Published in final edited form as:

J Am Coll Surg. 2013 June ; 216(6): 1181–1192. doi:10.1016/j.jamcollsurg.2013.02.012.

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

Thomas S. Lendvay, MD, FACS[†], Timothy C. Brand, MD, FACS[‡], Lee White, BSc (Hons), PhD^{****}, Timothy Kowalewski, PhD^{*}, Saikiran Jonnadula, MD^{**}, Laina Mercer, MS^{***}, Derek Khorsand, BSc(Hons)^{**}, Justin Andros, BSc(Hons)^{**}, Blake Hannaford, PhD^{*}, and Richard M. Satava, MD, FACS^{**}

[†] University of Washington School of Medicine, Department of Urology, Seattle, WA

^{*} University of Washington, Department of Electrical Engineering, Seattle, WA

^{**} University of Washington School of Medicine, Department of Surgery, Seattle, WA

^{***} Seattle Children's Research Institute, Core for Biomedical Statistics, Seattle, WA

^{****} University of Washington, Department of Bioengineering, Seattle, WA

[‡] Madigan Army Medical Center, United States Army, Tacoma, WA

Abstract

Background—Pre-operative 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 a brief virtual reality (VR) robotic warm-up would enhance robotic task performance and reduce errors.

Study Design—In a two-center randomized trial, fifty-one 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. Once successfully achieving performance benchmarks, surgeons were randomized to either receive a 3-5 minute VR simulator warm-up or read a leisure book for 10 minutes prior to 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.

© 2013 American College of Surgeons. Published by Elsevier Inc. All rights reserved.

Correspondence address: Thomas S. Lendvay, MD, FACS thomas.lendvay@seattlechildrens.org Seattle Children's Hospital 4800 sand Point Way, NE Seattle, WA 98105 Phone: (o) 206-987-1461, (c) 206-349-2212.

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

Disclosure Information: Nothing to disclose.

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

Results—Task time (-29.29sec, $p=0.001$, 95% CI -47.03,-11.56), path length (-79.87mm, $p=0.014$, 95% CI -144.48,-15.25), and cognitive errors were reduced in the warm-up group compared to 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 prior robotic and laparoscopic clinical experience, the more experienced surgeons($n=17$) demonstrated significant improvements from warm-up in task time (-53.5sec, $p=0.001$, 95% CI -83.9,-23.0) and economy of motion (0.63mm/sec, $p=0.007$, 95% CI 0.18,1.09), whereas improvement in these metrics was not statistically significantly appreciated in the less experienced cohort($n=34$).

Conclusions—We observed a significant performance improvement and error reduction rate 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.

Keywords

robotic; education; simulation; warm-up; rehearsal; da Vinci; surgery; virtual reality; laparoscopy; SurgTrak™

INTRODUCTION

Methods to improve surgical performance for trainees and practicing surgeons have become a national mission to mitigate surgical morbidity, reduce healthcare costs, accelerate learning curves, provide curricula for the introduction of new surgical technologies, and ensure that reductions in duty hours for trainees do not 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 significantly different time than the actual surgery on patients. But recent studies suggest that surgical simulation immediately before criterion surgical tasks may benefit performance. [11,12] This pre-surgical 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) improves performance in both surgical trainees and experienced minimally invasive surgeons.

High stakes professions like athletics and performing arts have long relied on the principles of the warm-up decrement (WUD, the decrease in performance after a period of rest) and the Activity Set hypothesis (the idea that to counter the WUD, 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 pre-surgical rehearsal though it is a high stakes profession drawing on intense psychomotor and cognitive efforts. The benefits of warm-up may be particularly important for robotic surgery due to 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 thus cognitive arousal is likely to greatly benefit from warm-up.

Do et al. was first to use a laparoscopic box trainer to study the effect of warm-up exercises on follow-up laparoscopic tasks and observed significant improvement in performance (25%) for both residents, irrespective of the post graduate year (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 versus a true warm-up effect and so Kahol et al. sought to address this in a laparoscopic VR simulation study. [18] Surgeons were randomized to either receive warm-up or no warm-up using a series of VR ring-transfer tasks that tested psychomotor, attentional, and visio-spatial skills. The results yielded a significant reduction in errors (33%). In addition, Kahol et al. showed that the warm-up effect was demonstrated in surgeons of all levels of expertise and generalized to dissimilar follow-up tasks, like an electrocautery task. In Kahol's study, both warm-up tasks and criterion tasks were in a virtual lab. But in 2010, Calatayud et al. showed that a VR simulation warm-up in the OR benefited residents performing laparoscopic cholecystectomies. Eight residents demonstrated by higher global performance scoring on an Objective Structured Assessment of Technical Skills (OSATS) tool. [11] In 2012, Lee et al. 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. Thus we sought to explore the role VR robotic warm-up has on similar and dissimilar robotic surgery tasks.

METHODS

Study Design

Experienced minimally invasive surgery faculty in General Surgery, Urology, and Gynecology from two medical centers underwent a validated robotic surgery proficiency curriculum on a VR robotic simulator and on the da Vinci surgical robot. Once successfully achieving performance benchmarks, each surgeon was randomized to either receive a 3-5 minute VR warm-up on the simulator or read a leisure book for 10 minutes prior to 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, technical and cognitive errors.

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 (UWMC) and Madigan Army Medical Center (MAMC) 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, musical and video-gaming experience, and minimally invasive surgery (MIS) experience; all play roles in surgical skill acquisition. Post-graduate year – 1-6, 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. Since the within-person dependence was not known precisely, 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, three 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 as 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 standard deviation.

Randomization—Permuted blocks randomization was used. Randomization was stratified by site (UWMC, MAMC) and surgical experience level (resident, 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 below)

Participant Flow [Figure 1.]—Once enrolled, each surgeon went through a robotic proficiency curriculum which included the 90-minute da Vinci (Intuitive Surgical Inc, Sunnyvale, CA) online didactics module to familiarize the surgeons with the da Vinci S/Si systems. After passing the tutorial, each surgeon went through a virtual reality (dV-Trainer simulator, MIMIC Technologies, Inc., Seattle, WA) and da Vinci dry lab robotics curriculum comprised of four progressively harder surgical skills modules on each platform, respectively. [Figures 2. and 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 the above plus adding motion to the task platform to test spatial relations capabilities. Proficiency benchmarks were established for each module based on performance by two experienced robotic surgeons (TSL, TCB) who have performed more than 150 robotic surgeries each. The benchmark required that each surgeon perform 2 consecutive task iterations within 120% of the mean task time of the two 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 less than 120% of the mean of the two benchmark surgeons performing the same task and with no ring drops or sequence errors. We chose two consecutive iterations of success to try to hone the legitimate proficiency of the surgeon for each task. We chose 120% of task time since 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 the ‘learning effect’ addressed: In order to mitigate the confounding effects of the learning curve throughout the study, each surgeon was required to reach proficiency benchmarks prior to 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. In order 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. [Figure 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 below)- to derive path length and economy of motion performance metrics. [21-23] Again, proficiency benchmarks had been obtained from the same two surgeons and 120% of the mean task times and zero-error rates through two consecutive iterations were required to advance to the next module. The modules included two Fundamentals of Laparoscopic Surgery (FLS) tasks being performed on the da Vinci robot – block transfer and intracorporeal suturing – as these have been repeatedly validated in laparoscopy curricula. [6,9]

Trial Sessions—After a surgeon reached proficiency, he/she was randomized to either the warm-up group or control group. Four trial sessions per surgeon were performed. The first three 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 surgeons performed the Pegboard Level 3 VR task once directly prior to performing the analogous da Vinci rotating rocking pegboard task. This generally took 3-5 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 prior to performing the da Vinci criterion task so as to minimize the likelihood that they were visually imagining the task to be performed since 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 felt that these also may prime the control surgeons.

During the fourth trial session, the warm-up and control pre-criterion process was the same as the first three 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, millimeters).
5. Economy of motion – Path length/Task Time (mm/sec).

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 pre-marked 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 by custom software. Video was recorded at 30 frames per second from the digital video imaging (DVI) output from the da Vinci Si/S master console using a DVI2USB® device (Epiphan Systems Incorporated, Ottawa, Ontario, Canada). Video was encoded using mpeg-4 compression to produce compact, manageable files.

Tool motion data was recorded at 30 Hertz (Hz). Tool position and orientation were captured with a 3D Guidance trakSTAR™ electromagnetic tracking system (Ascension Technology Corporation, Burlington, VT, USA). We retrofitted da Vinci training tools with rapid prototyped holders for the sensors on the proximal ends of the da Vinci instruments. [Figure 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, Alberta, Calgary, Canada). Peg touch errors from the rocking pegboard task were detected and the time of occurrence 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, Redmond, WA.) 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 1-3 was a test for the overall mean difference between the warm-up and the control groups. Because each surgeon contributed three observations to the data set, this test for continuous outcomes, was calculated using a repeated measures (RM) Analysis of Variance (ANOVA) model and the effect of experience level was investigated two different ways: by training level (resident vs. faculty) and by surgical experience (> 10 robotic and > 10 laparoscopic surgeries performed as the primary surgeon vs. ≤ to 10 cases in each modality). For binary

outcomes, repeated measures relative risk regression [33] was used to compare groups and test for interactions. Each surgeon only contributed one observation to the data for session 4 outcomes, thus 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, thirty-one from UWMC and twenty from MAMC 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 1-3, testing whether warm-up improved performance with similar VR and criterion tasks, we observed a statistically significant decrease in the task time (-29.29 seconds, $p=0.001$, 95% CI -47.03,-11.56) and path length (-79.87 mm, $p=0.014$, 95% CI -144.48,-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 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 ($p=0.087$) (Figure 6, Tables 2 and 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, 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/sec, $p=0.007$, 95% CI 0.18,1.09), task time (-53.5 sec, $p=0.001$, 95% CI -83.9,-23.0), and path length (-97 mm, $p=0.093$, 95% CI -210,16) favored the warm-up sub-group among the experienced cohort, whereas only path length (-75 mm, $p=0.063$, 95% CI -154, 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,-10) and task time (-31 sec, p=0.013, 95% CI -55.7,-6.4) were reduced in the resident warm-up group, whereas task time reduction only (-27.4 sec, p=0.039, 95% CI -53.5,-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 lab tasks. In our randomized study comparing warm-up and control groups of experienced and inexperienced surgeons, we demonstrated that pre-procedural warm-up does improve task performance and error reduction. This is a fundamental observation because, to date, the literature has established warm-up's potential role in conventional laparoscopy, but not in robotic surgery. Furthermore, 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 benefitting 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 et al. 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 may 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 et al. saw a warm-up benefit for a dissimilar intra-operative 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 may 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 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, may be additive to the warm-up benefit. Kahol et al. specifically emphasized that warm-up tasks need to not only 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 to the low frequency at which these errors occurred, such as peg touches. In order 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. Whereas less MIS-experienced surgeons not only require task priming, but may spend additional attentional capacity on performing the basic robotic manipulations (grasping, object transfer, camera and arm clutching). Gallagher et al. 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 regarding experience are significant because there are far more practicing robotic surgeons than there are robotic surgery trainees, and our findings may be relevant to hospital credentialing and maintenance of certification processes. When we divided the cohort by faculty vs. resident, our results were mixed. This may 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 which should be mentioned. First, although we randomized surgeons to one of two groups – warm-up or control – our proficiency curriculum may 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.

Second, we strove for intervals between sessions to never occur less than 24 hours apart or 24 hours from prior robotic surgery so that one robotic performance did not warm-up the prior one. However, we do not know if the 24 hr interval extrapolates to robotic surgery. In addition, participants ideally should have not 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 et al. showing 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 readouts of the quality of the signal and all our surgeons' sessions fell within the quality requirements of the tracking system so we feel 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. Prior to 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, [Figure 2.] 2) an optimal orientation holder for the sensors on the tools by testing multiple rapid prototyped interface elements prior to study launch [Figure 5.], 3) plumbs that dangled from set positions on the camera and instrument trocars down to the task module jig allowed us to set up the robot in identical port configurations between sessions, [Figure 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 which can be tracked by cameras within the OR such as used by Lee et al. 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 intra-operative tracking is possible because the markers are sterile, and although the electromagnetic tracker sensors are sterilizable, the transmitter needs to be within 1 meter of the sensors thus 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 (API) which has the capability of providing over 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 API does not capture video or tool contact data. [36]

Finally, our findings were unambiguous in a dry lab setting, yet the true test of robotic surgery VR warm-up will need to be in the operating room as Calatayud et al. and Mucksavage et al. did for conventional laparoscopy. [11,12] This fundamental research in the robotic dry lab setting, however, highlights the potential benefit employing pre-operative VR warm-up for patient robotic surgery for improving patient outcomes and reducing costs. Our experiment utilized 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 may 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 the use with robotics systems, since the software package that generates the virtual images can reside on any robotic system, and therefore the pre-operative warm-up would actually become part of the operative procedure. Pre-operative warm-up in open or laparoscopic surgery, on the other hand, requires an entirely separate simulator to be available in the operating room 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 computed tomography or magnetic resonance imaging scans) will also enable the surgeon to perform surgical rehearsal of the critical parts of the operation, such 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 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 lab setting. Further investigation is required to see if these results translate to the operating room. These data provide a foundation for future predictive validation studies assessing the role of robotic warm-up for improved patient outcomes, reduced operative cost, and pave the way for novel pre-procedural rehearsal investigation in all areas of surgery.

Acknowledgments

This study was supported by the Department of Defense US Army Medical Research and Materiel Command under award number W81XWH-09-1-0714 (PI: Lendvay). Views and opinions of, 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.

Abbreviations

VR	virtual reality
FLS	Fundamentals of Laparoscopic Surgery
PGY	post graduate year
UWMC	University of Washington Medical Center
MAMC	Madigan Army medical Center
OR	operating room
DVI	digital video input
WUD	warm-up decrement
ICC	interclass correlations
API	application programming interface

REFERENCES

1. Zhan C, Miller M. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *JAMA*. 2003; 290:1868–1874. [PubMed: 14532315]
2. Kohn, L.; JM, C.; Donaldson, M., editors. *To Err is Human: Building a Safer Health Care System*. National Academy Press; Washington, DC: 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. [PubMed: 19144931]
4. Batalden P, Leach D, Swing S, et al. General competencies and accreditation in graduate medical education. *Health Affairs*. 2002; 21:103–111. [PubMed: 12224871]
5. Healy GB. The college should be instrumental in adapting simulators to education. *Bull Am Coll Surgeons*. 2002; 11:10–11. [PubMed: 17387916]
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. [PubMed: 20103076]
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. [PubMed: 12368674]
9. Peters JH, Fried GM, Swanstrom LL, et al. the 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. [PubMed: 14694297]
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. [PubMed: 15650649]
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. [PubMed: 20485133]
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. [PubMed: 22050510]
13. Anshel MA. The effect of arousal on warm-up decrement. *Res Quarterly for Exerc and 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 Psychologica.* 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. *JSLs.* 2006; 10:297. [PubMed: 17212883]
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. [PubMed: 19228538]
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. [PubMed: 22192095]
20. Everitt, BS.; Howell, DC. *Encyclopedia of Statistics in Behavioral Science.* Vol. 3. Wiley; Chichester: 2005. Chapter contributed by Tom Snijders entitled “Power and Sample Size in Multilevel Linear Models.”; p. 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. [PubMed: 22819403]
22. White LW, Kowalewski T, Hannaford B, Lendvay TS. SurgTrak: 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 CG&A.* 2012; 32:71.
24. Hamilton CE, Mola WR. Warm-up effect in human maze learning. *J Exp Psychol.* 1953; 45:437–441. [PubMed: 13052886]
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. [PubMed: 19101255]
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. [PubMed: 21245669]
28. Pugh C. Warm-ups, mental rehearsals, and deliberate practice: Adopting the strategies of elite professionals. *J Surg Research.* 2012; 176:404–405. [PubMed: 21816430]

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. [PubMed: 11760752]
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 Informatics.* 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. [PubMed: 11708511]
32. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective assessment. *Surg Endosc.* 2003; 17:220–226. [PubMed: 12436236]
33. Zhou G. A Modified Poisson Regression approach to prospective studies with binary data. *Am J Epidemiol.* 2004; 159:702–706. [PubMed: 15033648]
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 [<http://www.bepress.com/uwbiostat/paper293>, checked 2/10/2013]. July, 2006
35. Jenison EL, Gil KM, Lendvay TS, Guy MS. Robotic surgical skills: Acquisition, maintenance, and degradation. *JSLs.* 2012; 16:218–228. [PubMed: 23477169]
36. Lin HC, Shafran I, Yuh D, Hager GD. Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Comp 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. [PubMed: 22640644]
38. Miller DC, Thorpe JA. SIMNET: The advent of simulator networking. *Proceedings 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.

Figure 1.
Patient flow diagram.

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.

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 lab modules from left to right. FLS block transfer, FLS intracorporeal suturing (this was the criterion task for sessions #4), Ring tower (The Chamberlain Group, Great Barrington, MA), Rotating rocking pegboard (this was the criterion task for sessions #1-3).

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

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

Table 1

Demographics and Baseline Characteristics by Intervention Group

Variable	Control (n=25)	Warm up (n=26)	p Value *
Age, y $\pm$ SD	35.32 $\pm$ 6.47	33.85 $\pm$ 5.82	0.40
Sex, n (%)			
Female	10 (40.0)	9 (34.6)	0.66
Male	15 (60.0)	17 (65.4)	
Musical instrument for >3 y, n (%)			
No	7 (28.0)	9 (34.6)	0.76
Yes	18 (72.0)	17 (65.4)	
Handedness, n (%)			
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, n (%)			
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, n (%)			
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, n (%)			
None	15 (60.0)	16 (61.5)	0.99
<2 $\times$ week	7 (28.0)	6 (23.1)	
2+ $\times$ week	3 (12.0)	4 (15.4)	
Laparoscopic cases, primary surgeon, n (%)			
None	1 (4.0)	0 (0.0)	0.55
10 or less	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, n (%)			
None	9 (36.0)	8 (30.8)	0.61
10 or less	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.

Table 2

Continuous Outcomes by Study Group (Sessions 1-3)

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, 0.47)	0.13
Task time	264.31	56.97	235.01	40.11	−29.29 (−47.03, −11.56)	0.001
Total peg touches	21.68	10.06	19.38	9.01	−2.29 (−6.71, 2.12)	0.31
Cognitive error	0.12	0.40	0.06	0.30	−0.06 (−0.17, 0.06)	0.34
Path length	1149.23	189.03	1069.37	132.97	−79.87 (−144.48, −15.25)	0.014

Each outcome was individually analyzed with repeated measures ANOVA.

Table 3

Binary Outcomes by Study Group (Sessions 1-3)

Proportion of sessions with error					
Error type	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

Each outcome was individually analyzed with relative risk (RR) regression.

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, 22.6)	0.70
Economy of motion	3.69	0.86	3.82	0.80	−0.14 (−0.61, 0.33)	0.56
Path length	401.4	114.4	401.5	134.8	0.0 (−71.2, 71.1)	0.99
Global tech. error, count [*]	0.44	0.58	0.12	0.33	0.32 (0.06, 0.59)	0.020

Each outcome was individually analyzed with a *t*-test.

^{*} Global Technical Error = composite of air knot, needle targeting errors by FLS (Entrance and Exit dots errors).

Table 5

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

Outcomes	<= 10 Robotic and <=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, 0.34)	0.90	4.31 (0.78)	4.94 (0.77)	0.63 (0.18, 1.09)	0.007
Task time, s	258.6 (43.7)	240.8 (39.6)	-17.8 (-39.2, 3.5)	0.10	272.9 (72.6)	219.4 (41.0)	-53.5 (-83.9, -23.0)	0.001
Peg touches, counts	24.2 (8.8)	20.7 (9.8)	-3.6 (-8.8, 1.7)	0.18	17.9 (10.7)	15.9 (6.1)	-2 (-9.4, 5.5)	0.60
Cognitive errors, counts	0.13 (0.45)	0.05 (0.30)	-0.08 (-0.22, 0.06)	0.27	0.10 (0.40)	0.10 (0.45)	0 (-0.21, 0.20)	0.96
Path length, mm/s	1,152 (174)	1,077 (140)	-75 (-154, 4)	0.06	1145 (213)	1049 (118)	-97 (-210, 16)	0.09

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

Table 6

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

Outcomes	Residents (n=27)		Faculty (n=24)		p Value	Difference (95% CI)	Warm up Mean (SD) (n=12)	Control Mean (SD) (n=12)	p Value	Difference (95% CI)	p Value
	Control Mean (SD) (n=13)	Warm up Mean (SD) (n=14)	Control Mean (SD) (n=12)	Warm up Mean (SD) (n=12)							
Economy of motion, mm/s	4.51 (0.64)	4.69 (0.64)	4.32 (0.68)	4.55 (0.14)	0.35	0.2 (-0.2, 0.6)	4.55 (0.14)	4.32 (0.68)	0.35	0.2 (-0.2, 0.6)	0.25
Task time, s	266.6 (51.9)	235.6 (40.5)	261.8 (62.7)	234.4 (9.4)	0.013	-31.0 (-55.7, -6.4)	234.4 (9.4)	261.8 (62.7)	0.013	-27.4 (-53.5, -1.4)	0.039
Peg touches, cnts	22.1 (10.3)	20.6 (8.3)	21.3 (9.9)	18.0 (2.3)	0.64	-1.5 (-7.6, 4.7)	18.0 (2.3)	21.3 (9.9)	0.64	-3.3 (-9.8, 3.2)	0.32
Cognitive errors, cnts	0.10 (0.38)	0.12 (0.47)	0.14 (0.47)	0.0 (0.00)	0.84	0.02 (-0.14, 0.17)	0.0 (0.00)	0.14 (0.47)	0.84	-0.14 (-0.30, 0.03)	0.10
Path length, mm/s	1,188 (180)	1,092 (140)	1,109 (192.7)	1043 (140.8)	0.029	-96 (-181, -10)	1043 (140.8)	1,109 (192.7)	0.029	-66 (-156, 23)	0.15

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