Association for Surgical Education

Fundamentals of Laparoscopic Surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial

Gideon Sroka, M.D., Liane S. Feldman, M.D., Melina C. Vassiliou, M.D., Pepa A. Kaneva, M.Sc., Raad Faye, M.D., Gerald M. Fried, M.D.*

Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, McGill University, Montreal, QC, Canada

Abstract

BACKGROUND: The purpose of this study was to assess whether training to proficiency with the Fundamentals of Laparoscopic Surgery (FLS) simulator would result in improved performance in the operating room (OR).

METHODS: Nineteen junior residents underwent baseline FLS testing and were assessed in the OR using a validated global rating scale (GOALS) during elective laparoscopic cholecystectomy. Those with GOALS scores ≤15 were randomly assigned to training (n = 9) or control (n = 8) groups. An FLS proficiency-based curriculum was used in the training group. Scoring on FLS and in the OR was repeated after the study period. Evaluators were blinded to randomization status.

RESULTS: Sixteen residents completed the study. There were no differences in baseline simulator (49.1 ± 17 vs 39.5 ± 16, P = .27) or OR scores (11.3 ± 2.0 vs 12.0 ± 1.8; P = .47). After training, simulator scores were higher in the trained group (95.1 ± 4 vs 60.5 ± 23, P = .004). OR performance improved in the control group by 1.8 to 13.8 ± 2.2 (P = .04), whereas the trained group improved by 6.1 to 17.4 ± 1.9 (P = .0005 vs control; P < .0001 vs baseline).

CONCLUSIONS: This study clearly demonstrates the educational value of FLS simulator training in surgical residency curricula.

© 2010 Elsevier Inc. All rights reserved.

KEYWORDS:
Fundamentals of Laparoscopic Surgery (FLS); Proficiency-based training; Operating room performance; Laparoscopic skill assessment; Simulation

In the past 2 decades since their introduction, video-laparoscopic techniques have been widely adopted by surgeons from various disciplines, including general and thoracic surgery, urology, and gynecology. Residency programs are facing the educational challenge of teaching unique skills required to safely perform laparoscopic surgery. In addition, the renewed emphasis on patient safety, ethical issues inherent in “practicing on patients,” and pressures to improve operating room (OR) efficiency have all become increasingly relevant. The traditional “see one, do one, teach one” approach is clearly not appropriate in this new environment.

Surgical training has a long tradition of incorporating simulation in the guise of cadaver, animal, and bench-top...
models to develop and practice open procedural skills. The opportunity to acquire the novel skills required for laparoscopic surgery in an environment that is efficient, effective, and does not jeopardize patient safety is also very appealing. The MISTELS program (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills), based on a physical laparoscopic trainer box, was developed with the aim of standardizing the teaching and evaluation of fundamental laparoscopic skills in a safe environment.1 MISTELS is an inexpensive, portable, and flexible system, and has been extensively validated.2,3 It has been shown that performance in the simulator improves progressively with practice4 and correlates with OR performance.5 MISTELS was incorporated as the manual skills component of the Fundamentals of Laparoscopic Surgery (FLS) program developed by the Society of American Gastrointestinal Endoscopic Surgeons (SAGES) and endorsed by the American College of Surgeons (ACS).

While the FLS simulator has become a de facto standard for evaluation of technical skill in laparoscopy,6 the question remains whether or not skills gained from training on the FLS simulator translate into better performance in the OR. Although transfer of learning from this simulator and other simulators to the OR has been evaluated,7–13 the studies have been hampered by lack of a standardized simulator and curriculum, inconsistent proctoring during training, and especially by the absence of a validated measure of laparoscopic OR performance.14 The Global Operative Assessment of Laparoscopic Skills (GOALS) was developed to evaluate intraoperative laparoscopic skills by direct observation. It was initially validated during dissection of the gallbladder from the liver bed, and has been shown to be a reliable and valid measure of technical skill in the clinical setting.15 The aim of this study was to assess the transfer of skills acquired by novices trained to proficiency on the FLS simulator to OR performance as measured by GOALS.

Methods

General surgery residents in postgraduate years (PGYs) 1 to 3 at McGill University with no previous FLS training were eligible for this Research Ethics Board-approved study (Project A03-E06-04A). After appropriate informed consent, 19 participants underwent baseline FLS testing, and were assessed in the OR using GOALS during dissection of the gallbladder from the liver bed during elective laparoscopic cholecystectomy. Those with GOALS scores ≤15 (n = 17) were randomly assigned to training (n = 9) or nontraining (n = 8) groups. The training group used the FLS simulator in a supervised proficiency-based curriculum, based on recommendations by Ritter and Scott.16 Both groups continued their regular residency training and subjects were asked to document their clinical laparoscopic experience throughout the study period. At the end of the study period (mean time between evaluations was 145 days) subjects were assessed again on the simulator and in the OR, using the same metrics. Evaluators of simulated and clinical performance were blinded to training group and to one another. Figure 1 summarizes the flow of participants through the study.

Ethical issues

Participants were assured that their evaluation would not change the way the procedure was done or the quality of the care patients received. Evaluations were restricted to dissection of the gallbladder from the liver bed and the attending surgeon could take over at any time. The data gathered were coded and all reporting was confidential and did not impact the resident’s official evaluation. Participants could choose to withdraw at any point during the study and they were made explicitly aware of this at the time of informed consent. To mitigate the potential ethical dilemma associated with restricting training opportunities thought to be beneficial from 1 group of residents, FLS training was offered to all those randomized to the nontraining group at the end of the study period.

Randomization

Randomization was done by drawing an assignment from a box with a 50% chance of being in the training or nontraining group, and was performed by an investigator not involved in the process of training and evaluations. Participants were asked to keep their randomization status confidential for the study period.

Simulator performance metrics

Laparoscopic proficiency training was performed using the manual skills portion of the FLS program,17 which includes a
CD-ROM of didactic material and the 5 MISTELS tasks that have been previously described in detail.\textsuperscript{3,4,18} Briefly, the simulator consists of a trainer box with an opaque cover, a built-in camera, and 2 trocars. It can be attached to any monitor with an s-video connection. The 5 tasks include peg transfer, circle cut, placement of a ligating loop, and simple suture tied with extra- and intracorporeal techniques.\textsuperscript{19} All participants went through an orientation, and viewed the video tutorial for all 5 tasks. Baseline scores were calculated after the first iteration of each task by an experienced FLS proctor using the standard FLS metrics. An overall summative evaluation (pass or fail) was provided after the completion of all tasks. The metrics reward efficiency (speed) but also penalize errors; higher scores indicate better performance. Final evaluations were performed in the exact same way as at baseline.

**Intraoperative laparoscopic performance assessment**

Intraoperative laparoscopic performance was evaluated with the GOALS. This global rating scale measures performance in 5 domains, 3 of which are specific for laparoscopic surgery (depth perception, bimanual dexterity and tissue handling), and 2 that are more generic (efficiency and autonomy). Each domain is scored on a Likert scale from 1 to 5, with anchoring descriptions at 1, 3, and 5. The total score ranges from 5 to 25, with higher scores indicating better performance. This tool has been shown to be reliable, valid, and feasible for evaluation of dissection of the gallbladder from the liver bed. We previously reported that junior residents (PGY 1–2) achieved a score of 12 (95% confidence interval [CI] 11–13), compared with 17 (95% CI 14–21) for intermediate level residents (PGY 3–4) and 22 (95% CI 20–24) for more experienced surgeons (PGY 5+).\textsuperscript{5} All participants were evaluated in the OR at baseline and at the end of the study by the attending surgeon and/or an evaluator blinded to their randomization status. The evaluators had previous experience in using GOALS through participation in the original validation studies. To maintain blinding, the importance of concealing the randomization status was repeatedly emphasized to the participants. The attending surgeon also evaluated the perceived difficulty of the dissection using a visual analog scale (VAS) from 0 (easiest) to 10 (most difficult).

**Setting**

The study took place within an accredited general surgery residency program. FLS training took place in the Steinberg-Bernstein Centre skills laboratory at the Montreal General Hospital. In addition to proctored sessions, the training group was given a keycard enabling unscheduled access to the skills laboratory at any time. OR assessments took place at 2 McGill University teaching hospitals.

**Training curriculum**

The training goal in the simulator was based on a study by Ritter and Scott.\textsuperscript{16} They used their own expert performance to set a specific time goal with “allowable” errors for each task. Proficiency was defined as performing each task within the specified time and error goals on at least 2 consecutive repetitions (with 10 additional nonconsecutive repetitions for tasks 1 and 5). Scott et al have also shown that this proficiency based curriculum is feasible for training novices, and allows sufficient skill acquisition for FLS certification in a mean of 9.7 hours of training.\textsuperscript{20} The criteria for proficiency testing was requested by the subject and confirmed by the proctor for the training group, and after at least 6 weeks for the nontraining group.

**Statistical analysis**

The primary outcome was the change in OR performance measured by GOALS. We determined the required power of the study using previously collected data on GOALS scores in 76 surgical residents. This demonstrated a mean score of 12.3 ± 2.8 for PGY 1–2 and 17.5 ± 3.6 for PGY 3–4. We defined 5 as a clinically meaningful difference in score. We calculated that 7 subjects in each group would give a power of 80% to detect a difference of 5 points in GOALS scores with an alpha of .05. Statistical analysis was conducted using SPSS version 11.0 (SPSS, Inc, Chicago, IL). Student $t$ test was used to compare mean FLS and GOALS scores in the training and nontraining groups. Paired $t$ tests were used to compare baseline and the final performance within each group. Multivariate analysis was used to assess the effect of simulator training on GOALS score after adjusting for baseline GOALS score. Data are expressed as mean ± SD. Demographic data (Table 1) are expressed as medians (interquartile range [IQR]). $P < .05$ was considered statistically significant.
Results

Of the 17 subjects who were randomized, 16 completed their final evaluations, 8 in each group. The groups were statistically similar at randomization (Table 1). There were more women in the control group (63% vs 25%) while the trained group had a longer time between OR assessments. Baseline FLS scores were similar in the trained and non-trained groups (49.1 ± 17 vs 39.5 ± 16, P = .27). At the final evaluation, FLS scores increased and the standard variations decreased in the trained group to 95.1 ± 4, compared with 60.5 ± 23 in the nontrained group (P = .004). At baseline, no participant had a score above the level required for FLS certification; at the second evaluation, all trained subjects would pass FLS, whereas only 3 of the 8 nontrained subjects had a passing score. All participants in the trained group met the predefined proficiency criteria after mean training time of 450 minutes (in 9 separate sessions), out of which 150 minutes were proctored, and the rest unsupervised.

Figure 2 summarizes the change in OR performance as measured by GOALS score, for the 2 groups. Baseline OR performance was similar (P = .47). Participants in the nontrained group improved their performance by 1.8 ± 2.1 points from 12.0 ± 1.8 to 13.8 ± 2.2 (P = .04), a clinically insignificant magnitude. In contrast, the trained group improved significantly by a mean of 6.1 ± 1.3 points from 11.3 ± 2.0 to 17.4 ± 1.9 (P = .0005 vs control; P < .0001 vs baseline). After adjusting for gender using multivariate analysis, group allocation remained significantly associated with the change in GOALS score (P = .001), with gender not significant (P = .16). There was no correlation of change in GOALS score with time between assessments (Pearson correlation = .13, P = .63).

Of the 5 individual domains evaluated by GOALS, simulator training was associated with greater improvements in the laparoscopy-specific domains (bimanual dexterity, tissue handling, depth perception) compared with the more generic domains (efficiency and autonomy) (Table 2). There was no difference in the attending surgeon’s assessment of the difficulty of the dissection for the trained and control groups at the baseline (2.5 vs 3, P = .15) or final evaluations (4.5 vs 2.5, P = .15).

Table 1 Comparison of simulator trained and control groups at enrollment and during the study period

<table>
<thead>
<tr>
<th></th>
<th>No simulator training (n = 8)</th>
<th>Simulator training (n = 8)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY 1/2/3</td>
<td>6/2/0</td>
<td>5/2/1</td>
<td>.58</td>
</tr>
<tr>
<td>Age (y)</td>
<td>27 (27–28)</td>
<td>27 (26.5–28.5)</td>
<td>.85</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>3/5</td>
<td>6/2</td>
<td>.13</td>
</tr>
<tr>
<td>Hand dominance (right/left)</td>
<td>7/1</td>
<td>7/1</td>
<td>1</td>
</tr>
<tr>
<td>Time between baseline and final evaluations (d)</td>
<td>113 (40–167)</td>
<td>162 (100–256)</td>
<td>.13</td>
</tr>
<tr>
<td>LC performed as primary during study (no.)</td>
<td>3.5 (2–5)</td>
<td>4.5 (3–7)</td>
<td>.21</td>
</tr>
<tr>
<td>LC participated as assistant during study (no.)</td>
<td>4.5 (4–6)</td>
<td>4.5 (3.5–8)</td>
<td>.92</td>
</tr>
<tr>
<td>Other laparoscopic cases performed/participated during study (no.)</td>
<td>2.5 (2–3.5)</td>
<td>2.5 (1–3.5)</td>
<td>.75</td>
</tr>
</tbody>
</table>

Data expressed as median (IQR). LC = laparoscopic cholecystectomy.

Table 2 Comparison of the difference (mean ± SD) in operating room performance in the domains assessed by GOALS from baseline to final assessment after simulator training compared with controls

<table>
<thead>
<tr>
<th></th>
<th>No simulator training (n = 8)</th>
<th>Simulator training (n = 8)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth perception</td>
<td>.5 ± .8</td>
<td>1.25 ± .7</td>
<td>.08</td>
</tr>
<tr>
<td>Bimanual dexterity</td>
<td>.5 ± 1.1</td>
<td>1.25 ± .6</td>
<td>.04</td>
</tr>
<tr>
<td>Efficiency</td>
<td>.4 ± 1.1</td>
<td>1.13 ± 1.0</td>
<td>.24</td>
</tr>
<tr>
<td>Tissue handling</td>
<td>.3 ± .7</td>
<td>1.13 ± 1.0</td>
<td>.04</td>
</tr>
<tr>
<td>Autonomy</td>
<td>.3 ± 1.0</td>
<td>.6 ± 1.1</td>
<td>.58</td>
</tr>
<tr>
<td>Total score</td>
<td>1.8 ± 2.1</td>
<td>6.1 ± 1.3</td>
<td>.0003</td>
</tr>
</tbody>
</table>

Each domain is scored from 1 (worst) to 5 (best) and the results summed to get a total score.
Comments

Technical skills are essential to the practice of surgery. There is great interest in the potential for teaching fundamental skills to novice surgeons in the simulator laboratory instead of the OR, both to address patient safety concerns and improve OR efficiency. Repetitive, goal-directed practice of psychomotor skills before the OR may allow some of these skills to become automated. Automaticity implies that when the trainee is then in the clinical environment, his or her attention can shift from the required manual skill to focus on the cognitive aspects of the tasks he or she is facing in the OR, namely, perception (recognition of the anatomy/pathology) and forecasting (looking ahead to the next step). However, incorporation of simulation into surgical training curricula also requires significant effort, time and money, and will only be fully established when its educational effectiveness is proven. This single-center randomized single-blinded trial demonstrates that training to proficiency using the FLS simulator curriculum improved OR performance during laparoscopic dissection of the gallbladder from the liver bed in novice surgeons compared with standard residency training. Training to proficiency required an average 2.5 hours of supervised training and 5 hours of unsupervised practice on the simulator. The magnitude of the difference in the improvement in OR performance between 2 groups appears clinically relevant, as the scores for the trained junior residents (PGY 1–2) was similar to mean deviation in scores from the baseline to final FLS assessment of the untrained junior residents (PGY 3–4).

Our results are consistent with other randomized studies demonstrating the transferability of skills learned in a variety of laparoscopic simulators to performance of laparoscopic cholecystectomy or its component steps in humans. In the only previous study using a physical box trainer, Scott et al. randomized 27 residents to a training group, who had ten 30-minute training sessions on the Southwestern Guided Endoscopic Module, or to a control group. Performance of laparoscopic cholecystectomy was measured using global rating scale domains modified from the Objective Structured Assessment of Technical Skill (OSATS), originally designed and validated for open bench model performance. The trained group showed greater improvements from baseline in 4 of the 8 domains. Two other studies confirmed transferability from a virtual reality psychomotor simulator (Minimally Invasive Surgical Trainer—Virtual Reality [MIST-VR], Mentice Inc, San Diego, CA) to the OR. Seymour et al. randomized 16 residents to MIST-VR training until a preset criterion level was reached, then assessed dissection of the gallbladder from the liver bed using a novel error scoring system that had not been previously validated and not performed at baseline. Fewer errors were made by the trained group. Grantcharov et al. randomized 16 residents to 10 repetitions of the MIST-VR tasks. Subjects were assessed during clipping of the cystic duct and artery and dissection of the gallbladder from the liver bed during laparoscopic cholecystectomy using an error score and economy of motion score, also modified from the OSATS. The simulator-trained group showed greater improvement from baseline and was slightly faster compared with controls. A systematic review of skills transfer after surgical simulation training concluded that for laparoscopic cholecystectomy, subjects who received simulator training before OR assessments performed better than subjects without such training. The authors emphasized the need for more standardized patient-based performance measures: there were large variations in the part of laparoscopic cholecystectomy that were evaluated, as well as in the different metrics that were used for the assessment. Several simulators were used, with diverse platforms, tasks, performance metrics and practice goals (time-based or criterion-based). None of the previous studies used a patient-based assessment that had been previously validated specifically for laparoscopic surgery.

The present study addressed several of these concerns. The FLS simulator was designed to build the fundamental skills required for laparoscopic surgery in novice surgeons. It is an inexpensive skills based simulator that does not model any 1 specific procedure, and as such has been used in general, urological and gynecologic surgery curricula. The FLS curriculum is standardized and readily available at relatively low cost. The performance metrics have been extensively validated with a passing score defined to differentiate competent from noncompetent laparoscopic surgeons. FLS certification, which requires passing a cognitive component as well as the manual skills test, will be required for general surgery board certification beginning in 2010. In the present study, baseline average score (44) was at the level expected for junior residents without simulator training. The criteria used to define proficiency were also evidence-based. The advantage of a proficiency rather than a time- or number-based goal relates to the observation that people learn at different rates. Using a proficiency goal ensures that this variability decreases with training, as evidenced by the narrowing of the standard deviation in scores from the baseline to final FLS assessment in the trained group. The time to achieve this proficiency level ranged from 5 hours to 12 hours.

We used a patient-based global rating assessment tool (GOALS) that is reliable and valid when used to evaluate laparoscopic dissection of the gallbladder from the liver bed, and is strongly correlated with FLS score. Others have provided evidence for construct validity for the GOALS when used to evaluate an entire laparoscopic cholecystectomy, as well as laparoscopic appendectomy. We chose dissection of the gallbladder alone to enable participation of junior-level residents, for whom simulation training is expected to provide the greatest benefit. The availability of GOALS data for specific levels of residency training allowed a sample size calculation to be made based on a clinically relevant difference, since an improvement by at least 5 points would represent the difference between novice and intermediate level residents in regular clinical
training. Requiring a GOALS score below 16 avoided the ceiling effect seen for this relatively straightforward OR procedure. Looking separately at the 5 domains included in the GOALS score, one can appreciate that FLS simulator training targeted the laparoscopic-specific domains of depth perception, bimanual dexterity and tissue handling. Our intervention had little to do with the cognitive part of the procedure, and it was not unexpected that the autonomy and efficiency domains were not affected by simulator training. An additional strength of the study was that the OR assessments were blinded to the training status of the participants.

The study has several limitations. This is a single-center study, in the center where the simulator and all the performance metrics were developed. The sample size was small, although this was supported by the sample size calculations. There was a longer time between the assessments in the simulator-trained group compared with controls, time during which the participants continued their usual training. The time between assessments did not correlate with change in OR performance. The time difference was related to the logistical issues inherent in asking the subjects to practice on the simulator in addition to their usual duties, with the simulator laboratory often located at a different hospital. Despite the discrepancies in time, the subjects reported similar participation in laparoscopic cases during the study period. This may be explained by the fact that these junior residents were not necessarily rotating through general surgery services during the study period. In addition, the improvement in OR performance seen in the trained group is greater than expected from usual training alone, as it was similar in magnitude to that reported between a junior and intermediate level resident. Gender composition differed between the groups, with more women randomized to the control group. However, in multivariate analysis, gender did not prove to be a significant predictor of change in OR performance. An additional limitation may be the reliance on a single OR performance metric (GOALS). While other measures of clinical performance (eg, time, errors) may also have provided useful data, we believe that the performances that were inefficient or error-prone were reflected in the lower GOALS scores.

In conclusion, we found a statistically and clinically significant improvement in OR laparoscopic performance in junior residents after proficiency based simulation training compared with untrained controls. This was achieved using an inexpensive widely available video trainer after 7.5 hours of simulator training.

References