

Objective measures for longitudinal assessment of robotic surgery training

Rajesh Kumar, PhD,^a Amod Jog, MSE,^a Balazs Vagvolgyi, MSE,^a Hiep Nguyen, MD,^b Gregory Hager, PhD,^a Chi Chiung Grace Chen, MD,^c and David Yuh, MD^d

Objectives: Current robotic training approaches lack the criteria for automatically assessing and tracking (over time) technical skills separately from clinical proficiency. We describe the development and validation of a novel automated and objective framework for the assessment of training.

Methods: We are able to record all system variables (stereo instrument video, hand and instrument motion, buttons and pedal events) from the da Vinci surgical systems using a portable archival system integrated with the robotic surgical system. Data can be collected unsupervised, and the archival system does not change system operations in any way. Our open-ended multicenter protocol is collecting surgical skill benchmarking data from 24 trainees to surgical proficiency, subject only to their continued availability. Two independent experts performed structured (objective structured assessment of technical skills) assessments on longitudinal data from 8 novice and 4 expert surgeons to generate baseline data for training and to validate our computerized statistical analysis methods in identifying the ranges of operational and clinical skill measures.

Results: Objective differences in operational and technical skill between known experts and other subjects were quantified. The longitudinal learning curves and statistical analysis for trainee performance measures are reported. Graphic representations of the skills developed for feedback to the trainees are also included.

Conclusions: We describe an open-ended longitudinal study and automated motion recognition system capable of objectively differentiating between clinical and technical operational skills in robotic surgery. Our results have demonstrated a convergence of trainee skill parameters toward those derived from expert robotic surgeons during the course of our training protocol. (*J Thorac Cardiovasc Surg* 2012;143:528-34)

Minimally invasive cardiothoracic operations have been facilitated by new surgical robotic technologies. Although more than 1700 surgical robotic systems were in clinical use worldwide¹ by mid-2010, the application of robotics to cardiothoracic surgery has not caught up with other surgical disciplines largely because of the steep learning curves in developing operational proficiency with surgical

robotic platforms,^{2,3} coupled with comparatively lower tolerances for technical error and delay. Specifically, the technical challenges presented in performing precise and complex reconstructive techniques with limited access and the longer cardiopulmonary bypass and aortic crossclamp times associated with robot-assisted cardiac operations²⁻⁴ have hampered widespread acceptance of robotics in the cardiothoracic surgical community. A smaller user base, the slow refinement of the technology, and, consequently, the slow accumulation of evidence of clinical benefit has also slowed the adoption of the new technology. Improved adoption and the use of robotic surgery technology will require improvements in both technology and training methods.

The traditional Halstedian principles of surgical training using a “see one, do one, teach one” apprenticeship model are not wholly applicable to surgical robotic training. To develop clinical proficiency, effective training and practice strategies to familiarize surgeons with new robotic technologies are required.^{2,3} However, currently practiced robotic training approaches lack uniform criteria for automatically assessing and tracking technical and operational skills. Establishing standard, objective, and automated skill measures, leading to effective training curricula and certification programs for robotic surgery will require (1) a significant cohort of robotic surgeons-in-training of similar

From the Department of Computer Science,^a Johns Hopkins University, Baltimore, Md; Department of Urology Children’s Hospital,^b Boston, Mass; Department of Obstetrics and Gynecology,^c Johns Hopkins Hospital, Baltimore, Md; and Section of Cardiac Surgery,^d Yale University School of Medicine, New Haven, Conn.

This work was supported in part by National Institutes of Health grant 1R21EB009143-01A1, National Science Foundation grants 0941362 and 0931805, and Johns Hopkins University internal funds. A collaborative agreement with Intuitive Surgical, Inc., with no financial support provided access to the Application Programming Interface data. The sponsors played no role in the study development, data collection, interpretation of data, or analysis.

Disclosures: The study sponsors played no role in the collection, analysis, or interpretation of the data. Dr Kumar was employed by Intuitive Surgical before joining Johns Hopkins 5 years previously. Intuitive Surgical is a publically traded company. The other authors (Drs Kumar, Jog, Tantillo, Vagvolgyi, Chen, Nguyen, and Hager) might own or have owned a small number of Intuitive securities (<\$10,000).

Received for publication July 4, 2011; revisions received Sept 11, 2011; accepted for publication Nov 7, 2011; available ahead of print Dec 15, 2011.

Address for reprints: Rajesh Kumar, PhD, Department of Computer Science, Johns Hopkins University, Hackerman Hall 117, 3400 North Charles Street, Baltimore, MD 21218 (E-mail: rajesh@jhu.edu).

0022-5223/\$36.00

Copyright © 2012 by The American Association for Thoracic Surgery

doi:10.1016/j.jtcvs.2011.11.002

Abbreviations and Acronyms

- API = application programming interface
- OSATS = objective structured assessment of technical skills
- SVM = support vector machine

skill who can be tracked longitudinally (eg, 1 year) during the acquisition of skills; (2) a set of standardized surgical tasks; (3) the ability to acquire and analyze large volumes of motion data; and (4) consistent “ground truth” assessment of the collected data by expert robotic surgeons.

Published research in robotic surgery training has been limited to the quantification of skill measures from *ab initio* training^{5,6} of relatively short duration. Previous efforts to objectively quantify the measures of skill on a limited number of trainees^{7,8} have also been predicated on comparing trainees of varying skill levels (eg, postgraduate year of training) with “expert” surgeons. These studies used the experimental tasks for both training and assessment. Robotic surgical systems require complex man–machine interactions, and robotic training assessment protocols cannot yet differentiate between the objective assessment of clinical task skills and machine operational and technical skills.

Consequently, we opted to take a new approach by developing a novel automated motion recognition system capable of objectively differentiating between operational and technical robotic surgical skills and longitudinally tracking trainees during skill development. We established multiple learning curves for each training step, provided a comparative analysis of skill development, and developed methods

for feedback to effectively address skill deficiencies. We also use our tasks as *benchmark evaluations*, not as training tasks. This is also the first longitudinal multicenter study involving robotic surgical training and constitutes the largest trainee cohort to date.

METHODS

The measurement of objective performance metrics in surgical training (ie, efficiency of hand movement) has previously required instrumented prototype devices that are not widely available, interfere with the surgical technique, and employ technologies that are not commonly available or easily integrated into conventional surgical instrumentation.⁹ As a novel “transparent” alternative, we have developed a new infrastructure to collect motion and video data from robotic surgical training that does not require any special instrumentation and holds the promise of creating a training environment that does not require on-site supervision by an expert surgeon. Although the methods described can be extended to use during human surgery, the present work primarily describes their application during controlled, structured environment during *ab initio* training.

Data Collection

Our motion data collection platform is based on the da Vinci surgical robotic system (Intuitive Surgical, Sunnyvale, Calif). Its Application Programming Interface (API)¹⁰ provides a robust motion data set containing 334 position and motion parameters. The API automatically streams the motion vectors, including joint angles, Cartesian position and velocity, gripper angle, and joint velocity and torque data for the master console manipulators, stereoscopic camera, and instruments over an Ethernet connection to an encrypted archival workstation. The API also streams several system events, including instrumentation changes, manipulator “clutching,” and visual field adjustments. The API can provide faster motion data acquisition rates (≤ 100 Hz) than those obtained with video recordings (typically ≤ 30 Hz). In addition, high-quality time-synchronized video can be acquired from the stereoscopic video system. Using the data collection framework (Figure 1, left), 334 system variables were sampled at 50 Hz, and stereoscopic video streams were collected at 30 Hz. These data were anonymized at the source, assigned a unique subject identifier, and

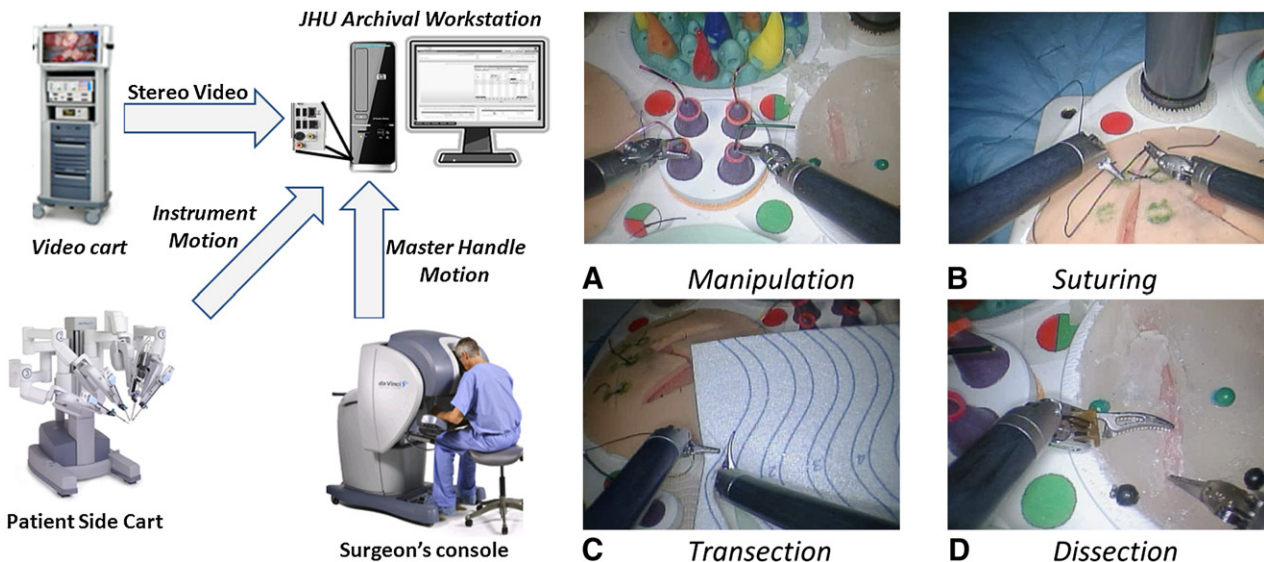


FIGURE 1. Archival system configuration with the da Vinci system (left) and inanimate training pods (right) for first robotic surgery training module.

archived in a database according to an approved institutional review board-approved protocol. For analysis, the archived data was also segmented into the task or system operation sequences. This process generated 20 to 25 GB of data per hour. No special training was required to operate the archival workstation, which can be left connected in place without affecting surgical or other training use.

Experimental Tasks

Training data were collected during all stages of training. Our training protocol was divided into different training modules:

Module I: System orientation skills. This training module is intended to familiarize the trainee with basic system and surgical skills, including master console clutching, camera control, manipulation scale change, retraction, suturing, tissue handling, bimanual manipulation, transection, and dissection. Because robotic trainees already practice these basic skills in current training regimens, we believed they were appropriate for benchmarking. On a monthly basis, we collected data from periodic benchmarking executions of 4 minimally invasive surgical skills taken from the Intuitive Surgical robotic surgery training practicum.¹¹ These tasks (Figure 1) are as follows:

Manipulation: The manipulation task tests the subject’s system operation skills. It requires the transfer of 4 rings from the center pegs of the task pod to the corresponding outer peg, followed by replacement of the rings to the inner pegs in sequence. The elementary task performance measures included task completion times and task errors (eg, dropped ring/peg, moving instruments outside of the field of view).

Suturing: The suturing task involves the repair of a linear defect with three 10-cm lengths of 3-0 Vicryl suture. The elementary task performance measures included task completion times and task errors (eg, dropped needles, broken sutures, inaccurate approximation).

Transection: The transection task involves cutting an S or circle pattern on a transection pod using curved scissors while stabilizing the pod with the third arm. The elementary task performance measures included task completion times and task errors (eg, cutting outside of the pattern).

Dissection: The dissection task requires dissection of a superficial layer of the pod to gain exposure to a buried vessel, followed by circumferential dissection to fully mobilize the vessel. The task completion times and errors (eg, damage to the vessel, incomplete mobilization, and excessive dissection) were measured.

These orientation laboratories typically produced 1 hour of training data. On successful acquisition of these basic skills, the trainees were graduated to module II.

Module II: Minimally invasive surgical skills. The minimally invasive surgical skills module is intended to familiarize the trainee with basic minimally invasive surgical skills, including port placement, instrument exchange, complex manipulation, and resolution of instrument collisions.

In the present report, we highlight the analysis of the first training module. Graduation between modules is based on the trainees reaching the expert skill levels or on completion of 6 months. We aim to continue to track our trainees to proficiency.

Recruitment and Status

The data from 12 subjects (8 trainees and 4 expert robotic surgeons) from 3 academic surgical training programs (Johns Hopkins, Boston Children’s, and Stanford) was analyzed in the present study. Additional training centers and subjects (target cohort of 24 subjects) are being added as approval is received from the institutional review boards and their training robots are activated for data collection by Intuitive Surgical. Our subjects

TABLE 1. Aggregate measures computed from longitudinal data

Task	Session 1	Session 2	Session 3	Session 4	Session 5
Task times (s)					
Expert					
Suturing	348	322			
Manipulation	238	238			
Transection	133	149			
Dissection	188	260			
Trainee					
Suturing	454	588	255	289	279
Manipulation	867	577	311	282	442
Transection	107	196	76	103	126
Dissection	363	291	191	492	200
Motion (m)					
Expert					
Suturing	13.0	10.3			
Manipulation	14.9	14.2			
Transection	1.8	1.2			
Dissection	3.2	6.6			
Trainee					
Suturing	12.9	15.0	6.1	6.1	6.8
Manipulation	27.8	17.8	15.1	16.5	21.1
Transection	1.7	1.6	0.5	1.1	1.1
Dissection	8.1	5.0	4.0	9.3	3.4
Events (count)					
Expert					
Suturing	8	2			
Manipulation	43	40			
Transection	3	2			
Dissection	0	2			
Trainee					
Suturing	0	0	2	6	4
Manipulation	98	61	61	50	89
Transection	1	1	1	5	3
Dissection	5	7	4	7	5

Experts performed each task twice to reduce variability. Sample task times (seconds), master handle motion distances (meters), and number of camera foot pedal events (counts) were detailed for tasks defined in first training module.

were stratified according to 4 skill levels: novice, beginner, intermediate, and expert. Novice trainees were defined as having no previous experience with the da Vinci robotic system. Beginner trainees possessed only limited dry-laboratory experience and no clinical experience with the da Vinci system. Intermediate trainees possessed limited clinical experience with the robotic system. Expert users were faculty surgeons with clinical robotic surgical practices. Performance data from each subject were collected at monthly intervals throughout their training period. Expert surgeons provided 2 executions of the training tasks to establish the skill metrics. Between benchmarking sessions, subjects (resident, fellow, or surgeons with 4–23 years of clinical experience), performed varying amounts of additional training or assisted in the operating room, as appropriate. Data from the subjects were archived periodically in data cartridges shipped to Johns Hopkins for additional analysis. In the present report, we analyzed 4 expert users and 8 other users with nonexpert skill levels.

Structured Assessment

To validate our framework’s construct, we applied Objective Structured Assessment of Technical Skills (OSATS)^{12,13} evaluations for each task execution. The OSATS global rating scale consists of 6 skill-related variables in operative procedures graded on a 5-point Likert-like scale (ie,

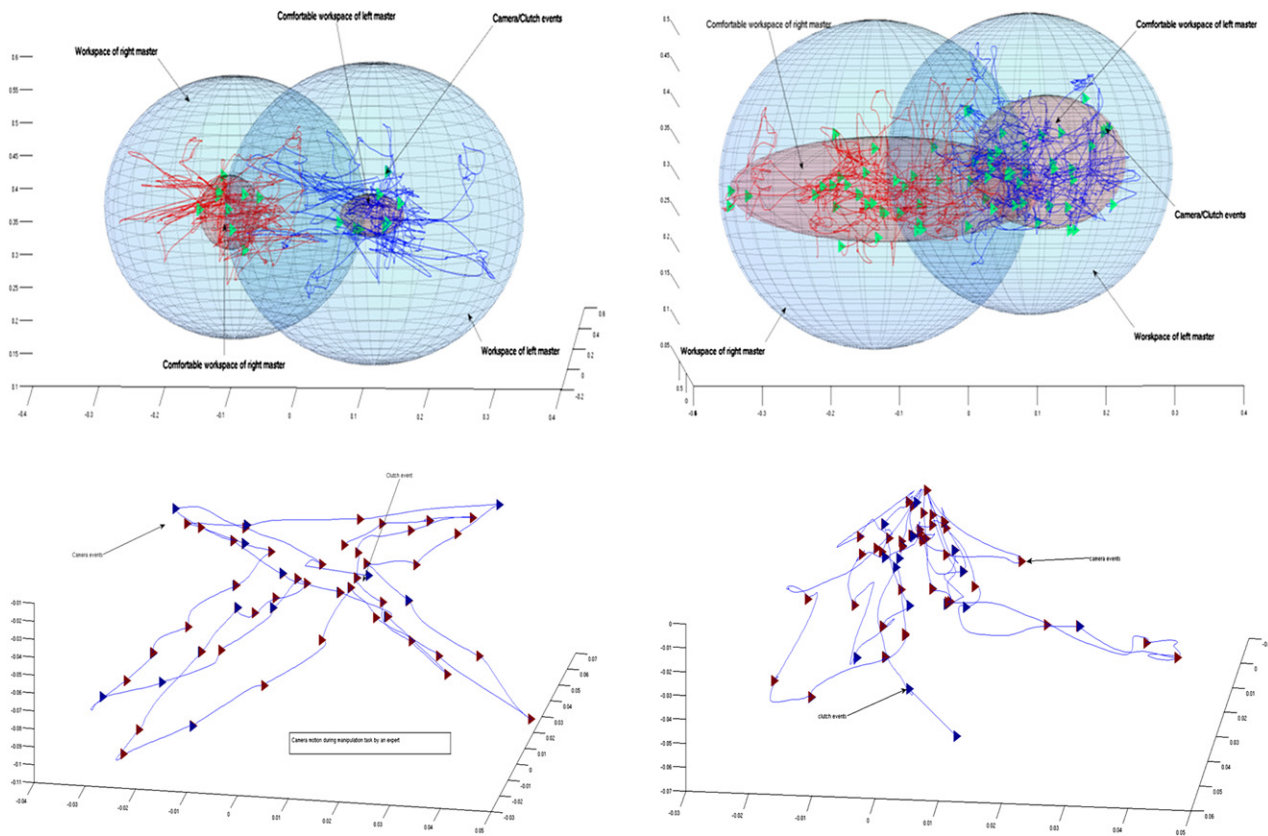


FIGURE 2. Master and camera workspaces used by experts (*left, top and bottom*) and novices (*right, top and bottom*) during performance of manipulation task. *Red ellipsoids* represent volumes of hand positions at end of master workspace adjustments (clutching). *Blue ellipsoids* represent master hand volumes. *Triangles* represent master (*green*) and camera (*red, blue*) clutch adjustments.

1–5). The middle and extreme points were explicitly defined. The 6 measured categories are respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of procedure, and knowledge of procedure. The “use of assistants” category is not generally applicable in the first training module and was therefore not evaluated. A cumulative score totaling the individual scores for each of the 6 categories was obtained (minimum score, 0; maximum score, 30). The OSATS evaluation constructs have been validated in terms of inter-rater variability and correlation with technical maturity^{13,14} and have been applied in evaluating technical facility with robot-assisted surgery.¹⁵

Automated Measures

At least 2 different types of automated measures can be computed from the longitudinal data we have acquired. The first include aggregated motion statistics, task measures, and associated longitudinal assessments (ie, learning curves). The second include measures computed using statistical analysis to compare the technical skills of the trainees to those of the expert surgeons.

Motion statistics and task measures. The computed elementary measures for the defined surgical task executions are listed in Table 1. Each of these measures is used to derive an associated learning curve over the longitudinally collected data.

Statistical classification of technical skill. Our group and collaborators^{16–19} have previously used the da Vinci API motion data to develop statistical methods for the automatic segmentation and analysis of basic surgical motions for the quantitative assessment of surgical skills. Lin and colleagues¹⁶ used linear discriminant analysis to reduce the motion parameters to 3 or 4 dimensions and Bayesian classification

to detect and segment basic surgical motions, termed “gestures.” Reiley and colleagues¹⁹ used a Hidden Markov Model approach for modeling the gestures. These studies reported that experienced surgeons perform surgical tasks significantly faster, more consistently, and more efficiently and with lower error rates.^{19,20} We summarized the assessments of robotic system operational skills using support vector machines (SVM) to cluster dimensionally reduced data, revealing different levels of competence. A SVM transforms the input data into a higher dimensional space using a kernel function. An optimization step then estimates the hyperplanes, separating the data into the 2 classes with maximal separation. For our binary expert versus nonexpert classification, the SVM classifier is directly applicable. A trained classifier allows held-out or unseen data to be labeled by an appropriate class label for computation of traditional sensitivity and specificity measures.

RESULTS

Structured Assessment

A clear separation can be seen between trainees according to their system operational skills and clinical background in Table 1, providing a validated “ground truth” for assessing our automated methods.

Workspace Management

Maintaining a compact operative workspace is an important robotic system operational skill. Expert robotic surgeons maintain an optimal field of view for a given

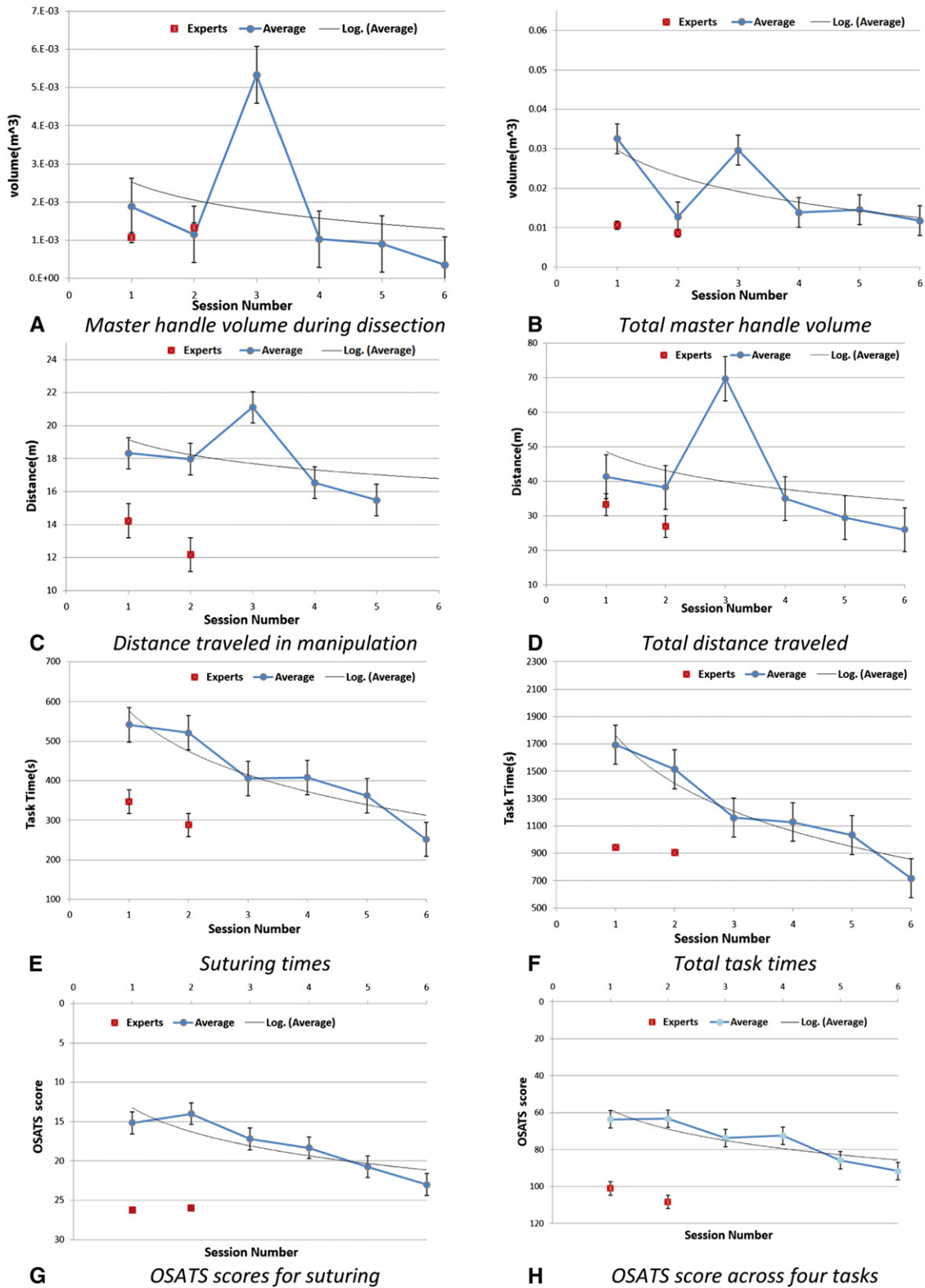


FIGURE 3. Learning curves determined by time, master handle distance, master handle volumes, and objective structured assessment of technical skills (OSATS) structured assessment measurements for individual tasks and all 4 tasks. Note that the OSATS score scale has been inverted and expert task metrics (red) appear in the bottom lower corner of the charts.

TABLE 2. Pairwise *t* tests comparing OSATS scores and training modules at 1-, 3-, and 5-month training intervals (2-tail *P* values, $\alpha = 0.05$)

Pairwise <i>t</i> test	Subjects (n)	OSATS/suturing time	OSATS/total time	OSATS/manipulation distance	OSATS/total task distance	OSATS/master handle volume in dissection	OSATS/total task distance volume
1	12	6E-5	2.8E-5	9.9E-4 (n = 8)	0.0014	1.5E-7	1.5E-7
3	6	0.0016	1.4E-4	0.0067	0.9303	8.4E-5	8.5E-5
5	3	0.0227	2.3E-4	0.0052	0.0043	8.4E-4	8.4E-4

OSATS, Objective structured assessment of technical skills.

operative task, keeping the instruments within the field of view at all times (Figure 2, bottom left); however, trainees tend to zoom out to a broad field of view that is not adjusted during task performance (Figure 2, bottom right).

The differences in workspace use between trainees and expert robotic surgeons performing the manipulation task are graphically illustrated in Figure 2, top. The trajectories represent master handle motion. The enclosing volumes represent the total volumes used and the positions of the master handles at the end of the master clutch adjustments. The workspace use evolved to become closer to the expert workspace use as trainees learned to adjust their workspaces more efficiently. Expert task executions also included regularly spaced camera clutch events to maintain the instruments in the field of view.

We used master handle motion for the computations, instead of the instrument tip motion reported in the published data. Master motion better measures the operational skills by capturing all the motion required to adjust the master workspace; this cannot be captured by instrument tip motion. We measured both the master distance and the volume in which the master handles are moved. An analysis of the instrument motion statistics and counts of other foot pedal actuations, instrument exchanges, and other system events will be reported separately.

Learning Curves

The learning curves derived from the task motion, times required to complete the defined surgical tasks, and corresponding learning curves determined from the expert OSATS structured assessments are shown in Figure 3. The analysis of variance ($F = 71.88 > 2.23$, $F = 51.02 > 2.37$, and $F = 71.4 > 2.57$ at $\alpha = 0.05$ at 1, 3, and 5 months, respectively) results were significant at the 5% level, indicating that the expected values for the task completion times, OSATS scoring, master motion distances, and master volumes differed significantly. Trainee performance improved with time, as indicated by the lower task completion times, smaller volumes, and shorter motion paths. These correlated with the improved OSATS scores. The expert measures displayed very small variability between the 2 executions.

The computed measures (eg, task completion times, total time, master motion distances, and master volumes) at 1-, 3-, and 5-month intervals correlated with the OSATS scores for the corresponding sessions ($P < .05$). For the suturing

task, at month 1, the mean OSATS score across 4 tasks (mean = 77.58; variance = 527.35, $N = 12$), and suturing time (mean = 466.29, variance = 39392.63, $N = 12$) were significantly greater than 0 [$t(11) = -6.27$, 2-tail $P = 6.07E-5$], providing evidence that task completion times correlated with our ground truth assessment. The *P* values for $\alpha = 0.05$ are listed in Tables 2 and 3.

Skill Assessment

For a portion of the data set (ie, 2 experts and 4 nonexperts), we clustered the motion data using principal component analysis to reduce the data dimensions for Cartesian instrument velocities signals. We then trained a binary SVM classifier on a portion of the data and used the trained classifier to perform expert versus nonexpert binary classification. This method correctly stratified our subjects according to their respective skill levels, with 83.33% accuracy for the suturing task and 76.25% accuracy for the manipulation task. Detailed automated analysis on this and expanded data sets will be reported separately.

DISCUSSION

Clinical skill measures should assess instrument–environment interactions. In our recent review,²¹ we described related work on the use of instrument motion in the surgical skill assessment. Although instrument motion was measured accurately using the sensors built into the robot, the interaction and effects of tools with the environment (ie, the patient or model) and additional tools such as needles and sutures was not captured in the kinematic motion data. At present, instrument motion has been primarily used as an indicator of “clinical” skill. In our approach, we focused on the “operational” skills for robotic surgery systems. Robotic surgery entails a complex human–machine interface. It is the complexity of this interface that creates the steep learning curves, even for laparoscopically trained surgeons.

TABLE 3. *P* values for 1-factor analysis of variation for all variables at 1-, 3-, and 5-month training intervals

ANOVA	Subjects (n)	<i>P</i> value	F	F critical
1	12	8.4E-24	47.7	2.22
3	6	7.8E-20	90.5	2.37
5	3	2.5E-15	472.1	2.85

ANOVA, Analysis of variance.

We have described a longitudinal study of robotic surgery trainees, including a preliminary assessment of structured expert assessment (OSATS) and automatically computed statistics and skill measures. Operational skill effects can be completely captured using the telemetry obtained from the robotic system. With appropriate tasks and measures, separate learning curves can be derived. In particular, we noted very strong agreement between the structured assessment of task performance using OSATS and computed master workspace measures (ie, distance, volume, and time). Additional computed measures, not described in the present study, include camera motion effects, instrument motion measures, learning curves determined from the system events, learning curves determined from abnormal events, and reactions to abnormal events.

We performed a longitudinal analysis to develop our learning curves. This is an essential exercise toward the development of both training curricula and metrics that are discriminative of operational skill. As noted, the measures of skill determined by master manipulation showed large differences between experts and nonexperts and convergence of nonexperts toward the experts as the training progressed. *Ab initio* training, in which the operational skills and system orientation are most important, is only the first step in robotic surgery training. Additional training modules will add port constraints, instrument collisions, and more complex tasks. The immediate application of these methods is in an *ab initio* proficiency-based training regimen (compared with time limited) for robotic surgery trainees, and the first applications of the developed methods in the training program (Division of Head and Neck Surgery, Johns Hopkins) are currently underway.

The present analysis used only a portion of our data and reported only some of the measures computed. Additional larger studies involving larger data sets and alternative methods are underway. In ongoing work, we have measured the response times to errors and their development curves as additional skill measures. Finally, automated skill classification accuracies greater than 80% highlight the quality of our data. In ongoing work, we are using alternative supervised and unsupervised multiclass classification for both operational and surgical task skills. The effect of the amount of the varying times and types of training between the benchmarking sessions is also under investigation.

The authors thank Dr Myriam Curet and Dr Catherine Mohr for their insightful comments and help with this work. The support received is also gratefully acknowledged.

References

1. Intuitive Surgical Inc. Intuitive Surgical Web site. Available at: <http://www.intuitivesurgical.com/products/faq/index.aspx>. Accessed April 2011.
2. Novick RJ, Fox SA, Kiaii BB, Stitt LW, Rayman R, Kodera K, et al. Analysis of the learning curve in telerobotic, beating heart coronary artery bypass grafting: a 90 patient experience. *Ann Thorac Surg*. 2003;76:749-53.
3. Rodriguez E, Chitwood WR Jr. Outcomes in robotic cardiac surgery. *J Robot Surg*. 2007;1:19-23.
4. Chitwood WR Jr. Current status of endoscopic and robotic mitral valve surgery. *Ann Thorac Surg*. 2005;79:S2248-53.
5. Ro CY, Toumpoulis IK, Ashton RC Jr, Imielinska C, Jebara T, Shin SH, et al. A novel drill set for the enhancement and assessment of robotic surgical performance. *Stud Health Technol Inform*. 2005;111:418-21.
6. Judkins TN, Oleynikov D, Stergiou N. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surg Endosc*. 2009;23:590-7.
7. Sarle R, Tewari A, Shrivastava A, Peabody J, Menon M. Surgical robotics and laparoscopic training drills. *J Endourol*. 2004;18:63-7.
8. Narazaki K, Oleynikov D, Stergiou N. Objective assessment of proficiency with bimanual inanimate tasks in robotic laparoscopy. *J Laparosc Advanc Surg Techn*. 2007;17:47-52.
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. DiMaio S, Hasser C. The da Vinci Research Interface, Systems and Architectures for Computer Assisted Interventions workshop (MICCAI 2008). Available at: <http://www.midasjournal.org/browse/publication/622>. Accessed April 2011.
11. Mohr C, Curet M. The Intuitive Surgical System Skill Practicum. Sunnyvale, CA: Intuitive Surgical, Inc; 2008.
12. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84:273-8.
13. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med*. 1996;71:1363.
14. Darzi A, Smith S, Taffinder N. Assessing operative skill. *BMJ*. 1999;318:887.
15. Hernandez JD, Bann SD, Munz Y, Moorthy K, Datta V, Martin S, et al. Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da Vinci system. *Surg Endosc*. 2004;18:372-8.
16. Lin HC, Shafran I, Yuh D, Hager GD. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. *Comput Aid Surg*. 2006;11:220-30.
17. Verner L, Oleynikov D, Hotmann S, Zhukov L. Measurements of level of surgical expertise using flight path analysis from Da Vinci robotic surgical system. *Stud Health Technol Inform*. 2003;94:373-8.
18. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc*. 2003;17:1525-9.
19. Reiley CE, Hager GD. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. *Proc Med Image Comput Comput-Assisted Intervent (MICCAI)*. 2009;435-42.
20. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World J Surg*. 2001;25:1478-83.
21. Reiley CE, Lin HC, Yuh DD, Hager GD. A review of methods for objective surgical skill evaluation. *Surg Endosc*. 2011;25:356-66.