
Virtual Reality Robotic Surgery Warm-Up Improves Task Performance in a Dry Laboratory Environment: A Prospective Randomized Controlled Study

Thomas S Lendvay, MD, FACS, Timothy C Brand, MD, FACS, Lee White, BSc(Hons), PhC, Timothy Kowalewski, PhD, Saikiran Jonnadula, MD, Laina D Mercer, MS, Derek Khorsand, BSc(Hons), Justin Andros, BSc(Hons), Blake Hannaford, PhD, Richard M Satava, MD, FACS

BACKGROUND: Preoperative simulation warm-up has been shown to improve performance and reduce errors in novice and experienced surgeons, yet existing studies have only investigated conventional laparoscopy. We hypothesized that a brief virtual reality (VR) robotic warm-up would enhance robotic task performance and reduce errors.

STUDY DESIGN: In a 2-center randomized trial, 51 residents and experienced minimally invasive surgery faculty in General Surgery, Urology, and Gynecology underwent a validated robotic surgery proficiency curriculum on a VR robotic simulator and on the da Vinci surgical robot (Intuitive Surgical Inc). Once they successfully achieved performance benchmarks, surgeons were randomized to either receive a 3- to 5-minute VR simulator warm-up or read a leisure book for 10 minutes before performing similar and dissimilar (intracorporeal suturing) robotic surgery tasks. The primary outcomes compared were task time, tool path length, economy of motion, technical, and cognitive errors.

RESULTS: Task time (-29.29 seconds, $p = 0.001$; 95% CI, -47.03 to -11.56), path length (-79.87 mm; $p = 0.014$; 95% CI, -144.48 to -15.25), and cognitive errors were reduced in the warm-up group compared with the control group for similar tasks. Global technical errors in intracorporeal suturing (0.32 ; $p = 0.020$; 95% CI, $0.06-0.59$) were reduced after the dissimilar VR task. When surgeons were stratified by earlier robotic and laparoscopic clinical experience, the more experienced surgeons ($n = 17$) demonstrated significant improvements from warm-up in task time (-53.5 seconds; $p = 0.001$; 95% CI, -83.9 to -23.0) and economy of motion (0.63 mm/s; $p = 0.007$; 95% CI, $0.18-1.09$), and improvement in these metrics was not statistically significantly appreciated in the less-experienced cohort ($n = 34$).

CONCLUSIONS: We observed significant performance improvement and error reduction rates among surgeons of varying experience after VR warm-up for basic robotic surgery tasks. In addition, the VR warm-up reduced errors on a more complex task (robotic suturing), suggesting the generalizability of the warm-up. (J Am Coll Surg 2013;216:1181-1192. © 2013 by the American College of Surgeons)

Finding methods to improve surgical performance for trainees and practicing surgeons has become a national mission to mitigate surgical morbidity, reduce health

care costs, accelerate learning curves, provide curricula for the introduction of new surgical technologies, and ensure that reductions in duty hours for trainees do not

Disclosure Information: Nothing to disclose.

This study was supported by the Department of Defense US Army Medical Research and Materiel Command under award number W81XWH-09-1-0714 (principal investigator: Lendvay). The views, opinions, and endorsement by the author(s) do not reflect those of the US Army or the Department of Defense.

The Seattle Children's Core for Biomedical Statistics is supported by the Center for Clinical and Translational Research at Seattle Children's Research Institute and grant UL1RR025014 from the NIH National Center for Research Resources.

Abstract presented at the American College of Surgeons 98th Annual Clinical Congress, Surgical Forum, Chicago, IL, October 2012.

Received November 21, 2012; Revised February 10, 2013; Accepted February 13, 2013.

From the Departments of Urology (Lendvay) and Surgery (Jonnadula, Khorsand, Andros, Satava), University of Washington School of Medicine, Departments of Bioengineering (White) and Electrical Engineering (Kowalewski, Hannaford), University of Washington, Core for Biomedical Statistics, Seattle Children's Research Institute (Mercer), Seattle, and Madigan Army Medical Center, US Army, Tacoma (Brand), WA.

Correspondence address: Thomas S Lendvay, MD, FACS, Seattle Children's Hospital, 4800 Sand Point Way, NE, Seattle, WA 98105. email: thomas.lendvay@seattlechildrens.org

Abbreviations and Acronyms

FLS = Fundamentals of Laparoscopic Surgery
ICC = interclass correlations
MIS = minimally invasive surgery
OR = operating room
VR = virtual reality

compromise surgical education.¹⁻³ Surgical simulation methods are mandated by some surgical professional boards^{4,5} and the merits of surgical simulation have been validated both in and out of the operating room (OR).⁶⁻¹⁰ Most surgical simulation is carried out in dry and animate laboratories at a very different time than actual surgery on patients. But recent studies suggest that surgical simulation immediately before criterion surgical tasks can benefit performance.^{11,12} This presurgical rehearsal, or warm-up, promises to boost surgical performance. Now that high-fidelity simulator curricula exist for robotic surgery, we hypothesized that virtual reality (VR) robotic surgical warm-up for similar (basic skills) and dissimilar (complex task, intracorporeal suturing) tasks improves performance in both surgical trainees and experienced minimally invasive surgeons.

High-stakes professions, such as athletics and performing arts, have long relied on the principles of the warm-up decrement (ie, the decrease in performance after a period of rest) and the Activity Set hypothesis (ie, the idea that to counter the warm-up decrement, some activity to elevate the arousal and readiness of the subject is required to boost performance) to optimize performance readiness.¹³⁻¹⁶ Yet, surgery does not involve a prescribed warm-up or presurgical rehearsal, although it is a high-stakes profession drawing on intense psychomotor and cognitive efforts. The benefits of warm-up can be particularly important for robotic surgery because of the increased information presented to the surgeon through the visual monitor, as visual cues must be processed to derive forces applied by the tools (synesthesia) and cognitive arousal is likely to benefit greatly from warm-up.

Do and colleagues were the first to use a laparoscopic box trainer to study the effect of warm-up exercises on follow-up laparoscopic tasks, and they observed a considerable improvement in performance (25%) for both residents, irrespective of PGY level, and a medical student control group ($p < 0.0001$).¹⁷ The study was not able to discriminate the effects of the learning curve vs a true warm-up effect and so Kahol and colleagues sought to address this in a laparoscopic VR simulation study.¹⁸ Surgeons were randomized to receive either warm-up or no warm-up using a series of VR ring-transfer tasks that tested psychomotor, attentional, and visuospatial skills.

The results yielded a substantial reduction in errors (33%). In addition, Kahol and colleagues showed that the warm-up effect was demonstrated in surgeons of all levels of expertise and generalized to dissimilar follow-up tasks, such as an electrocautery task. In Kahol and colleagues' study, both warm-up tasks and criterion tasks were in a virtual laboratory. But in 2010, Calatayud and colleagues showed that a VR simulation warm-up in the OR benefited residents performing laparoscopic cholecystectomies.¹¹ Eight residents demonstrated higher global performance scoring on an Objective Structured Assessment of Technical Skills tool.¹¹ In 2012, Lee and colleagues showed that brief reality-based laparoscopic suturing and VR task warm-up immediately before the colon mobilization in laparoscopic nephrectomies performed by senior urology residents yielded higher global assessment scoring and reduced task time.¹⁹ All of these studies have been performed with conventional laparoscopy, yet there have been no studies looking at the value of surgical warm-up in robotic surgery. We sought to explore the role VR robotic warm-up has on similar and dissimilar robotic surgery tasks.

METHODS**Study design**

Residents and experienced minimally invasive surgery faculty in General Surgery, Urology, and Gynecology from 2 medical centers underwent a validated robotic surgery proficiency curriculum on a VR robotic simulator and on the da Vinci surgical robot (Intuitive Surgical Inc). Once successfully achieving performance benchmarks, each surgeon was randomized to either receive a 3- to 5-minute VR warm-up on the simulator or read a leisure book for 10 minutes before performing similar and dissimilar (intracorporeal suturing) robotic surgery tasks. Three serial trial sessions were performed with similar warm-up and criterion tasks, followed by a dissimilar warm-up to test generalizability. The primary outcomes analyzed and compared were task time, tool path length, economy of motion, and technical and cognitive errors (Fig. 1).

Participant recruitment

Institutional Review Board approval (#35096) was granted to recruit surgical residents and faculty from the Departments of Urology, General Surgery, and Gynecology at the University of Washington Medical Center and Madigan Army Medical Center to get a representation of both civilian academic and military sector training programs. After acquiring informed consent, each enrollee filled out a demographics questionnaire. The question domains included level of training, handedness,

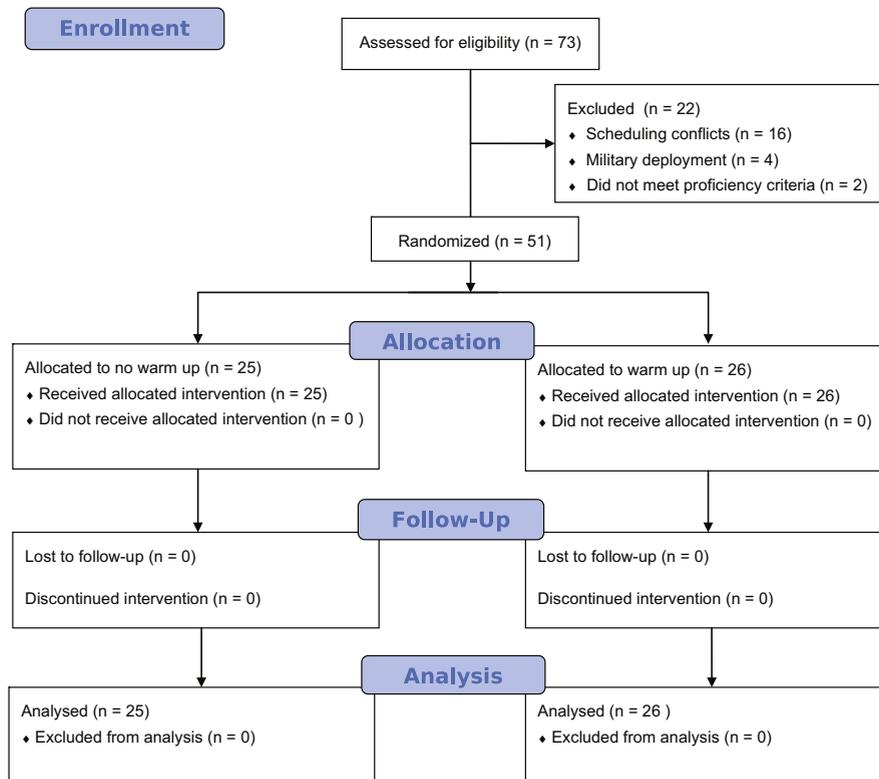


Figure 1. Patient flow diagram.

musical instrument and video-gaming experience, and minimally invasive surgery (MIS) experience; all play roles in surgical skill acquisition. All PGY1–6, surgical fellows, and faculty who were experienced in MIS were recruited. All subjects participated in a proficiency curriculum.

Statistical power/sample size calculation

The statistical power in a repeated measures design was driven by the number of independent subjects in the study, the number of serial observations on each individual, and the degree of within-person dependence among observations contributed by the same individual. Because the within-person dependence was not precisely known, interclass correlations (ICCs) between 0.5 and 0.8 were explored, which covers the typical range for studies involving repeated measurements on the same person.²⁰ With 51 participants, 3 observations per individual, and assuming an ICC of 0.8, we calculated 95% statistical power for detecting an overall difference between the warm-up and control groups, if the warm-up factor describes at least an additional 20% of the total variation (0.20 increase in R^2). An ICC of 0.8 provides a conservative estimate because it implies observations

within subjects will be highly correlated. The statistical power is even higher for smaller ICC values. We did not have preliminary measurements on the path-length metric, however, for the purposes of power assessment, all that matters is the spread of the group means relative to the within-group SD.

Randomization

Permuted blocks randomization was used. Randomization was stratified by site (University of Washington Medical Center and Madigan Army Medical Center) and surgical experience level (resident and faculty). Randomization assignments were provided to each site in sealed, fully opaque envelopes, so that upcoming study group assignments could not be anticipated by study staff or potential enrollees. Randomization occurred at the time the surgeon completed their proficiency curriculum (described later).

Participant flow

Once enrolled, each surgeon went through a robotic proficiency curriculum that included the 90-minute da Vinci online didactics module to familiarize the surgeons with the da Vinci S/Si systems. After passing the tutorial, each surgeon went through a VR (dV-Trainer simulator;



Figure 2. Experimental set-up. Demonstration of proficiency modules in their respective jigs and the plumb lines draping down onto the jig to ensure standard robotic arm positioning.

MIMIC Technologies, Inc.) and da Vinci dry laboratory robotics curriculum composed of 4 progressively harder surgical skills modules on each platform, respectively (Figs. 2, 3). The proficiency curriculum was generated based on incorporating progressively more complex technical skills, such as object transfer, followed by camera and instrument clutching, followed by all these plus adding motion to the task platform to test spatial relations capabilities. Proficiency benchmarks were established for each module based on performance by 2 experienced robotic surgeons (TSL, TCB) who have each performed >150 robotic procedures. The benchmark required that each surgeon perform 2 consecutive task iterations within 120% of the mean task time of the 2 experienced surgeons with a zero error rate respective to each module. For example, in the VR Pegboard Level 1 module, a surgeon would have to do as many iterations of the task until 2 consecutive iterations yielded a task time <120% of the mean of the 2 benchmark surgeons performing the same task and with no ring drops or sequence errors. We chose 2 consecutive iterations of success to try to hone the legitimate proficiency of the surgeon for each task. We chose 120% of task time because we did not think it necessary for every surgeon to reach experienced surgeon times to demonstrate proficiency at a particular task. And we did not want to rely solely on task time as the primary benchmark criterion because fast yet error-prone performance is not desirable in surgery.

Concern for learning effect was addressed. To mitigate the confounding effects of the learning curve throughout the study, each surgeon was required to reach proficiency benchmarks before the trial sessions on the da Vinci robot. The intention was to obtain some proficiency equity among the surgeons and familiarity with instrument/camera clutching and manipulation. To equalize the up-front learning of each surgeon irrespective of randomization designation (simulator warm-up or no warm-up), we believed that each surgeon must be given the exact same opportunity to learn the manipulations of both the robot and the simulator to lessen the chance that the warm-up group will have the added benefit of using the simulator at each warm-up trial session.

Once VR robotic simulator proficiency was met, surgeons tested to proficiency on the da Vinci robot through 4 task modules (Fig. 4). Construct validation of the da Vinci curriculum was demonstrated through the use of retrofitted da Vinci training instruments capable of tracking tool motions and errors—SurgTrak (described later)—to derive path length and economy of motion performance metrics.²¹⁻²³ Again, proficiency benchmarks had been obtained from the same 2 surgeons and 120% of the mean task times and zero error rates through 2 consecutive iterations were required to advance to the next module. The modules included 2 Fundamentals of Laparoscopic Surgery (FLS) tasks being performed on the da Vinci robot—block transfer and intracorporeal

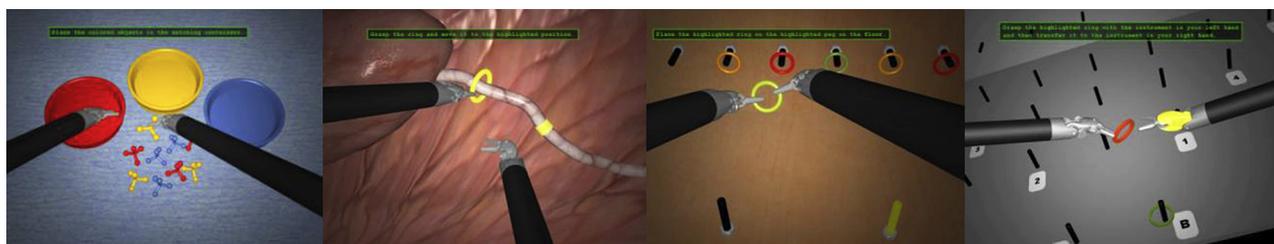


Figure 3. MIMIC dV-Trainer VR simulation modules from left to right. Pick and Place, Ring Walk Level 1, Pegboard Level 1, Pegboard Level 3.

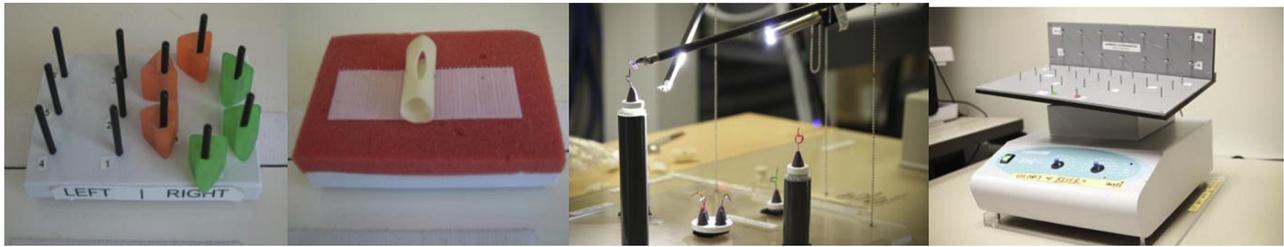


Figure 4. da Vinci dry laboratory modules from left to right. Fundamentals of Laparoscopic Surgery (FLS) block transfer, FLS intracorporeal suturing (this was the criterion task for session 4), Ring Tower (The Chamberlain Group), Rotating Rocking Pegboard (this was the criterion task for sessions 1 to 3).

suturing—because these have been repeatedly validated in laparoscopy curricula.^{6,9}

Trial sessions

After a surgeon reached proficiency, he or she was randomized to either the warm-up group or control group. Four trial sessions per surgeon were performed. The first 3 tested performance with or without warm-up on the da Vinci rocking pegboard criterion task. Each of these sessions was separated by a minimum of 24 hours so that one session did not warm-up the surgeon for the next session.²⁴⁻²⁶ In addition, the surgeon could not have performed the trial session if they had done any robotic clinical practice within 24 hours of the session for the same reason. The warm-up group performed the Pegboard Level 3 VR task once directly before performing the analogous da Vinci rotating rocking pegboard task. This generally took 3 to 5 minutes to complete and, unlike in the proficiency curriculum, it was not mandatory for them to perform the VR task with a zero error rate. The controls spent 10 minutes reading a leisure book immediately before performing the da Vinci criterion task so as to minimize the likelihood that they were visually imagining the task to be performed because visual imagery warm-up has been shown to prime surgeons.^{27,28} They could not read any scientific manuscripts or surf the web because we believed that these could also prime the control surgeons.

During the fourth trial session, the warm-up and control precriterion process was the same as the first 3 sessions, but the criterion task became the FLS intracorporeal suturing task to assess whether warm-up generalized to more complex and dissimilar tasks.

Objective performance metrics

Based on existing surgical curricula validation studies, we chose the following performance metrics to track on the simulator and the da Vinci robot.^{10,29-32}

1. Total task time (seconds).

2. Cognitive errors (total count): rings placed on incorrect pegs, incorrect sequence of pegs.
3. Technical errors (total count): dropped rings, peg touches.
4. Tool path length (total distance traveled for instruments [mm]).
5. Economy of motion: path length/task time (mm/s).

During the FLS intracorporeal suturing module, additional performance metrics were assessed based on FLS validation of the knot-tying exercise.⁹

1. Error: breaking the suture.
2. Error: not placing the suture through the premarked entrance and exit spots.
3. Error: gap left in suture knot (air knot).

SurgTrak tool motion tracking and video capture

To capture the objective performance metrics, we developed a system consisting of video recording and surgical tool motion recording combined with custom software. Video was recorded at 30 frames per second from the digital video imaging output of the da Vinci Si/S master console using a DVI2USB device (Epiphan Systems Incorporated). Video was encoded using mpeg-4 compression to produce compact, manageable files. Tool motion data were recorded at 30 Hz. Tool position



Figure 5. Retrofitted da Vinci training instrument with sensor housing on back end.

and orientation were captured with a 3D Guidance trakSTAR electromagnetic tracking system (Ascension Technology Corporation). We retrofitted da Vinci training tools with rapid prototyped holders for the sensors on the proximal ends of the da Vinci instruments (Fig. 5).

These data enabled us to compute path length and economy of motion metrics for each task performance. Grasper pose and electric contact between the tool tips and the pegboard posts were recorded using a PhidgetInterfaceKit 8/8/8 (Phidgets Incorporated). Peg touch errors from the rocking pegboard task were detected and the time of occurrence was recorded by our software. Data streams from the video recording, position recording, and error recording were united using software running on a Windows 7 (Microsoft)-based laptop computer.^{22,23} Errors on the ring tower, FLS block transfer task, and FLS suturing task were documented in real-time by study personnel and double-checked by video review.

Statistical methods

Demographic and clinical characteristics measured at baseline were summarized by treatment group and compared with Fisher's exact test for categorical variables and *t*-test for continuous variables. The primary comparison for sessions 1 to 3 was a test for the overall mean difference between the warm-up and control groups. Because each surgeon contributed 3 observations to the dataset, this test for continuous outcomes was calculated using a repeated measures ANOVA model and the effect of experience level was investigated by training level (resident vs faculty) and surgical experience (>10 robotic and >10 laparoscopic operations performed as the primary surgeon vs ≤10 cases in each modality). For binary outcomes, repeated measures relative risk regression³³ was used to compare groups and test for interactions. Each surgeon contributed only one observation to the data for session 4 outcomes, *t*-tests were used to test for a significant difference between study groups, and the effect of experience level was investigated with linear regression models with an interaction term. Session 4 binary outcomes and tests for interactions were modeled with relative risk regression.³⁴ Data were analyzed using R Version 2.11.1.

RESULTS

Seventy-three surgeons were assessed for eligibility, with 22 not completing the proficiency curriculum due to scheduling conflicts, military deployment during the study, or inability to meet the proficiency criteria within the study time period. Fifty-one participants, 31 from the University of Washington Medical Center and 20 from Madigan Army Medical Center, were randomized and completed the study (warm-up, *n* = 26; control,

n = 25). Once randomized, no surgeon dropped out. In each demographic category, the surgeons were well matched between the groups, including between faculty and resident participants (Table 1).

For sessions 1 to 3, testing whether warm-up improved performance with similar VR and criterion tasks, we observed a statistically significant decrease in the task

Table 1. Demographics and Baseline Characteristics by Intervention Group

Variable	Control (<i>n</i> = 25)	Warm-up (<i>n</i> = 26)	<i>p</i> Value*
Age, y, mean ± SD	35.32 ± 6.47	33.85 ± 5.82	0.40
Sex, <i>n</i> (%)			
Female	10 (40.0)	9 (34.6)	0.66
Male	15 (60.0)	17 (65.4)	
Musical instrument for >3 y, <i>n</i> (%)			
No	7 (28.0)	9 (34.6)	0.76
Yes	18 (72.0)	17 (65.4)	
Handedness, <i>n</i> (%)			
Ambidextrous	0 (0.0)	1 (3.8)	0.36
Left	2 (8.0)	0 (0.0)	
Right	23 (92.0)	25 (96.2)	
Training year, <i>n</i> (%)			
PGY1	1 (4.0)	0 (0.0)	0.34
PGY2	0 (0.0)	2 (7.7)	
PGY3	4 (16.0)	9 (34.6)	
PGY4	3 (12.0)	1 (3.8)	
PGY5	3 (12.0)	1 (3.8)	
PGY6	2 (8.0)	1 (3.8)	
Faculty	12 (48.0)	12 (46.2)	
Subspecialty, <i>n</i> (%)			
Urology	14 (56.0)	14 (53.8)	0.61
General surgery	7 (28.0)	5 (19.2)	
OBGYN	4 (16.0)	7 (26.9)	
Recent video game use, <i>n</i> (%)			
None	15 (60.0)	16 (61.5)	0.99
<2×/week	7 (28.0)	6 (23.1)	
2+×/week	3 (12.0)	4 (15.4)	
Laparoscopic cases, primary surgeon, <i>n</i> (%)			
None	1 (4.0)	0 (0.0)	0.55
≤10	3 (12.0)	3 (11.5)	
11–25	3 (12.0)	1 (3.8)	
25+	18 (72.0)	22 (84.6)	
Robotic cases, primary surgeon, <i>n</i> (%)			
None	9 (36.0)	8 (30.8)	0.61
≤10	6 (24.0)	10 (38.5)	
11–25	3 (12.0)	1 (3.8)	
25+	7 (28.0)	7 (26.9)	

*Comparison of surgeons by group. All categorical variables were compared with Fisher's exact test and age was compared with a *t*-test.

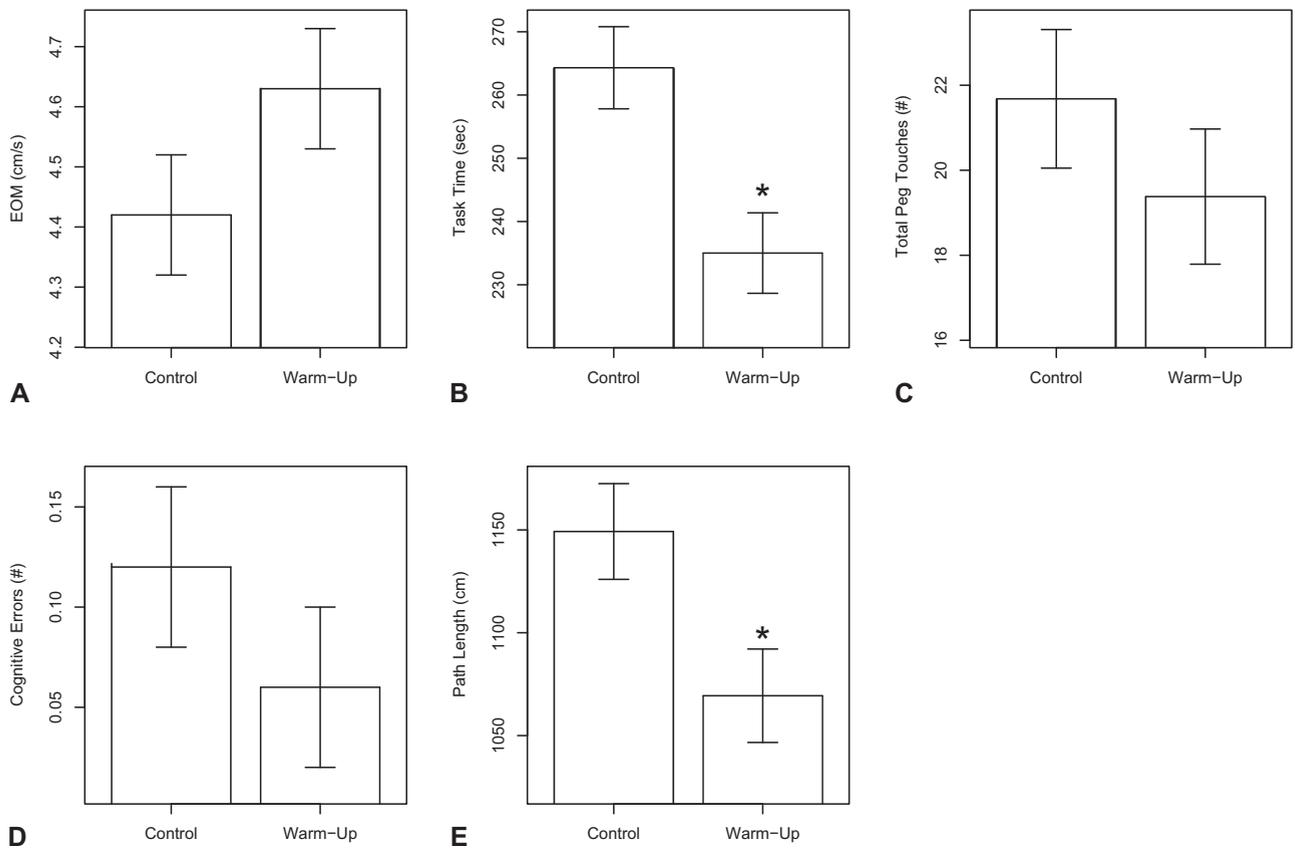


Figure 6. Control vs warm-up. (A) Economy of motion; (B) task time; (C) peg touch errors; (D) cognitive errors; and (E) tool path length.

time (-29.29 seconds; $p = 0.001$; 95% CI, -47.03 to -11.56) and path length (-79.87 mm; $p = 0.014$; 95% CI, -144.48 to -15.25). Economy of motion favored the warm-up group but was not statistically significant. Technical errors—dropping rings or touching the pegs with the instruments—did not show statistically significant differences, yet cognitive error reduction favored the warm-up group, but was not statistically significant. The proportion of sessions with errors of placing the rings on incorrect pegs (sequence errors) favored the warm-up group, but because of the wide

confidence interval, this was neither statistically significant nor conclusive ($p = 0.087$; Fig. 6; Tables 2, 3).

For session 4, testing whether a dissimilar VR task can warm-up surgeons for a more complex task (FLS intracorporeal suturing) task, we observed no significant improvements in task time, economy of motion, or path length for the warm-up group. However, when we assessed global technical errors for the suturing (needle entrance, exit errors, and air knot errors, collectively), we observed a near 4-fold reduction in the proportion of sessions with these errors ($p = 0.020$). Individually,

Table 2. Continuous Outcomes by Study Group (Sessions 1 to 3)

Outcomes	Control		Warm-up		Difference (95% CI)	p Value
	Mean	SD	Mean	SD		
Economy of motion	4.42	0.66	4.63	0.66	0.21 (-0.06 to 0.47)	0.13
Task time, s	264.31	56.97	235.01	40.11	-29.29 (-47.03 to -11.56)	0.001
Total peg touches	21.68	10.06	19.38	9.01	-2.29 (-6.71 to 2.12)	0.31
Cognitive error	0.12	0.40	0.06	0.30	-0.06 (-0.17 to 0.06)	0.34
Path length, mm	1,149.23	189.03	1,069.37	132.97	-79.87 (-144.48 to -15.25)	0.014

Outcomes were individually analyzed with repeated measures ANOVA.

Table 3. Binary Outcomes by Study Group (Sessions 1 to 3)

Error type	Proportion of sessions with error				
	Control	Warm-Up	RR	95% CI	p Value
Ring drops	0.320	0.333	0.96	0.58–1.59	0.87
Air transfer	0.040	0.051	0.78	0.19–3.14	0.73
Out of order (sequence)	0.080	0.013	6.24	0.77–50.76	0.09

All outcomes were individually analyzed with relative risk (RR) regression.

each error was reduced in the warm-up group, but the differences were not statistically significant (Table 4).

When we assessed the effect that MIS experience (>10 laparoscopic and >10 robotic cases as primary surgeon vs ≤10 cases in each modality as primary surgeon) had on the warm-up effect, we observed that the warm-up effect was more pronounced with experience. Economy of motion (0.63 mm/s; $p = 0.007$; 95% CI, 0.18–1.09), task time (–53.5 seconds; $p = 0.001$; 95% CI –83.9 to –23.0), and path length (–97 mm; $p = 0.093$; 95% CI, –210 to 16) favored the warm-up subgroup among the experienced cohort, and only path length (–75 mm; $p = 0.063$; 95% CI, –154 to 4) favored the warm-up group in the inexperienced cohort, and not to a statistically significant degree (Table 5).

When the groups were divided based on resident ($n = 27$) vs faculty ($n = 24$) level, the results were mixed. Path length (–96 mm; $p = 0.029$; 95% CI, –181 to –10) and task time (–31 seconds, $p = 0.013$; 95% CI, –55.7 to –6.4) were reduced in the resident warm-up group, and task time reduction only (–27.4 seconds; $p = 0.039$; 95% CI, –53.5 to –1.4) reached statistical significance in the faculty warm-up group. Path length and economy of motion only favored, but not statistically significantly, the warm-up group (Table 6).

DISCUSSION

We hypothesized that robotic surgery VR warm-up would enhance technical and cognitive performance on da Vinci dry laboratory tasks. In our randomized study comparing warm-up and control groups of experienced and inexperienced surgeons, we demonstrated that

preprocedural warm-up does improve task performance and error reduction. This is a fundamental observation because, to date, the literature has established a warm-up's potential role in conventional laparoscopy, but not in robotic surgery. In addition, laparoscopic warm-up has been shown to decrease operative times in experienced surgeons in the OR,¹² a finding consistent with our observations of warm-up benefiting experienced performers. Many of our tracked performance metrics favored the warm-up group. Task time, path length, economy of motion, error reduction—all surrogates for surgical technical ability—were significantly improved.

We also hypothesized, as Kahol and colleagues showed, that a dissimilar warm-up task can generalize a warm-up benefit or elevate criterion task performance.¹⁸ We observed a statistically significant reduction in the proportion of sessions with global technical errors in suturing, such as air knots and inaccurate needle targeting. The value of this finding is that the ideal warm-up curricula might not need to look like the planned robotic surgery tasks. We did not observe, however, significant improvements in standard technical performance metrics, such as task time or path length, in the generalizability session. It is possible that robotic suturing is so highly technical that psychomotor practice of actual suturing is still the best warm-up task for suturing. When looking at warmed-up urology residents, Lee and colleagues saw a warm-up benefit for a dissimilar intraoperative task of taking down the white line of Toldt in a nephrectomy, but did not see a benefit once the case got to suturing up of the white line at the end of the case. This was explained by the fact that suturing during the nephrectomy was at the end of the case and all surgeons might have experienced the maximal amount of warm-up from all the steps leading up to the end of the case.¹⁹

Similar to the enhancement seen in laparoscopy, we demonstrated a reduction in not only technical errors, but cognitive errors. This suggests that warm-up curricula recruiting not only simple psychomotor centers of the brain, but also spatial relations centers, can be additive to the warm-up benefit. Kahol and colleagues specifically emphasized that warm-up tasks need to not only

Table 4. Continuous Outcomes by Study Group (Session 4)

Outcomes	Control (n = 25)		Warm up (n = 26)		Difference (95% CI)	p Value
	Mean	SD	Mean	SD		
Task time, s	111.2	29.3	107.6	37.8	3.6 (–15.4 to 22.6)	0.70
Economy of motion	3.69	0.86	3.82	0.80	–0.14 (–0.61 to 0.33)	0.56
Path length, mm	401.4	114.4	401.5	134.8	0.0 (–71.2 to 71.1)	0.99
Global technical error, count*	0.44	0.58	0.12	0.33	0.32 (0.06 to 0.59)	0.020

All outcomes were individually analyzed with a t -test.

*Composite of air knot, needle targeting errors by Fundamentals of Laparoscopic Surgery (Entrance and Exit dots errors).

Table 5. Effect of Experience on Performance Metrics (Warm-Up vs Control)

Outcomes	≤10 Robotic and ≤10 laparoscopic cases (n = 34)			>10 Robotic and >10 laparoscopic cases (n = 17)		
	Control, mean (SD) (n = 15)	Warm-up, mean (SD) (n = 19)	Difference (95% CI) p Value	Control, mean (SD) (n = 10)	Warm-up, mean (SD) (n = 7)	Difference (95% CI) p Value
Economy of motion	4.49 (0.56)	4.51 (0.57)	0.02 (-0.3 to 0.34) 0.90	4.31 (0.78)	4.94 (0.77)	0.63 (0.18 to 1.09) 0.007
Task time, s	258.6 (43.7)	240.8 (39.6)	-17.8 (-39.2 to 3.5) 0.10	272.9 (72.6)	219.4 (41.0)	-53.5 (-83.9 to -23.0) 0.001
Peg touches, counts	24.2 (8.8)	20.7 (9.8)	-3.6 (-8.8 to 1.7) 0.18	17.9 (10.7)	15.9 (6.1)	-2 (-9.4 to 5.5) 0.60
Cognitive errors, counts	0.13 (0.45)	0.05 (0.30)	-0.08 (-0.22 to 0.06) 0.27	0.10 (0.40)	0.10 (0.45)	0 (-0.21 to 0.20) 0.96
Path length, mm/s	1,152 (174)	1,077 (140)	-75 (-154 to 4) 0.06	1,145 (213)	1,049 (118)	-97 (-210 to 16) 0.09

The mean (SD) and estimated difference between warm up and control for the 5 continuous outcomes measured in sessions 1 to 3 in the study overall and broken up by robotic/laparoscopic case experience.

stimulate psychomotor centers, but also spatial relations and short-term memory centers.¹⁸ In our study, however, some errors were not affected by warm-up, in part because of the low frequency at which these errors occurred, such as peg touches. To observe statistical significance with this metric, a larger sample size would have been needed, however, it remains unclear whether peg touches are a clinically valid surrogate of precision.

An interesting and unexpected finding was that when the MIS experience of the surgeon was the cohort discriminator, warm-up seemed to benefit the more experienced surgeon. This could be explained by unequal proficiency in robotic skills. We attempted to create a rigorous proficiency curriculum to level baseline robotic skills. And, although all surgeons had met our defined proficiency benchmarks, this most likely did not assure equivalent skills. So we hypothesize that experienced surgeons derive a performance boost from warm-up because they only have to be familiarized with the specific task to do better; they do not have to focus on basic manipulations of the robot itself. Less MIS-experienced surgeons not only require task priming, but may spend additional attentional capacity on performing the basic robotic manipulations (eg, grasping, object transfer, camera and arm clutching). Gallagher and colleagues¹⁰ demonstrated that novice surgeons expend a large proportion of their fixed attentional capacity on performing basic technical skills and experienced performers do not have such high demands on simple psychomotor skills. Experienced surgeons can invest more attention to decision making.¹⁰ These findings about experience are important because there are far more practicing robotic surgeons than there are robotic surgery trainees, and our findings can be relevant to hospital credentialing and maintenance of certification processes. When we divided the cohort by faculty vs resident, our results were mixed. This might reflect that not all faculty in our group were experienced robotic surgeons because we did not require as an inclusion criterion that “MIS experience” meant robotic surgery experience. Some of these faculty members had robust conventional laparoscopic experience, but no robotic experience.

There were some key limitations to our study that should be mentioned. First, although we randomized surgeons to 1 of 2 groups—warm-up or control—our proficiency curriculum might not have leveled the proficiency between the groups. Although our groups were very well matched, another design for this study to minimize group skill differences would have been to have each surgeon be their own control.

Table 6. Effect of Training Level on Performance Metrics (Warm-up vs Control)

Outcomes	Residents (n = 27)				Faculty (n = 24)			
	Control, mean (SD) (n = 13)	Warm-up, mean (SD) (n = 14)	Difference (95% CI)	p Value	Control, mean (SD) (n = 12)	Warm-up, mean (SD) (n = 12)	Difference (95% CI)	p Value
Economy of motion, mm/s	4.51 (0.64)	4.69 (0.64)	0.2 (-0.2 to 0.6)	0.35	4.32 (0.68)	4.55 (0.14)	0.2 (-0.2 to 0.6)	0.25
Task time, s	266.6 (51.9)	235.6 (40.5)	-31.0 (-55.7 to -6.4)	0.013	261.8 (62.7)	234.4 (9.4)	-27.4 (-53.5 to -1.4)	0.039
Peg touches, counts	22.1 (10.3)	20.6 (8.3)	-1.5 (-7.6 to 4.7)	0.64	21.3 (9.9)	18.0 (2.3)	-3.3 (-9.8 to 3.2)	0.32
Cognitive errors, counts	0.10 (0.38)	0.12 (0.47)	0.02 (-0.14 to 0.17)	0.84	0.14 (0.47)	0.0 (0.00)	-0.14 (-0.30 to 0.03)	0.10
Path length, mm/s	1,188 (180)	1,092 (140)	-96 (-181 to -10)	0.029	1,109 (192.7)	1,043 (140.8)	-66 (-156 to 23)	0.15

The mean (SD) and estimated difference between warm up and control for the 5 continuous outcomes measured in sessions 1 to 3 by training level, where faculty are defined as PGY >6 and residents ≤6.

Second, we strove for intervals between sessions to never occur <24 hours apart or 24 hours from earlier robotic surgery so that one robotic performance did not warm-up the earlier one. However, we do not know if the 24-hour interval extrapolates to robotic surgery. In addition, participants ideally should not have had longer than 2 weeks between sessions, but this was not logistically feasible in some circumstances. Many of our surgeons were on active clinical services and rotated through services that altered the consistency of their intervals. Recognizing the work of Jenison and colleagues, which showed that after 4 weeks of rest, robotic surgery skills degrade, we strove to minimize the number of intervals that exceeded this threshold.³⁵ We did not, however, adjust surgeons' data based on intervals between sessions.

We have validated portions of the proficiency curriculum using this tracking methodology, but there is potential for varied signal integrity throughout the sessions. The proprietary Ascension software provided us with real-time read-outs of the quality of the signal and all our surgeons' sessions fell within the quality requirements of the tracking system, so we believe that we captured accurate data. In addition, task time and error recognition were not dependent on the tracker data. Signal quality between the transmitter and the sensors on the instruments can be affected by the amount of ferrous material and components generating their own electromagnetic fields. Before enrolling participants, we tested the optimal positioning of the sensors, the transmitter, and the various dry modules to minimize signal distortion. We standardized the positioning of the arms of the robot in relation to the task modules and the transmitter by creating:

1. A jig that housed each module in a fixed position relative to the transmitter (Fig. 2);
2. An optimal orientation holder for the sensors on the tools by testing multiple rapid prototyped interface elements before study launch (Fig. 5);
3. Plumbs that dangled from set positions on the camera and instrument trocars down to the task module jig that allowed us to set up the robot in identical port configurations between sessions (Fig. 2); and
4. Calibration software that tested for sufficient data inputs from all systems before each task iteration commenced.

Alternative instrument tracking methods could include optical fiducials that can be tracked by cameras within the OR, such as those used by Lee and colleagues for their intra-operative laparoscopic study.¹⁹ They tracked surgeon arm and hand movements by affixing sterile markers on the gowns and gloves of the surgeons and used high-resolution cameras to detect precise movements. The

advantage of this method is that intraoperative tracking is possible because the markers are sterile, and although the electromagnetic tracker sensors are sterilizable, the transmitter needs to be within 1 m of the sensors, prohibiting its practicality in the OR. The disadvantage with optical tracking is that this method requires clear line of sight, which is not always possible in the OR. Perhaps a preferable method would be to capture data directly from the da Vinci application programming interface, which has the capability of providing >100 data elements of the instruments' movements in real-time, but such access is limited to a few centers through contractual agreements with Intuitive Surgical, Inc, and the application programming interface does not capture video or tool contact data.³⁶

Finally, our findings were unambiguous in a dry laboratory setting, yet the true test of robotic surgery VR warm-up will need to be in the OR, as Calatayud and colleagues and Mucksavage and colleagues did for conventional laparoscopy.^{11,12} This fundamental research in the robotic dry laboratory setting, however, highlights the potential benefit using preoperative VR warm-up for patient robotic surgery to improve patient outcomes and reduce costs. Our experiment used the MIMIC dV-Trainer, which is a desktop platform that has the same VR modules as the current Intuitive backpack simulator that drives VR simulation modules at the da Vinci console. So our findings might be easily translatable into the OR due to the parity between our VR curriculum and what is available today in the OR on the da Vinci Si system. This is a decided advantage for use with robotics systems because the software package that generates the virtual images can reside on any robotic system and, therefore, the preoperative warm-up would actually become part of the operative procedure. Preoperative warm-up in open or laparoscopic surgery, on the other hand, requires an entirely separate simulator to be available in the OR for the surgeon to practice the warm-up. Likewise, in future-generation robotic systems, not only will a warm-up module be included in the robotic system, but downloading patient-specific images (from CT or MRI scans) will also enable the surgeon to perform surgical rehearsal of the critical parts of the operation, so that any errors can be discovered and avoided during the actual operation. The value of mission rehearsal has proven to be of great value in many other domains, such as the military and aviation, and has the potential to greatly increase patient safety in surgery as well.³⁷⁻³⁹

CONCLUSIONS

A brief VR robotic simulation warm-up improves robotic surgery task performance and reduces errors for experienced and inexperienced robotic surgeons in a dry

laboratory setting. Further investigation is required to see if these results translate to the OR. These data provide a foundation for future predictive validation studies assessing the role of robotic warm-up for improved patient outcomes and reduced operative cost, and pave the way for novel preprocedural rehearsal investigation in all areas of surgery.

Author Contributions

Study conception and design: Lendvay, Brand, White, Kowalewski, Mercer, Hannaford, Satava

Acquisition of data: Lendvay, Brand, White, Kowalewski, Jonnadula, Khorsand, Andros

Analysis and interpretation of data: Lendvay, Brand, White, Kowalewski, Jonnadula, Mercer, Khorsand, Andros, Hannaford

Drafting of manuscript: Lendvay, Brand, White, Kowalewski, Jonnadula, Mercer, Satava

Critical revision: Lendvay, Brand, White, Kowalewski, Jonnadula, Mercer, Khorsand, Andros, Hannaford, Satava

Acknowledgment: The authors thank Farrah Leland for assistance with grant preparation.

REFERENCES

1. Zhan C, Miller M. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *JAMA* 2003;290:1868–1874.
2. Kohn L, JM C, Donaldson M, eds. *To Err is Human: Building a Safer Health Care System*. Washington, DC: National Academy Press; 2000.
3. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009;360:491–499.
4. Batalden P, Leach D, Swing S, et al. General competencies and accreditation in graduate medical education. *Health Affairs* 2002;21:103–111.
5. Healy GB. The college should be instrumental in adapting simulators to education. *Bull Am Coll Surgeons* 2002;11:10–11.
6. Sroka G, Feldman L, Vassiliou M, et al. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial. *Am J Surg* 2010;199:115–120.
7. Lendvay T, Casale P, Sweet R, Peters C. Initial validation of a virtual-reality robotic simulator. *J Robotic Surg* 2008;2:145–149.
8. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002;236:458–464.
9. Peters JH, Fried GM, Swanstrom LL, et al, SAGES FLS Committee. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 2004;135:21–27.

10. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005;241:364.
11. Calatayud D, Arora S, Aggarwal R, et al. Warm-up in a virtual reality environment improves performance in the operating room. *Ann Surg* 2010;251:1181.
12. Mucksavage P, Lee J, Kerbl D, et al. Preoperative warming up exercises improve laparoscopic operative times in an experienced laparoscopic surgeon. *J Endourol* 2012;26:765–768.
13. Anshel MA. The effect of arousal on warm-up decrement. *Res Q Exerc Sport* 1985;56:1–9.
14. Anshel MH, Wrisberg CA. Reducing warm-up decrement in the performance of the tennis serve. *J Sport Exerc Psychol* 1993;15:290–303.
15. Wrisberg CA, Salmoni AW, Schmidt RA. Warm-up effects in the learning of discrete motor skills. *Acta Psychol* 1975;39:311–320.
16. Nascon J, Schmidt RA. The activity-set hypothesis for warm-decrement. *J Motor Behav* 1971;3:1–16.
17. Do AT, Cabbad M, Kerr A, et al. A warm-up laparoscopic exercise improves the subsequent laparoscopic performance of OB-GYN residents: A low-cost laparoscopic trainer. *JLS* 2006;10:297.
18. Kahol K, Satava R, Ferrara J, et al. Effect of short-term pretrial practice on surgical proficiency in simulated environments: a randomized trial of the “preoperative warm-up” effect. *J Am Coll Surg* 2009;208:255–268.
19. Lee J, Mucksavage P, Kerbl D, et al. Laparoscopic warm-up exercises improve performance of senior-level trainees during laparoscopic renal surgery. *J Endourol* 2012;26:545–550.
20. Snijders T. Power and sample size in multilevel linear models. In: Everitt BS, Howell DC, eds. *Encyclopedia of Statistics in Behavioral Science*, Vol. 3. Chichester: Wiley; 2005: 1570–1573.
21. Tausch TJ, Kowalewski TM, White LW, et al. Content and construct validation of a robotic surgery curriculum using an electromagnetic instrument tracker. *J Urol* 2012;188: 919–923.
22. White LW, Kowalewski T, Hannaford B, Lendvay TS. Surg-Trak: affordable motion tracking and video capture for the da Vinci surgical robot. *Proc 2011 Meeting SAGES* 2012;1: 204.
23. Schroeder D, Keefe D, Kowalewski T, et al. Visualizing surgical training databases: exploratory visualization, data modeling, and formative feedback for improving skill acquisition. *IEEE Comp Graph Appl* 2012;32:71.
24. Hamilton CE, Mola WR. Warm-up effect in human maze learning. *J Exp Psychol* 1953;45:437–441.
25. Anshel MH, Wrisberg CA. The effect of arousal and focused attention on warm-up decrement. *J Sport Behav* 1988;11: 18–31.
26. Stefanidis D, Walters KC, Mostafavi A, et al. What is the ideal interval between training sessions during proficiency-based laparoscopic simulator training? *Am J Surg* 2009;197:126–129.
27. Arora S, Aggarwal R, Sirimanna P, et al. Mental practice enhances surgical technical skills: a randomized controlled study. *Ann Surg* 2011;253:265–270.
28. Pugh C. Warm-ups, mental rehearsals, and deliberate practice: adopting the strategies of elite professionals. *J Surg Res* 2012; 176:404–405.
29. Gallagher AG, Richie K, McClure N, et al. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World J Surg* 2001;25:1478–1483.
30. Gunther S, Rosen J, Hannaford B, et al. The red DRAGON: a multi-modality system for simulation and training in minimally invasive surgery. *Stud Health Technol Inform* 2007; 125:149–154.
31. Figert PL, Park AE, Witzke DB, et al. Transfer of training in acquiring laparoscopic skills. *J Am Coll Surg* 2001;193:533.
32. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective assessment. *Surg Endosc* 2003;17:220–226.
33. Zhou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159: 702–706.
34. Lumley T, Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. UW Biostatistics Working Paper Series. Working Paper 293, July 2006. Available at: <http://www.bepress.com/uwbiostat/paper293>. Accessed February 10, 2013.
35. Jenison EL, Gil KM, Lendvay TS, Guy MS. Robotic surgical skills: acquisition, maintenance, and degradation. *JLS* 2012; 16:218–228.
36. Lin HC, Shafran I, Yuh D, Hager GD. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. *Comput Aided Surg* 2006;11:220–230.
37. Makiyama K, Nagasaka M, Inuiya T, et al. Development of a patient-specific simulator for laparoscopic renal surgery. *Int J Urol* 2012;19:829–835.
38. Miller DC, Thorpe JA. SIMNET: the advent of simulator networking. *Proc IEEE* 1995;83:1114–1123.
39. Proctor MD, Bauer M, Lucario T. Helicopter flight training through serious aviation gaming. *J Def Mod Simul* 2007;4: 277–294.