



# HHS Public Access

Author manuscript

*Surgery*. Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

*Surgery*. 2016 November ; 160(5): 1400–1413. doi:10.1016/j.surg.2016.05.004.

## A Marker-less Technique for Measuring Kinematics in the Operating Room

Lane L. Frasier, MD<sup>1</sup>, David P. Azari, MS<sup>2</sup>, Yue Ma, PhD<sup>1</sup>, Sudha R. Pavuluri Quamme, MD MS<sup>1</sup>, Robert G. Radwin, PhD<sup>2,3</sup>, Carla M. Pugh, MD PhD<sup>1</sup>, Thomas Y. Yen, PhD<sup>2,3</sup>, Chia-Hsiung Chen, PhD<sup>4</sup>, and Caprice C. Greenberg, MD MPH<sup>1,2</sup>

<sup>1</sup>Wisconsin Surgical Outcomes Research (WiSOR) Program, Department of Surgery, University of Wisconsin-Madison, Madison, WI

<sup>2</sup>Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI

<sup>3</sup>Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI

<sup>4</sup>Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI

### Abstract

**Background**—Quantitative analysis of technical skill relies largely on specially-tagged instruments or tracers on surgeons' hands, often in simulated settings. We investigated a novel, marker-less technique for evaluating technical skill during open surgeries, and differentiating tasks and surgeon experience level.

**Methods**—We recorded the operative field via in-light camera for open operations. Sixteen cases yielded 138 video clips of suturing and tying tasks 5 seconds in length. Video clips were categorized based on surgeon role (attending, resident) and task sub-type (suturing tasks: Body Wall, Bowel Anastomosis, Complex Anastomosis; tying tasks: Body Wall, Superficial Tying; Deep Tying). We tracked a region of interest on the hand to generate kinematic data. Nested multi-level modeling addressed the non-independence of clips obtained from the same surgeon.

**Results**—Interaction effects for suturing tasks were seen between role and task categories for average speed ( $p=0.04$ ), standard deviation of speed ( $p=0.05$ ), and average acceleration ( $p=0.03$ ). There were significant differences across task categories for standard deviation of acceleration ( $p=0.02$ ). Significant differences for tying tasks across task categories were observed for

---

**Corresponding Author:** Caprice Greenberg, 600 Highland Avenue, BX 7375 Clinical Science Center, Madison, WI 53792-3284, Phone: 608-262-0395, Fax: 608-263-2354.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

#### Financial Disclosures:

Caprice Greenberg serves as a consultant for Johnson & Johnson's Human Performance Institute. Multimedia Video Task Analysis(TM) and MVTA(TM) are registered trademarks of the Wisconsin Alumni Research Foundation at the University of Wisconsin-Madison. The software was developed in the laboratory of Dr. Robert Radwin sponsored by donations and gifts of industrial and government members of the Ergonomics Analysis and Design Consortium.

maximum speed ( $p=0.02$ ), standard deviation of speed (0.04), and average ( $p=0.02$ ), maximum ( $p<0.01$ ), and standard deviation ( $p=0.03$ ) of acceleration.

**Conclusions**—We demonstrated the ability to detect kinematic differences in performance using marker-less tracking during open surgical cases. Suturing task evaluation was most sensitive to differences in surgeon role and task category and may represent a scalable approach to provide quantitative feedback to surgeons about technical skill.

---

## Introduction

Like all physicians, surgeons must develop clinical knowledge and the ability to make sound medical decisions, but they must also develop the technical skills needed to operate. There is ongoing interest in measuring surgical technical performance, especially after Birkmeyer and colleagues demonstrated that surgeons with the best technical skills, as rated by their peers, had the lowest rates of patient complications following bariatric surgery.<sup>1</sup> Because technical errors, defined as an error of manual technique that occurs during an operation, compose nearly 75% of the adverse events related to surgery,<sup>2-5</sup> measuring surgical technical skill is a vital step toward reducing adverse events and improving patient outcomes.

Birkmeyer's approach evaluated the technical skill of surgeons using a global rating scale. These types of subjective assessment measures are common, but suffer from variability between raters and require significant time investment from the raters, limiting scalability. Early approaches to objective measurement of surgical technical skill centered on dexterity analysis systems<sup>6-8</sup> using sensors mounted on the hands or on special instruments<sup>9-14</sup> which provided quantitative output such as path length, time taken, force/torque ratios, and the number of movements needed to complete a given task. The use of these specialized measurement systems has largely limited this work to benchtop assessment and simulated operative cases.<sup>7</sup>

While this work has advanced our understanding and assessment of the psychomotor properties of surgical technical skill, these methods have limited generalizability outside of academic institutions and simulation settings, and are not easily applied to the analysis of actual, open surgery. A review by Reiley et al (2011) found that the majority of quantitative assessment systems for surgical technical skills were limited to simulated environments and/or minimally invasive approaches due to the inability to effectively track surgeons' hands in a sterile environment.<sup>7</sup> Similarly, a 2014 review of vascular surgery skill assessment by Mitchell et al found that objective measures of technical skill were obtained in simulated settings or limited to metrics such as procedure time.<sup>15</sup> More recently, Duran et al. describe the development and validation of the Fundamentals of Endovascular Surgery assessment tool, which uses global rating scales, error metric assessment, and position and movement of an endovascular catheter tip in a simulated model.<sup>16</sup>

Novel marker-less video tracking methods based on cross correlation template-matching algorithms were developed in our co-author's lab (Radwin) by Chen et. al,<sup>17</sup> to trace the trajectory of a selected region of interest over successive video frames. This algorithm has been adjusted to account for challenging video conditions, including poor lighting or

resolution and blurred motion. We evaluated whether the marker-less video tracking system could be used to determine motion kinetics associated with various surgical technical tasks, and could measure the differences in performance between surgeons of varying skill levels.

## Methods

### Case Selection

After approval from our institutional review board, the inpatient operating room schedule was screened for eligible cases. These included open colorectal, complex upper gastrointestinal, hepatobiliary, surgical oncology, transplant, vascular, thoracic, and cardiac operations, scheduled in an OR equipped with the necessary video recording equipment. Emergency cases, cases not requiring general anesthesia, and those performed after hours and on weekends were excluded.

Attending surgeons for eligible cases were contacted about possible participation. If they were agreeable, written informed consent was obtained. Once the attending surgeon consented, the participating surgical resident was also approached about participation, with written informed consent obtained prior to case recording. Residents from any post-graduate year (PGY) were eligible for participation; ultimately, residents from the PGY 3 and 5 classes were approached on a case-by-case basis as they most commonly assisted the surgical attendings for the types of cases targeted for recording. If the surgical resident declined to participate, recording proceeded, but only the surgical attending's hands were analyzed. For future eligible cases involving surgeon(s) who had already consented to participate, electronic or verbal confirmation of continued involvement was obtained for each case.

### Data Collection

When an approved case began, the research team remotely activated recording of the OR's in-light camera, which streams to a secure hospital computer enabled with an AXIS video encoder (Axis Communications, Lund Sweden). This method of recording was chosen as in-light cameras are already available, require no additional equipment to be added to the room, potentially impacting workflow and personnel movement, and introduce no risk of contaminating the sterile field as would be possible with a pole-mounted or head-mounted cameras or video glasses. Since the objective of this work is to evaluate the feasibility of a scalable, automated assessment, we chose to demonstrate feasibility using technology that already exists in increasing numbers of operating rooms and would not require any additional work on the part of the operative team. Since the in-light camera only captured the operative field without audio, no patient details or protected health information was ever recorded, and the patient's identity was not visible in the video recording. Likewise, no members of the operative team other than the surgeon(s) were ever visible via the in-light camera and no audio was recorded. The institution's standard surgical consent form contains language providing consent for filming and recording for the purposes of performance improvement, education, and research. Therefore, written informed consent was not obtained other members of the operating room team, and no additional consent was required from the patient. At the conclusion of each case, participating surgeon(s) completed a

questionnaire which included information such as the operation performed, role of the participant (attending versus resident), dominant hand, and location of the participant throughout the case relative to the patient. An estimated twenty cases were needed to demonstrate feasibility; twenty-two cases were recorded, capturing footage from ten surgeons (six attendings, four residents).

## Data Analysis

Operative cases were then evaluated using Multimedia Video Task Analysis (MVTA)<sup>TM</sup> software (Wisconsin Alumni Research Foundation, Madison, WI), developed in our co-author's lab (Radwin) for conducting human factors time studies of video recordings. In MVTA, categories of interest, called records, are created by the user and listed along the left side of the analysis window (Figure 1). Each record can contain multiple events, also created by the user and seen along the right side of the analysis window, which can be marked whenever they occur. Each record was defined to house all events for either the attending or resident surgeon. We created events for suturing and tying surgical tasks and identified when surgeon participants were performing one of these tasks. To ensure that surgeons' hands were attributed to the correct participant, these clips were identified by a member of the research team who is a general surgery resident and familiar with the operating room and procedures being performed (LLF). The entire operation was scanned for clips, and changes in surgeon positioning were easily identified using context clues such as the position and appearance of the surgeons' heads. Figure 1 presents a screen shot of a resident surgeon performing a suturing task. To ensure that clips captured a reasonable fraction of a given task, analysis was limited to clips >5 seconds in length for which the surgeons' hands were clearly visible. This eliminated 42 clips (35 due to inadequate task clip length and 7 due to occlusion of view of the hand) resulting in 138 usable clips from 14 separate operations.

Marker-less tracking software tracked a region of interest (ROI) on the surgeon's hands over successive video frames in each clip. Identification of a ROI on a surgeon's dominant hand can be seen in Figure 2. Our group has previously used marker-less tracking software to evaluate descriptive statistics of dominant vs non-dominant hand movement during reduction mammoplasty operations.<sup>18</sup> A unique ROI was defined at the start of every video clip, identifying a portion of the hand (generally the index finger or thumb) that remained in view for the entire clip. Because kinematic data are obtained by measuring the changes in the ROI over time, the actual location of the ROI does not matter as long as it follows the hand throughout a clip. The position, speed, and acceleration of the ROI for each clip were quantified across successive frames. Each video was recorded at 30 frames per second. Because of varied positioning of the in-light camera throughout an operation, the videos could not be distance calibrated. To address this, in-frame video pixel measurements of the surgeon hand breadth were used to calibrate the kinematic record of each clip. Hand dimensions have been shown to provide acceptable calibration for hand speed estimates<sup>19</sup> and the proximal interphalangeal joint breadth was scaled to the population means of males (23.0 mm) or females (19.9 mm), depending on the gender of the surgeon. Proximal interphalangeal joint breadth was used because of its small coefficient of variation of 0.071 for males and 0.064 for females, based on anthropomorphic measurements from the US Army.<sup>20</sup> Surgeon hand measurements in pixels were averaged across three frames for each

video clip; these calibration measurements were averaged for every unique camera-surgeon position relationship. Pixel-millimeter calibrations were re-calculated for every time the camera or patient moved positions. The accuracy of measurement was limited by the size of each pixel in the video frame.

Prior research has demonstrated the sensitivity of suturing<sup>9, 11, 12, 21</sup> and tying<sup>11</sup> tasks in differentiating skill level during open benchtop assessments. Initial exploration of our data confirmed these findings, as attending surgeons had higher observed mean speed and acceleration measures for both suturing and tying tasks. However, on initial analysis we recognized the potential for significant confounding by depth and tissue type. In review of the literature, we identified recent work that similarly demonstrated varying kinematics when participants performed interrupted suture tasks on a variable tissue simulator. Specifically, on a model simulating friable tissue, participants had significantly increased idle time,<sup>22</sup> suture time, and path length<sup>23</sup> compared to arterial or fascial tissue models.

For these reasons, we further classified each suturing and tying task video clip into one of three task categories. This was performed by a member of the research team familiar with surgical technical tasks (LLF). Suturing tasks were categorized as S1: Body Wall (including skin or fascial closure, suturing on hernia mesh, and/or sewing in surgical lines and drains; n=28); S2: Bowel Anastomosis (which included maturation of an ostomy at the level of the skin; n=20); and S3: Complex Anastomosis (which included non-bowel intra-abdominal anastomoses including hepatobiliary and vascular anastomoses; n=12). Tying tasks were categorized as T1: Body Wall (including skin or fascial closure, tying of hernia mesh, and/or tying surgical lines and drains; n=20); T2: Superficial Tying (including tying of subcutaneous and superficial peritoneal and thoracic structures and ostomy maturation; n=46); and T3: Deep Tying (which included retroperitoneal and deep intra-abdominal and thoracic structures; n=12). These categories roughly correlated with increasing complexity, and separated tasks based on the primary tissue being manipulated. We hypothesized that motion kinematics would demonstrate significant differences when working with different tissues or completing more versus less complex components of an operation.

The average, maximum and standard deviation of calibrated speed and acceleration measures were then analyzed using SAS version 9.4 mixed procedure (SAS Institute Inc., Cary, NC). To account for the non-independence of multiple video clips utilized from the same case, and multiple cases obtained from the same surgeon, three-level nested models were applied. Data were compared across surgeon role (attending versus resident) and across task category. Initially, a two-factor model was tested using main effects for surgeon role and task category as well as an interaction effect between the two. If the interaction effect was not statistically significant, the interaction term was excluded and a second two-factor model was tested using only the main effects (surgeon role and task category). Ad hoc pairwise comparisons were performed for a significant interaction or task category main effect. The Tukey-Kramer test was used to control for the inflation of type I error rate associated with multiple comparisons. Given no theoretical hypothesis on the direction of the comparisons, two-tailed tests were adopted.

## Results

Twenty-two cases were initially recorded. Recorded cases represented a broad spectrum of surgical operations and included open hernia repair, complex colorectal, hepatobiliary, surgical oncology, and thoracic cases. Of these, 16 yielded a total of 138 usable clips from 6 attendings and 3 residents. Videos were excluded due to lack of visible suturing and tying clips of suitable length ( $n=5$ ) or poor visual quality ( $n=3$ ). The number of cases recorded and clips obtained for each surgical task for attending and resident participants are listed in Table 1. The number of cases recorded and clips obtained for individual surgeon participants, ranging from 1–5 cases, are shown in Table 2.

### Observed Means: Suturing

Sixty suturing clips were analyzed. Observed means are summarized in Table 3. Attending surgeons had higher means for all kinematic measures evaluated overall. When assessed for each task separately, this finding persisted for S1 (Body Wall) and S3 (Complex Anastomosis), but not for S2 (Bowel Anastomosis) tasks. When evaluating suturing categories, S1 tasks tended to have higher means overall than S2 or S3 tasks; however, this pattern did not persist when attending and resident surgeons were evaluated as separate groups.

### Observed Means: Tying

Seventy-eight tying tasks were evaluated, with observed means summarized in Table 4. When tying tasks were assessed overall, attending surgeons demonstrated higher means across all measures. T1 (Body Wall) tasks had higher observed means than T2 (Superficial Tying), and T2 had higher observed means than T3 (Deep Tying) tasks for all kinematic measures except maximum acceleration when assessed overall; however, this pattern did not persist when attending and resident surgeons were evaluated independently. When task category was also assessed by role, attending surgeons demonstrated higher means for all kinematic measures for T2 and T3 tasks, but lower means for all measures of T1 tasks compared to residents.

### Model 1: Predicted Main Effects with Inclusion of an Interaction Effect

Interaction effects in the nesting model (Table 5) for suturing tasks were seen for several measures related to speed and acceleration, including average speed ( $p=0.04$ ), standard deviation of speed ( $p=0.05$ ) and average acceleration ( $p=0.03$ ). When comparing suturing tasks, attending surgeons had significantly higher average speed, standard deviation of speed, and average acceleration for S1 (Body Wall) compared to S2 (Bowel Anastomosis) tasks ( $p<0.01$ ,  $p=0.04$ , and  $<0.01$ , respectively). Additionally, attending surgeons had significant higher average speed ( $p<0.01$ ) and acceleration ( $p<0.01$ ) for S1 compared to S3 (Complex Anastomosis) tasks. These differences were not seen for resident comparisons. Additionally, when performing S2 tasks, attending surgeons had significantly higher average speed ( $p=0.03$ ), standard deviation of speed ( $p=0.05$ ), and average acceleration ( $p=0.03$ ) compared to residents. No differences between attending and resident surgeons were seen for S2 or S3 tasks). Significant suturing comparisons are displayed in Figure 3.

For tying tasks, no interaction effects were identified for any measure evaluated. In other words, while observed means for kinematics were different between attendings and residents, including several T1 (Body Wall) tasks in which attendings had lower mean kinematic measures compared to residents, these were not statistically significant. Figure 4 displays significant tying comparisons.

### **Model 2: Predicted Main Effects without an Interaction Effect**

In the absence of interaction effects, the remaining kinematic measures were evaluated using a two-factor model (role and task category, Table 6). As interaction effects were previously identified for several suturing measures, results from the two-factor model are not reported for these measures, as they should not be interpreted under these conditions.

For suturing tasks, attending surgeons were not consistently different than residents for average speed, average acceleration, or standard deviation of acceleration. However, when comparing task categories, S1 (Body Wall) had higher standard deviation of acceleration compared to S3 (Complex Anastomosis) tasks ( $p=0.04$ ).

For tying tasks, there were no significant differences between attending and resident surgeons for any kinematic measure. However, we did identify differences between task categories: T1 tasks had significantly higher predicted maximum acceleration compared to T2 tasks ( $p<0.01$ ) and significantly higher predicted maximum speed, standard deviation of speed, average and maximum acceleration, and standard deviation of acceleration compared to T3 tasks ( $p = 0.03, 0.05, 0.02, 0.2, \text{ and } 0.03$ , respectively).

## **Discussion**

This study has demonstrated the feasibility of obtaining kinematic data for two common surgical tasks during open surgery, without the necessity of specialized tracking devices or limiting analysis to laparoscopic cases. Previous work evaluating descriptive differences in dominant versus non-dominant hand movements in attending surgeons versus residents<sup>18</sup> demonstrated the feasibility of this approach for open surgical cases. Here, we have extended this work and sought to discern differences based on surgeon role and task relationship to the various stages of surgery, with the ultimate aim of identifying high-yield sections of open operations amenable to differentiating skill level and providing feedback metrics to surgeons. Our identification of statistically significant differences across several kinematic measures for two tasks encountered in a majority of surgical procedures represents a measurable step forward in a scalable methodology to assess surgical skill under a variety of actual surgical conditions and toward generalization of the ability to measure technical skill. These tasks were evaluated within the context and flow of an entire operative case, rather than evaluating discrete tasks without an operative context, as is seen with simulation benchtop models.

We identified significant differences in speed and acceleration metrics for suturing and tying tasks based on the type of task being performed. Our findings are consistent with prior work in which these tasks consistently differentiated surgeons based on experience level.<sup>11</sup> We also build on recent work which identified differences in idle time,<sup>22</sup> path length, and suture

time<sup>23</sup> for suturing tasks performed on models simulating more or less complex tissues. However, our approach is novel in that it consistently identified differences between attendings and residents, and between task categories, during a wide variety of actual operations and without restricting task parameters as is commonly used in highly-controlled bench-top assessments.

Even more exciting was our identification of interaction effects between surgeon role and task category for suturing. In other words, attending surgeons' suturing kinematics are significantly different when working on different tissue types and when comparing attending and resident surgeons working on the same tissue type. This represents a critical step forward in assessment of surgical technical skill and suggests that assessment of suturing technical skill is most sensitive to surgeon experience and complexity of surgery, and represents a promising target for further quantification of technical skill. .

Interestingly, increasing task complexity did not always correlate with stepwise decreases in speed and acceleration metrics; several S2 (bowel anastomosis) task metrics were lower than S3 (complex anastomosis) measures, indicating that surgeons were slower at bowel anastomosis suturing tasks than when sewing more complex anastomoses. This was seen when assessing tasks overall and also when assessed by roles (attending and resident surgeons).

Additionally, residents had higher maximum speed, maximum acceleration, standard deviation of speed and standard deviation of acceleration for S2 tasks compared to attending surgeons. These findings seem to indicate that increasing technical skill does not translate solely to increased speed and acceleration. For example, decreased acceleration seen in S2 tasks performed by attendings may represent a smoother, steady pace compared to a less skilled operator who demonstrates pauses and rapid changes in motion which would lead to increased maximum speed and acceleration.

Overall, our findings highlight the unstudied relationships between technical elements (suturing, tying) and the larger context of an operation. Ongoing work is needed to understand the relationships between the kinematic data evaluated here and other assessments of technical skill. It may be that kinematic data assessment is most appropriate for evaluation of certain tasks, such as abdominal fascial closure, while other tasks such as a complex hepatobiliary anastomosis are more appropriately evaluated with a global assessment score or a combination of several metrics.

Of note, S1 (Body Wall) suturing tasks were most likely to have interaction or category effects compared to S2 (Bowel Anastomosis) or S3 (Complex Anastomosis) tasks. This may be due to the higher number of S1 clips available for analysis, as all operations analyzed required closure of an open abdominal wound, whereas S2 and S3 clips could only be obtained for specific types of operations.

There are several limitations to consider. While the in-light camera represents a non-invasive method of video capture, the captured images are dependent on where the operating surgeons focus the boom light, which is not necessarily where they are working. Surgeons' hands moved in and out of the video frame or were obscured by a surgeon's head leaning



over the operative field, reducing the data available for technical analysis. These limitations could be resolved with use of a wide-angle lens on a boom separate from the in-light camera, which is becoming increasingly available. The research team could set the angle and location of the camera boom prior to case start and begin recording remotely, minimizing interference with the operative team and ensuring high-quality data capture. Wearable technologies like GoPro® and mobile video glasses have also been successfully employed by our group and others.<sup>18, 24</sup> Additionally, some components of our current methodology, such as clip identification were time-intensive and not easily scalable. Ongoing evaluation of the kinematic differences between task sub-categories (e.g., T1 vs T2, S1 vs S3) could lead to the development of software capable of recognizing the kinematic patterns associated with suturing and tying tasks and sub-tasks, identifying and flagging clips for quick confirmation by the surgeon. Likewise, our distance calibration was time-consuming and based on population measures of hand size. Since the in-light camera location and distance from the surgeons' hands were constantly changing throughout the operation, it was not practical to make precise distance calibrations. Kinematic measures relative to hand dimensions were considered a pragmatic approach to approximating distances, given the relatively low precision of marker-less video tracking of the different hands, and are considered sources of random error. The kinematic measures are therefore approximations and contain some measurement errors, possibly contributing to additional variability and noise in the statistical analysis. Further accuracy could be obtained by calibrating based on surgeon glove size, obtaining standardized, one-time measurements of the surgeons' hands, or including a standardized reference such as a ruler in the operative field.

Power analysis was not conducted prior to the study as it was not possible to determine the number of clips that we would have per case. Sample sizes were set by prior experience in this type of analysis. Post hoc power analysis focused on testing the role difference between attending versus resident surgeons in the kinematics for tying tasks, given that no significant role difference was found for any of the kinematic measures. Power analysis was conducted using Optimal Design 3.01,<sup>25</sup> a power analysis software program that takes into account the clustered data structure. Power was influenced by multiple factors, including type I error rate, sample sizes (i.e., number of surgeons, number of cases performed per surgeon, and number of video clips included per case), intra-class correlations (ICCs) at surgeon level and at case level, and effect size (i.e., the role difference in a kinematic measure in standardized metric). The analysis showed that the standardized role difference observed in the current study (0.11 to 0.47) was smaller than the minimum effect size (0.85 to 1.26) that could be detected with sufficient power (.80) for all six kinematic measures, which could have led to the statistical insignificance. Future studies with increased sample sizes and improved procedures reducing unusable video clips to be excluded from analysis would help increase power and allow a greater possibility in detecting small role difference.

Video recording of operations is likely to become increasingly commonplace given growing availability and sophistication of video technology and medicine's cultural shift toward increased transparency.<sup>26</sup> This type of marker-less tracking could offer the ability to analyze data and provide feedback to surgeons for a wide variety of open or laparoscopic cases and practice settings, allowing surgeons to easily obtain information on their technical performance as part of ongoing skill development. Specifically, as we further develop our

understanding of quantitative skill measurement, kinematic data could provide objective feedback to learners about the kinematics of tasks during new skills and procedures with a measure of their progress toward mastery.

This approach combines the use of video, which allows for self-observation and reflection, and the provision of objective, numerical feedback. Ultimately, this methodology could provide a high-throughput, scalable method of providing objective data to surgeons regarding their technical performance in open operations. In time, identifications of kinematic patterns associated with various stages of technical proficiency, as well as the kinematics associated with “expert” status could be determined (and, as noted elsewhere, may not correlate solely with high speed and acceleration). These data could provide surgeons with benchmarks with which to compare their data over time. Potential applications include assessment of surgeons during skill acquisition as well as maintenance, and could be useful for residency programs and for surgeons re-entering the operation room after time away for medical or personal reasons.

Prior research indicates that the number of hand movements<sup>9–11</sup> and time taken<sup>9–12, 21</sup> decreases for a given task with increased surgeon experience. Fewer hand movements, synonymous with increased efficiency, result in smooth, steady hand motion and cycles of motion without hesitation. These concepts are conceptually parallel to higher peak speed and acceleration. Decreased time taken, synonymous with increased speed, is consistent with our findings of increased mean speed and acceleration for suturing tasks. However, increased efficiency alone is insufficient to truly evaluate technical skill. A surgeon may move quickly but with poor results – a stitch that pulls through due to poor placement or a dropped knot throw due to moving too quickly. Quality assessment was not performed during this evaluation as we sought to demonstrate feasibility. Future work must include correlating these kinematic measures with other evaluations such as global assessment or cosmetic and functional outcomes to ensure that surgeons are not sacrificing quality for speed.

Finally, significant attention in education and human factors has focused on the concept of expertise and the circumstances surrounding experts’ transitions from automated, routine behaviors to more deliberate, effortful evaluation and actions.<sup>27</sup> This transition may occur deliberately, at pre-identified stages in the operation related to patient- or procedure-specific characteristics, or with intra-operative identification of an unexpected difficulty or roadblock to proceeding further. This change in cognitive processes has been labeled ‘slowing down’, but there are no data exploring how this transition affects kinematic movements during an operation. We feel that this represents a critical avenue of future investigation. Such transitions as described by the reviewer are included in hidden Markov models. The work we are doing now can inform development of such multivariable and multi-dimensions models. Further work is needed to determine whether and how the kinematics of highly-demanding portions of the case can vary from more routine, automated portions, and how they might change with expected versus unexpected task complexity. Description of markers identifying these high-intensity periods could provide surgeons with another tool for thoughtful reflection and self-assessment, and allow surgeons to anticipate when a transition to deliberate, effortful activity may be needed in future cases. It is very possible that the transitions in speed or acceleration will be more predictive of performance than absolute

measures. Furthermore, this methodology could identify changes in motion kinematics as a marker of potentially difficult or high-risk periods of an operation. This could eventually allow for rapid processing of large volumes of operative video and selective manual review of points at highest risk for safety compromise.

Next steps for this work include comparison of technical kinematic data with global assessment such as the Objective Structured Assessment of Technical Skill (OSATS) and identification of kinematic patterns associated with varying degrees of proficiency; development of scalable measurement of surgeons' hands, and further assessment of the relationships between skill, speed and acceleration, and case complexity.

## Acknowledgments

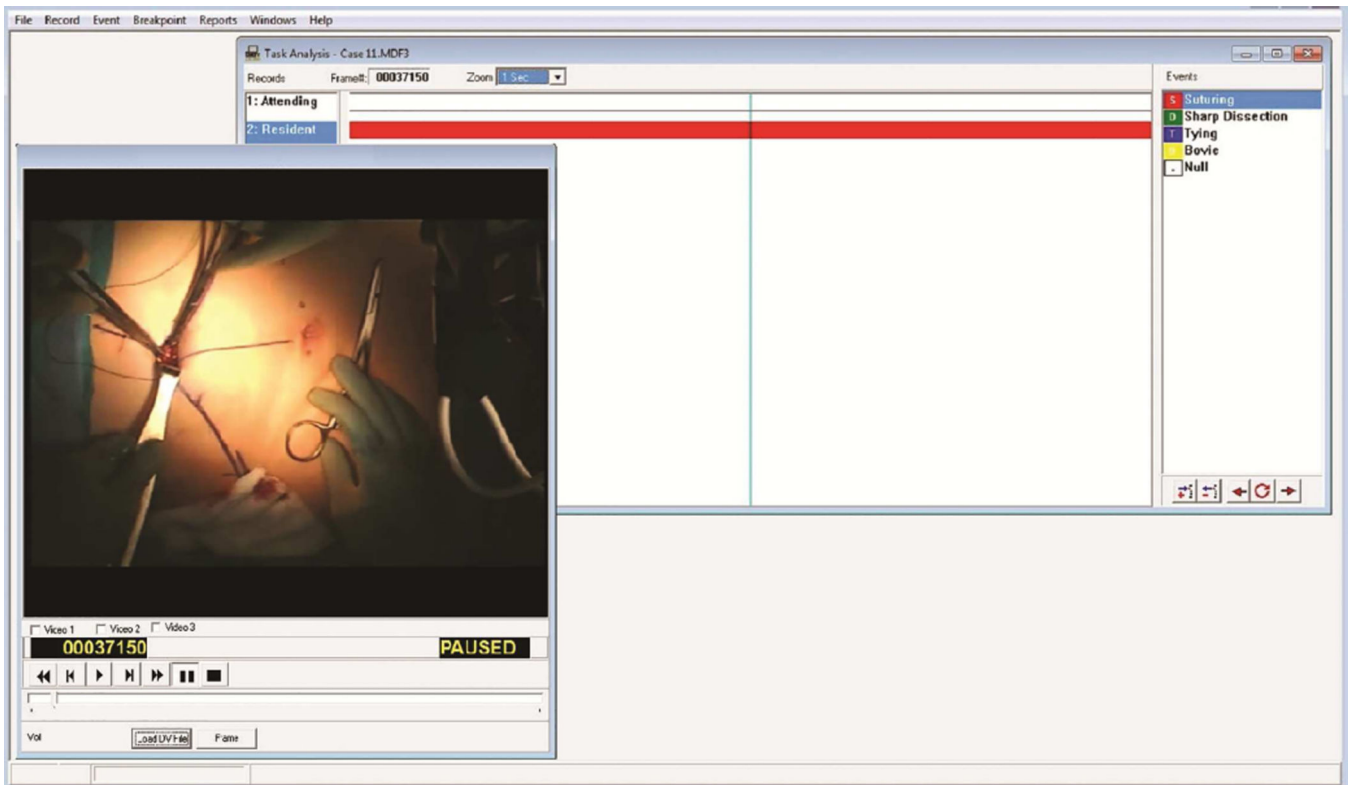
The authors would like to acknowledge Linda Yang, BS, for her work on this project.

**Sources of Financial Support:** This project was funded by the UW Institute for Clinical and Translational Research (CTSA) grant UW #PRJ67FC. Lane Frasier is currently supported by AHRQ F32 HS022403. She previously received support via NIH/ National Cancer Institute T32 CA90217 and the AAS Research Fellowship Award. David Azari, Chia-Hsiung Chen and Robert Radwin received support via NIH 1R21 EB01458301.

## References

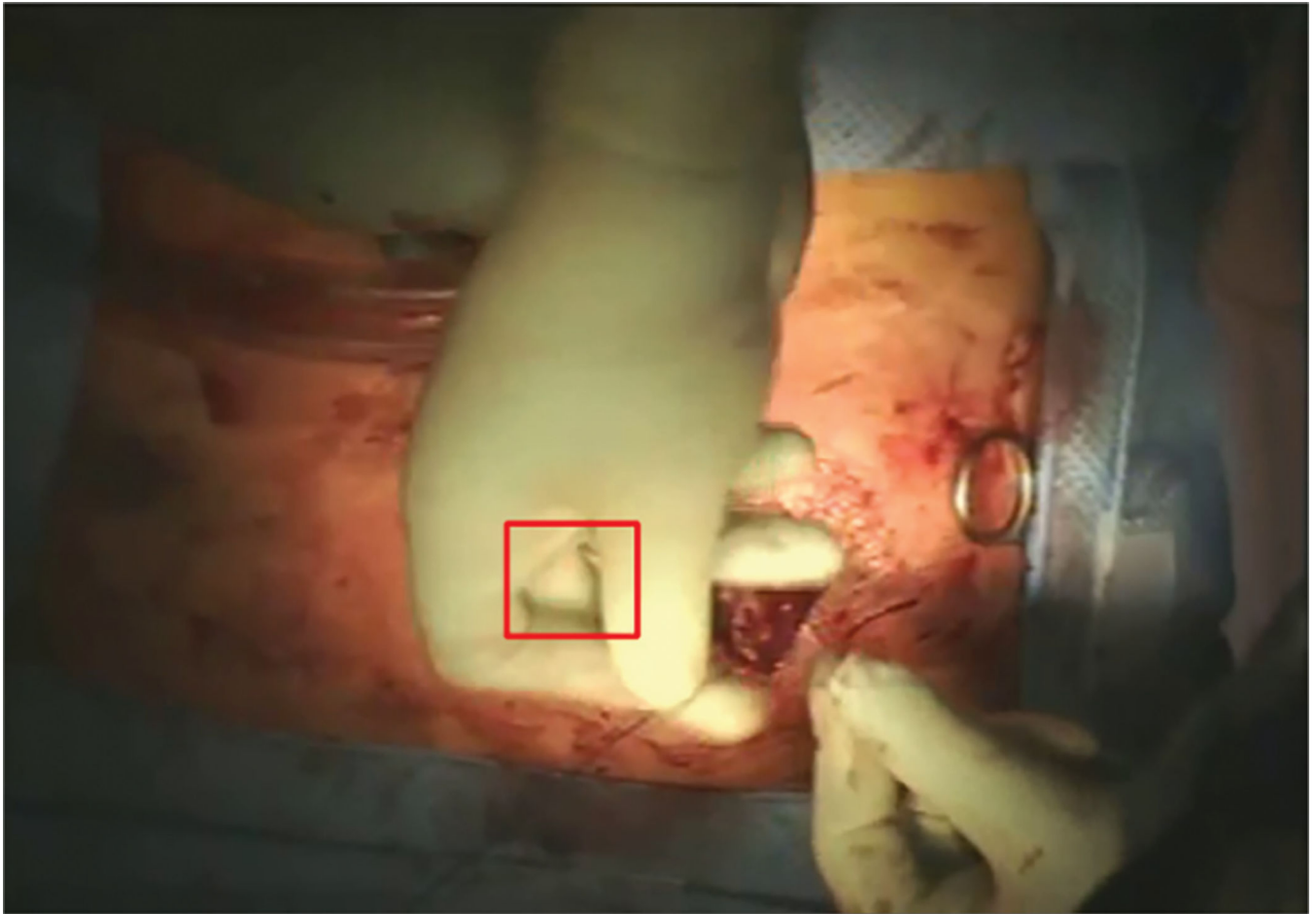
1. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med.* 2013; 369:1434–1442. [PubMed: 24106936]
2. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med.* 1991; 324:370–376. [PubMed: 1987460]
3. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med.* 1991; 324:377–384. [PubMed: 1824793]
4. Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery.* 2003; 133:614–621. [PubMed: 12796727]
5. Rogers SO Jr, Gawande AA, Kwaan M, Puopolo AL, Yoon C, Brennan TA, et al. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery.* 2006; 140:25–33. [PubMed: 16857439]
6. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *Brit Med J.* 2003; 327:1032–1037. [PubMed: 14593041]
7. Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. *Surg Endosc.* 2011; 25:356–366. [PubMed: 20607563]
8. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna G. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg.* 2011; 202:469. [PubMed: 21798511]
9. 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; 193:479–485. [PubMed: 11708503]
10. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. *Am J Surg.* 2002; 184:70–73. [PubMed: 12135725]
11. Bann SD, Khan MS, Darzi AW. Measurements of surgical dexterity using motion analysis of simple bench tasks. *World J Surg.* 2003; 27:390–394. [PubMed: 12658479]
12. Khan MS, Bann SD, Darzi AW, Butler PEM. Assessing surgical skill using bench station models. *Plast Reconstr Surg.* 2006; 120:793–800.

13. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Pappasavvas P, Dosis A, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. *Ann Surg.* 2007; 245:992–999. [PubMed: 17522527]
14. Brydges R, Classen R, Larmer J, Xeroulis G, Dubrowski A. Computer-assisted assessment of one-handed knot tying skills performed within various contexts: a construct validity study. *Am J Surg.* 2006; 192:109–113. [PubMed: 16769286]
15. Mitchell EL, Arora S, Moneta GL, Kret MR, Dargon PT, Landry GJ, et al. A systematic review of assessment of skill acquisition and operative competency in vascular surgery training. *J Vasc Surg.* 2014; 59:1440–1455. [PubMed: 24655750]
16. Duran C, Estrada S, O'Malley M, Sheahan MG, Shames ML, Lee JT, et al. The model for Fundamentals of Endovascular Surgery (FEVS) successfully defines the competent endovascular surgeon. *J Vasc Surg.* 2015; 62:1660–1666. [PubMed: 26598123]
17. Chen, CH.; Hu, YH.; Radwin, RG. A motion tracking system for hand activity assessment. 2014 IEEE China Summit & International Conference. China: Signal and Information Processing (ChinaSIP); 2014. p. 320-324.
18. Glarner CE, Hu Y-Y, Chen C-H, Radwin RG, Zhao Q, Craven MW, et al. Quantifying technical skills during open operations using video-based motion analysis. *Surgery.* 2014; 156:729–734. [PubMed: 24962187]
19. Akkas, O.; Azari, DP.; Chen, CH.; Yen, HH.; Armstrong, TJ.; Ulin, SS. An equation for estimating hand activity level based on measured hand speed and duty cycle. Chicago, IL: Human Factors and Ergonomics Society Annual Meeting; 2014.
20. Greiner, TM. Hand anthropometry of US Army Personnel (No. TR-92/011. Natick, MA: Army Natick Research Development and Engineering Center; 1991.
21. Khan MS, Bann SD, Darzi A, Butler PEM. Use of suturing as a measure of technical competence. *Ann Plast Surg.* 2003; 50:304–309. [PubMed: 12800910]
22. D'Angelo A-LD, Rutherford DN, Ray RD, Laufer S, Kwan C, Cohen ER, et al. Idle time: an underdeveloped performance metric for assessing surgical skill. *Am J Surg.* 2015; 209:645–651. [PubMed: 25725505]
23. D'Angelo A-LD, Rutherford DN, Ray RD, Mason A, Pugh CM. Operative skill: quantifying surgeon's response to tissue properties. *J Surg Res.* 2015; 198:294–298. [PubMed: 26003012]
24. Karam MD, Thomas GW, Koehler DM, Westerlind BO, Lafferty PM, Ohrt GT, et al. Surgical coaching from head-mounted video in the training of fluoroscopically guided articular fracture surgery. *J Bone Joint Surg.* 2015; 97:1031–1039. [PubMed: 26085538]
25. Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush S. Optimal Design Version 3.0. 2013
26. Makary MA. The power of video recording taking quality to the next level. *JAMA.* 2013; 309:1591–1592. [PubMed: 23546452]
27. Moulton C-A, Regehr G, Lingard L, Merritt C, MacRae H. Slowing down to stay out of trouble in the operating room: remaining attentive in automaticity. *Acad Med.* 2010; 85:1571–1577. [PubMed: 20881677]



**Figure 1. Multi-media Video Task Analysis**

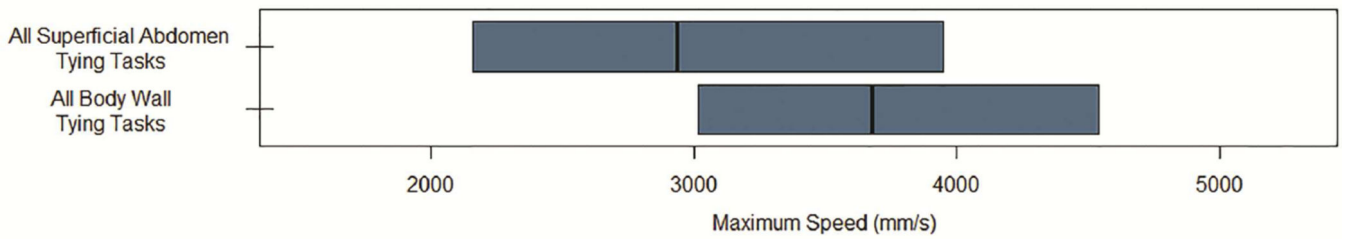
Categories of interest, called records, are created by the user and listed along the left-hand side of the analysis window. Here, attending and resident surgeon activities are identified and tracked separately. Each record can contain multiple events, also created by the user and seen along the right-hand side of the analysis window, which can be marked whenever they occur. For each record, we created events for common surgical tasks (suturing, tying, electrocautery and sharp dissection) and identified when surgeon participants were performing one of these tasks. In this screenshot, the resident surgeon is performing a suturing task.



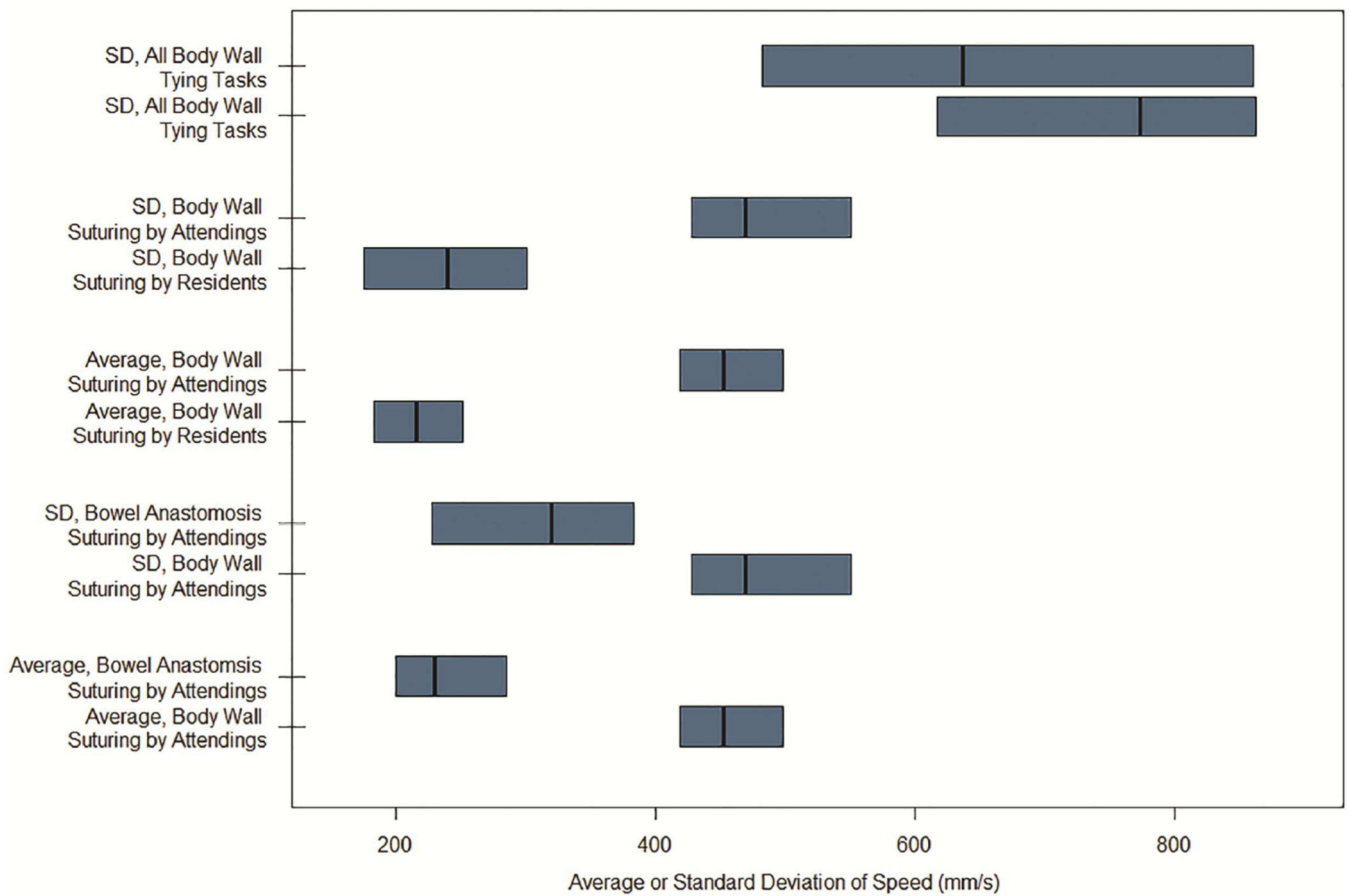
**Figure 2. Marking a Region of Interest**

In this image, a Region of Interest (ROI) shown in the square box, is manually marked on a surgeon's hand. The ROI is tracked over successive frames to measure position, speed, and acceleration of the surgeon's hand.

### Significant Maximum Speed Comparisons



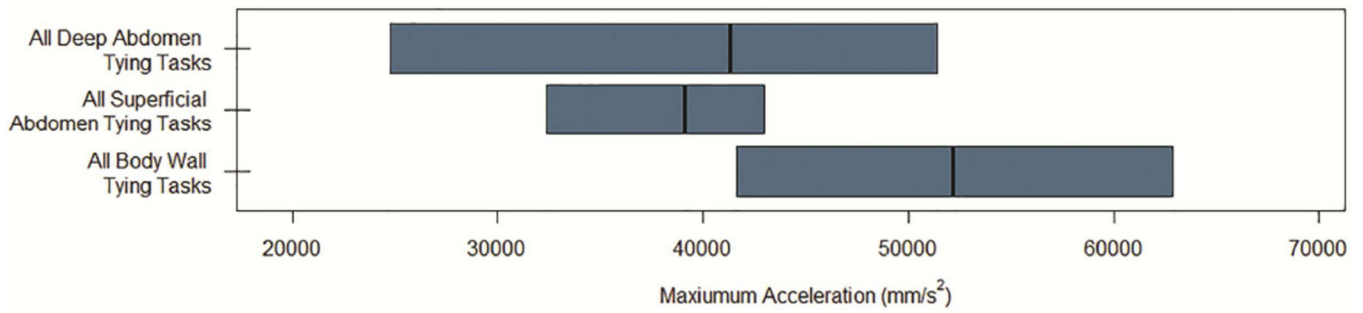
### Significant Speed Comparisons



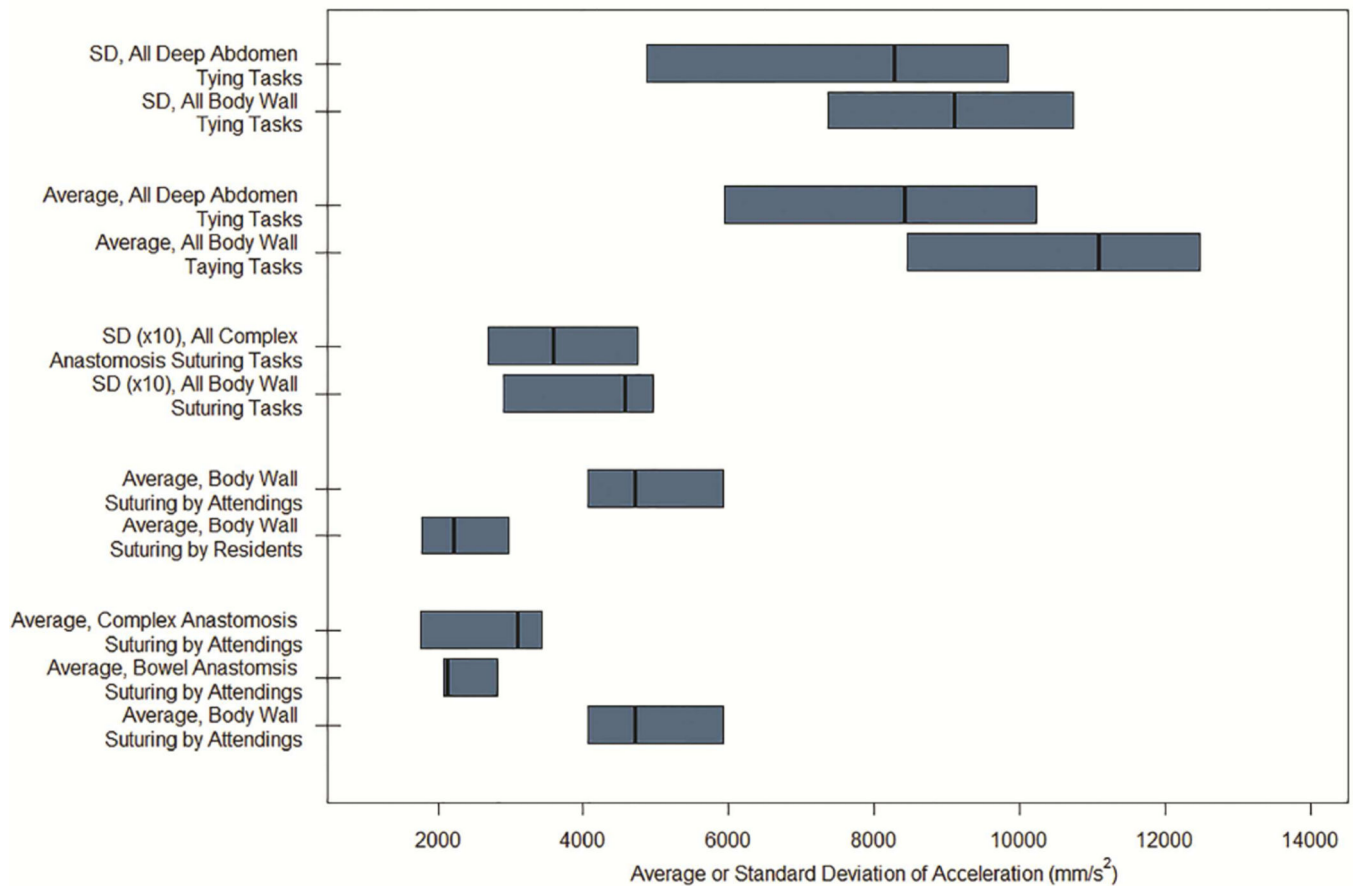
**Figure 3. Significant Comparisons for Speed Kinematics**

Median and interquartile ranges shown for maximum (A) and average or standard deviation (B) of speed. Only comparisons reaching statistical significance ( $p < 0.05$ ) displayed. SD, Standard Deviation; S1, Body Wall Suturing Tasks; S2, Bowel Anastomosis Suturing Tasks; S3, Complex Anastomosis Suturing Tasks; T1, Body Wall Tying Tasks; T2 Superficial Tying Tasks; T3, Deep Tying Tasks.

### Significant Maximum Acceleration Comparisons



### Significant Acceleration Comparisons



**Figure 4. Significant Comparisons for Acceleration Kinematics**

Median and interquartile ranges shown for maximum (A) and average or standard deviation (B) of acceleration. Only comparisons reaching statistical significance ( $p < 0.05$ ) displayed. SD, Standard Deviation; S1, Body Wall Suturing Tasks; S2, Bowel Anastomosis Suturing Tasks; S3, Complex Anastomosis Suturing Tasks; T1, Body Wall Tying Tasks; T2 Superficial Tying Tasks; T3, Deep Tying Tasks. Standard deviation for All Complex



Anastomosis Suturing Tasks and All Body Wall Suturing Tasks are increased (10x) to match x-axis.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**  
**Summary of Video Recording and Analysis by Surgical Task and Surgeon Role**

	Attending	Resident	Total
Number of Cases Recorded *	22	6	22
Number of Cases with Usable Video Clips *	14	6	14
Number of Usable Video Clips	87	51	138
Number of Suturing Clips	39	21	60
S1: Body Wall	19	9	28
S2: Bowel Anastomosis	11	9	20
S3: Complex Anastomosis	9	3	12
Number of Tying Clips	48	30	78
T1: Body Wall	12	8	20
T2: Superficial Abdomen	27	19	46
T3: Deep Abdomen	9	3	12

\* Some cases contain data from both attending and resident surgeons; sum of row may be greater than total

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**  
**Summary of Video Recording and Analysis by Surgeon Participant**

Surgeon ID	Cases Recorded*	Cases with Usable Elements*	Number of Video Clips		Total
			Suturing	Tying	
A	1	1	7	5	12
B	9	5	10	23	33
C	4	4	10	10	20
D	4	2	11	7	18
E	3	1	0	2	2
F	1	1	1	1	2
G	2	2	4	9	13
I	3	3	13	16	29
J	1	1	4	5	9
Total	22	14	60	78	138

\* Some cases contain data from both attending and resident surgeons; sum of column may be greater than total

**Table 3**  
**Observed Mean Kinematics for Suturing Tasks (mean ± SD, n=60)**

Role	n	Speed (mm/s)			Acceleration (mm/s <sup>2</sup> )		
		Average	Maximum	Standard Deviation	Average	Maximum	Standard Deviation
Attending	39	386.70 ± 172.87	2205.69 ± 787.27	421.53 ± 172.85	3700.68 ± 1665.78	22863.61 ± 9905.88	4038.68 ± 1804.09
Resident	21	219.22 ± 60.81	1852.34 ± 920.64	299.94 ± 131.08	2233.81 ± 678.08	18833.46 ± 12416.91	2903.34 ± 1409.52
<b>Task Category</b>							
S1	28	402.88 ± 174.42	2268.77 ± 837.74	431.01 ± 187.26	4093.13 ± 1677.65	23690.76 ± 10788.75	4209.39 ± 1851.40
S2	20	227.60 ± 63.39	1865.40 ± 754.71	320.88 ± 124.33	2237.20 ± 588.06	19726.32 ± 10531.25	3085.44 ± 1231.21
S3	12	321.00 ± 174.65	2007.32 ± 974.91	354.38 ± 162.29	2657.05 ± 1244.83	19109.66 ± 11748.95	3242.23 ± 1966.40
<b>Role by Task Category</b>							
Attending S1	19	488.21 ± 138.49	2552.49 ± 595.20	507.89 ± 160.36	4889.51 ± 1341.74	26728.61 ± 8663.26	4871.29 ± 1641.38
Attending S2	11	240.67 ± 67.77	1663.27 ± 554.39	307.07 ± 101.78	2349.31 ± 630.08	16988.19 ± 6938.71	2910.16 ± 1023.74
Attending S3	9	350.88 ± 194.48	2199.87 ± 1058.19	379.12 ± 182.65	2842.58 ± 1391.98	21873.03 ± 12376.00	3660.23 ± 2128.25
Resident S1	9	222.76 ± 75.68	1733.13 ± 1046.84	268.72 ± 129.70	2411.88 ± 873.05	17277.52 ± 12481.36	2812.04 ± 1506.45
Resident S2	9	211.62 ± 57.27	2112.44 ± 918.55	337.75 ± 152.27	2100.18 ± 535.39	23060.69 ± 13434.32	3299.68 ± 1481.98
Resident S3	3	231.38 ± 20.40	1429.67 ± 284.59	280.14 ± 19.82	2100.45 ± 390.84	10819.56 ± 3015.83	1988.23 ± 65.39

SD, standard deviation; S1, Body Wall; S2, Bowel Anastomosis; S3, Complex Anastomosis

**Table 4**

**Observed Mean Kinematics for Tying Tasks (mean ± SD, n=78)**

Role	n	Speed (mm/s)			Acceleration (mm/s <sup>2</sup> )		
		Average	Maximum	Standard Deviation	Average	Maximum	Standard Deviation
Attending	48	898.64 ± 274.92	3424.19 ± 1037.92	732.74 ± 194.47	10396.13 ± 3396.44	44891.04 ± 13359.27	8672.15 ± 2334.65
Resident	30	735.73 ± 234.61	3317.62 ± 839.43	632.09 ± 169.46	8044.72 ± 2856.61	39177.53 ± 13418.01	6953.70 ± 2169.33
<b>Task Category</b>							
T1	20	905.46 ± 259.89	3627.93 ± 1080.69	730.58 ± 183.58	10934.73 ± 3297.05	50283.28 ± 14377.00	8850.89 ± 2095.39
T2	46	832.75 ± 285.60	3341.46 ± 850.67	685.22 ± 186.46	9218.41 ± 3434.74	40033.44 ± 11550.77	7770.36 ± 2439.40
T3	12	732.59 ± 204.34	3135.31 ± 1148.43	666.84 ± 224.26	8134.52 ± 2639.30	40241.01 ± 15756.85	7534.97 ± 2636.73
<b>Role by Task Category</b>							
Attending T1	12	887.95 ± 261.95	3506.38 ± 1191.98	704.44 ± 182.40	10912.39 ± 3541.25	49500.31 ± 14624.64	8638.78 ± 2122.77
Attending T2	27	938.81 ± 304.08	3359.54 ± 989.15	739.55 ± 205.69	10612.57 ± 3635.43	42463.57 ± 12598.42	8705.58 ± 2575.97
Attending T3	9	792.38 ± 175.65	3508.55 ± 1078.59	750.01 ± 193.00	9058.47 ± 2237.31	46027.78 ± 13615.30	8616.35 ± 2061.89
Resident T1	8	931.71 ± 272.39	3810.26 ± 934.86	769.78 ± 190.44	10968.24 ± 3130.05	51457.73 ± 14910.18	9169.05 ± 2154.89
Resident T2	19	668.02 ± 171.77	3315.78 ± 628.21	608.02 ± 122.85	7237.24 ± 1828.85	36580.11 ± 9101.75	6441.37 ± 1454.50
Resident T3	3	553.23 ± 204.70	2015.60 ± 305.07	417.36 ± 55.85	5362.69 ± 1710.02	22880.67 ± 4611.22	4290.84 ± 428.43

SD, standard deviation; T1, Body Wall; T2, Superficial Tying; T3, Deep Tying

**Table 5**

**Model 1: Evaluation of Interaction Effects**

Model predicted differences (MPD) are listed only for significant measures ( $p < 0.05$ ).

	Speed			Acceleration		
	Average	Maximum	SD	Average	Maximum	SD
	MPD (mm/s)	MPD (mm/s)	MPD (mm/s)	MPD (mm/s <sup>2</sup> )	MPD (mm/s <sup>2</sup> )	MPD (mm/s <sup>2</sup> )
<b>Suturing (n=60)</b>						
Interaction Effect						
Task Category						
Attending: S1 vs S2	<b>243.08</b>	--	<b>200.27</b>	<b>2435.49</b>	--	--
Attending: S1 Vs S3	<b>214.22</b>	--	166.65	<b>2279.00</b>	--	--
Attending: S2 vs S3	-28.87	--	-33.62	-156.49	--	--
Resident: S1 vs S2	3.70	--	-79.15	250.06	--	--
Resident: S1 vs S3	-15.70	--	-27.60	228.66	--	--
Resident: S2 vs S3	-19.41	--	51.55	-21.40	--	--
<b>Role</b>						
S1: Attending v Resident	<b>293.33</b>	--	<b>264.18</b>	<b>2565.25</b>	--	--
S2: Attending v Resident	53.95	--	-15.24	379.82	--	--
S3: Attending v Resident	63.41	--	69.94	514.90	--	--
<b>Tying (n=78)</b>						
Interaction Effect						
Task Category						
Attending: T1 vs T2	--	0.42	--	--	0.56	--
Attending: T1 Vs T3	--	--	--	--	--	--
Attending: T2 vs T3	--	--	--	--	--	--
Resident: T1 vs T2	--	--	--	--	--	--
Resident: T1 vs T3	--	--	--	--	--	--
Resident: T2 vs T3	--	--	--	--	--	--
<b>Role</b>						
T1: Attending v Resident	--	--	--	--	--	--

	Speed			Acceleration			SD	P-value
	Average		Maximum	Average		Maximum		
	MPD (mm/s)	P-value	MPD (mm/s)	MPD (mm/s <sup>2</sup> )	P-value	MPD (mm/s <sup>2</sup> )		
T2: Attending v Resident	--	--	--	--	--	--	--	--
T3: Attending v Resident	--	--	--	--	--	--	--	--

MPD, Model predicted difference; SD, standard deviation; S1, Body Wall; S2, Bowel Anastomosis; S3, Complex Anastomosis; T1, Body Wall; T2, Superficial Tying; T3, Deep Tying

**Table 6**

**Model 2: Main Effects of Role and Task Category**

Model predicted differences (MPD) are listed only for significant measures ( $p < 0.05$ ).

Comparators	Average			Speed			Acceleration			SD				
	MPD (mm/s)	P-value	Maximum	MPD (mm/s)	P-value	Maximum	MPD (mm/s <sup>2</sup> )	P-value	Maximum	MPD (mm/s <sup>2</sup> )	P-value	Maximum	MPD (mm/s <sup>2</sup> )	P-value
Suturing (n=60)														
Role	NR	NR	0.22	NR	NR	NR	NR	NR	0.25	NR	NR	0.25	NR	0.07
Attending vs Resident	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Task Category	NR	NR	0.11	NR	NR	NR	NR	NR	0.14	NR	NR	0.14	NR	<b>0.02</b>
S1 v S2	--	--	--	--	--	--	--	--	--	--	--	--	--	0.16
S1 v S3	--	--	--	--	--	--	--	--	--	--	--	--	--	<b>0.04</b>
S2 v S3	--	--	--	--	--	--	--	--	--	--	--	--	--	0.83
Tying (n=78)														
Role	0.53	0.53	0.22	0.80	0.80	0.46	0.80	0.80	0.74	0.80	0.80	0.74	0.80	0.44
Attending vs Resident	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Task Category	0.16	0.16	<b>0.02</b>	0.04	0.04	<b>0.02</b>	0.04	0.04	< <b>0.01</b>	0.04	0.04	< <b>0.01</b>	0.04	<b>0.03</b>
T1 v T2	--	--	581.18	0.06	0.11	1359.93	0.20	0.20	1121.00	0.20	0.20	1192.66	0.20	0.10
T1 v T3	--	--	<b>904.50</b>	<b>0.03</b>	<b>0.05</b>	<b>3067.20</b>	<b>0.02</b>	<b>0.02</b>	<b>12950.00</b>	<b>0.02</b>	<b>0.02</b>	<b>2022.90</b>	<b>0.02</b>	<b>0.03</b>
T2 v T3	--	--	323.33	0.55	0.60	1707.27	0.20	0.20	1739.77	0.20	0.20	830.5	0.20	0.48
Role by Task Category	0.42	0.42	0.69	0.22	0.22	0.56	0.22	0.22	0.51	0.22	0.22	0.51	0.22	0.18

MPD, Model predicted difference; SD, standard deviation; S1, Body Wall; S2, Bowel Anastomosis; S3, Complex Anastomosis; T1, Body Wall; T2, Superficial Tying; T3, Deep Tying; NR, not reported (due to previously identified interaction effects)