# Predictive Validity of a Training Protocol Using a Robotic Surgery Simulator

*Patrick Culligan, MD,\* Emil Gurshumov, MD,\* Christa Lewis, DO,\* Jennifer Priestley, PhD,†
Jodie Komar, BSN,\* and Charbel Salamon, MD\**

**Background:** Robotic surgery simulation may provide a way for surgeons to acquire specific robotic surgical skills without practicing on live patients.

**Methods:** Five robotic surgery experts performed 10 simulator skills to the best of their ability, and thus, established expert benchmarks for all parameters of these skills. A group of credentialed gynecologic surgeons naive to robotics practiced the simulator skills until they were able to perform each one as well as our experts. Within a week of doing so, they completed robotic pig laboratory training, after which they performed supracervical hysterectomies as their first-ever live human robotic surgery. Time, blood loss, and blinded assessments of surgical skill were compared among the experts, novices, and a group of control surgeons who had robotic privileges but no simulator exposure. Sample size estimates called for 11 robotic novices to achieve 90% power to detect a 1 SD difference between operative times of experts and novices ($\alpha = 0.05$).

**Results:** Fourteen novice surgeons completed the study—spending an average of 20 hours (range, 9.7–38.2 hours) in the simulation laboratory to pass the expert protocol. The mean operative times for the expert and novices were 20.2 (2.3) and 21.7 (3.3) minutes, respectively ($P = 0.12$; 95% confidence interval, −1.7 to 4.7), whereas the mean time for control surgeons was 30.9 (0.6) minutes ($P < 0.0001$; 95% confidence interval, 6.3–12.3). Comparisons of estimated blood loss (EBL) and blinded video assessment of skill yielded similar differences between groups.

**Conclusions:** Completing this protocol of robotic simulator skills translated to expert-level surgical times during live human surgery. As such, we have established predictive validity of this protocol.

**Key Words:** da Vinci, Morristown Protocol, robotic simulator, robotic surgery

(*Female Pelvic Med Reconstr Surg* 2014;20: 48–51)

The last 3 years have seen a dramatic increase in utilization of robotic assisted surgery in the United States and around the world.[1] Despite recent studies that raised concerns about the cost-effectiveness and overall comparative effectiveness of robotic assisted surgery,[2,3] more and more surgeons from an increasing variety of specialties are seeking to incorporate robotics in their practices.[4] After first gaining popularity as a tool for minimally invasive prostatectomy, the da Vinci Surgical System (Intuitive Surgical, Sunnyvale, Calif) is now being used in the

fields of gynecologic oncology, general gynecology, general and colorectal surgery, head and neck surgery, and cardiothoracic surgery. Possible advantages offered by the robotic system include the 3-dimensional high-definition view provided by a stable surgeon-controlled camera and the wristed motion of the instruments. The onslaught of new surgeons seeking to use the da Vinci system raises questions about the current training and credentialing practices for robotic surgery. Even among otherwise experienced surgeons, a steep learning curve tends to exist when they begin to incorporate robotic assistance.[5] Shortening this learning curve remains an important goal—to shorten operative times and improve patient outcomes.[6] One possible solution to the problem of providing robotic surgical training to an ever-growing population of attending and resident surgeons is virtual reality simulation.

The first commercially available virtual reality simulator for the da Vinci Surgical System was recently produced by a joint venture between Intuitive Surgical and Mimic Technologies (Seattle, Wash). This platform allows the user to sit at the actual da Vinci surgeon console and see 3-dimensional images while performing the simulated surgical skills.

Initial studies of this device and software established their face, content, and construct validity,[7–9] but no studies have determined whether mastery of the simulator can translate into actual surgical proficiency on a live human (ie, predictive validity).

Therefore, our objective was to determine whether completion of a rigorous robotic surgical simulation protocol could provide novice robotic surgeons with actual advanced robotic surgery skills in an operating room setting.

## MATERIALS AND METHODS

This study was approved by the Atlantic Health System institutional review board (R11-01-018) and was listed on www.clinicaltrials.gov (NCT01618994) before initiation of the study. This was an investigator-initiated study. Although the study was supported by Intuitive Surgical, this support primarily amounted to supplying a surgical simulator that was only used for the study and was returned upon study completion. Intuitive Surgical had no control over study methods or reporting of results.

We conducted our study in 2 main phases. The first phase involved establishment of our training protocol using a commercially available robotic surgery simulation platform (da Vinci Skills Simulator; Intuitive Surgical). Once established, we named this set of exercises and benchmark the "Morristown Protocol." The second phase of the study involved validation of the Morristown Protocol as a way to enhance the training of novice robotic surgeons.

At the time of our study, the da Vinci Skills Simulator included 24 exercises and scenarios designed to help surgeons improve their proficiency with the robotic console controls. At the initiation of this study, there were 5 gynecologic surgeons who were averaging at least 75 gynecologic robotic cases per year at our center. We gathered these 5 surgeons and asked

each to perform all of the simulator exercises and to categorize each one as "definitely helpful," "possibly helpful," or "not very helpful" for novice robotic surgeons in training. The resultant 5 lists were compared and found to be quite similar—allowing for selection of the 10 simulator skills that constituted our protocol. All 10 skills had been categorized as "definitely helpful" by all of our experts. The mean number of robotic cases performed by our group of experts in the 12 months before our study was 142 (range, 77–246).

Then, over the course of several weeks, all 5 expert robotic surgeons performed these 10 simulator skills to the best of their ability, and all attempts were recorded. Each surgeon was allowed to perform each simulator skill as many times as they wanted until they were satisfied that they had performed them to the best of their ability. We did not want the experts to know each other's scores during this process, so only 1 expert at a time was allowed in the simulation laboratory. Each time an expert used the simulator for a given skill, the computer recorded multiple parameters including "economy of motion," "excessive instrument force," "instrument collisions," "instruments out of view," "master workspace range," "time to complete exercise," and "overall score." The best attempts by each surgeon for each of the 10 skills were compiled to determine "optimal expert performance" for each parameter of each skill. In doing so, some parameters were deemed "not applicable" for certain skills. For example, in the skill "Matchboard 3," none of the experts could optimally complete that task without having instruments out of view for at least a portion of the exercise. Therefore, "instruments out of view" score was deemed irrelevant to that skill. A full list of the 10 simulator skills and all "optimal expert performance" parameters making up the Morristown Protocol are listed in Table 1.

After establishing the Morristown Protocol, we recruited our study surgeons. Eligible candidates were board-certified OB-GYN attending physicians with full laparoscopic hysterectomy privileges at Morristown Medical Center but no prior robotic surgery training or experience. As study participants, they agreed to the following series of events:

They would complete the online orientation to the da Vinci Surgical System.

They would complete all 10 simulator skills at the "expert level" as per the Morristown Protocol (working at their own pace—spending as much time as necessary to pass each skill).

Within a week of completing the Morristown Protocol, they would complete the standard da Vinci pig laboratory training required of all new robotic surgeons.

Within a week of completing the pig laboratory, they would perform their first-ever robotic assisted surgery (a supracervical hysterectomy) as the main outcome measure for our study using the dual-console da Vinci system (supervised by one of the senior authors P.C. or C.S. on the other console).

We established a standardized hysterectomy technique for the purposes of the study so that each surgical step would be done in the same order throughout the protocol. Operative time was defined as the time from first grasp of the uterine fundus with the robotic single-tooth tenaculum to amputation of the uterus at the level of the internal cervical os. Blood loss was measured by carefully suctioning all visible blood from the field immediately after the hysterectomy. By design, study hysterectomies were performed on nonenlarged uteri, and all of these cases were videotaped to allow for review by masked observers. The outcome measures used to evaluate each surgery were operative time, blood loss, and "surgical skill" as measured by the global assessment tool for evaluation of intraoperative laparoscopic skills (GOALS) tool.[10] These masked assessments of surgical skill were performed by 2 urogynecology fellows (C.L. and E.G.). Because our main clinical question centered on the degree to which a group of novice robotic surgeons compared to experts after protocol completion, operative times, GOALS scores, and blood loss for the experts and novices were compared using independent and 1-sample $t$ tests.

To develop our sample size estimates, our robotic experts performed robotic supracervical hysterectomies via our standardized protocol. Our null hypothesis was that there would be no significant difference between expert and novice operative times. We defined a significant difference in operative times to be greater that 1 SD from mean expert time. The mean operative time for study hysterectomies by our expert group was 20.2 (2.31) minutes, so our sample size estimates called for 11 study surgeons to achieve our 90% power, assuming noninferiority, with $\alpha = 0.05$.

We tracked the actual time on the simulator required for each study surgeon to pass the Morristown Protocol, and we measured the number of attempts each of them made before

**TABLE 1.** Details of the Morristown Protocol (Scores Reflected as Percentile and Raw Score)

| Exercise | Overall Score | Drops Score (Raw No. Drops) | Economy of Motion Score (Raw Score) | Excessive Instrument Force Score (Raw Score) | Instrument Collisions Score (Raw Score) | Instruments Out of View Score (Raw Score) | Master Workspace Range Score (Raw Score) | Time to Complete Exercise Score (Actual Time in Seconds) |
|---|---|---|---|---|---|---|---|---|
| Peg board 2 | 91 | 100 (0) | 90 (231) | 100 (0) | 100 (0) | 100 | N/A | 98 (71) |
| Matchboard 2 | 93 | 100 (0) | 94 (257) | 100 (4.6) | 100 (0) | 100–(3.1) | N/A | 99–(82) |
| Suture sponge 2 | 88 | 100 (0) | 80 (255) | 100 (0) | 100 (3) | 100–(2.3) | 82 (6.5) | 95 (142) |
| Tubes | 92 | 100 (0) | 85 (344) | 0–0 | 95 (4) | 100–(0.5) | 90 (7) | 97 (157) |
| Ring walk 3 | 92 | | 95 (249) | 100 (0) | 100 (3) | 100–(3.3) | N/A | 98 (110) |
| Matchboard 3 | 95 | 100 (0) | 97 (477) | 100 (13) | 100 (3) | N/A | 100 (12) | 100 (158) |
| Camera targeting 2 | 93 | 100 (0) | 93 (225) | 100 (0) | 100 (3) | N/A | N/A | 97 (71) |
| Energy dissection 1 | 90 | N/A | 96 (100) | 100 (0) | 100 (1) | 100 (0) | 100 (7) | 99 (89.4) |
| Energy dissection 2 | 86 | N/A | 85 (133) | 100 (0) | 100 (0) | 100 (0) | N/A | 100 (94) |
| Energy switching 1 | 94 | N/A | 91 (155) | 100 (0) | 100 (0) | 100 (0) | 91 (11.4) | 100 (61) |

N/A indicates not applicable.

passing each of the 10 skills. On the basis of the number of attempts required for passing, we identified the most challenging of the 10 skills. We then plotted the number of attempts made before passing that skill by study surgeon and visually looked for a natural cutoff value that would allow us to dichotomize the group as "high performers" (who required the least amount of simulator hours to pass) versus low performers (who required significantly more simulator hours to pass). We performed this analysis in hopes of giving future simulator users a "quick and dirty" way to predict their future time requirements should they use our protocol.

Finally, we recruited a small convenience sample of gynecologic surgeons with robotic privileges at Morristown Medical Center. These 4 "control surgeons" (who had never been exposed to the da Vinci Skills Simulator) also performed supracervical hysterectomies for the purposes of the study. These surgeons had performed enough cases to be granted unsupervised robotic-gynecologic privileges to perform hysterectomies. We compared operative times, EBL, and GOALS scores between these control surgeons and the novice group using $t$ tests. Uterine specimen weights were compared between the 3 groups using analysis of variance.

## RESULTS

To allow for the possibility of attrition, 14 study surgeons were recruited. The demographics of each group of surgeons are listed in Table 2. All 14 study surgeons completed the entire Morristown Protocol, spending an average of 20 hours (range, 9.7–38.2 hours) on the simulator before doing so.

The mean hysterectomy operative times for the expert and study surgeon groups were 20.2 (2.31) and 21.7 (3.3) minutes, respectively [$P = 0.12$; 95% confidence interval (CI), −1.7 to 4.7]. The mean operative time for the control surgeon group was 30.9 (0.6) minutes, which was significantly longer compared to study surgeons' times ($P < 0.0001$; 95% CI, 6.3-12.3). The mean EBL for the expert and study surgeons was 25 and 25.4 mL, respectively ($P = 0.34$; 95% CI, −0.8 to 1.6). The mean EBL for control surgeons was significantly higher than for study surgeons at 31.25 mL ($P < 0.0001$; 95% CI, 0.4–11.4).

The mean expert and study surgeon GOALS scores were 50 and 34.7, respectively ($P < 0.0001$; 95% CI, 3.8–8.6). The mean GOALS score for our control surgeons was 31.1, which was almost significantly worse than our study surgeon group ($P = 0.07$; 95% CI, 0.4–4.2). The specimen weights of the uteri

for all 3 groups were quite similar. The mean weights for the uteri of the study surgeons, expert surgeons, and control surgeons were 46 (24), 51 (21), and 57 (16) g, respectively ($P = 0.85$). The most challenging skill was "Matchboard 3." Surgeons who passed this skill in 25 or less attempts (n = 4; mean, 16.67 attempts) were classified as high performers and were compared to the remaining low performers (n = 10; mean, 61.50 attempts). The mean overall times spent on the simulator to pass the whole protocol for the high and low performers were 11.6 and 23.3 hours, respectively ($P = 0.001$).

## DISCUSSION

This study established the predictive validity of the Morristown Protocol—a challenging robotic surgery simulation training curriculum. A group of robotic surgery novices learned to effectively use the da Vinci Surgical System via virtual reality simulation and translate that knowledge into actual skills in the operating room setting. The performance of our study surgeons (relative to our control surgeons) suggests that completion of this protocol may shorten the robotic surgery learning curve. Our study surgeons completed the simulator protocol at their own pace—with wide variation in hours required to do so. Nevertheless, their performance of the study hysterectomies was uniformly excellent. Also, performance on the skill "Matchboard 3" could potentially be used to predict the general amount of time a given surgeon would need to spend on the simulator to pass the Morristown Protocol. Further study will be required before this particular skill could actually be used as a predictor.

The primary strengths of our study were the establishment of true expert benchmarks for the virtual reality simulator and the incorporation of a real-life surgical outcome to establish predictive validity. Robotic assisted supracervical hysterectomy proved to be a useful outcome measure, because we were able to standardize the steps of this moderately complex surgical procedure. Our results were particularly compelling when one considers that our outcome measures were collected during our study surgeons' first-ever robotic surgical experience on a live human. Other strengths of our study were the incorporation of a control surgeon group and our use of the GOALS scores[10] of surgical skill.

Our study limitations included the relatively small number of experts and control surgeons as well as our complete focus on technical robotic skills as opposed to inclusion of cognitive laparoscopic training. Our requirement that study surgeons be board certified in OB-GYN and credentialed to perform laparoscopic hysterectomy limited the generalizability of our findings. Further study would be required to determine the usefulness of our protocol for training residents and fellows.

Nevertheless, hospital credentialing committees could incorporate the Morristown Protocol as a required step in the initial training process or for recredentialing of low-volume robotic surgeons. Many surgeons complete their initial robotic training, obtain full robotic privileges, but subsequently perform very few robotic assisted cases. These situations can be challenging for hospital credentials committees that may be reluctant to revoke surgeons' privileges simply based on low numbers of cases. Often these low-volume robotic surgeons face difficulty when trying to schedule robotic cases, because block robotic surgery time has already been given to surgeons with higher volumes. Credentials committees could require either a certain number of cases per year or completion of our protocol before renewing robotic privileges. Ideally, the da Vinci Skills Simulator could be accessed by staff surgeons any time day or night—thus, enabling any surgeon willing to put the time in an avenue for recertification.

**TABLE 2.** Demographics of Surgeon Groups

| | Study Surgeons (n = 14) | Control Surgeons (n = 4) | Significance, *P* |
|---|---|---|---|
| Male, n | 3 | 3 | |
| Female, n | 11 | 1 | |
| Age, y | 49.1 (8.5) | 53.5 (11.1) | 0.40 |
| Time since med school, y | 23.8 (8.7) | 27 (12.3) | 0.56 |
| Time since OB-GYN residency, y | 19.5 (8.7) | 22.8 (11.9) | 0.55 |
| Time since ABOG certification, y | 17 (8.4) | 19.3 (12.6) | 0.68 |
| Duration of advanced laparoscopic privileges, y | 11.3 (9.3) | 12.8 (5.2) | 0.77 |

ABOG indicates American Board of Obstetrics and Gynecology.

Our protocol could also be used by any surgical residency program as a required hurdle for any residents wishing to learn robotic techniques. No surgical residency training programs require residents to learn robotic techniques, so receiving this training could be thought of as a privilege conferred only to residents interested enough to complete the simulator protocol on their own time.

Another possible use of our simulator protocol would be to make its completion a prerequisite for pig laboratory training. Pig laboratory training (offered by Intuitive Surgical) is a requirement for new robotic surgeons, but due to an ever-growing number of surgeons seeking to learn robotics, there is a waiting list for the training. Perhaps those willing to complete our simulator protocol could be given higher priority to receive pig laboratory training.

In conclusion, we established a challenging simulator training protocol using the da Vinci Skills Simulator—the completion of which seemed to shorten the learning curve for novice robotic surgeons. Our use of a complex operation (actually performed on a live human) as our end point of interest was particularly unique and established the predictive validity of our training curriculum. Further studies using the Morristown Protocol as a component of resident or attending surgeon training in robotics are warranted.

## ACKNOWLEDGMENTS

## REFERENCES

1. Wright JD, Ananth CV, Lewin SN, et al. Robotically assisted vs. laparoscopic hysterectomy among women with benign gynecologic disease. *JAMA* 2013;309(7):689–698.

2. Wright JD, Burke WM, Wilde ET, et al. Comparative effectiveness of robotic versus laparoscopic hysterectomy for endometrial cancer. *J Clin Oncol* 2012;30(8):783–791.

3. Barbash GI, Glied SA. New technology and health care costs: the case of robotic assisted surgery. *N Engl J Med* 2010;363(8):701–704.

4. Griffen FD, Sugar JG. The future of robotics: a dilemma for general surgeons. *Bull Am Coll Surg* 2013;98(7):9–15.

5. Steinberg PL, Merguerian PA, Bihrle W III, et al. The cost of learning robotic-assisted prostatectomy. *Urology* 2008;72:1068–1072.

6. Raman JD, Dong S, Levinson A, et al. Robotic radical prostatectomy: operative techniques, outcomes and learning curve. *JSLS* 2007;11:1.

7. Hung AJ, Zehnder P, Patil MB, et al. Face, content and construct validity of a novel robotic surgery simulator. *J Urol* 2011;186:1019–1025.

8. Kelly DC, Margules AC, Kundavaram CR, et al. Face, content, and construct validation of the da Vinci Skills Simulator. *Urology* 2012;79(5):1068–1072.

9. Finnegan KT, Meraney AM, Staff I, et al. da Vinci Skills Simulator construct validation study: correlation of prior robotic experience with overall score and time score performance. *Urology* 2012;80(2):330–335.

10. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 2005;190:107–113.