



HHS Public Access

Author manuscript

Ann Surg. Author manuscript; available in PMC 2023 October 01.

Published in final edited form as:

Ann Surg. 2022 October 01; 276(4): 701–710. doi:10.1097/SLA.0000000000005595.

Do Individual Surgeon Preferences Affect Procedural Outcomes?

Hossein Mohamadipanah, PhD,

Department of Surgery, Stanford University;

Calvin A. Perumalla, MS, PhD,

Department of Surgery, Stanford University;

LaDonna E. Kears, MD,

Department of Surgery, Stanford University;

Su Yang, BS,

Department of Surgery, Stanford University;

Brett J. Wise, BS,

Department of Surgery, Stanford University;

Cassidi K. Goll, BA,

Department of Surgery, Stanford University;

Anna K. Witt, BS,

Department of Surgery, Stanford University;

James R. Korndorffer, MD, MHPE,

Department of Surgery, Stanford University;

Carla M. Pugh, MD, PhD

Department of Surgery, Stanford University

Structured Abstract:

Objectives: Surgeon preferences such as instrument and suture selection and idiosyncratic approaches to individual procedure steps have been largely viewed as minor differences in the surgical workflow. We hypothesized that idiosyncratic approaches could be quantified and shown to have measurable effects on procedural outcomes.

Methods: At the ACS Clinical Congress, experienced surgeons volunteered to wear motion tracking sensors and be videotaped while evaluating a loop of porcine intestines to identify and repair two pre-configured, standardized enterotomies. Video annotation was used to identify individual surgeon preferences and motion data was used to quantify surgical actions. Chi-square analysis was used to determine whether surgical preferences were associated with procedure outcomes (bowel leak).

Results: Surgeons' (N = 255) preferences were categorized into four technical decisions. Three out of the four technical decisions (repaired injuries together, double layer closure, corner-stitches versus no corner-stitches) played a significant role in outcomes, $p < 0.05$. Running versus interrupted did not affect outcomes. Motion analysis revealed significant differences in average operative times (leak-6.67 min vs. no leak-8.88 min, $p = 0.0004$) and work effort (leak-path length=36.86 cm vs. no leak-path length=49.99 cm, $p = 0.001$). Surgeons who took the riskiest path but did not leak had better bimanual dexterity (leak=0.21/1.0 vs. no leak=0.33/1.0, $p = 0.047$) and placed more sutures during the repair (leak=4.69 sutures vs. no leak=6.09 sutures, $p = 0.03$).

Conclusion: Our results show that individual preferences affect technical decisions and play a significant role in procedural outcomes. Future analysis in more complex procedures may make major contributions to our understanding of contributors to procedure outcomes.

Mini Abstract:

Surgeon preferences such as instrument and suture selection and idiosyncratic approaches to individual procedure steps have been largely viewed as minor differences in the surgical workflow. However, when quantifying these differences using video and sensor data with experienced surgeons, we found that individual preferences affect technical decisions and play a significant role in procedural outcomes.

Introduction:

The surgical workflow is often the primary focus of many process improvement measures in surgery. Researchers and quality improvement leaders have evaluated surgical teams, operative procedure times as well as operating room turnover times.^{1,2} In addition, numerous research studies have reproducibly shown that technical skills have a significant impact on patient outcomes in the clinical environment.^{3,4} For years it has been presumed that mastery of technical skills by practicing surgeons is one of the most obvious, modifiable targets for improving workflow efficiency. In addition, it is also accepted that non-technical skills such as surgical decision making also play a major role in the surgical workflow however there is a paucity of research linking surgical decisions and outcomes. One of the reasons for this is that cognitive drivers for surgical actions and decision-making remain elusive and difficult to quantify.⁵ In the quest to achieve holistically accurate surgical performance and outcome metrics, several approaches have been explored including the use of standardized, global, and procedure-specific performance assessments combined with video-based assessments, and peer-to-peer review.

For video-based assessment, one study used videos of surgeons performing a laparoscopic gastric bypass. The videos were reviewed by surgeon peers and were rated using two performance assessment tools: 1) Objective Structured Assessment of Technical Skills [OSATS], and 2) Global Evaluative Assessment of Robotic Skills [GEARS].³ The scores were shown to be correlated with patient outcomes including risk-adjusted complication rates, as well as patient readmission, and reoperation rates. On the other end of the spectrum for technical skills measurement, the emergence of sensor technology has enabled tracking and quantification of surgical actions such as economy of motion and dexterity. Initial validity evidence for this type of performance data was shown over twenty years ago in a

publication from Imperial College which found that sensor-based metrics can discriminate between novice and experienced surgeons.^{6,7} Today more researchers are exploring this type of technology and this field of work has aided in quantifying surgical technical skills in new ways.⁸⁻¹⁰

In addition to assessment of technical skills, the role of intra-operative judgment and decision making and its effect on patient outcomes has recently been acknowledged as a critical element to consider with evaluating patient outcomes and the surgical workflow.¹¹⁻¹³ Another element that contributes to the surgical workflow includes personal or idiosyncratic approaches. These approaches stem from alliance or blind adoption of the approach of the surgical mentor.¹⁴ These trained, handed-down approaches become protocolized, automated, and subsequently handed down to the next generation of trainees.

Previous studies have stratified and characterized the different types of intraoperative decision-making strategies but data continues to lag regarding the effect of these strategies on patient outcomes.¹⁵ In a study by Hashimoto et al., procedural maps were developed for a laparoscopic cholecystectomy based on residents' and surgeons' perceptions of decision points, procedural steps, and other related factors. During their analysis, the maps were shown to have the ability to discriminate between novice and expert decision-making.¹⁶ Simulation-based studies have also explored the relationship between intra-operative decision-making and procedural outcomes.¹⁷

Despite this early work exploring surgical workflow and outcomes, surgeon preferences have not been previously suggested or known to have a considerable impact on procedural outcomes. As such, efforts to understand the effect of preferences and idiosyncratic approaches on procedural or patient outcomes are limited. We hypothesize that surgeons' idiosyncratic approaches can be quantified and shown to have a measurable effect on procedural outcomes.

Methods:

Study Design

Settings and Participants—The study was conducted over a 3-day period in conjunction with the 2019 American College of Surgeons (ACS) Clinical Congress. A convenience sample of practicing and retired surgeons attending the 2019 ACS Clinical Congress in San Francisco, CA were recruited via word of mouth and formal advertisement through ACS marketing channels, including an ACS bulletin publication¹⁸ in the months preceding the event. The study took place in the exhibit hall, where 10 simulated operating stations were set up in a 1,500 sq. ft. exhibit booth. This study was approved by the Stanford University Institutional Review Board.

Research Protocol—This simulation-based study was conducted using a cross-sectional observation study design. Figure 1 illustrates the major steps and components included in our data collection procedure. When participants arrived at the ACS exhibit, they gave informed consent to participate in the study and completed a short, written survey. For practicing surgeons, survey data included age, gender, handedness, surgical specialty,

years in practice, level of clinical workload, percent of time operating, administrative responsibilities, and level of experience in repairing enterotomies. For retired surgeons, survey data included age, gender, handedness, surgical specialty, years in practice, administrative responsibilities, level of experience in repairing enterotomies, years retired from surgery, and retirement activities. Subsequently, participants were brought to a simulated operating room table (Figure 1a) and were fitted with electromagnetic motion tracking sensors on each hand (Figure 1b). The research team had prepared customized lab coats by sewing low-profile, flexible braided sleeving and Velcro loops onto the coat to keep the sensor wires from interfering with surgical movement. Three electromagnetic sensors were placed on each hand - one on the second phalanx of the forefinger, one on the first phalanx of the thumb, and one on the inner wrist. There was a total of six sensors per participant. The sensors were secured with 3M Transpore tape and standard, sterile gloves were placed over the sensing components. Once surgeons were outfitted with sensors, they were situated in front of the enterotomy repair simulation (Figure 1c). A researcher, acting as an operative assistant, read each participant an introductory narrative to explain the surgical task. Participants were told that they had just finished a difficult two-hour lysis of adhesions and that their objective was to run the bowel (Figure 1d) and repair any enterotomies they discovered (Figure 1e). The goal was to progress in the procedure as they would in a real-life scenario. They were allotted 15-minutes for the task.

Two cameras were placed at each OR station in a standardized fashion to collect video recordings of the encounter. One camera captured an overhead view and the other captured a frontal view of the operative field. As surgeons progressed through the procedure, they could rely on their operative assistants to hand them tools or hold down parts of the bowel. The operative assistants were considered to be at the level of a medical student and were situated directly behind the frontal view camera. Operative assistants asked surgeons to put their tools down when the 15-minute timer went off and those who completed their repair before the timer went off were asked to verbalize when they were finished. After participants completed their repair, researchers took pictures from a top-down view using a template to achieve a standardized database of photos for offline review.

Materials—The enterotomy repair simulation consisted of a segment of porcine bowel pre-configured with two standardized enterotomies. The large enterotomy was a 1cm long defect in the bowel created by removing a portion of bowel near the antimesenteric border of the porcine intestines, Figure 1d. To produce the small injury, the bowel was punctured using surgical scissors at a distance of .5cm away from the larger injury. This process was repeated for every piece of bowel used. The pre-injured segment of porcine bowel was placed on top of a 15" × 9" tray that contained imagery of bowel to mimic the anatomy of the abdominal cavity. The porcine bowel was oriented the same way for each participant and covered in 15 cc of artificial blood. In addition to the bowel tray (simulated abdomen), participants were presented with a 13" × 17" surgical instrument tray with needle holders, forceps, scissors, scalpels, mosquito forceps, and an assortment of sutures. Each tool tray was equipped with tracings of each instrument such that tools would be presented in the same fashion for each participant.

Performance Evaluation—Participants' enterotomy repairs were evaluated with a standardized leak test (Figure 1f). Researchers performed the leak test by tying off one end of the bowel segment and inserting a tube into the other end. The tube was secured tightly with a zip tie and liquid was perfused into the bowel using a motorized fluid pump with standardized pressure. Researchers took note of any leaks in the bowel repair and saved this information into a secure database. In addition, the leak test was also captured on surgical video from both cameras. After the leak test, the operative field was then cleaned and reset for the next participant.

Video Analysis - Recording Technical Decisions—For each participant that fit the inclusion criteria, the videos that captured the procedure were viewed and the technical decisions that were made were noted and entered into a codebook. Leak test results were also entered into the research codebook which was stored in Red-Cap®.

Video coding strategies were protocolized. Video reviewers were trained that when a participant made a “technical decision” it meant that he or she chose one of two mutually exclusive operative strategies. For example, one of the operative strategies that the participants made was whether to repair the bowel by joining the two injuries together or by repairing them separately. Choosing one operative strategy automatically prohibits the participant from choosing the other. The effect of this technical decision on outcomes was observed by calculating the percentage of participants who had unsuccessful outcomes (bowel leak) in both groups thus corresponding to the two mutually exclusive strategies. From this data, we calculated the leak rate. Chi-Square analysis was used to compare leak rates for the different operative strategies at each of the major technical decision points.

To fully understand the impact of each technical decision, participants were first stratified according to the specific operative strategy chosen that resulted in the best possible outcome or lowest leak rate. Participants who chose the operative strategy associated with the highest leak rate were said to have chosen a riskier pathway since a higher proportion of participants in their group led to an unsuccessful outcome. This group was further stratified by additional operative strategies that resulted in successful outcomes for the participants in that group i.e., low leak rate. This analysis process was repeated multiple times until no more technical decisions could be used to stratify the participant group.

Data Analysis - Inclusion Criteria—Participants who had missing or corrupted data (survey, video, motion) were not included in the analysis. Resident participants were also excluded from our analysis as our primary goal was to focus on the technical skills and decision-making of the practicing physician. In addition, we excluded participants who leaked because of a missed enterotomy as our goal was to focus on the technical and cognitive decisions of those who found both enterotomies and repaired them. Additionally, the analyzed dataset only contained participants who were at the level of an attending surgeon.

Data Analysis - Strategy and Statistical Methods—Our goal was to quantify and characterize the technical skills and decisions of experienced surgeons. The wearable technology and video data allowed for several data points to achieve this goal. In our

prior work evaluating practicing clinicians performing simulated breast examinations on sensor-enabled mannequins, we found that participant stratification based on procedure outcomes (as opposed to years of experience) provided the most useful analysis when categorizing procedural skills.¹⁹ As such, our goal for this research study was to first define the successful outcome of an enterotomy repair. For this study, this was determined by implementing a leak test. All participants were stratified based on whether their repair leaked or did not leak instead of number of years of experience. Prior work using sensor metrics largely focused on characterizing the difference in technical skills metrics (dexterity, velocity, etc.) when comparing students and experienced surgeons. This approach assumes that a greater number of years in practice correlates with superior technical skills. Our approach focuses on the technical skills of those who achieve the best outcomes.

Once we identified those who leaked and did not leak, we then analyzed the video and motion data with the goal of characterizing technical skills and decisions that could be associated with outcomes.

Motion Analysis: Quantifying Technical Dexterity—Motion data for each participant was also analyzed. For each participant, time-series data of the X, Y, and Z motion positions were analyzed. Motion metrics including idle time, path length, bimanual dexterity, and working volume were calculated. These metrics are known to be associated with surgical skill and dexterity.⁹ For each operative strategy, the motion metrics were generated during the suturing phase of the procedure. Motion during the time participants were running the bowel was out of the scope of this study. The suture phase was annotated as the start of when a participant first touched the bowel tissue with a suture to when they are observed to finish their last suture by either cutting the suture tail (running technique) or not throwing more sutures and then cutting the last suture tail (interrupted). Within each operative strategy, participants were grouped as either having leaked or not leaked. A comparison to find any significant differences in motion metrics between these two groups within each operative strategy and technical decision was evaluated using the Kruskal-Wallis test or one-way ANOVA depending on whether the data followed normal distribution which is determined by a Kolmogorov-Smirnov test.

Additionally, motion metrics for suture actions only (*excluding knot tying and other actions between suture placements*) were generated. The suture action time frame is annotated with the start being when a participant touches the suture to the tissue and ends when they pull the suture out of the tissue. A participant can have multiple suture action time frames which is equal to the number of suture throws they have. For each of these suture action time frames, motion metrics were generated and then averaged over the entire number of suture throws that occurred during the repair. With these average motion metrics for each of the three final pathways, operative strategies for the interrupted approach, participants would be grouped within their own operative strategy as either having leaked or not leaked. A comparison to find any significant differences in the average motion metrics between these two groups within each operative strategy was evaluated using the Kruskal-Wallis test or one-way ANOVA depending on if the samples came from a standard normal distribution which is determined by a Kolmogorov-Smirnov test

Cognitive Decision Analysis—The number of suture throws was also annotated for each operative strategy. Within each of these operative strategies, the leak and no leak subgroup was compared to see if there were any significant differences in the number of suture throws using a chi-square test. Additionally, for each operative strategy, the correlation between their motion metrics during the suture phase and the number of suture throws was analyzed using a spearman, Pearson, and Kendall test.

Results:

Participant Demographics:

Demographics for the surgeon participants (N=255) included attending surgeons (n = 201), resident physicians (n = 40), and retired surgeons (n = 14), Table 1. Regarding gender statistics, participants identified as male (n = 179), female (n = 65), or unidentified (n = 11). The age range from 41–50 years had the largest number of participants (n = 62). When evaluating number of years in practice, the highest number of participants had 0–10 years of experience (n = 65) although a significant number of participants had 11–20 years in practice (n = 60).

Video Analysis:

Video analysis focused on surgeons who repaired both enterotomies. A total of 107 surgeons were excluded from the analysis based on the following criteria: retired surgeons (n = 14); missed an enterotomy (n = 39) or had incomplete data (n = 54). When analyzing leak rate for the participants we found that a total of 120 (81%) of the 148 participants, who repaired both injuries, did not leak.

Video analysis revealed four major technical decisions, Figure 2. For Technical Decision 1: participants decided to either repair the large and small enterotomies separately or as a single injury by making an incision between the two injuries. For Technical Decision 2: participants decided to repair the large injury with a double layer or a single layer closure. For Technical Decision 3: participants used an interrupted suture repair or a running suture repair. For Technical Decision 4: Participants either placed a corner stitch while repairing the injury or did not place a corner stitch. Three out of the four technical decisions resulted in significant differences in leak rate.

Motion Analysis:

Motion analysis results were analyzed in two parts: 1) the entire enterotomy repair phase for all participants who repaired both enterotomies (n=148) and 2) a focused sub-analysis of the suturing phase for those who performed an interrupted repair on the riskiest pathway during the third technical decision (n = 38). The entire enterotomy repair phase accounted for all motion that occurred after a participant first picks up a suture to begin the enterotomy repair and ends when the repair is complete, and they no longer have an instrument in their hand. This analysis allowed for an in-depth comparison of the motor actions executed for the four technical decisions. Our analysis included *idle time*, *bimanual dexterity*, *path length*, *duration*, and *working volume*, Table 2. The results showed that surgeons who executed

more actions (i.e. long path length) and took more time (i.e. longer duration) were less likely to leak.

To further investigate the meaning of these results, we looked at the data to see if there was a correlation between path length and number of sutures placed during the enterotomy repair as well as duration and the number of sutures placed. After reviewing the motion data for all participants who did not miss an enterotomy (n=148), our results showed that the number of suture throws and path length was positively correlated, $r(146)=0.54, p<0.01$. Additionally, the duration of the enterotomy repair and the number of suture throws were also shown to be positively correlated, $r(146)=0.59, p<0.01$. This correlation pattern is consistent for each of the technical decision steps.

Additional analysis showed that participants who followed the riskiest pathway and had the strongest correlation between *path length* and *number of suture throws* were those who followed the decisions leading to using *no corner stitches* but did not leak (n=22) $r(21)=0.67, p<0.0$, Figure 2. Indicating that those who leaked may have had shorter path lengths and placed less suture during the repair. Participants who had the strongest correlation between *number of suture throws* and *duration* were those who did use *corner stitches* and did not leak (n=16), $r(15)=0.78, p<0.01$, Figure 2. Indicating that those in this pathway who leaked may have spent less time suturing and also placed less suture during the repair.

The enterotomy repair phase analysis investigated motion only during the times in which a participant was actively passing a suture through the bowel. Motion metrics from the previous phase were also looked at with the inclusion of the number of suture throws, Table 3. When looking at the results, the two metrics that came to be significant were the number of suture throws and bimanual dexterity.

When independently analyzing the number of suture throws for those participants who did interrupted suturing, the results showed that participants who placed more sutures in the bowel had a lower leak rate. [Riskiest Path leak=4.69 sutures vs. no leak=6.09 sutures, $p=0.03$]. Analyzing the number of suture throws provided major insights into the cognitive and technical decisions of the participants. Additionally, it was found that bimanual dexterity was significantly higher for those that did not leak compared to those who did however this was true only for the riskiest pathway leading to the *no corner stitch* operative strategy (leak=0.21/1.0 vs. no leak=0.33/1.0, $p=0.047$). Figure 3 shows representative motion plots for the right and left hand of a participant who leaked compared to one who did not leak. Note that the left hand of the surgeon who did not leak has a longer path length and larger working volume, indicating greater self-assistance or bimanual dexterity.

Discussion:

This study evaluated the effects of idiosyncratic, surgeon preferences on technical decision-making and procedural outcomes. The results show that well-known surgeon preferences such as single or double-layer closure or use of corner stitches, significantly affected procedure outcomes. These results support a paradigm shift in how we think about surgical

workflows and procedure outcomes. Previous studies on surgical outcomes largely focused on surgeons' technical skills. Birkmeyer et al., noted that surgeons who were rated in the lowest tier of technical skills performance had worse outcomes compared to surgeon rated in the highest tier of technical skills performance.³ In addition, Stulberg et al., ascribed a 26% variance in risk-adjusted complication rates to differences in surgeons' individual technical skills.⁴ The possibility of adding another, objective variable to our understanding and empirical evaluation of surgical outcomes has huge implications for patient care and surgical training and evaluation.

Quantification of surgical preferences and the effects on surgical outcomes was made possible due to advances in video-based assessment techniques and wearable sensor technology. While this study did not use machine learning (ML) and artificial intelligence (AI) algorithms to analyze the video data, video annotation techniques that support ML and AI were used. Structured, well-defined human annotation techniques are the key to discovery and comparison of differences and nuances in the surgical workflow. Annotation techniques that focus on segmentation of surgical videos into procedure steps allow for focused, subtask analysis of events that are pertinent and specific to a procedure step.²⁰ Our study used a multi-step, sub-task analysis and was able to categorize and quantify moment-to-moment technical decisions and approaches that have traditionally been thought not to influence surgical outcomes. Sub-task analysis of surgical videos holds the promise for significantly advancing our ability to fully implement video-based assessment in the surgical profession. Accumulation of video libraries and continued advancement of AI and ML algorithms are the most significant limiting factors to achieving automated surgical performance assessment using video.

Data from the wearable motion tracking sensors also made a major contribution to the study results by quantifying surgeon's technical skills. In essence, the motion results revealed that if you are on the riskiest pathway for technical decisions and preferences that mastery in technical skill may prevent poor outcomes. The specific metric that showed a significant difference between surgeons who leaked while on the riskiest pathway and those who did not was bimanual dexterity. Surgeons who did not leak were more ambidextrous throughout the enterotomy repair. Another critical finding in this study was that surgeon preference for closer sutures was also associated with a significantly lower leak rate. Surgeons who preferred the interrupted suturing technique, and did not leak, were found to place an average of 1–1.5 additional suture throws during the enterotomy repair. This finding was consistent irrespective of the surgeon's technical decision pathway.

The combination of motion metrics and advanced video-based assessment techniques helped to confirm our hypothesis that idiosyncratic surgeon preferences could be quantified and shown to have measurable effects on procedural outcomes. Our data analysis approach abandoned the traditional research methods that sought validity evidence for motion metrics by comparing medical student or novice performance metrics to experienced surgeons. While these studies were foundational and necessary in paving the path for use of motion data to quantify surgical skill, the most important work ahead is to use these metrics to facilitate a greater understanding of patient outcomes. Use of simulation-based assessments and mass data collection efforts with experienced surgeons are extremely valuable in

quantifying the variety of approaches surgeons make when faced with a standard clinical scenario. Direct application of motion metrics and video-based assessment in the operating room is the ultimate goal however, wide variance in patient anatomy may contribute to some of the technical differences observed in the data. A recent study by Hung et al., used motion and video-based assessment of surgeon performance during a real-life robotic prostatectomy. The subtask analysis focused on needle driving during the vesicourethral anastomosis. Suturing and needle driving were deconstructed into specific surgical gestures. Surgeons with unstructured needle driving protocols were found to have increased operative time, more needle driving attempts, and increased tissue trauma.²¹ In another study by Hung et al, motion data during robotic prostatectomy was linked to significant differences in the rate of post-operative erectile dysfunction.²²

Study limitations include those inherently associated with all simulation-based studies, a lack of wholistic realism that may affect suspension of disbelief for surgeon participants. Use of animal tissue verses silicon models helps to decrease these issues but there are still notable differences in the integrity of porcine bowel compared to human bowel. Another limitation relates to the allotted time to do the repair and lack of an opportunity to for the surgeon participants to engage in the standard manual leak test that is often performed any bowel repair. In the operating room if an enterotomy repair was subjected to a manual leak test and fluid or air seen exiting the repair the surgeon would simply place another stitch. Despite this simulation and time-based nuance in our study, the surgeons were engaged in the repair and appeared to approach the task as they normally would in the operating room. As such, this study observed real surgical decisions, found a variety of differences amongst experienced surgeons and, ultimately showed that idiosyncratic preferences could be quantified and linked to procedural outcomes.

The implications of this study relate broadly to surgical training and patient outcomes. Surgical training is still grounded in the basic tenets of the apprenticeship model with progressive transfer of patient care responsibilities and graded autonomy in the OR.²³ Surgical preferences and idiosyncrasies are handed down by tradition and modeled by trainees who become an amalgam of their surgical instructors. Advances in technology and surgical performance metrics offer new insight for the practicing surgeon and the trainee and will continue to make strides in closing the gap between the surgical process and outcomes.

Conclusions:

The well-known saying, “there’s more than one way to skin a cat” has been used to describe the countless differences in surgical preferences and techniques, all of which are motivated towards the same goal: excellent patient outcomes.¹¹ For many surgeons, each case is a matter of balancing one’s creative merit with known techniques. While having an exhaustive set of surgical techniques in one’s arsenal is common practice, the results of this study provide quantifiable metrics confirming that some approaches are better than others.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

National Institutes of Health (NIH) 5R01DK12344502, American College of Surgeons Foundation

Paper29: 29. Do Individual Surgeon Preferences Affect Procedural Outcomes?

*Hossein Mohamadipannah, *Calvin Perumalla, *Su Yang, *Brett Wise, *Cassidi Goll, *LaDonna Kearse, *Anna Witt, James R. Korndorffer, Carla Pugh

Surgery, Stanford University School of Medicine, Stanford, CA

Dr. David Hoyt (Orange, CA):

Good morning. The effectiveness of surgery is often based on personal experience, legacy and it is not always on evidence. We learn from master surgeon mentors and adopt their techniques. We model our decision-making on our training, our bias and with whom we are influenced. Surgical technique is lacking in prospective randomized trials. They are difficult to do because of their complexity and expense and we often think that we know what works without having to do formal trial. Coronary artery bypass grafting has never been subjected to a randomized clinical trial. Intestinal anastomosis is an essential skill with over 100 years of practice. The right technique should be clearly defined yet simple concepts such as combining two enterotomies into one, one versus two layer closure, the use of a corner stitch are still variables in practice. These concepts have been discussed in hundreds if not thousands of morbidity and mortality conferences without consensus and persistent variation in practice. So what? Does it matter? The answer is simple how something is done matters if it leads to a difference in complications as this study has shown, in this case an anastomotic leak. With some notable exceptions direct comparison of a technique with specific outcomes is not frequently studied. The current study laboratory protocol offers a way to evaluate procedural details to allow us to come to consensus, find best practice and teach this to new trainees. Dr. Pugh and her colleagues have developed tools to assess skill metrics which can be used to assess deterioration of skills, the effectiveness of practice and the safety during the credentialing of new procedures. The current study adds to this whole body of effort in a whole new dimension while allowing the assessment of one technique versus alternatives in the hands of many surgeons. How you do something appears to matter. Evidence based surgical skills have been defined by the commission on cancer and are currently being assessed in the verification program. We will see if changes in practice occur knowing that currently there is still great variability such practice as mesorectal excision which leads to changes in actual clinical outcomes. I have several questions for the authors, first what do you see as the limitations of this assessment technique if compared to assessment with real patients? Are the two really similar? Is a lot of human tissue more forgiving than pig intestine? Does how surgeons perform in ATLS moulage predict how they will function in the emergency department? Second, is the surgical community really ready to adopt evidence-based surgery and be transparent with self evaluation know when you need to change your practice. I did intestinal surgery for a long time and was never taught to use a corner stitch except for pyloroplasty. Should I have done something differently? Finally

we know from the assessment of cost that the variability can be critical and patients want better assurance that they will have high value care. If we learn someone is doing something that is inferior, how do we teach self correction and what are the implications for surgical leadership in credentialing in surgical education? This study and the techniques Dr. Pugh and her colleagues have presented are the future of surgical practice, technical assessment and new clinical outcomes. It presents an opportunity to assess many legacy techniques and compare surgeon variation using this in vitro model and like any good study opens up many more questions than answers. Congratulations Dr. Pugh on a wonderful study and presentation and I appreciate the opportunity to review the manuscript. Thank you very much.

Response from Carla Pugh:

Thank you Dr. Hoyt. To your first question, the main limitation with our study and our approach (using a loop of porcine intestines) relates to complexity and comparing that to what would happen in the operating room. The anatomy is not as complete, pig intestines do differ, some of them are thinner, but the scenario does fit well within the range of clinical variation that we do see in the OR and, in real-life, we do tend to adjust what we do based on what is presented to us in terms of the tissues. Other realism differences relate to bleeding. This was a pro-section so the blood vessels were not attached so this was less complex. We also did not present operative distractions. So, there are some differences, but when you look back at many of the research that has been done over the past 15–20 years looking at simulation or virtual reality there is a translation in terms of what people do in a simulation environment to what you see in an operating room and that has been shown in multiple studies over the past 20 years. With respect to your question on adapting evidence based surgery and if our practitioners are really ready to move forward with this, we did survey the participants in this study. There were 255 surgeons and obviously there is some bias because those who came to the booth had some interest and wanted to participate, but it was notable that between 87 to 95% of them wanted to contribute their data to our database and they were very interested in comparing their performance to someone else. I think what we learned was that people want the information and they want to do self-assessment however, they are just not ready for the information to be used against them before we really understand what it means. There are definitely early adapters that want to compare themselves and they saw this as information that we currently don't have and they thought it would be useful.

With respect to your specific question regarding doing intestinal surgery and should you have done something differently from your perspective, I love that question because I think a lot of people look at our work and they want to say "oh you are now trying to get us to be cookie cutter surgeons" meaning that we all do the exact same thing. Well, it turns out I don't think that is the answer. In this study there were surgeons that were on the riskiest pathway, but not all of them leaked. So, what are those factors in that pathway that would increase your leak rate? If you have that information, you know to put in more sutures, and make sure that you have the technical skill in terms of bimanual dexterity for example. So, I think instead of having cookie-cutter surgeons, if you are on that risky pathway and that is your preferred pathway you need to know what the risks are. For your last question relating

to the variability in how we teach self-correction, again I really think this is an information access issue and right now we just don't have this information, we never had this information and this is a new finding that your preferences do affect outcomes. I think people want to know that. The opportunity is there for us to start to collect this data, build databases and obviously our goal is to get to the operating room so we can increase our ability to share tips and tricks and understand the risks of what we do. Thank you.

Dr. Robert Cerfolio (New York, NY):

Yes, thank you very much, Robert Cerfolio from New York University Langone. Congratulations you have now delved into the important issue of dogma versus data - or preferences, which you want to call idiosyncrasies and I would call them dogma compared to the actual objective data. So the question is and it gets to your last point - are we willing to change our behavior when presented with clear clean data or do we all spend time shooting holes in the data. So as a surgeon and as a past COO of a large academic health care center and administrator the problem is it's the latter, you can't convince an experienced surgeon that they are dogma is not correct. They, and most unfortunately, are too often not receptive to change even when the data is irrefutable. So let me ask you a pointed question this example of the piano fingers examining the breast versus the rolling fingers have you been able to use that data to change the piano physician to the rubbing physician?

Response from Carla Pugh:

So I love that question because that is actually not the goal. So we don't want to change you to a completely different approach. If you have perfected that piano fingers technique to what you thought you were observing when you learned it. What we know is that there were people who did piano fingers and were accurate. So I would not change a piano fingers to a rubbing. I would show those persons who do piano fingers and do it inaccurately that they need to apply more force.

Dr. Robert Cerfolio (New York, NY):

Politely, I differ- that is the goal. The goal should be for all of us to get better and to want to get better and check our ego at the door and recognize our own biases, to give our patients better care at a lower cost and provide higher value care and better patient experience to everyone. If you have prospective randomized level A evidence why would anyone not change? The reason is and I dealt with it as COO for 5 years and as a physician leader for 25 years, it's we too often lack the awareness of our own inherent bias and ego, spend our effort shooting holes in the data but too little on self-reflection and process improvements that we can make. So you can move the goalpost, you changed piano finger to now "more force" - you can make it any variable you want, it is irrelevant, the goal is to change their behavior to get patients higher quality care and better value from their treating physicians. Maybe you did not make them go from piano fingers to rubbing but you changed what they did - my question is you did you measure who changed their behavior. I hope you did as that is the holy grail of leadership. I have bad news for everybody in the room and that is whether you want to do it or not, forced change in our behavior is coming from either

insurance companies, or hospital administrators like me, we are going to make you change. I spent most of our time as a hospital administrator using objective metrics like the Efficiency Quality Index, the EQI that we have written a lot about, to give physicians useful, accurate, and clinically meaningful data and metrics to help them to make changes to get better and to adopt behaviors that their better performing colleagues were already doing, to get out of their silo, to put aside their ego and to want to get better, because patients get better care when they do. It is the value proposition. This is the way to do it, so I congratulate you on this. Thank you for the wonderful presentation.

Dr. Peter Angelos (Chicago, IL):

Dr. Pugh congratulations on a really outstanding study. Two questions, number one how do you think that someone like me who says, “Well I’ve been doing it this way for a long time and it seems to work, how can I learn what might be a better strategy for doing these operations?” The second question is about your title. It currently refers to “preferences.” However, I am curious is it really preferences or just habits that are completely unexamined like how much we may use our non-dominant hand? Congratulations and I hope you keep pursuing this line of research.

Response from Carla Pugh:

Thank you Dr. Angelos. Yes. So, better strategies. I think that it is again looking at the outcomes. I do think that people when they see their approach and that they are different than others I think it really is shown clearly in the sensor data. Someone may say, “I placed that stitch and it looks great”, but this other person is doing it and they are more efficient and they get less leaks. I think it is more comparing someone without comparing apples to apples and I think surgeons are less likely to change their behavior if they think that you’re trying to compare them to someone who was trained differently or does something completely different. In terms of habit versus preferences or idiosyncratic approaches, I think they are all sort of one in the same. We are still trying to find the right language for this and that is the point. We’ve never really talked about this before and we don’t have a structured language regarding how to talk about this, but if we continue to do the work and publish, I think then we will adopt this and it will become the norm that we want to look at what are those differences, who am I more like and are they more efficient than I am and then what are the outcomes. Outcomes are not only the ones we are looking at now in terms of morbidity, mortality, it’s really efficiency, errors and other differences in technique. So, I think we can add metrics to the current outcomes that we talk about as well. Thank you so much.

Dr. David Harrington (Providence, RI):

Great paper, thank you for presenting it. One of the myths I have always believed was that great surgeons have beautiful economies of motion in the operating room - they waste no motion. But a quick calculation in the back of the room shows that myth appears to have been debunked by your data. The length of the hand pathways divided by time is the same between leak and no leak and therefore presumably between a non-master surgeon

and a master surgeon. Do you agree that your data shatters this very fond myth of mine? Is economy of motion a hallmark of a mater surgeon?

Response from Carla Pugh:

Not at all and it's been interesting presenting this work. I've heard more than one surgeon say the best way to do a Whipple really fast is to go slow. Meaning, do it right the first time. And, in those moments where it counts. That is one of the things we are seeing in the data when you look at the motion patterns of fellows versus twenty-year veterans in certain procedures. When it comes to that important stitch in a pancreaticojejunostomy or closing an atrial appendage the experienced surgeons are moving slow and they get it right the first time, they don't redo a stitch, they don't poke that tissue with the needle more than once and the fellows are quick and they start over if they don't like their suture. And the experienced surgeons are more quick on the other parts so there is something to it in terms of when you are slowing down and really being perfect and then when you are efficient.

Dr. Kelly Hunt (Houston, TX):

Thank you for that wonderful paper. As part of the American College of Surgeons Cancer Research Program we developed evidence-based standards for cancer operations and we called them critical elements based on evidence or expert opinion. One of the things that we were able to do under Dr. Hoy's leadership was to get some of those critical elements (operative standards) into the Commission on Cancer as quality measures as part of the accreditation process. One of the things that we found is that we had to figure out a way to document whether or not those critical elements were followed during the operative procedure and so we had to then move towards synoptic reporting in order to be certain that you could audit those operations and be sure they were performed accordingly. We also found that implementation was a little more challenging based on practice, years of experience and so forth so I just wonder if you could speak toward your work and how you would see that implemented and whether synoptic reporting would be a component of that documentation?

Response from Carla Pugh:

Thank you, I am in 100% agreement, this is a way forward in that space. This is the digital operative report. Right now you look at our operative reports and we verbalize what we do, we talk about our instruments and all of our different procedure steps, but it's not in enough detail to get what you are talking about. Even looking at a video you are not sure why people are making certain decisions whereas velocity and path length are much more specific in terms of what it is that people are doing and it helps us understand those detailed steps. So, I think that with this type of approach you can get more specific, quantified differences in terms of what people are doing in each segment of an operation. Thank you.

References:

1. Tschan F, Keller S, Semmer NK, et al. Effects of structured intraoperative briefings on patient outcomes: multicentre before-and-after study. *Br J Surg.* 2022;109(1):136–144. doi:10.1093/bjs/zxab384
2. Cerfolio RJ, Ferrari-Light D, Ren-Fielding C, et al. Improving Operating Room Turnover Time in a New York City Academic Hospital via Lean. *Ann Thorac Surg.* 2019;107(4):1011–1016. doi:10.1016/j.athoracsur.2018.11.071 [PubMed: 30629927]
3. 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. doi:10.1056/NEJMsa1300625 [PubMed: 24106936]
4. Stulberg JJ, Huang R, Kreutzer L, et al. Association Between Surgeon Technical Skills and Patient Outcomes. *JAMA Surg.* 2020;155(10):960–968. doi:10.1001/jamasurg.2020.3007 [PubMed: 32838425]
5. Flin R, Youngson G, Yule S. How do surgeons make intraoperative decisions? *Qual Saf Health Care.* 2007;16(3):235–239. doi:10.1136/qshc.2006.020743 [PubMed: 17545353]
6. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg.* 2001;193(5):479–485. doi:10.1016/S1072-7515(01)01041-9 [PubMed: 11708503]
7. DATTA V, MACKAY S, DARZP A, GILLIES D. Motion Analysis in the Assessment of Surgical Skill. *Comput Methods Biomech Biomed Engin.* 2001;4:515–523. doi:10.1080/10255840108908024
8. Hung A, Chen J, Jarc A, Hatcher D, Hooman D, Gill I. Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study. *J Urol.* 2017;199(1):296–304. doi:10.1016/j.juro.2017.07.081 [PubMed: 28765067]
9. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective Assessment. *Surg Endosc Interv Tech.* 2003;17(2):220–226. doi:10.1007/s00464-002-8869-8
10. Oropesa I, Sánchez-González P, Lamata P, et al. Methods and Tools for Objective Assessment of Psychomotor Skills in Laparoscopic Surgery. *J Surg Res.* 2011;171(1):e81–e95. doi:10.1016/j.jss.2011.06.034 [PubMed: 21924741]
11. Coselli J, Preventza O. More Than One Way to Skin a Cat. *J Thorac Cardiovasc Surg.* 2015;149(6):e96–7. doi:10.1016/j.jtcvs.2015.02.031 [PubMed: 26060016]
12. Way LW, Stewart L, Gantert W, et al. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. *Ann Surg.* 2003;237(4):460–469. doi:10.1097/01.SLA.0000060680.92690.E9 [PubMed: 12677139]
13. Pugh CM, DaRosa DA. Use of cognitive task analysis to guide the development of performance-based assessments for intraoperative decision making. *Mil Med.* 2013;178(10 Suppl):22–27. doi:10.7205/MILMED-D-13-00207 [PubMed: 24084302]
14. Lubowitz JH, Provencher MT, Brand JC, Rossi MJ. The Apprenticeship Model for Surgical Training Is Inferior. *Arthrosc J Arthrosc Relat Surg.* 2015;31(10):1847–1848. doi:10.1016/j.arthro.2015.07.014
15. Pugh CM, Santacaterina S, DaRosa DA, Clark RE. Intra-operative decision making: More than meets the eye. *J Biomed Inform.* 2011;44(3):486–496. doi:10.1016/j.jbi.2010.01.001 [PubMed: 20096376]
16. Hashimoto DA, Axelsson GC, Jones CB, et al. Surgical procedural map scoring for decision-making in laparoscopic cholecystectomy. *Am J Surg.* 2018;217(2):356–361. doi:10.1016/j.amjsurg.2018.11.011 [PubMed: 30470551]
17. Mohamadipannah H, Perrone KH, Peterson K, et al. Sensors and Psychomotor Metrics: A Unique Opportunity to Close the Gap on Surgical Processes and Outcomes. *ACS Biomater Sci Eng.* Published online March 22, 2020. doi:10.1021/acsbmaterials.9b01019
18. Pugh C, Goll C, Witt A, Mohamadipannah H, Wise B. The Surgical Metrics Project: What was achieved, and where is it headed? *Bull Am Coll Surg.* July 2020(July 2020).
19. Laufer S, Cohen ER, Kwan C, et al. Sensor Technology in Assessments of Clinical Skill. *N Engl J Med.* 2015;372(8):784–786. doi:10.1056/NEJMc1414210

20. 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(3):414–421. doi:10.1097/SLA.0000000000003460 [PubMed: 31274652]
21. Chen J, Oh PJ, Cheng N, et al. Use of Automated Performance Metrics to Measure Surgeon Performance during Robotic Vesicourethral Anastomosis and Methodical Development of a Training Tutorial. *J Urol.* 2018;200(4):895–902. doi:10.1016/j.juro.2018.05.080 [PubMed: 29792882]
22. Hung AJ, Chen J, Che Z, et al. Utilizing Machine Learning and Automated Performance Metrics to Evaluate Robot-Assisted Radical Prostatectomy Performance and Predict Outcomes. *J Endourol.* 2018;32(5):438–444. doi:10.1089/end.2018.0035 [PubMed: 29448809]
23. Polavarapu H, Kulaylat A, Hamed O. 100 years of surgical education: The past, present, and future. *Bull Am Coll Surg.* Published online July 1, 2013.

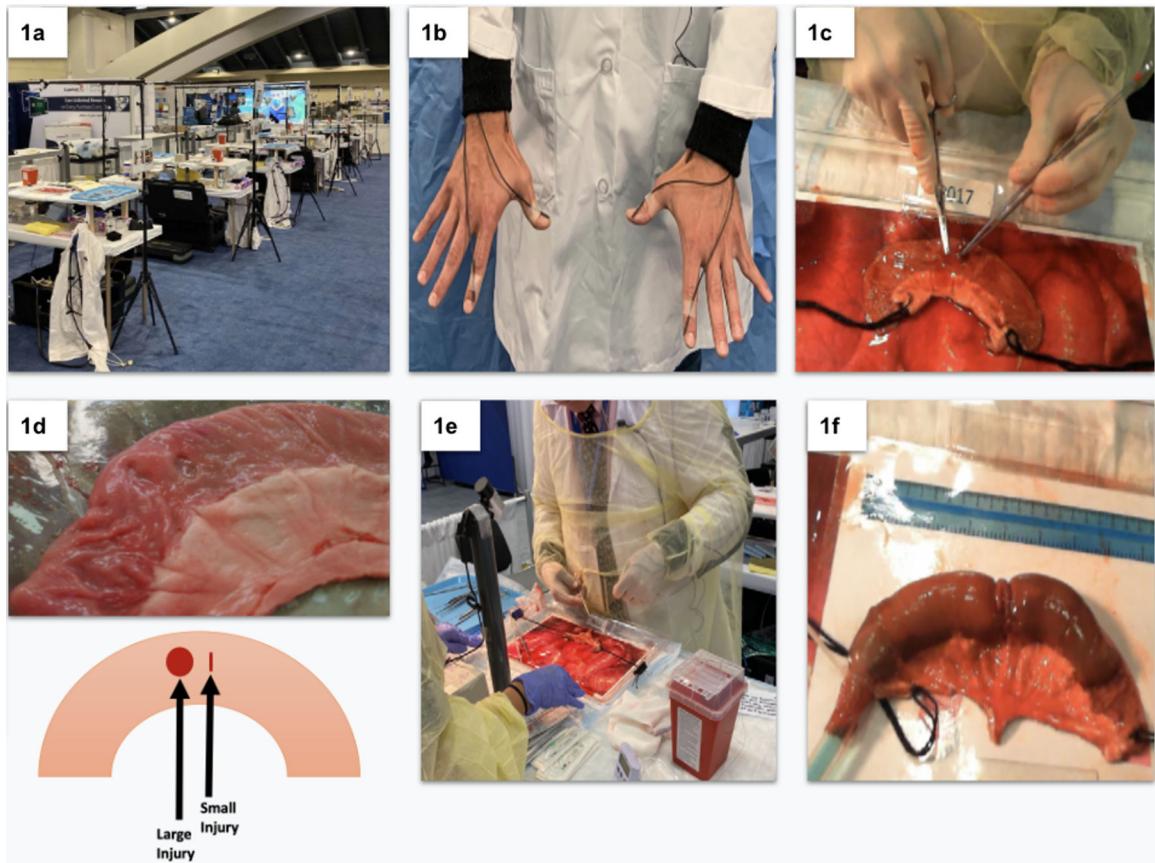


Figure 1:
a) ACS exhibit. **b)** EM sensors on participant's right and left hands. **c)** participant repairing an enterotomy. **d)** Bowel segment with standardized injuries. **e)** participant and surgical assistant view. **f)** Bowel segment being filled with blue liquid for leak test.

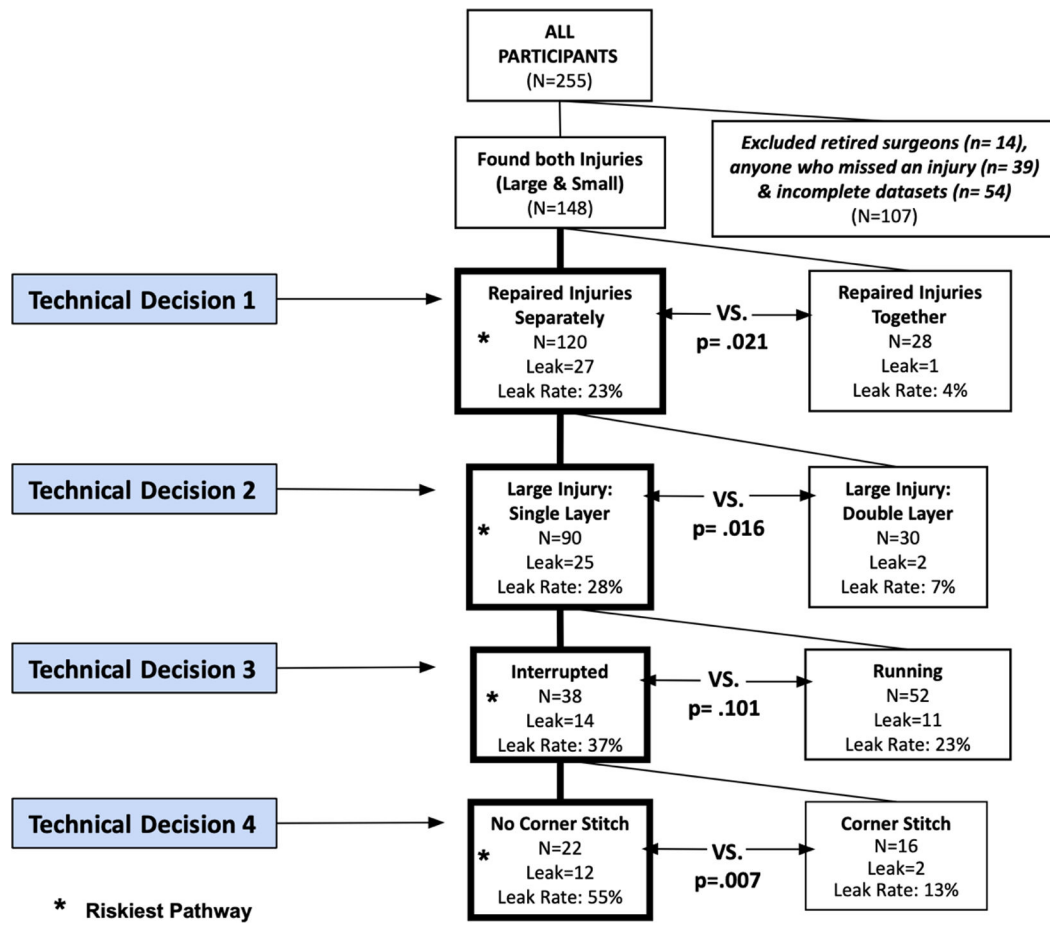


Figure 2: Enterotomy repair broken down into a decision tree and noting the riskiest pathway.

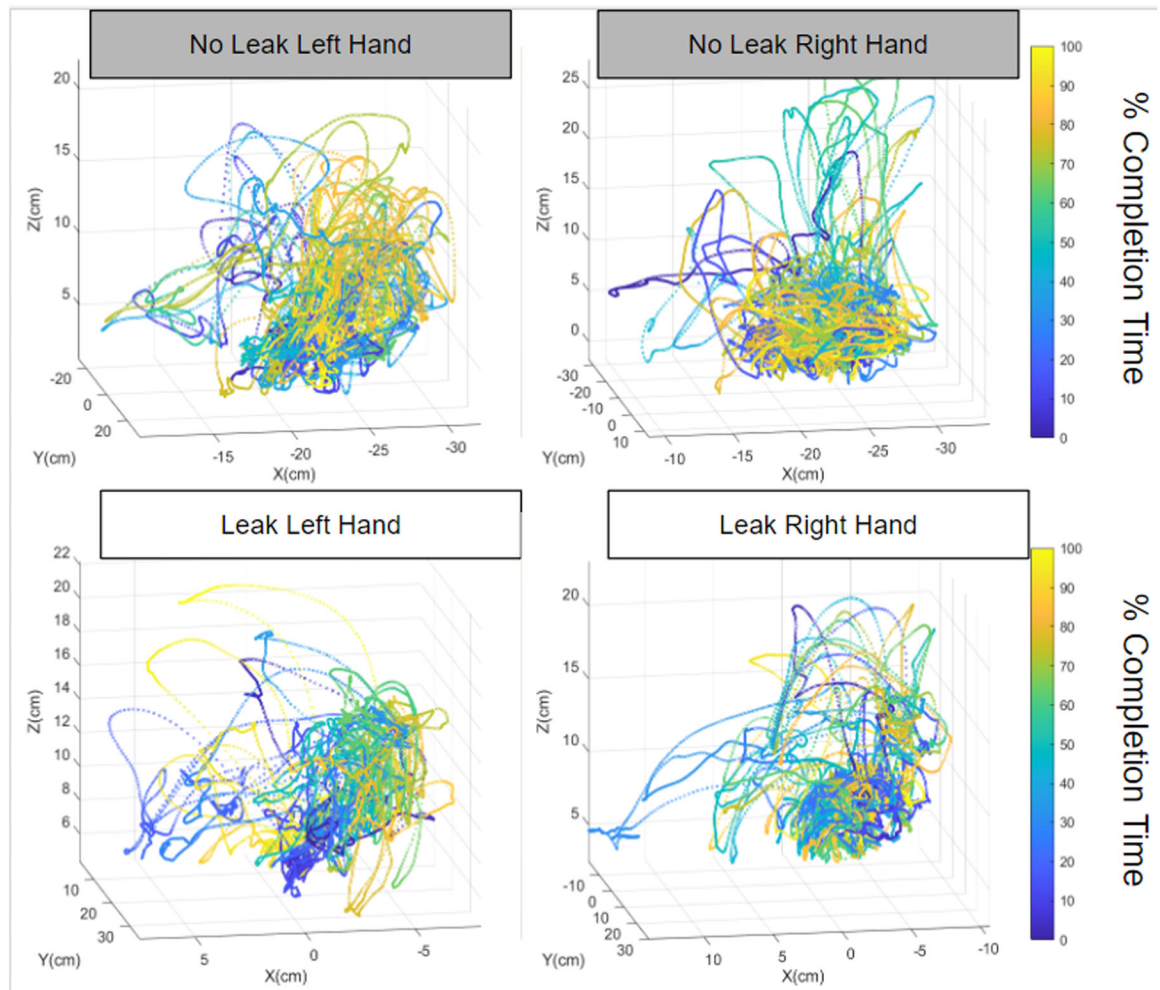


Figure 3. Motion Plots for the left and right hand of two participants. Note the longer path length and larger working volume for the left hand of the participant that did not leak, underscoring bimanual dexterity.

Table 1:

Gender, Age and Specialty for all participants. Years in practice for attending physicians.

Gender:		Age:		Surgical Specialty (attending only):		Years in Practice (attending only):	
Female:	65	21–30:	27	General Surgery:	108	0–10:	65
Male:	179	31–40:	57	Colorectal:	13	11–20:	60
*	11	41–50:	62	Surgical Oncology:	11	21–30:	47
		51–60:	48	Trauma	10	31 and up:	18
		61–70:	33	Min. Invasive Surgery:	13	*	11
		71–80:	9	Other:	35		
		80 and up:	2	*	11		
		*	17				

*=
no response

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Motion metrics for each operative strategy during the repair phase for those who leaked and did not leak.

	Repair Phase																	
	Repair Injuries Separately (Op. #1a) (n = 120)			Repair Injuries Together (Op. #1b) (n = 28)			Large Injury: Single Layer (Op. #2a) (n = 90)			Large Injury: Double Layer (Op. #2b) (n = 30)			Running (Op. #3a) (n = 52)			Interrupted (Op. #3b) (n = 38)		
	Leak n = 27	NoLeak n = 93	P	Leak n = 1	NoLeak n = 27	P	Leak n = 25	NoLeak n = 65	P	Leak n = 2	NoLeak n = 28	P	Leak n = 11	NoLeak n = 41	P	Leak n = 14	NoLeak n = 24	
Duration (min)	6.70 ± 2.50	8.50 ± 2.80	.0025	3.90 ± 0.00	10.1 ± 4.10	.094	6.50 ± 2.40	7.80 ± 2.50	.016	9.90 ± 2.80	10.2 ± 2.70	.96	7.30 ± 2.33	7.98 ± 2.80	.52	5.91 ± 2.33	7.46 ± 2.03	
Idle Time 0.05 (min)	0.700 ± 0.5	0.900 ± 0.7	.11	0.570 ± 0	0.940 ± 0.6	.66	0.680 ± 0.5	0.84 ± 0.7	.32	0.660 ± 0.4	1.06 ± 0.6	.27	0.710 ± .48	0.890 ± .68	.55	0.660 ± .56	0.760 ± .67	
Idle Time 0.5 (min)	0.290 ± 0.33	0.380 ± 0.42	.31	0.280 ± 0	0.340 ± 0.38	.85	0.300 ± 0.3	0.350 ± 0.4	.71	0.140 ± 0.19	0.440 ± 0.4	.24	0.280 ± .31	0.360 ± .39	.83	0.310 ± .36	0.330 ± .47	
Path Length (meters)	37.4 ± 15.0	46.6 ± 17.1	.0053	22.8 ± 0	61.6 ± 42	.19	35.5 ± 13.4	41.3 ± 13.1	.028	60.9 ± 17.2	59.0 ± 18.9	.73	37.9 ± 12.1	39.8 ± 12.5	.52	33.6 ± 14.5	43.8 ± 13.8	
Working Volume (meters)	0.079 ± 0.011	0.083 ± 0.014	.18	0.085 ± 0	0.092 ± 0.05	.85	0.07 ± 0.01	0.08 ± 0.01	.26	0.08 ± 0.01	0.08 ± 0.01	.80	0.08 ± .01	0.08 ± .01	.55	0.076 ± .01	0.079 ± .02	
Bimanual Dexterity	0.27 ± 0.09	0.28 ± 0.07	.50	0.14 ± 0	0.29 ± 0.06	.09	0.26 ± 0.09	0.26 ± 0.06	.90	0.3 ± 0.14	0.3 ± 0.08	.80	0.24 ± .06	0.26 ± .05	.35	0.28 ± .09	0.26 ± .07	

Table 3:

Motion analysis for the suture action phase for participants who leaked and did not leak when using the interrupted suturing technique.

	Suture Phase								
	Interrupted (n=38)			Interrupted - No Corner Stitch (n=22)			Interrupted Corner Stitch (n=16)		
	Leak n = 14	NoLeak n = 24	P	Leak n = 12	NoLeak n = 10	P	Leak n = 2	NoLeak n = 14	P
Duration (sec.)	9.41 ± 4.67	7.45 ± 3.60	0.16	9.91 ± 4.85	8.75 ± 3.59	0.52	6.34 ± 1.88	6.43 ± 3.41	0.97
Idle Time 0.05 (sec)	6.10 ± 3.30	4.77 ± 3.22	0.13	6.48 ± 3.31	5.43 ± 3.84	0.49	3.84 ± 3.03	4.26 ± 2.68	0.84
Path Length (meters)	0.156 ± 0.136	0.141 ± 0.139	0.42	0.160 ± 0.145	0.212 ± 0.181	0.46	0.131 ± 0.074	0.0861 ± 0.055	0.27
Working Volume (meters)	0.0094 ± 0.005	0.011 ± 0.010	0.42	0.0098 ± 0.0053	0.0162 ± 0.013	0.13	0.0073 ± 0.0002	0.0084 ± 0.0043	0.75
Bimanual Dexterity	0.268 ± 0.183	0.325 ± 0.155	0.19	0.214 ± 0.117	0.334 ± 0.153	.047	0.596 ± 0.223	0.319 ± 0.162	0.11
Suture Throws (# of throws)	5.00 ± 1.65	6.04 ± 1.74	0.04	4.69 ± 1.31	6.09 ± 1.70	0.03	7.0 ± 2.83	6.0 ± 1.84	0.57