A marker-less technique for measuring kinematics in the operating room

Lane L. Frasier, MD,[a] David P. Azari, MS,[b] Yue Ma, PhD,[a] Sudha R. Pavuluri Quamme, MD, MS,[a] Robert G. Radwin, PhD,[b,c] Carla M. Pugh, MD, PhD,[a] Thomas Y. Yen, PhD,[b,c] Chia-Hsiung Chen, PhD,[d] *and* **Caprice C. Greenberg, MD, MPH,[a,b]** *Madison, WI*

Background. *Often in simulated settings, quantitative analysis of technical skill relies largely on specially tagged instruments or tracers on surgeons' hands. We investigated a novel, marker-less technique for evaluating technical skill during open operations and for 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 duration. Video clips were categorized based on surgeon role (attending, resident) and task subtype (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, multilevel modeling addressed the nonindependence of clips obtained from the same surgeon.*
Results. *Interaction effects for suturing tasks were seen between role and task categories for average speed (P = .04), standard deviation of speed (P = .05), and average acceleration (P = .03). There were significant differences across task categories for standard deviation of acceleration (P = .02). Significant differences for tying tasks across task categories were observed for maximum speed (P = .02); standard deviation of speed (P = .04); and average (P = .02), maximum (P < .01), and standard deviation (P = .03) of acceleration.*
Conclusion. *We demonstrated the ability to detect kinematic differences in performance using marker-less tracking during open operative cases. Suturing task evaluation was most sensitive to differences in surgeon role and task category and may represent a scalable approach for providing quantitative feedback to surgeons about technical skill. (Surgery 2016;■:■-■.)*

From the Wisconsin Surgical Outcomes Research Program, Department of Surgery,[a] the Department of Industrial and Systems Engineering,[b] the Department of Biomedical Engineering,[c] and the Department of Electrical and Computer Engineering,[d] University of Wisconsin-Madison, Madison, WI

LIKE ALL PHYSICIANS, SURGEONS must develop clinical knowledge and the ability to make sound medical decisions as well as the technical skills needed to operate. There is ongoing interest in measuring operative technical performance, especially after Birkmeyer et al[1] demonstrated that surgeons with the best technical skills, as rated by their peers, had the lowest rates of patient complications following bariatric operations. Because technical errors, defined as an error of manual technique that occurs during an operation, account for nearly 75% of the adverse events related to

Supported by the University of Wisconsin Institute for Clinical and Translational Research grant UW #PRJ67FC. Lane Frasier is currently supported by Agency for Healthcare Research and Quality F32 HS022403. She previously received support via National Institutes of Health/National Cancer Institute T32 CA90217 and the Association for Academic Surgery Research Fellowship Award. David Azari, Chia-Hsiung Chen, and Robert Radwin received support via National Institutes of Health 1R21 EB01458301.

Caprice Greenberg serves as a consultant for Johnson & Johnson's Human Performance Institute. Multimedia Video Task Analysis and MVTA 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, and development was sponsored by the donations and gifts of industrial and government members of the Ergonomics Analysis and Design Consortium.

Accepted for publication May 3, 2016.

Reprint requests: Caprice C. Greenberg, MD, MPH, Department of Surgery and Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI 53792-3284. E-mail: greenberg@surgery.wisc.edu.

0039-6060/$ - see front matter

http://dx.doi.org/10.1016/j.surg.2016.05.004

operations,[2-5] measuring operative technical skill is a vital step toward reducing adverse events and improving patient outcomes.

Birkmeyer's approach used a global rating scale to evaluate the technical skill of surgeons. These types of subjective assessment measures are common but suffer from variability between raters and require significant time investment from the raters, limiting their scalability. Early approaches to objective measurement of operative technical skill centered on dexterity analysis systems,[6-8] using sensors mounted on the hands or on special instruments,[9-14] 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.[7]

While this work has advanced our understanding and assessment of the psychomotor properties of operative technical skill, these methods have limited generalizability outside of academic institutions and simulation settings, and they are not easily applied to the analysis of actual, open operations. A review by Reiley et al[7] found that the majority of quantitative assessment systems for operative 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. Similarly, a 2014 review of vascular surgery skill assessment by Mitchell et al[15] found that objective measures of technical skill were obtained in simulated settings or limited to metrics such as procedure time. More recently, Duran et al[16] described 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.

Novel marker-less, video-tracking methods based on cross-correlation, template-matching algorithms were developed by Chen et al[17] in our coauthor's (RGR) laboratory 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 operative 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 (OR) 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 postgraduate 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 surgeons who have 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 via the OR's in-light camera, which streams to a secure hospital computer enabled with a video encoder (AXIS video encoder; Axis Communications, Lund, Sweden). This method of recording was chosen, as in-light cameras are readily available, require no additional equipment with the potential to affect the OR's workflow and personnel movement, and introduce no risk of contaminating the sterile field as would be possible with a pole-mounted or head-mounted camera or video glasses. Because the objective of this work was to evaluate the feasibility of a scalable, automated assessment, we chose to demonstrate feasibility using technology that already exists in increasing numbers of ORs and would not require any additional work on the part of the operative team.

Because 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

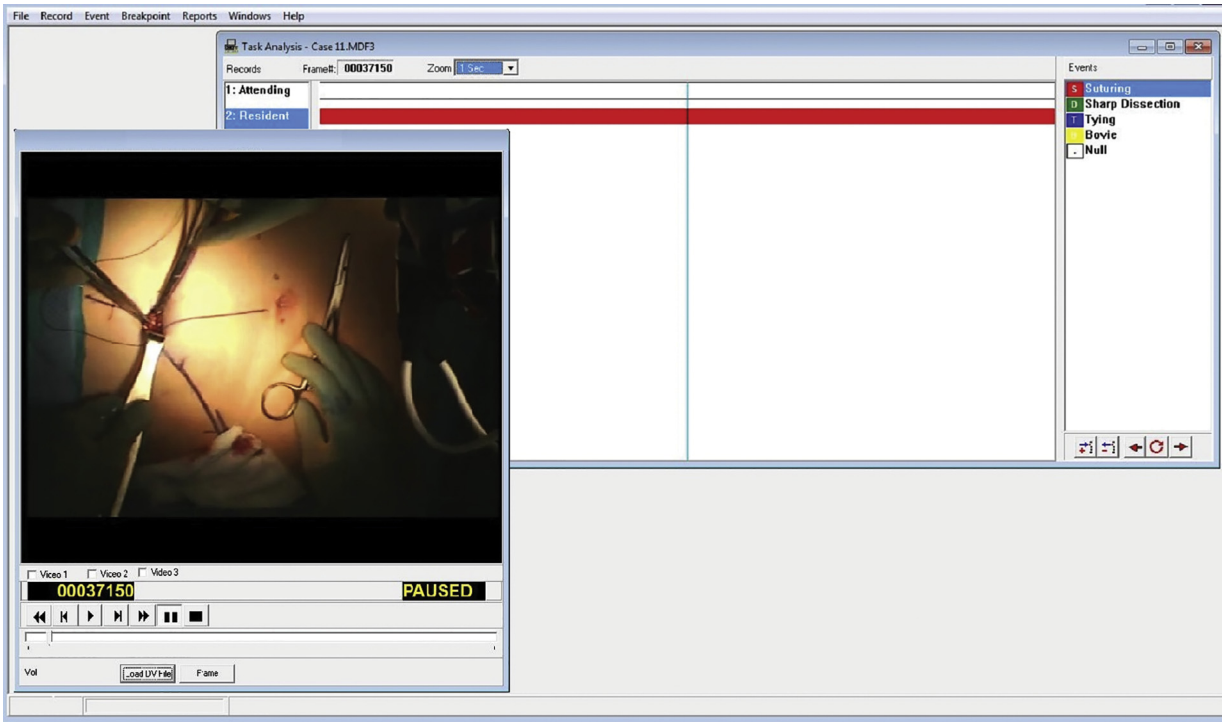

Fig 1. Multimedia video task analysis. Categories of interest, called records, are created by the user and listed along the left-hand side of the analysis window. Here, attending and resident surgeon activities are identified and tracked separately. Each record can contain multiple events, also created 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 operative tasks (suturing and tying) and identified when surgeon participants were performing one of these tasks. In this screenshot, the resident surgeon is performing a suturing task. (Color version of this figure is available online.)

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 from other members of the OR team, and no additional consent was required from the patient.

At the conclusion of each case, participating surgeons 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 20 cases were needed to demonstrate feasibility; 22 cases were recorded, capturing footage from 10 surgeons (6 attendings, 4 residents).

Data analysis. Operative cases were then evaluated using video analysis software (Multimedia Video Task Analysis [MVTA]; Wisconsin Alumni Research Foundation, Madison, WI), developed in our coauthor's (RGR) laboratory 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 (Fig 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 the resident surgeon.

We created events for suturing and tying 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 (LLF) who is a general surgery resident and familiar with the OR and procedures being performed. 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.

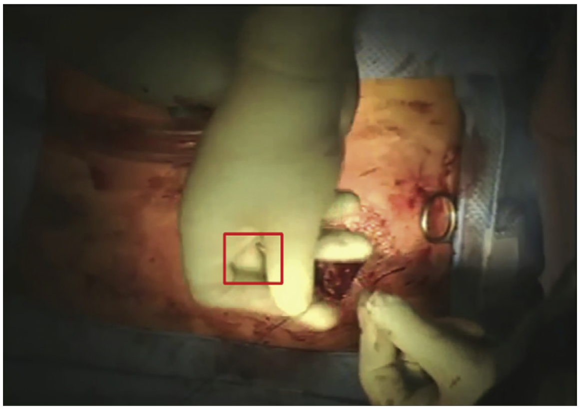

Fig 2. Marking a region of interest. In this image, an ROI, shown in the *square box*, is manually marked on a surgeon's hand. The ROI is tracked over successive frames to measure position, speed, and acceleration of the surgeon's hand. (Color version of this figure is available online.)

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 duration 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 followed a region of interest (ROI) on the surgeon's hands over successive video frames in each clip. Identification of an ROI on a surgeon's dominant hand can be seen in Fig 2. Our group has previously used marker-less tracking software to evaluate descriptive statistics of dominant versus nondominant hand movement during reduction mammoplasty operations.[18] 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's handbreadth were used to calibrate the kinematic record of each clip. Hand dimensions have been shown to provide acceptable calibration for hand speed estimates,[19] and the proximal interphalangeal joint breadth was scaled to the population means of men (23.0 mm) or women (19.9 mm), depending on the sex of the surgeon.

Proximal interphalangeal joint breadth was used because of its small coefficient of variation of 0.071 for men and 0.064 for women, based on anthropomorphic measurements from the US Army.[20] Surgeon hand measurements in pixels were averaged across 3 frames for each video clip; these calibration measurements were averaged for every unique camera–surgeon position. Pixel-millimeter calibrations were recalculated 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[9,11,12,21] and tying[11] 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. On initial analysis, however, 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,[22] suture time, and path length[23] compared with arterial or fascial tissue models.

For these reasons, we further classified each suturing and tying task video clip into 1 of 3 task categories. This was performed by a member of the research team familiar with operative 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 skin level; $n = 20$); and S3: complex anastomosis (which included nonbowel 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; $n = 12$).

Table I. Summary of video recording and analysis by operative 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 attending and resident surgeons; sum of the attending and resident columns may be greater than total.

Table II. Summary of video recording and analysis by surgeon participant

Surgeon ID	*Cases recorded**	*Cases with usable elements**	*Number of video clips: Suturing*	*Tying*	*Total*
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;
**Surgeon H consented to participate in the study, but did not perform any cases in operating rooms with the necessary recording equipment and thus, was not recorded.

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 software (version 9.4 MIXED procedure; SAS Institute, Inc, Cary, NC). To account for the nonindependence of multiple video clips from the same case and multiple cases obtained from the same surgeon, 3-level nested models were applied. Data were compared across surgeon role (attending versus resident) and across task category.

Initially, a 2-factor model was tested using main effects for surgeon role and task category as well as an interaction effect between the 2. If the interaction effect was not statistically significant, the interaction term was excluded and a second 2-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, 2-tailed tests were adopted.

RESULTS

Twenty-two cases were initially recorded. Recorded cases represented a broad spectrum of operations and included open hernia repair, complex colorectal, hepatobiliary, surgical oncology, and thoracic cases. Of these, 16 yielded 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 operative task for attending and resident participants are listed in Table I. The number of cases recorded and clips obtained for individual surgeon participants, ranging from 1–5 cases, are shown in Table II.

Observed means: Suturing. Sixty suturing clips were analyzed. Observed means are summarized in Table III. 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) tasks 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 IV. When tying tasks were assessed overall, attending surgeons demonstrated higher means across all measures. When assessed overall, T1 (body wall) tasks had higher observed means than T2 (superficial tying) tasks, and T2 tasks had higher observed means than T3 (deep tying) tasks for all kinematic measures except maximum acceleration; 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

Table III. Observed mean kinematics for suturing tasks (mean ± SD, n = 60)

		Speed (mm/s)			Acceleration (mm/s²)		
	n	Average	Maximum	Standard deviation	Average	Maximum	Standard deviation
Role							
Attending	39	386.70 ± 172.87	2,205.69 ± 787.27	421.53 ± 172.85	3,700.68 ± 1,665.78	22,863.61 ± 9,905.88	4,038.68 ± 1,804.09
Resident	21	219.22 ± 60.81	1,852.34 ± 920.64	299.94 ± 131.08	2,233.81 ± 678.08	18,833.46 ± 12,416.91	2,903.34 ± 1,409.52
Task category							
S1	28	402.88 ± 174.42	2,268.77 ± 837.74	431.01 ± 187.26	4,093.13 ± 1,677.65	23,690.76 ± 10,788.75	4,209.39 ± 1,851.40
S2	20	227.60 ± 63.39	1,865.40 ± 754.71	320.88 ± 124.33	2,237.20 ± 588.06	19,726.32 ± 10,531.25	3,085.44 ± 1,231.21
S3	12	321.00 ± 174.65	2,007.32 ± 974.91	354.38 ± 162.29	2,657.05 ± 1,244.83	19,109.66 ± 11,748.95	3,242.23 ± 1,966.40
Role by task category							
Attending S1	19	488.21 ± 138.49	2,552.49 ± 595.20	507.89 ± 160.36	4,889.51 ± 1,341.74	26,728.61 ± 8,663.26	4,871.29 ± 1,641.38
Attending S2	11	240.67 ± 67.77	1,663.27 ± 554.39	307.07 ± 101.78	2,349.31 ± 630.08	16,988.19 ± 6,938.71	2,910.16 ± 1,023.74
Attending S3	9	350.88 ± 194.48	2,199.87 ± 1,058.19	379.12 ± 182.65	2,842.58 ± 1,391.98	21,873.03 ± 12,376.00	3,660.23 ± 2,128.25
Resident S1	9	222.76 ± 75.68	1,733.13 ± 1,046.84	268.72 ± 129.70	2,411.88 ± 873.05	17,277.52 ± 12,481.36	2,812.04 ± 1,506.45
Resident S2	9	211.62 ± 57.27	2,112.44 ± 918.55	337.75 ± 152.27	2,100.18 ± 535.39	23,060.69 ± 13,434.32	3,299.68 ± 1,481.98
Resident S3	3	231.38 ± 20.40	1,429.67 ± 284.59	280.14 ± 19.82	2,100.45 ± 390.84	10,819.56 ± 3,015.83	1,988.23 ± 65.39

SD, Standard deviation.

Table IV. Observed mean kinematics for tying tasks (mean ± SD, n = 78)

		Speed (mm/s)			Acceleration (mm/s²)		
	n	Average	Maximum	Standard deviation	Average	Maximum	Standard deviation
Role							
Attending	48	898.64 ± 274.92	3,424.19 ± 1,037.92	732.74 ± 194.47	10,396.13 ± 3,396.44	44,891.04 ± 13,359.27	8,672.15 ± 2,334.65
Resident	30	735.73 ± 234.61	3,317.62 ± 839.43	632.09 ± 169.46	8,044.72 ± 2,856.61	39,177.53 ± 13,418.01	6,953.70 ± 2,169.33
Task category							
T1	20	905.46 ± 259.89	3,627.93 ± 1,080.69	730.58 ± 183.58	10,934.73 ± 3,297.05	50,283.28 ± 14,377.00	8,850.89 ± 2,095.39
T2	46	832.75 ± 285.60	3,341.46 ± 850.67	685.22 ± 186.46	9,218.41 ± 3,434.74	40,033.44 ± 11,550.77	7,770.36 ± 2,439.40
T3	12	732.59 ± 204.34	3,135.31 ± 1,148.43	666.84 ± 224.26	8,134.52 ± 2,639.30	40,241.01 ± 15,756.85	7,534.97 ± 2,636.73
Role by task category							
Attending T1	12	887.95 ± 261.95	3,506.38 ± 1,191.98	704.44 ± 182.40	10,912.39 ± 3,541.25	49,500.31 ± 14,624.64	8,638.78 ± 2,122.77
Attending T2	27	938.81 ± 304.08	3,359.54 ± 989.15	739.55 ± 205.69	10,612.57 ± 3,635.43	42,463.57 ± 12,598.42	8,705.58 ± 2,575.97
Attending T3	9	792.38 ± 175.65	3,508.55 ± 1,078.59	750.01 ± 193.00	9,058.47 ± 2,237.31	46,027.78 ± 13,615.30	8,616.35 ± 2,061.89
Resident T1	8	931.71 ± 272.39	3,810.26 ± 934.86	769.78 ± 190.44	10,968.24 ± 3,130.05	51,457.73 ± 14,910.18	9,169.05 ± 2,154.89
Resident T2	19	668.02 ± 171.77	3,315.78 ± 628.21	608.02 ± 122.85	7,237.24 ± 1,828.85	36,580.11 ± 9,101.75	6,441.37 ± 1,454.50
Resident T3	3	553.23 ± 204.70	2,015.60 ± 305.07	417.36 ± 55.85	5,362.69 ± 1,710.02	22,880.67 ± 4,611.22	4,290.84 ± 428.43

SD, Standard deviation.

and T3 tasks but lower means for all measures of T1 tasks compared with residents.

Model 1: Predicted main effects with inclusion of an interaction effect. Interaction effects in the nesting model (Table V) for suturing tasks were seen for several measures related to speed and acceleration, including average speed (P = .04), standard deviation of speed (P = .05), and average acceleration (P = .03). When comparing suturing tasks, attending surgeons had significantly higher average speed, standard deviation of speed, and average acceleration for S1 (body wall) compared with S2 (bowel anastomosis) tasks (P < .01, P = .04, and P < .01, respectively). Additionally, attending surgeons had significantly higher average speed (P < .01) and acceleration (P < .01) for S1 tasks compared with 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 = .03), standard deviation of speed (P = .05), and average acceleration (P = .03) compared with residents. No differences between attending and resident surgeons were seen for S2 or S3 tasks. Significant suturing comparisons are displayed in Fig 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 with 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 2-factor model (role and task category, Table VI). As interaction effects were previously identified for several suturing measures, results from the 2-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 from residents for average speed, average acceleration, or standard deviation of acceleration. However, when comparing task categories, S1 (body wall) tasks had higher standard deviation of acceleration compared with S3 (complex anastomosis) tasks (P = .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 with T2 tasks (P < .01) and significantly higher predicted maximum speed, standard deviation of speed, average and maximum acceleration, and standard deviation of acceleration compared with T3 tasks (P = .03, .05, .02, .2, and .03, respectively).

DISCUSSION

This study has demonstrated the feasibility of obtaining kinematic data for 2 common operative tasks during open operations, without necessitating specialized tracking devices or limiting analysis to laparoscopic cases. Previous work evaluating descriptive differences in dominant versus nondominant hand movements in attending surgeons versus residents[18] demonstrated the feasibility of this approach for open operative cases. Here, we have extended this work and sought to discern differences based on surgeon role and task relationship to the various stages of an operation, 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 2 tasks encountered in a majority of operative procedures represents a step forward in a scalable methodology to assess operative skill under a variety of actual operative 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.

Based on the type of task performed, we identified significant differences in speed and acceleration metrics for suturing and tying tasks. Our findings are consistent with prior work in which these tasks consistently differentiated surgeons based on experience level.[11] We also build on recent work that identified differences in idle time,[22] path length, and suture time[23] for suturing tasks performed on models simulating more or less complex tissues. Our approach, however, is novel in that it consistently identified differences between attendings and residents and between task categories, during a wide variety of actual operations, without restricting task parameters commonly used in highly controlled benchtop assessments.

Even more exciting was our identification of interaction effects between surgeon role and task category for suturing. In other words, attending

Table V. Model 1: Evaluation of interaction effects

	Speed						Acceleration					
	Average		*Maximum*		*SD*		*Average*		*Maximum*		*SD*	
	MPD (mm/s)	P *value*	*MPD (mm/s)*	P *value*	*MPD (mm/s)*	P *value*	*MPD (mm/s^2)*	P *value*	*MPD (mm/s^2)*	P *value*	*MPD (mm/s^2)*	P *value*
Suturing (n = 60)												
Interaction effect		.04		.12		.05		.03		.09		.12
Task category												
Attending: S1 versus S2	243.08	<.01	—	—	200.27	.04	2,435.49	<.01	—	—	—	—
Attending: S1 versus S3	214.22	<.01	—	—	166.65	.10	2,279.00	<.01	—	—	—	—
Attending: S2 versus S3	−28.87	1.00	—	—	−33.62	1.00	−156.49	1.00	—	—	—	—
Resident: S1 versus S2	3.70	1.00	—	—	−79.15	.93	250.06	1.00	—	—	—	—
Resident: S1 versus S3	−15.70	1.00	—	—	−27.60	1.00	228.66	1.00	—	—	—	—
Resident: S2 versus S3	−19.41	1.00	—	—	51.55	1.00	−21.40	1.00	—	—	—	—
Role												
S1: attending versus resident	293.33	.03	—	—	264.18	.05	2,565.25	.03	—	—	—	—
S2: attending versus resident	53.95	.98	—	—	−15.24	1.00	379.82	.98	—	—	—	—
S3: attending versus resident	63.41	.99	—	—	69.94	.99	514.90	.99	—	—	—	—
Tying (n = 78)												
Interaction effect	—	.42	—	.69	—	.22	—	.56	—	.51	—	.18
Task category												
Attending: T1 versus T2	—	—	—	—	—	—	—	—	—	—	—	—
Attending: T1 versus T3	—	—	—	—	—	—	—	—	—	—	—	—
Attending: T2 versus T3	—	—	—	—	—	—	—	—	—	—	—	—
Resident: T1 versus T2	—	—	—	—	—	—	—	—	—	—	—	—
Resident: T1 versus T3	—	—	—	—	—	—	—	—	—	—	—	—
Resident: T2 versus T3	—	—	—	—	—	—	—	—	—	—	—	—
Role												
T1: attending versus resident	—	—	—	—	—	—	—	—	—	—	—	—
T2: attending versus resident	—	—	—	—	—	—	—	—	—	—	—	—
T3: attending versus resident	—	—	—	—	—	—	—	—	—	—	—	—

MPDs are listed only for significant measures ($P < .05$).
MPD, Model predicted difference; *SD*, standard deviation.

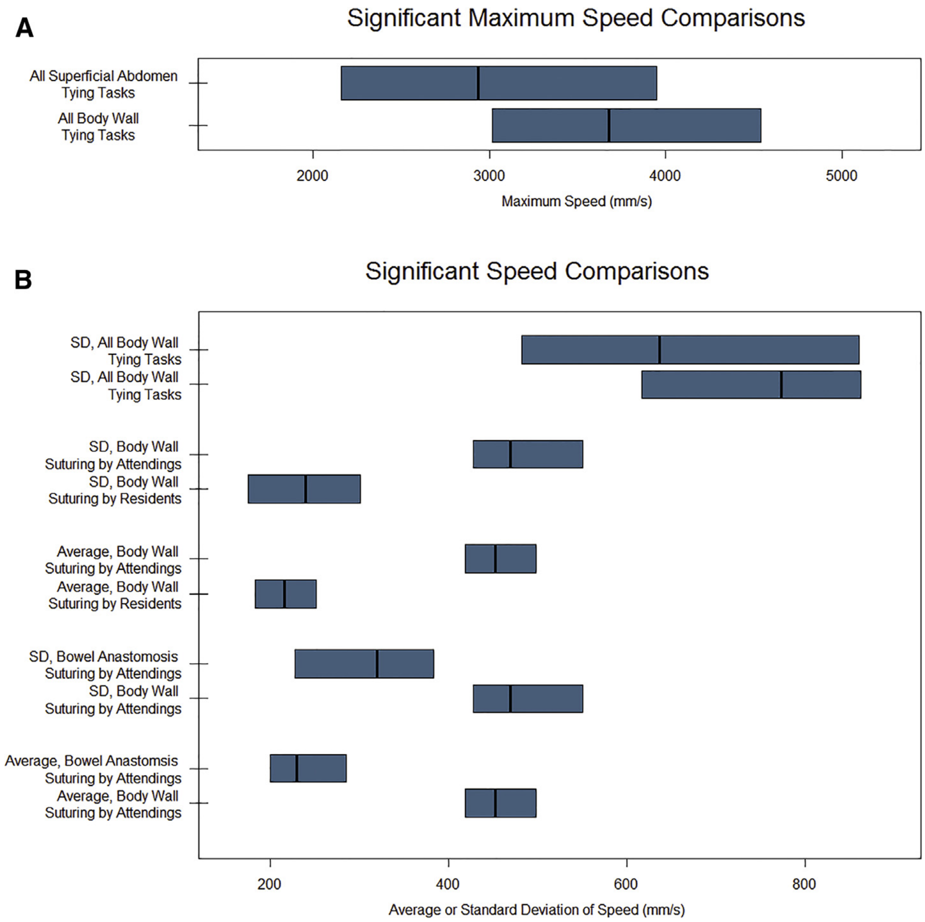

Fig 3. Significant comparisons for speed kinematics. Median and interquartile ranges shown for maximum (*A*) and average or standard deviation (*B*) of speed. Only comparisons reaching statistical significance ($P < .05$) are displayed. (Color version of this figure is available online.)

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 operative technical skill, suggests that assessment of suturing technical skill is most sensitive to surgeon experience and complexity of the operation, 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 they were at sewing more complex anastomoses. This was seen when assessing tasks overall and also when assessing 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 with attending surgeons. These findings seem to indicate that increasing technical skill does not translate solely to increased speed and acceleration. For example, the decreased acceleration seen in S2 tasks performed by attendings may represent a smoother, steadier pace compared with 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

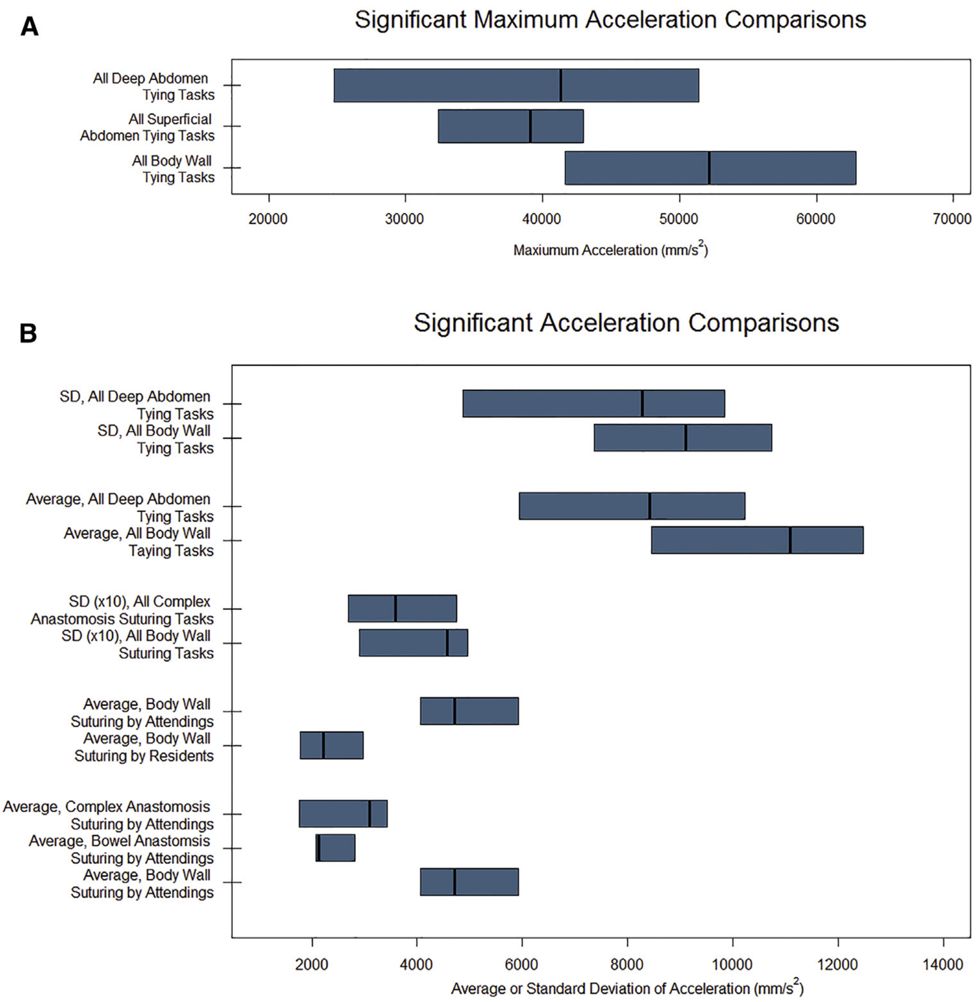

Fig 4. Significant comparisons for acceleration kinematics. Median and interquartile ranges shown for maximum (*A*) and average or standard deviation (*B*) of acceleration. Only comparisons reaching statistical significance ($P < .05$) are displayed. Standard deviation for all complex anastomosis suturing tasks and all body wall suturing tasks are increased (10×) to match x-axis. (Color version of this figure is available online.)

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 with 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 noninvasive 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, a setup that 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 (GoPro, Inc., San Mateo, CA) and mobile video glasses, have also been successfully employed by our group and others.[18,24]

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 subcategories (eg, T1 versus T2, S1 versus S3) could lead to

Table VI. Model 2: Main effects of role and task category

	Speed						*Acceleration*					
	Average		*maximum*		*SD*		*Average*		*maximum*		*SD*	
Comparators	*MPD (mm/s)*	P *value*	*MPD (mm/s)*	P *value*	*MPD (mm/s)*	P *value*	*MPD* (mm/s^2)	P *value*	*MPD* (mm/s^2)	P *value*	*MPD* (mm/s^2)	P *value*
Suturing (*n* = 60)												
Role		NR		.22		NR		NR		.25		.07
Attending versus resident	—	—	—	—	—	—	—	—	—	—	—	—
Task category		NR		.11		NR		NR		.14		.02
S1 versus S2	—	—	—	—	—	—	—	—	—	—	1,115.46	.16
S1 versus S3	—	—	—	—	—	—	—	—	—	—	1,543.54	.04
S2 versus S3	—	—	—	—	—	—	—	—	—	—	428.09	.83
Tying (*n* = 78)												
Role		.53		.22		.80		.46		.74		.44
Attending versus resident	—	—	—	—	—	—	—	—	—	—	—	—
Task category		.16		.02		.04		.02		<.01		.03
T1 versus T2	—	—	581.18	.06	91.85	.11	1,359.93	.20	1,121.00	<.01	1,192.66	.10
T1 versus T3	—	—	904.50	.03	145.67	.05	3,067.20	.02	12,950.00	.02	2,022.90	.03
T2 versus T3	—	—	323.33	.55	53.82	.60	1,707.27	.20	1,739.77	.91	830.5	.48
Role by task category		.42		.69		.22		.56		.51		.18

MPDs are listed only for significant measures ($P < .05$).
MPD, Model predicted difference; *SD*, standard deviation; *NR*, not reported (due to previously identified interaction effects).

the development of software capable of recognizing the kinematic patterns associated with suturing and tying tasks and subtasks, 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 were 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 attained 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 (Optimal Design is open source sponsored by the William T. Grant Foundation, New York, NY),[25] 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 (ie, number of surgeons, number of cases performed per surgeon, and number of video clips included per case), intraclass correlations at surgeon level and at case level, and effect size (ie, 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 (0.80) for all 6 kinematic measures, which could have led to the statistical insignificance. Future studies with increased sample sizes and improved procedures to exclude unusable video clips 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.[26] 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 obtain information easily 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 learners with objective feedback 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 and 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 OR after time away for medical or personal reasons.

Prior research indicates that the number of hand movements[9-11] and time taken[9-12,21] decreases for a given task with increased surgical 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.

Increased efficiency alone, however, 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.[27] This transition may occur deliberately, at preidentified stages in the operation related to patient- or procedure-specific characteristics, or with intraoperative 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 are included in hidden Markov models. The work we are doing now can inform development of such multivariable and multidimensions models.

Further work is needed to determine whether 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.

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

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

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;369:1434-42.
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;324:370-6.
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;324:377-84.
4. Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. Surgery 2003;133:614-21.
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.
6. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. Brit Med J 2003;327:1032-7.
7. Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. Surg Endosc 2011;25:356-66.
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.
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;193:479-85.
10. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. Am J Surg 2002;184:70-3.
11. Bann SD, Khan MS, Darzi AW. Measurements of surgical dexterity using motion analysis of simple bench tasks. World J Surg 2003;27:390-4.
12. Khan MS, Bann SD, Darzi AW, Butler PE. Assessing surgical skill using bench station models. Plast Reconstr Surg 2006;120:793-800.
13. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 2007;245:992-9.
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;192:109-13.
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;59:1440-55.
16. Duran C, Estrada S, O'Malley M, Sheahan MG, Shames ML, Lee JT, et al. The model for Funadmentals of Endovascular Surgery (FEVS) successfully defines the competent endovascular surgeon. J Vasc Surg 2015;62:1660-6.
17. Chen CH, Hu YH, Radwin RG. A motion tracking system for hand activity assessment. In: Signal and Information Processing (ChinaSIP). IEEE China Summit & International Conference; 2014 Jul 9-13; Xi'an, China. Piscataway (NJ): IEEE; 2014:320-4.
18. Glarner CE, Hu YY, Chen CH, Radwin RG, Zhao Q, Craven MW, et al. Quantifying technical skills during open operations using video-based motion analysis. Surgery 2014;156:729-34.

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. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2014;58:1600–4.
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 PE. Use of suturing as a measure of technical competence. Ann Plast Surg 2003;50:304-9.
22. D'Angelo AL, 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;209:645-51.
23. D'Angelo AL, Rutherford DN, Ray RD, Mason A, Pugh CM. Operative skill: Quantifying surgeon's response to tissue properties. J Surg Res 2015;198:294-8.
24. Karam MD, Thomas GW, Koehler DM, Westerlind BO, Lafferty PM, Ohrt GT, et al. Surgical coaching from head-mounted video in the tarining of fluoroscopically guided articular fracture surgery. J Bone Joint Surg 2015;97:1031-9.
25. Raudenbush, SW, Spybrook, J, Congdon, R, Liu, X, Martinez, A, Bloom, H, Hill, C. (2011). Optimal Design Plus Empirical Evidence (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.
26. Makary MA. The power of video recording taking quality to the next level. JAMA 2013;309:1591-2.
27. Moulton CA, 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;85:1571-7.