	Skill Data Source	Evaluators	Post-Operative Outcomes
Varban, et al. (2021)	Michigan Bariatric Surgery Collaborative, Michigan, U.S.A.	Laparoscopic sleeve gastrectomy	25	3607 surgeries among 3,088 patients	2015 – 2016	Modified OSATS	Self-selected representative video	371 reviews for 33 videos performed by 25 surgeons	30-day postoperative reoperation, readmission, and ED visits. Surgical complications: SSI, infection, abscess, leak, bowel obstruction requiring reoperation, blood transfusion, reoperation, splenectomy. Medical complications: pneumonia, respiratory failure, renal failure, VTE, AMI, cardiac arrest, death.

Abbreviations: OSATS, Objective Structured Assessment of Technical Skills; COSATS, Colorectal Objective Structured Assessment of Technical Skills; CD, Clavien-Dindo; ED, emergency department; SSI, surgical site infection; VTE, venous thromboembolism, AMI, acute myocardial infarction; GEARS, Global Evaluative Assessment of Robotic Skills; GERT, Generic Error Rating Tool.

articles[42–44] reported technical skill scores among individuals with and without select outcomes (**Figure 4**).

Meta-analytic pooling of the associations between the highest vs. lowest quartile of surgeon technical skill and the outcome of reoperation resulted in a summary OR of 0.44 (95% CI, 0.23, 0.83), with low heterogeneity across the three studies ($I^2 \leq 0.01\%$, $\tau^2 < 0.01$, $P = .48$). Meta-analytic pooling of the association between the highest vs. lowest quartile of surgeon technical skill and hemorrhage resulted in a summary OR of 0.66 (95% CI, 0.65, 0.68), with low heterogeneity across the two studies ($I^2 \leq 0.01\%$, $\tau^2 < 0.01$, $P = .99$). The association between the highest vs. lowest quartile of surgeon technical skill and obstruction was 0.33 (95% CI, 0.30, 0.35) with low heterogeneity across the two studies ($I^2 \leq 0.01\%$, $\tau^2 < 0.01$, $P = .97$). The association between the highest vs. lowest quartile of surgeon technical skill and medical complication was 0.23 (95% CI, 0.19, 0.27) with low heterogeneity across the two studies ($I^2 \leq 0.01\%$, $\tau^2 < 0.01$, $P = .95$). The association between the highest vs. lowest quartile of surgeon technical skill and the outcomes of readmission, ED visit, mortality, leak, infection, VTE, cardiac complication, and pulmonary complication did not reach statistical significance. Funnel plots are displayed for each outcome (**Figure 3**). Egger's test indicated no significant publication bias (p range, 0.36 – 0.75) six of the seven studies reported either precalculated ORs or information to derive ORs of the association between surgeon technical skill and any postoperative complication. Meta-analytic pooling of the odds ratios yielded a summary OR of 0.37 (95% CI, 0.21, 0.66) with moderate heterogeneity across studies ($I^2 = 55\%$, $\tau^2 = 0.31$, $P = .05$). A forest plot of studies that reported on the association between surgeon technical skill and postoperative complications is shown in **Figure 4**. Egger's test indicated no significant publication bias ($P = .86$).

DISCUSSION

Across outcomes measured there is a consistent, albeit not always statistically significant, association between surgical technical skills and clinical outcomes. Unfortunately, very few articles met eligibility criteria, which limits the interpretation and indicates a significant opportunity for expanded, structured research. Nonetheless, our results are consistent with other studies that explore surgeon technical performance and clinical outcomes, in particular, a recently published systematic review by Balvardi et al.[46]

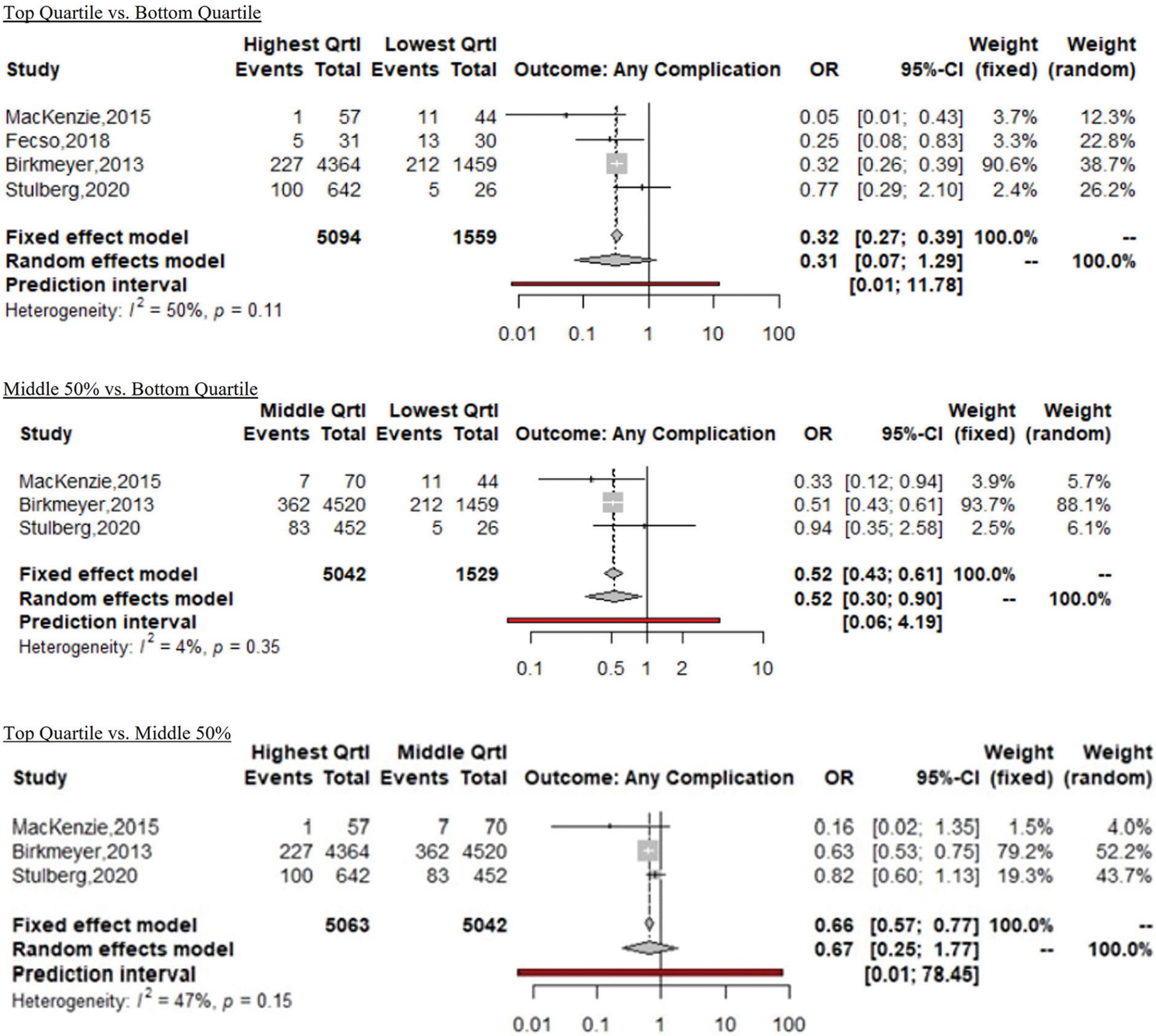

Figure 2. Forest plots for "any complication" as defined by authors, for highest quartile compared to lowest quartile of technical skill and for middle quartile compared to lowest quartile of technical skill.

Measuring Surgeon Performance

It's important to recognize that our study restricted the definition of surgeon technical performance to the most direct assessment of surgical technical skill, as epitomized by the OSATS or GEARS scales. These measure technical skill by direct observation of video of the surgeon's performance, including but not limited to respect for tissue, flow of the operation, time and motion, knowledge of the procedure, knowledge of the instruments, efficiency, bimanual dexterity, etc. The published literature, however, is replete with examples of surgeon technical skill that use different measures of performance. Indirect measures include experience, quality, and outcomes. Common measures of surgeon experience conflate technical skills with experience and include evaluation based on residency,[47,48] years of experience in practice,[49] count (or recent frequency) of surgeries performed,[50,51] or specialty.[49,52] Surgical quality measures include evaluation of a procedure's end result, typically evaluated by review of medical records, operative narrative, or post hoc imaging results.[53–55] Finally, surgeon performance is routinely measured using intraoperative and postoperative outcomes, such as successful completion of procedure components,

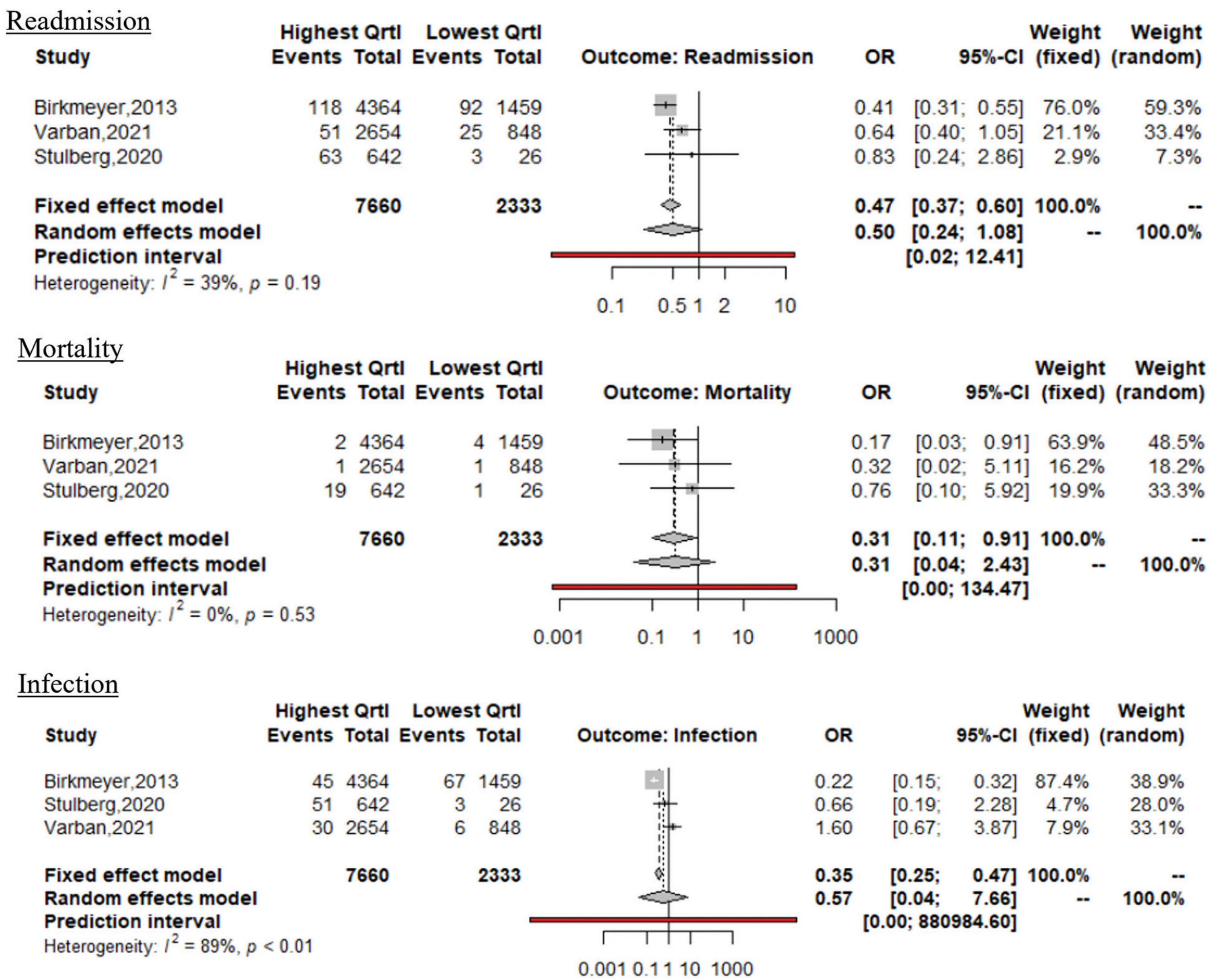

Figure 3. Forest plots of individual clinical outcomes, comparing top and bottom quartile of surgeon technical skill.

operative time, hospital length-of-stay, complications, reoperations, and readmissions.[16,56–58]

This variability complicates the evaluation of peer-reviewed literature relating surgeon performance with peri- and postoperative clinical outcomes. This complexity is illustrated by a recently published systematic review investigating the association between surgeon technical performance and patient outcomes in surgery,[59] which includes articles that measured surgeon technical skill by completion of procedural tasks (e.g. "exploration of Cooper's ligament"),[53,60,61] evaluation of surgical outcomes based on operative reports or post hoc imaging,[62,63] and assignment of surgical errors following medical records review.[64,65] Though these varying concepts of surgeon technical performance are clearly interrelated, they are not equivalent and should be carefully considered when determining the root cause of intra- and postoperative patient outcomes.

Direct, Observational Assessment of Surgical Technical Skills

A significant challenge in comparing surgeon technical skill in relation to outcomes is the variability in measurement of technical skill. In recent years, standardized and validated instruments for measuring surgeon technical skills[66] have been developed. Though not widely adopted in routine surgical practice, these measures have proved valuable for research and quality improvement efforts. Measures of technical skill can be separated into those that measure nonrobotic surgery skills, such as OSATS,

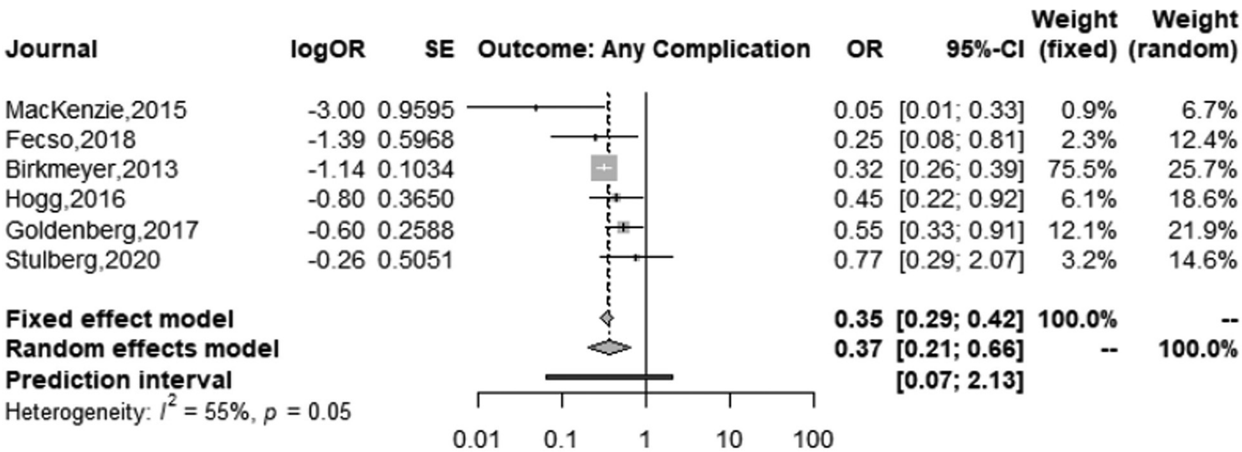

Figure 4. Forest plots of composite "any outcome" measure.

which measures respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants and knowledge of the specific procedure, and those that measure technical skills during robotic surgery, such as GEARS, which measures domains such as depth perception, bimanual dexterity, efficiency, force sensitivity, and robotic control. While standardized and validated, problems with these assessments exist. For example, flow of operation is intended to be a global measure of the procedural flow. However, a significant amount of subjectivity is injected when one must score a procedure where some of the procedure flowed smoothly, and other aspects did not. Additionally, these instruments include items and response options that apply arbitrary anchors, confusing measurement focus, and subjective scoring guidelines that provide little guidance on differentiating scores along the scale continuum.

Video in Training, Certification, and Ongoing Learning

Recent calls-to-action focus on the importance of integrating VBA to evaluate surgeon skill and to support a continual learning model.[28,34,67] Incorporation of these videos into surgical training and quality improvement programs avoids the cost and resourcing challenges of real-time measurement via observation and is of growing importance for surgical training programs.[68] An interesting perspective is provided by Blencowe et al.[69] who reported on a comparison between video and direct observation of several surgical procedures. The evaluations were not associated with outcomes (bariatric surgery) but included interviews with surgeons who agreed that there is significant variability, lack of meaningful standards, etc. in surgical procedures. This perspective is consistent with attempts to incorporate VBA into surgical qualifications. In Japan, for example, the endoscopic surgical skill qualification system includes the scoring of an operative video among other certification criteria.[70]

Routine videorecording of surgeries will be essential for accelerating the training of residents, as well as implementing and supporting ongoing learning models to support surgeons through a career, such as the American Board of Surgery's continuous certification program. Despite the fact that VBA accelerates a trainee's acquisition of skills when compared to standard mentoring approaches,[30,31] there is a poor understanding of available technology, its power, and ease of use.[71] In addition, novel surgical techniques and new surgical devices evolve rapidly, and VBA can be used to ensure safe implementation, track their utilization and associated outcomes, and provide a platform for ongoing surgical skill evaluation related to their use.

To improve surgical outcomes, the American College of Surgeons (ACS) implemented the National Surgical Quality Improvement Program, a voluntary but nationally recognized surgical quality improvement program, measuring clinical quality beginning with intraoperative outcomes and continuing through 30-days postprocedure.[72] Building on this surgical quality reporting and training capability, the ACS is embracing a concept called entrustable professional activities (EPAs).[73] In the surgical context, EPAs represent a way "to translate the broad concepts of competency into everyday practice".[74] VBA is ripe to fulfill this goal for surgical technical competency, as discrete phases of procedures can be scored with objective procedure-specific assessments (OPSAs) focused on safe procedural conduct — competency — a major goal of EPAs. EPAs are not about identifying "exemplar technical skill." The authors' viewpoint is that EPAs should be constructed to define safe vs. unsafe practice, and as such, associated scales are less subjective and more reliable.

In addition to OPSAs, the degree of procedural difficulty should ideally be captured, giving depth to the meaning of a score in the context of EPAs. For example, a resident scoring at a staff level of technical competence for an easy laparoscopic cholecystectomy, may be unsafe in a hard case of laparoscopic cholecystectomy. A global operative difficulty score, similar to the System for Improving and Measuring Procedural Learning[75] construct, needs to be incorporated into VBA.[15] The combination of OPSA/EPAs and global operative difficulty goes beyond accurate assessment, enabling the identification of technical improvement opportunities specific to the phase of a procedure, as well as the creation of rich teaching libraries based upon skill level, case difficulty, and even the phase of the procedure.

Surgical specialties must come to the realization that VBA scores, regardless of the assessment scale, lose resonance in the absence of the patient's associated outcome, at least until there is definitive evidence that a specific score, in fact, results in a consistently best outcome. As this review has shown, there is significant inconsistency in the reporting of complications, how long the patient has been followed, and how the complication is categorized. This is a significant opportunity for improvement, as it is possible to not only standardize the approach to VBA, but also to patient follow-up and outcome measurement.

Our Recommendations

With the collective goal of eliminating variation in surgeon technical skill as a contributing factor to surgical outcomes, the authors propose the following four requirements for defining a comprehensive surgical quality infrastructure:

1. **Risk Factor Assessment and Mitigation.** Routine assessment to identify and mitigate pre-, intra-, and postoperative risk factors for suboptimal surgical outcomes, including but not limited to surgeon technical skill. Ideally, this function would be automated, with real-time, data-driven smart alerts.
2. **Ongoing Surgical Learning.** Transitioning VBA from generic, non-specific assessments of technical skill to objective procedure-specific assessments linked to EPAs. Reducing subjectivity, improving reliability and accuracy, and measuring discrete, critical phases of individual procedures will result in the ability to accelerate resident training based upon data, supports safe implementation of new procedures and technologies, and establish benchmarks for safe procedural conduct.
3. **Routinely Assigning a Case Difficulty Score.** Routine assessment of the global operative difficulty establishes a data-driven methodology upon which to advance residents to higher case difficulties, based on their OPSAs/EPAs scores, by procedural difficulty. It also provides additional context for understanding outcomes based upon standardized case difficulty.
4. **Outcomes Evaluation.** Implement standardized procedures, terminology, and definitions for documenting and collecting surgical outcomes. This will enable the ability to understand the meaning of the OPSA/EPA scores in the context of the patient's outcome, as well as establish comparability across patient and provider populations for the purpose of defining optimal practice. At a minimum, especially for publication, we suggest 100% of the patient population be followed for a minimum of 30-days postoperatively, with all complications being categorized using a procedurally-adapted National Coordinating Council for Medication Error Reporting and Prevention and Clavien-Dindo classifications.

Development and implementation of this recommended system will enable — finally — apples-to-apples comparisons that have data-driven validity for technical skill and outcomes between residents, staff surgeons, departments, specialties, organizations, systems, countries, and continents. This, or a comparable system, is required for driving continuous quality improvement through shared learnings.

LIMITATIONS

Interpretation of these results is hampered by several limitations. First, only seven articles met the eligibility criteria and among those, four had modest sample sizes. This is particularly important because the results are heavily weighted to Birkmeyer,[35] which had the largest sample size, by far. Second, there was substantial variability among a limited number of outcomes measures included in the eligible articles. This, coupled with the relatively short follow-up period, puts a limit on the potential generalizability and value of the association (although highlights the importance of the four requirements of a comprehensive surgical quality infrastructure). One powerful association, noted in MacKenzie et al.[40] and Curtis[76] indicate the potential impact of surgeon performance on long-term outcomes. Though only the MacKenzie article was eligible for inclusion, both articles include lymph node count and resection margins as outcomes. Both measures are significant predictors of cancer recurrence. In other words, not only are near-term outcomes linked to technical skill scores, but early results strongly suggest long term outcomes are too, at least in oncology. Third, the eligible studies are dominated by gastrointestinal procedures, specifically foregut surgery. The contribution of surgeon's technical skill to clinical outcomes, likely varies by surgery type and complexity. Future research should focus on both a broader array of procedure types and a consistent, standardized set of clinical outcomes, both short- and long-term.

CONCLUSIONS

Our systematic literature review and meta-analysis indicates that surgeon technical skill is a significant predictor of clinical outcomes. However, despite the development and validation of numerous scoring instruments to assess surgeon technical skills, there are surprisingly few articles that evaluate the association between skill and outcomes. Within the limited number of articles that do study this association, determining significance is hampered by low sample sizes and lack of consistency in how outcomes and complications were defined. The authors recommend a thoughtful approach for the development of a comprehensive surgical quality infrastructure that could significantly reduce the challenges identified by this study.

References:

1. Ibrahim AM, Ghaferi AA, Thumma JR, Dimick JB. Variation in outcomes at bariatric surgery centers of excellent. *JAMA Surg*. 2017;152(7):629–636.

2. Sheetz KH, Ibrahim AM, Nathan H, Dimick JB. Variation in surgical outcomes across networks of the highest rated US hospitals. *JAMA Surg*. 2019;154(6):510–515.

3. Cullen DJ, Apolone G, Greenfield S, Guadagnoli E, Cleary P. ASA physical status and age predict morbidity after three surgical procedures. *Ann Surg*. 1994;220(1):3–9.

4. Turrentine FE, Wang H, Simpson VB, Jones RS. Surgical risk factors, morbidity, and mortality in elder patients. *J Am Coll Surg*. 2006;203(6):865–877.

5. Finks JF, English WJ, Carlin AM, et al. Predicting risk for venous thromboembolism with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative. *Ann Surg*. 2012; 255(6):1100–1104.

6. Valle JA, Glorioso TJ, Bricker R, et al. Association of coronary anatomical complexity with clinical outcomes after percutaneous or surgical revascularization in the Veterans Affairs clinical assessment reporting and tracking program. *JAMA Cardiol*. 2019;4(8):727–735.

7. Kim W, Song KY, Lee HJ, Han SU, Hyung WJ, Cho GS. The impact of comorbidity on surgical outcomes in laparoscopy-assisted distal gastrectomy: A retrospective analysis of multicenter results. *Ann Surg*. 2008;248(5):793–799.

8. Latkauskas T, Rudinskaite G, Kurtinaitis J, et al. The impact of age on post-operative outcomes of colorectal cancer patients undergoing surgical treatment. *BMC Cancer*. 2005;5:153.

9. Hyer JM, White S, Cloyd J, et al. Can we improve prediction of adverse surgical outcomes? Development of a surgical complexity score using a noval machine learning technique. *J Am Coll Surg*. 2020;230(1):43–52.e1.

10. Lin HS, Watts JN, Peel NM, Hubbard RE. Frailty and post-operative outcomes in older surgical patients: a systematic review. *BMC Geriatr*. 2016;16(1):1–12.

11. Makary MA, Segev DL, Pronovost PJ, Syin D, et al. Frailty as a predictor of surgical outcomes in older patients. *J Am Coll Surg*. 2010;210(6):901–908.

12. Nathens AB, Jurkovich GJ, Maier RV, et al. Relationship between trauma center volume and outcomes. *JAMA*. 2001; 285(9):1164–1171.

13. Markar SR, Penna M, Karthikesalingam A, Hashemi M. The impact of hospital and surgeon volume on clinical outcome following bariatric surgery. *Obes Surg*. 2012;22(7): 1126–1134.

14. Schrag D, Panageas KS, Riedel E, et al. Surgeon volume compared to hospital volume as a predictor of outcome following primary colon cancer resection. *J Surg Oncol*. 2003;83(2):68–78.

15. Dimick JB, Cowan JA Jr, Upchurch GR Jr, Colletti LM. Hospital volume and surgical outcomes for elderly patients with colorectal cancer in the United States. *J Surg Res*. 2003;114(1):50–56.

16. Reames BN, Bacal D, Krell RW, Birkmeyer JD, Birkmeyer NJO, Finks JF. Influence of median surgeon operative duration on adverse outcomes in bariatric surgery. *Surg Obes Relat Dis*. 2015;11(1):207–213.

17. Harrison T, Robinson P, Cook A, Parker MJ. Factors affecting the incidence of deep wound infection after hip fracture surgery. *J Bone Joint Surg Br*. 2012;94(2):237–240.

18. Aletti GD, Dowdy SC, Podratz KC, Cilby WA. Relationship among surgical complexity, short-term morbidity, and overall

survival in primary surgery for advanced ovarian cancer. *Am J Obstet Gynecol*. 2007;197(6):676.e1–e1.

19. Paruch JL, Merkow RP, Bentrem DJ, et al. Impact of hepatectomy surgical complexity on outcomes and hospital quality rankings. *Ann Surg Oncol*. 2014;21(6):1773–1780.

20. Mavros MN, Bohnen JD, Ramly EP, Velmahos GC, et al. Intraoperative adverse events: risk adjustment for procedure complexity and presence of adhesions is crucial. *J Am Coll Surg*. 2015;221(2):345–353.

21. Finks JF, Kole KL, Yenumula PR, et al. Predicting risk of serious complications with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative. *Ann Surg*. 2011; 254(4):633–640.

22. Schmidt CM, Turrini O, Parikh P, et al. Effect of hospital volume, surgeon experience, and surgeon volume on patient outcomes after pancreaticoduodenectomy: a single-institution experience. *Arch Surg*. 2010;145(7):634–640.

23. Trinh QD, Bjartell A, Freedland SJ, et al. A systematic review of the volume-outcome relationship for radical prostatectomy. *Eur Urol*. 2013;64(5):786–798.

24. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273–278.

25. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1):107–113.

26. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol*. 2012;187(1):247–252.

27. de Montbrun SL, Roberts PL, Lowry AC, et al. A novel approach to assessing technical competence of colorectal surgery residents: the development and evaluation of the Colorectal Objective Structured Assessment of Technical Skill (COSATS). *Ann Surg*. 2013;258(6):1001–1006.

28. Feldman LS, Pryor AD, Gardner AK, et al. SAGES video-based assessment (VBA) program: a vision for life-long learning for surgeons. *Surg Endosc*. 2020;34(8):3285–3288.

29. Ritter EM, Gardner AK, Dunkin BJ, Schultz L, Pryor AD, Feldman L. Video-based assessment for laparascopic fundoplication: initial development of a robust tool for operative performance assessment. *Surg Endosc*. 2020;34(7):3176–3183.

30. Augestad KM, Butt K, Ignjatovic D, Keller DS, Kiran R. Video-based coaching in surgical education: a systematic review and meta-analysis. *Surg Endosc*. 2020;34(2):521–535.

31. Soucisse ML, Boulva K, Sideris L, Drolet P, Morin M, Dubé P. Video coaching as an efficient teaching method for surgical residents – a randomized controlled trial. *J Surg Educ*. 2017; 74(2):365–371.

32. Greenberg CC, Dombrowski J, Dimick JB. Video-based surgical coaching: an emerging approach to performance improvement. *JAMA Surg*. 2016;151(3):282–283.

33. Grenda TR, Pradarelli JC, Dimick JB. Using surgical video to improve technique and skill. *Ann Surg*. 2016;264(1):32–33.206

34. Prebay ZJ, Peabody JO, Miller DC, Ghani KR. Video review for measuring and improving skill in urological surgery. *Nat Rev Urol*. 2019;16(4):261–267.

35. 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.

36. Stulberg JJ, Huang R, Kreutzer L, et al. Association between surgeon technical skills and patient outcomes. *JAMA Surg*. 2020; 155(10):960–968.

37. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Open Med*. 2009;3(3):e123–e130.

38. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*. 2000;283(15):2008–2012.

39. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol*. 2010;25(9):603–605.

40. Mackenzie H, Ni M, Miskovic D, et al. Clinical validity of consultant technical skills assessment in the English National Training Programme for Laparoscopic Colorectal Surgery. *Br J Surg*. 2015;102(8):991–997.

41. Varban OA, Thumma JR, Finks JF, Carlin AM, Ghaferi AA, Dimick JB. Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy: a video-based study. *Ann Surg*. 2021;273(4):766–771.

42. Fecso AB, Bhatti JA, Stotland PK, Quereshy FA, Grantcharov TP. Technical performance as a predictor of clinical outcomes in laparoscopic gastric cancer surgery. *Ann Surg*. 2019;270(1): 115–120.

43. Goldenberg MG, Goldenberg L, Grantcharov TP. Surgeon performance predicts early continence after robot-assisted radical prostatectomy. *J Endourol*. 2017;31(9):858–863.

44. Hogg ME, Zenati M, Novak S, et al. Grading of surgeon technical performance predicts postoperative pancreatic fistula for pancreaticoduodenectomy independent of patient-related variables. *Ann Surg*. 2016;264(3):482–491.

45. Fecso AB, Kuzulugil SS, Babaoglu C, Bener AB, Grantcharov TP. Relationship between intraoperative non-technical performance and technical events in bariatric surgery. *Br J Surg.* 2018; 105(8):1044–1050.

46. Balvardi S, Kammili A, Hanson M, et al. The association between video-based assessment of intraoperative technical performance and patient outcomes: a system review. *Surg Endosc.* 2022;36(11):7938–7948.

47. Van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med.* 2012;10(1):65–11.

48. Hope WW, Hooks IIW, Kilbourne SN, Adams A, Kotwall CA, Clancy TV. Assessing resident performance and training of colonoscopy in a general surgery training program. *Surg Endosc.* 2013;27(5):1706–1710.

49. Bilimoria KY, Phillips JD, Rock CE, Hayman A, Prystowsky JB, Bentrem DJ. Effect of surgeon training, specialization, and experience on outcomes for cancer surgery: a systematic review of the literature. *Ann Surg Oncol.* 2009;16(7):1799–1808.

50. Nathan H, Cameron JL, Choti MA, Schulick RD, Pawlik TM. The volume-outcomes effect in hepato-pancreato-biliary surgery: Hospital versus surgeon contributions and specificity of the relationship. *J Am Coll Surg.* 2009;208(4):528–538.

51. Bolling SF, Li S, O'Brien SM, Brennan JM, Prager RL, Gammie JS. Predictors of mitral valve repair: clinical and surgeon factors. *Ann Thorac Surg.* 2010;90(6):1904–1911; discussion 1912.

52. Sahni NR, Dalton M, Cutler DM, Birkmeyer JD, Chandra A. Surgeon specialization and operative mortality in United States: retrospective analysis. *BMJ.* 2016;354:i3571.

53. Bacha EA, Larrazabal LA, Pigula FA, et al. Measurement of technical performance in surgery for congenital heart disease: the stage I Norwood procedure. *J Thorac Cardiovasc Surg.* 2008;136(4):993–997.

54. Nathan M, Karamichalis JM, Liu H, et al. Intraoperative adverse events can be compensated by technical performance in neonates and infants after cardiac surgery: a prospective study. *J Thorac Cardiovasc Surg.* 2011;142(5):1098–1107.

55. Ellis SG, Burke MN, Murad MB, et al. Predictors of successfully hybrid-approach chronic total coronary artery occlusion stenting: an improved model with novel correlates. *JACC Cardiovasc Interv.* 2017;10(11):1089–1098.

56. Nathan M, Karamichalis JM, Liu H, et al. Surgical technical performance scores are predictors of late mortality and unplanned reinterventions in infants after cardiac surgery. *J Thorac Cardiovasc Surg.* 2012;144(5):1095–1101.e7.

57. Driessen SR, Van Zwet EW, Haazebroek P, et al. A dynamic quality assessment tool for laparoscopic hysterectomy to measure surgical outcomes. *Am J Obstet Gynecol.* 2016;215(6):754.e1–754.e8.

58. Shuhaiber J, Gauvreau K, Thiagarjan R, et al. Congenital heart surgeon's technical proficiency affects neonatal hospital survival. *J Thorac Cardiovasc Surg.* 2012;144(5):1119–1124.

59. Fecso AB, Szasz P, Kerezov G, Grantcharov TP. The effect of technical performance on patient outcomes in surgery: a systematic review. *Ann Surg.* 2017;265(3):492–501.

60. Arvidsson D, Berndsen FH, Larsson LG, et al. Randomized clinical trial comparing 5-year recurrence rate after laparoscopic versus Shouldice repair of primary inginual hernia. *Br J Surg.* 2005;92(9):1085–1091.

61. Karamichalis JM, Thiagarajan RR, Liu H, Mamic P, Gauvreau K, Bacha EA. Stage I Norwood: optimal technical performance improves outcomes irrespective of preoperative physiologic status or case complexity. *J Thorac Cardiovasc Surg.* 2010; 139(4):962–968.

62. Docquier PL, Manche E, Autrique JC, Geulette B. Complications associated with gamma nailing. A review of 439 cases. *Acta Orthop Belg.* 2002;68(3):251–257.

63. Frank RM, McGill KC, Cole BJ, et al. An institution-specific analysis of ACL reconstruction failure. *J Knee Surg.* 2012; 25(2):143–149.

64. Rogers SO Jr, Gawande AA, Kwaan M, et al. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery.* 2006;140(1):25–33.

65. Somville FJ, van Sprundel M, Somville J. Analysis of surgical errors in malpractice claims in Belgium. *Acta Chir Belg.* 2010; 110(1):11–18.

66. McMullan RD, Urwin R, Sunderland N, Westbrook J. Observational tools that quantify nontechnical skills in the operating room: a systematic review. *J Surg Res.* 2020;247: 306–322.

67. Van de Graaf FW, Lange MM, Spakman JI, et al. Comparison of systematic video documentation with narrative operative report in colorectal cancer surgery. *JAMA Surg.* 2019;154(5):381–389.

68. Maertens H, Aggarwal R, Moreels N, Vermassen F, Van Herzeele I. A proficiency based stepwise endovascular curricular training (PROSPECT) program enhances operative performance in real life: a randomised controlled trial. *Eur J Vasc Endovasc Surg.* 2017;54(3):387–396.

69. Blencowe NS, Blazeby JM, Donovan JL, Mills N. Novel ways to explore surgical interventions in randomised controlled trials: applying case study methodology in the operating theatre. *Trials.* 2015;16:589.

70. Tanigawa N, Lee SW, Kimura T, et al. The Endoscopic Surgical Skill Qualification System for gastric surgery in Japan. *Asian J Endosc Surg.* 2011;4(3):112–115.

71. Esposito AC, Yoo PS, Lipman JM. Video coaching: a national survey of surgical residency program directors [published online ahead of print, 2021 Dec 21]. *J Surg Educ*. 2021 (21):S1931–7204.

72. Cohen ME, Liu Y, Ko CY, Hall BL. Improved surgical outcomes for ACS NSQIP hospitals over time. *Ann Surg*. 2016; 263(2):267–273.

73. Greenberg JA, Minter RM. Entrustable professional activities: the future of competency-based education in surgery May already be here. *Ann Surg*. 2019;269(3):407–408.

74. Lindeman B, Petrusa E, Phitayakorn R. Entrustable professional activities (EPAs) and applications to surgical training. Available at: https://www.facs.org/education/division-of-education/publications/rise/articles/entrustable. Accessed January 1, 2022.

75. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment SIMPL (System for Improving and Measuring Procedural Learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016;73(6):e118–e130.

76. Curtis NJ, Foster JD, Miskovic D, et al. Association of surgical skill assessment with clinical outcomes in cancer surgery. *JAMA Surg*. 2020;155(7):590–598.

Appendix

Appendix Table 1. MeSH Search Tearms for use in Meta-Analysis of Surgeon Skill and Clinical Outcomes

TOPIC	CONVENTIONAL TERMS	MeSH TERMS	APPLICABLE SUBHEADINGS	BUILDER TERMS
LEVEL OF SKILL	SURGEON SKILL OR LEVEL OF EXPERIENCE	PROFESSIONAL COMPETENCE	N/A	"Professional Competence"[Mesh]
	SURGEON SKILL OR LEVEL OF EXPERIENCE	CLINICAL COMPETENCE	N/A	"Clinical Competence"[Mesh]
CLINICAL OUTCOMES	CLINICAL OUTCOMES	TREATMENT OUTCOME	N/A	"Treatment Outcome"[Mesh]
SURGERY	SURGERY	SURGICAL PROCEDURES, OPERATIVE	Adverse Effects, Complications, Surgery, Mortality	("Surgical Procedures, Operative/adverse effects"[Mesh] OR "Surgical Procedures, Operative/complications"[Mesh] OR "Surgical Procedures, Operative/mortality"[Mesh] OR "Surgical Procedures, Operative/surgery"[Mesh])
SURGICAL TECHNICAL QUALITY	SURGICAL QUALITY OR SURGERY QUALITY	N/A	N/A	Objective Structured Assessment of Technical Skills Global Evaluative Assessment of Robotic Global Operative Assessment of Laparoscopic Skills Arthroscopic Surgical Skill Evaluation Tool Crowd-Sourced Assessment of Technical Skills Colorectal Objective Structured Assessment of Technical Skills General Surgery Objective Structured Assessment of Technical Skill

Appendix Table 2. Criteria for the Newcastle-Ottawa Scale regarding star allocation to assess quality of studies

Criteria	Acceptable (star awarded)		Unacceptable (star not awarded)
Cohort Studies†			
Representativeness of exposed cohort	Representative group of participating surgeons		Self-selected, volunteer sample of surgeons
Selection of non-exposed cohort	Same setting and source population as exposed cohort		Different setting from exposed cohort
Ascertainment of exposure	Operative videos evaluated by more than one independent rater		Operative videos evaluated by one rater only or raters with knowledge of surgeon identity.
Comparability	Adjusted for patient-level risk of poor outcomes.		No reported patient-level risk adjustment
	Adjusted for clustering of outcomes from multiple patients per surgeon.		No reported adjustment for clustering of patients by surgeon.
Assessment of outcomes	Secure records or independent assessment		Self-reported, not reported
Adequacy of follow up	Complete follow up or missing follow-up outcomes unlikely to introduce bias.		No statement regarding missing data or completeness of follow up
Case-Control Studies			
Case definition adequate	Case status confirmed by independent validation		Case status defined by self-report
Case representativeness	Consecutive or obviously representative case series		Potential for selection biases
Selection of controls	Controls derived from same population		Selected form independent population
Definition of controls	Same inclusion criteria related to previous or history of outcome(s) of interest		Different inclusion/exclusion criteria related to history of/prior events of outcome(s) of interest
Comparability of cases & controls	Study controls for risk factors for peri- and post-operative outcomes		No adjustment for risk factors.
Ascertainment of exposure	Operative videos evaluated for each procedure		Surgeon self-selected representative video
Same method of ascertainment for cases and controls	Yes		No
Non-response rate	Same rate for both groups		

† "Demonstration that outcome of interest was not present at start of study" was excluded as an assessment criterion.

Appendix Table 3a. Quality assessment of studies using a modified Newcastle-Ottawa Quality Assessment scale* for cohort studies.

Study	Selection				Comparability	Exposure			
	Adequate case definition	Representativeness of cases	Selection of controls	Definition of controls		Ascertainment of Exposure	Same method of ascertainment for cases and controls	Non-response rate	Total (8 *)
Goldenberg (2017)	*	*	*	*	*	*	*	*	******** (8)
Hogg (2016)	*	*	*	*	*	*	*	*	******** (8)
Fecso (2019)	*	*	*	*	*	*	*	*	******** (8)

Appendix Table 3b. Quality assessment of studies using a modified Newscastle-Ottawa Quality Assessment scale for case-control Studies.

Study	Selection			Comparability	Outcome		
	Representativeness of exposed cohort	Selection of non-exposed cohort	Ascertainment of exposure		Assessment of outcomes	Adequacy of follow up	Total (7 *)
Birkmeyer (2013)	-	*	*	* *	*	*	****** (6)
Mackenzie (2015)	*	*	*	* -	*	*	****** (6)
Stulberg (2020)	-	*	*	* *	*	*	****** (6)
Varban (2021)	-	*	*	* *	*	*	****** (6)