The American Journal of Surgery 193 (2007) 797–804

Laparoscopy

Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies

Gunnar Ahlberg, M.D., Ph.D.[a,b,*], Lars Enochsson, M.D., Ph.D.[c,d], Anthony G. Gallagher, Ph.D.[f], Leif Hedman, Ph.D.[d], Christian Hogman, M.Sc.[b], David A. McClusky III, M.D.[e], Stig Ramel, M.D., Ph.D.[a], C. Daniel Smith, M.D., F.A.C.S.[g], Dag Arvidsson, M.D., Ph.D.[c]

[a]*Department of Surgery, Karolinska University Hospital, SE-171 76, Sweden*
[b]*Center for Advanced Medical Simulation, Karolinska University Hospital, SE-171 76, Sweden*
[c]*Department of Surgery, Karolinska University Hospital, SE-141 52, Huddinge, Sweden*
[d]*Center for Advanced Medical Simulation, Karolinska University Hospital, SE-141 52, Huddinge, Sweden*
[e]*Department of Surgery, Emory University, Atlanta, GA 30322, USA*
[f]*Department of Psychology, Queens University, University Road, Belfast BT7 1NN, Northern Ireland, United Kingdom*
[g]*Emory Endosurgery Unit, Emory University School of Medicine, Atlanta, GA 30322, USA*

Manuscript received November 12, 2005; revised manuscript June 20, 2006

Abstract

Background: Virtual reality (VR) training has been shown previously to improve intraoperative performance during part of a laparoscopic cholecystectomy. The aim of this study was to assess the effect of proficiency-based VR training on the outcome of the first 10 entire cholecystectomies performed by novices.

Methods: Thirteen laparoscopically inexperienced residents were randomized to either (1) VR training until a predefined expert level of performance was reached, or (2) the control group. Videotapes of each resident's first 10 procedures were reviewed independently in a blinded fashion and scored for predefined errors.

Results: The VR-trained group consistently made significantly fewer errors (P = .0037). On the other hand, residents in the control group made, on average, 3 times as many errors and used 58% longer surgical time.

Conclusions: The results of this study show that training on the VR simulator to a level of proficiency significantly improves intraoperative performance during a resident's first 10 laparoscopic cholecystectomies.

Keywords: Virtual reality; Proficiency-based; Training; Simulation; Laparoscopic

There is no doubt that the laparoscopic revolution in the late 1980s was disruptive technology in the sense that the paradigm shifted toward this new method in the surgical field and the impact was immediately felt on a global level. After the initial descriptions of laparoscopic cholecystectomy, first made by Mühe in 1985 and later by Mouret in 1987, developments in the surgical field were ignited [1]. The laparoscopic approach has gained enormous importance and currently is the gold standard for several abdominal procedures, including cholecystectomy, and few abdominal surgical procedures have escaped the laparoscopic approach [1].

The operating room has been the classroom for surgical training throughout the history of surgery. Laparoscopic surgery, however, imposes specific difficulties such as loss of 3-dimensional visualization, loss of tactile feedback, and counterintuitive instrument movement. Automation in navigating laparoscopic instruments and overcoming these difficulties probably are best achieved outside the operating room (OR) and, therefore, the conventional apprenticeship model does not fit well with the development of such skills. Furthermore, a clear learning curve in laparoscopic surgery has been described as the period before acceptable technical proficiency is reached, during which time the risk for com-

* Corresponding author. Tel.: +468-517-796-26; fax: +468-33-15-87.
E-mail address: gunnar.ahlberg@karolinska.se

0002-9610/07/$ – see front matter
doi:10.1016/j.amjsurg.2006.06.050

plications clearly is increased [2–4]. The educational vaccination against unnecessary iatrogenic illness is, according to the Kohn et al [5] publication "To Err is Human," structured training curricula with validated certification models.

Surgical care is one of the most complex systems in medicine and increasing complexity implies an increase in the chance for error. Medical errors have caused substantial ethical and economic issues. It also has been stated that surgical care accounts for about half of the adverse events when errors are reviewed [5,6]. Technical proficiency is an essential ingredient in surgical competence. Despite its importance to surgeons, technical proficiency historically has been poorly evaluated. The introduction of disruptive technology with respect to surgical procedures such as laparoscopy requires an equally disruptive paradigm shift in surgical training that would match the development in OR technology. Currently, at no point throughout the surgical training programs has there been a compulsory or objective assessment of technical skills. The present evaluation of trainee's skills can be regarded only as subjective, often unstructured, and possibly biased. Although well established, there is an obvious lack of objectivity and the term *good hands* is no longer acceptable in grading technical skills [7–12].

Simulation technology has gained considerable interest during recent years and one clear benefit of this technology, compared with other training systems such as box trainers and animal models, is the ability to obtain objective measurements of technical skills, which makes the creation of certification levels more easily achievable. Several studies have shown that simulators can be used as assessment tools for laparoscopic skills [13], but surprisingly few studies have been able to empirically show skills transfer from virtual reality (VR) to the OR environment [14,15]. In both of these cited studies only a part of the laparoscopic cholecystectomy (LC) was assessed (ie, dissection of the gallbladder from the liver-bed portion) and trainees only completed 1 LC posttraining.

The concept of proficiency-based training has been introduced recently [14,16]. It is based on real-world skills because it was established on the basis of experienced laparoscopic surgeons' performance on a simulator. The strength of this approach includes the guarantee of a minimum and a more homogeneous skill set across trainees.

The LapSim (Surgical Science Inc., Göteborg, Sweden) system was described as being able to transfer the basic skills acquired in the simulator to a porcine model using surgical novices as trainees [17]. Since then the system has been developed further and currently comprises a Calot's triangle dissection module, adding procedural skills training to the system.

The aim of this study was to investigate whether proficiency-based training on the LapSim VR system improves objectively assessed, intraoperative performance during the initial learning of LC, and if that improvement is persistent over time.

Materials and Methods

Subjects

Thirteen surgical residents (6 men, 7 women) from 9 different institutions in Sweden participated in this study. The introduction to laparoscopy training in the Swedish education system varies and the subjects therefore differed with respect to their experience from postgraduate year 1 to 2. They all had experience in assisting with laparoscopic procedures but had no previous experience in performing an LC. All residents were scheduled to start with laparoscopic training and each supervisor approved participation in the study. All subjects were matched with respect to sex and randomly assigned, with a sealed-envelope method, to either the simulator-training group or to the control group.

Apparatus

The LapSim is a VR simulator for laparoscopic skills training. The system consists of a software program run on a dual Xeon 1.8-GHz processor (Intel Corporation, Santa Clara, CA) using Windows XP (Microsoft Corporation, Redmond, WA). The computer is equipped with 256 MB of internal RAM, a NVIDIA Quadro2 EX graphics card (NIVIDIA Corporation, Santa Clara, CA), a 15-inch monitor, and a virtual laparoscopic interface manufactured by Immersion Inc. (San Jose, CA), including 2 laparoscopic instruments. Force feedback is optional. In this study we used the 2.0 version of the system including LapSim basic skills and LapSim dissection programs, without force feedback. The basic skills program offers a set of 9 training modules, ranging from camera navigation to more advanced skills such as coordination, clip application, lifting and grasping, cutting, and suturing. All modules can be configured individually with respect to difficulty and preferences. Metrics parameters such as time, errors, tissue tension and damage, patterns of movement, and bleeding are all measured for each attempt. In the assessment mode, requirements for every task can be defined. In the dissection program, different anatomic variations of the hepatoduodenal ligament are available and a variety of instruments can be used for dissection training. The system has a high degree of realism in terms of its graphics and tissue-instrument interaction.

Psychometric and simulator tests

Before randomization, all subjects were tested in a series of psychometric tests and in the simulator to control for confounding variables in the design and to ensure that no extreme individual differences existed. The Vandenburg and Kuse mental rotation test was used as a high-level visuospatial function test [18]. The forward and backward number series test, and the number and letter combination series test from Wechsler Adult Intelligence Scale, third edition (WAIS III) (The Psychological Corporation, San Antonio, TX) [19] were used to assess verbal working memory. Further, a series of 5 tasks in the simulator were used for baseline skills assessment: grasping, lift grasp, cutting right, cutting left, and clip application. All subjects were tested without prior practice. An experienced laparoscopic surgeon demonstrated each task before each subject performed the same task individually. No task-specific requirements were used in the pretest procedure. In addition, each subject filled in a questionnaire to control for variables associated with flow (ie, their attitude toward the simulator).

Table 1
Comparison between study subjects (pretrial test) and laparoscopic experts in the 5 test tasks in the simulator

Metrics for lift and grasp	Students					Experts					
	Median	Mean	Max	Min	Variance	Median	Mean	Max	Min	Variance	*P* level
Total time, s	110	116	182	62	1262	40	38	54	25	129	.001357
Left instrument path length, m	1.13	1.36	1.68	1.14	.033	1.00	.90	1.13	.63	.061	.001357
Left instrument angular path, °	318	331	470	253	3425	225	221	286	175	2115	.001357
Right instrument path length, m	1.23	1.19	1.43	.96	.029	.97	.89	1.09	.69	.032	.001903
Right instrument angular path, °	291	295	366	239	1886	201	219	270	180	1317	.001357
Tissue damage, #	2.0	2.9	9.0	.0	5.91	1.0	1.2	4.0	.0	2.70	.076127
Max damage, mm	4.7	5.4	15.9	.0	23.30	1.9	2.1	6.3	.0	6.69	.346546

Examples from the lift and grasp task are shown.
Max = maximum; Min = minimum.

Construct validity

To be able to show that the metrics used in the simulator reflect the user's actual level of technical skills and thereby prove construct validity, 5 experts in laparoscopy were tested, all with extensive experience in advanced laparoscopy (>300 LCs, fundoplications, and spleenectomies). The experts were tested in the same series of 5 tasks used for baseline assessment, with identical settings and instructions used for the study subjects. A significant difference between experts and novices was shown (Table 1).

Simulator training

An expert level of proficiency was defined on the simulator by assessing the 5 laparoscopic surgeons in 6 different tasks: suturing with and without easy grip function, lift and grasp, clipping, and ultrasonographic dissection with both hands. Thus, we were able to define the expert level of proficiency on the simulator by calculating the median values for every parameter in each task from all experts. This proficiency level created requirements for the 6 tasks in examination mode. For the initial training we used these 6 tasks at 3 different levels of difficulty. Thus, in total, 18 tasks were used in the basic skills program without requirements for passing. The subjects assigned to simulator training then practiced under supervision and received feedback given by the simulator as well as oral feedback given by the supervisor after each completed task until they showed proficiency on each of the 6 examination tasks at least twice. The training was performed under supervision of a laparoscopic expert in a distributed fashion consisting of 1 hour training sessions followed by 15 minutes of recovery with a maximum of 8 sessions per day during 1 week, with as many 1-hour training sessions as each trainee individually needed to reach the proficiency level.

Trainees also were able to use the dissection module for familiarization with the anatomy in the triangle of Calot but because of the early version of the dissection program available at the time, this training was performed without performance requirements. All simulator training was performed under the supervision of an experienced laparoscopic surgeon. No further training in the simulator was allowed after reaching the proficiency level.

Surgical procedures

Patients with a history of uncomplicated gallstone disease who were admitted for LC were included in the study. All subjects performed their first 10 individual full procedure LCs under the supervision of an experienced laparoscopic surgeon at each respective center. The first procedure should be performed within 2 weeks either after study commencement (control group), or after reaching the proficiency level in the simulator (training group), and all procedures should be performed within 6 months. The supervisors were blinded with respect to the subjects' training status. The supervisors were instructed to take over the procedure at any time if the dissection was performed in an unsafe way. Each procedure was videotaped and a protocol was filled out stating, on a minute-by-minute basis, which parts of the procedure were performed by the apprentice and at which point the supervisor took over one or both of the instruments. All subjects and supervisors viewed an instructional video of an LC performed in a standardized, 2-handed, dissection fashion and all defined errors in the assessment form were shown and discussed.

Performance assessments

Surgery numbers 1, 5, and 10 were assessed for each subject. If a patient was shown to suffer from severe chronic or acute cholecystitis during the surgery then an adjacent procedure was assessed instead. The procedure was divided into 3 different phases: exposure of the cystic duct and artery, clip placement followed by division of the cystic duct and artery, and gallbladder excision. Video assessments were performed by 2 observers, both experienced in laparoscopic surgery, who were blinded concerning the subjects' training status, the location of the procedure, and by whom the surgery was performed. The procedures were scored on a minute-by-minute basis using the scoring matrix described in Table 2. This methodology has been described previously [14,20,21]. The reviewers were trained thoroughly in the assessment model by using unrelated recordings until they reached an interrater reliability greater than 90% before the series, measured as an index of concordance.

Approval for the study design was obtained from the regional ethical committee at the Karolinska Institute.

Table 2
Assessed surgical error definitions

Exposure errors	
1	Lack of progress: absolutely no progress is made for an entire minute; each minute spent dealing with the consequences of a predefined error represents a lack of progress and should be evaluated as suc
2	Burn nontarget tissue: any application of electrocautery to nontarget tissue
3	Nontarget structure injury: there is a perforation or tear of a nontarget structure (ie, liver, bowel, common duct) with or without associated bleeding or bile leakage; injury in this instance does not include electrocautery along the surface because this would be classified as burning nontarget tissue
4	Instrument out of view: a dissecting instrument with cautery capability is placed outside the field of view of the telescope such that the tip is unviewable; an instrument will be considered to have cautery capabilities at the moment that cautery is applied to tissue and at all times thereafter until the instrument is changed; hook instruments are an exception in that they are considered to always have cautery capabilities; no error will be attributed to an incident of an instrument out of view as the result of a sudden telescope movement
5	Attending takeover: the supervising attending surgeon takes the dissecting instrument or retracting instrument from the resident and performs a component of the procedure; the error occurs throughout the entire period the attending has control and each interval during this period will evaluated as such; the error ends once the resident resumes control of the instrument(s)
6	Gallbladder injury: there is a gallbladder wall perforation with or without leakage of bile
7	Cystic duct injury: there is a perforation or tear of the cystic duct indicated by leakage of bile
8	Inappropriate dissection: dissection is conducted such that either (1) tearing of tissue occurs during dissection within the triangle of Calot, or (2) the plane of dissection within the triangle extends to include areas along the common bile duct
9	Incorrect angle of gallbladder retraction: retraction of the gallbladder is provided such that dissection proceeds within an inadequately distracted or closed triangle of Calot before fenestration of the triangle window
10	Dropped retraction: retracted tissue is dropped suddenly; errors are counted only if regrasping and retraction along a similar angle subsequently are required
Clipping and tissue division errors	
General errors (applies to both clipping and transection)	
1	Attending takeover: for definition see number 5, exposure
Clipping application errors	
2	Clip overlap: clip placed on previously placed clip
3	Clip spacing error: less than 1-mm spacing between distal porta and proximal gallbladder clips
4	Poor clip orientation: clip placed $>10°$ from perpendicular as oriented to cystic duct or cystic artery after clip applier is removed
5	Partial closure: partial closure of cystic artery or duct (clip does not completely cross structure)
6	Poor application: clip is applied such that it is (1) incompletely closed, (2) scissored, or (3) requires regrasp manipulation
7	Poor visualization: clip applied without visualization of the tip of the clip applier
8	Nontarget tissue clipped: Clip applied to nontarget tissue caught in closed tip outside the cystic duct or cystic artery
9	Clip drop: clip comes out of clip applier before application; application occurs once any part of the structure being clipped is within the jaws of the clip applier
Tissue division errors	
10	Inappropriate division: cut less than 1 mm to nearest porta side clip
11	Clip cutting: scissors closure on clip
12	Nontarget injury: cut nontarget tissue
Dissection errors	
1	Lack of progress: for definition see number 1, exposure
2	Burn nontarget tissue: for definition see number 2, exposure
3	Instrument out of view: for definition see number 4, exposure
4	Attending takeover: for definition see number 5, exposure
5	Gallbladder injury: there is a gallbladder wall perforation with or without leakage of bile
6	Liver injury: there is a liver capsule and parenchyma penetration, or capsule stripping with or without associated bleeding
7	Incorrect plane of dissection: the dissection is conducted outside the recognized plane between the gallbladder and the liver (ie, in the submucosal plane of the gallbladder, or subcapsular plan of the liver)
8	Tearing tissue: uncontrolled tearing of tissue with the dissecting or retracting instrument

Statistical methods

All analyses were performed using SAS software (version 9.1.3; SAS institute, Inc., Cary, NC). Graphs and descriptive statistics presented in this article were constructed using Statistica software (version 7; StatSoft, Inc, Tulsa, OK). For the primary (total number of errors) and secondary (total time) outcomes, linear mixed-effects models were used to evaluate the expected mean difference between the simulator-trained and traditionally trained students.

A covariance structure of compound symmetry was found to be the most appropriate for taking into account the correlation within a particular student on different laparoscopies performed. The choice of covariance structure was evaluated by Akaike's information criteria [22]. Because of heterogeneous variances for the training groups, different variance parameters were estimated.

Because of the small number of subjects, univariate models were applied to make inferences about the possible confounders regarding the different psychologic and background variables. Converted procedures were treated as if values were missing completely at random. The Cook's distance and scatter plots of the predicted values versus the raw residuals were used as methods of detecting outliers.

The interrater reliability was calculated as the Cohen's κ coefficient and through analysis of concordance.

$$\frac{\textit{Observation event agreements}}{\textit{Total number of observations}}$$

Results

A total of 120 cholecystectomies were registered within the frame of this study. Eleven trainees performed 10 procedures each and 1 trainee in each group performed 5 each for local logistical reasons. For these latter 2 trainees, surgeries 1 and 5 were assessed. In total, 37 procedures were reviewed. Three observed procedures in the control group were converted to open surgeries.

There was no significant difference between the 2 groups concerning baseline parameters (ie, sex, age, postgraduate year, and previous laparoscopic experience [assisting]). Further, the 2 groups did not differ significantly in their visiospatial assessment (mental rotation test), working memory assessment (number and letter repetition), or in any of the 5 simulator test tasks (Table 3). In addition, the groups did not differ concerning flow experience during simulator testing.

All subjects in the simulator-trained group managed to reach a proficiency level on the simulator after a median of 360 task trials (range, 161–594). Each trainee had to pass all tasks in examination mode twice for proficiency to have been deemed. Two independent experts in laparoscopy, both blinded concerning the subjects' training status, surgical center, and procedure number, performed the video assessment. The interrater reliability between the 2 observers was .98 throughout the assessment, calculated as concordance or in median .80 (95% confidence interval), when the κ coefficient was calculated. In 1 surgery in the control group a severe acute cholecystitis was immediately obvious when the assessment started and surgery number 6 was therefore reviewed instead.

Figs. 1 to 3 show the frequency of errors for the VR-trained group and the control group during the exposure part of the procedure (Fig. 1), the clipping and dissection portion (Fig. 2), and dissection of the gallbladder from the liver bed

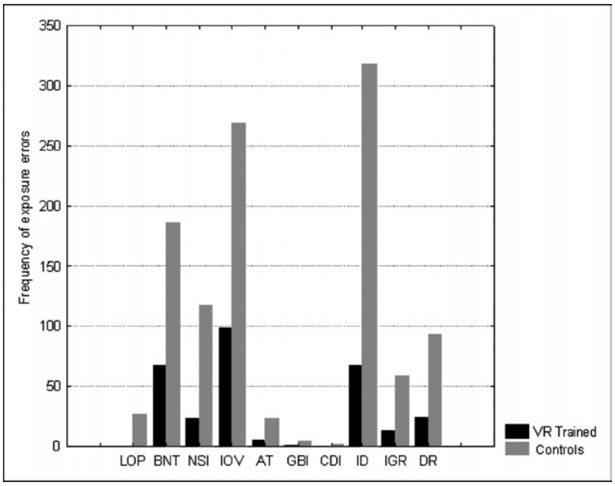

Fig. 1. Frequency of error for each error type during exposure of the triangle of Calot. LOP, lack of progress; BNT, burn nontarget tissue; NSI, nontarget structure injury; IOV, instrument out of view; AT, attending take over; GBI, gallbladder injury; CDI, cystic duct injury; ID, inappropriate dissection; IGR, incorrect angle of gallbladder retraction; DR, dropped retraction. In all error categories a greater number of errors were observed in the control group (□) than in the VR group (■).

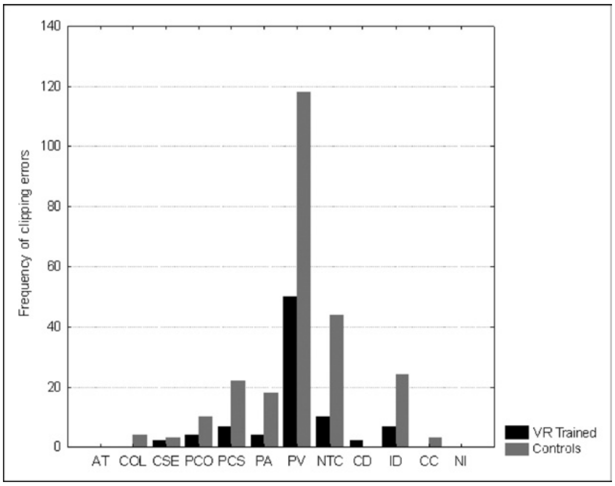

Fig. 2. Frequency of error for each error type during clipping and division. AT, attending takeover; COL, clip overlap; CSE, clip spacing error; PCO, poor clip orientation; PCS, partial closure; PA, poor application; PV, poor visualization; NTC, nontarget tissue clipped; CD, clip drop; ID, inappropriate division; CC, clip cutting; NI, nontarget injury. In all error categories except CD a greater number of errors were observed in the control group (□) than in the VR group (■).

Table 3
Baseline assessment parameters for all subjects described as median (minimum-maximum)

	Trained	Control
	Median (min-max)	Median (min-max)
Baseline assessment parameters		
Age, y	32.0 (30.0–34.0)	32.5 (29.0–45.0)
Laparoscopic assisting experience	15.0 (10.0–25.0)	18.0 (10.0–30.0)
Visuospatial assessment	8.0 (6.0–13.0)	13.5 (2.0–16.0)
Working memory assessment	26.0 (17.0–32.0)	26.0 (23.0–35.0)
Sex, male/female	3/4	3/3
Simulator test (lift and grasp)		
Total time, s	110.3 (61.9–150.9)	114.0 (73.4–182.4)
Left instrument path length, m	1.3 (1.2–1.6)	1.4 (1.1–1.7)
Left instrument angular path, °	317.8 (253.2–397.9)	314.7 (276.4–470.2)
Right instrument path length, m	1.2 (1.0–1.4)	1.2 (1.0–1.4)
Right instrument angular path, °	290.8 (240.0–365.8)	274.9 (238.5–358.9)
Tissue damage, #	2.0 (.0–3.0)	4.0 (2.0–9.0)
Max damage, mm	4.7 (.0–7.4)	5.2 (.7–15.9)

Similar results were seen in all 5 test tasks in the simulator, here as an example depicted in the lift and grasp task. There was no significant difference between the 2 groups concerning baseline parameters.

Min = minimum; max = maximum.

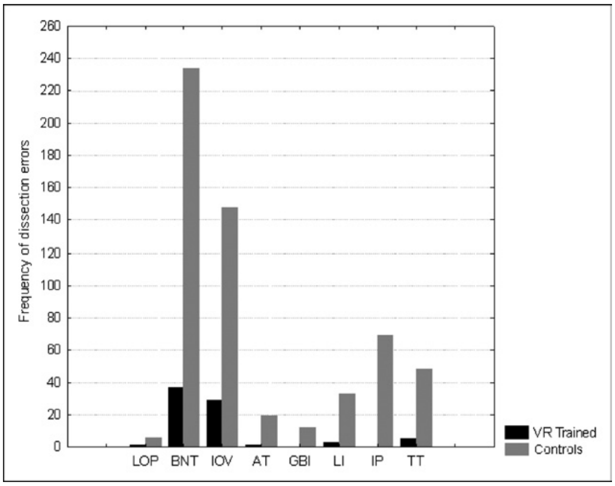

Fig. 3. Frequency of error for each error type during gallbladder dissection. LOP, lack of progress; BNT, burn nontarget; IOV, instrument out of view; AT, attending take over; GBI, gallbladder injury; LI, liver injury; IP, incorrect plane of dissection; TT, tearing tissue. In all error categories a greater number of errors were observed in the control group (□) than in the VR group (■).

(Fig. 3). In all categories except clip drop, the control group made more errors in comparison with the VR group. Indeed, for the majority of error categories the control group made more errors.

Fig. 4 shows the mean number of errors in the control group and the VR-trained group for LC procedures 1, 5, and 10. The VR-trained group made significantly fewer objectively assessed, intraoperative errors during the exposure portion of the procedure (means: control group, 53.4; VR-trained, 15; $P < .04$), clipping and tissue division (means: control group, 7.1; VR-trained, 1.9; $P < .008$), and dissection (means: control group, 29.5; VR trained, 11.5; $P < .03$). The control group also showed considerable variability in performance in comparison with the VR-trained group. In fact, the variation in total errors between the 2 groups with respect to the entire procedure was 8 times higher in the control group, as compared with the VR-trained group (Table 4).

In addition, when errors in exposure, clipping/division, and dissection were looked on separately, the difference remained. The mean number of total errors were, on average, 3 times higher in the control group, as compared with the simulator-trained group, in each part of the entire procedure (Table 4). There were no clear signs of reduction of errors over time when all subjects were assessed, nor was there any statistical difference over time within the groups with respect to error reduction.

Fig. 5 shows the means and SD of the number of objectively assessed, intraoperative errors for the entire LC for each group. The control group made considerably more errors and showed greater variability in performance, as evident from the larger SDs. Indeed, overall the VR-trained group made significantly fewer errors throughout the procedure. Based on the means and the SDs (with the assumption of unequal variance) of these differences, the statistical power was .999.

The surgical time was, on average, 58% longer in the control group when compared with the simulator-trained group. However, this difference did not reach statistical significance ($P < .0586$). In total, 3 conversions among the assessed surgeries were noted, all in the control group. One extreme observation was seen in the control group, with nearly 300 errors for an entire procedure performed by the oldest trainee (age, 45 y). This observation was excluded because it can be quite legitimately treated as a statistical outlier because it was more than 8 SDs from the mean when the value was excluded.

Comments

In this study we objectively assessed, in a prospective blinded fashion, the intraoperative errors of 13 residents who performed 120 LCs. Half of these procedures were performed by residents who had first shown laparoscopic technical proficiency on the LapSim VR simulator before performing surgery. The results of this study indicate that the VR-trained group performed to a significantly higher degree during their first 10 LCs, as compared with the control group. Subjects in the control group made, on average, 3 times as many errors as the VR-trained group and, although not statistically significant, displayed 58% longer surgical times as compared with subjects in the VR-trained group.

The VR-trained group also showed a more homogeneous performance across the procedures that were assessed. The control group consistently showed greater variability throughout the series. This helps to highlight one of the goals of an effective training program that, whether it be VR dependent or not, the goal should be to train surgeons to perform well and consistently within the group and not just as individual pockets of excellence. This assessment method, with error recognition on a minute-by-minute basis, has been described previously and was sufficient when comparing behavior within the 2 groups. In the present study the interrater reliability analysis showed an interrater reliability greater than .9, which indicates a high degree of agreement between the 2 observers, showing high reliability. Our results are, therefore, consistent with the findings of Seymour et al [14].

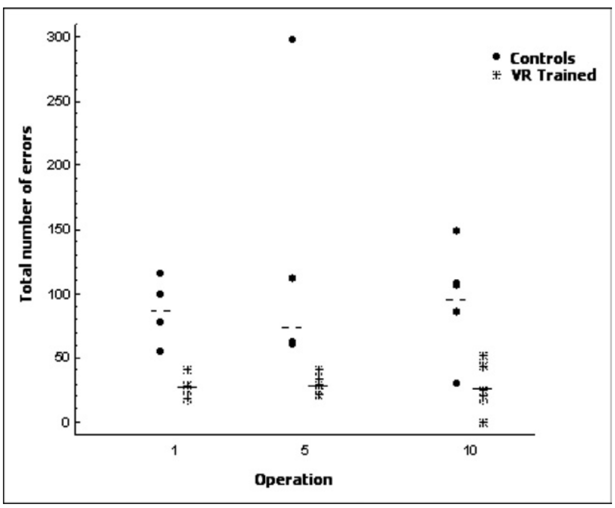

Fig. 4. Horizontal bars indicate the mean number of objectively assessed intraoperative errors for the control group (●) and the VR-trained group (∗) for LC numbers 1, 5, and 10.

Table 4
Mean values concerning total errors for the entire procedure and for the 3 different parts of the procedure together with variance and *P* values

	Trained*	Mean errors	95% confidence interval		Variance	*P* value
			Lower	Upper		
Errors for the entire procedure	0	86.2	58.18	114.12	916.68	
	1	28.4	23.51	33.32	118.69	.0037
Exposure errors	0	53.4	16.70	90.13	623.31	
	1	15.0	11.16	18.79	68.44	.0402
Clipping and tissue division errors	0	7.1	3.95	10.25	41.11	
	1	1.9	.93	2.87	5.57	.0080
Dissection errors	0	29.5	13.99	45.01	61.50	
	1	11.5	8.82	14.08	28.77	.0310

* 0, controls; 1, VR-trained group.

Objective assessment of technical skills is of vital interest when creating certification algorithms for residents before practice on patients. In this study we chose a specific group of residents, highly motivated and with the opportunity to start with LC but without prior experience. The fact that there were no differences in either the baseline performance or flow test results between the groups indicates that they were equally motivated.

The effect of simulator training, with respect to the transfer of skills from the simulated environment to the OR, is of crucial interest if this technology is to reach a compulsory status in surgical training and education. In laparoscopic training only a few previous studies have been able to show skills transfer to an actual surgical task. The breakthrough in this research was a study published by Seymour et al [14] in the *Annals of Surgery* 2002 in which skills acquired in the Minimal Invasive Surgical Trainer-Virtual Reality (MIST-VR) (Mentice, Göteborg, Sweden) system were shown to transfer to a real laparoscopic task in the OR. The dissection of the gallbladder from the liver in LC was videotaped and reviewed in an objective study design based on error analysis. Another study by Grantcharov et al [15], published in the *British Journal of Surgery* in 2004, also showed skills transfer with the Minimal Invasive Surgical Trainer-Virtual Reality system to the OR with another video assessment model based on subjective scoring. Both of the earlier-described studies used residents with prior experience in LC, and simulator training was shown to improve their performance.

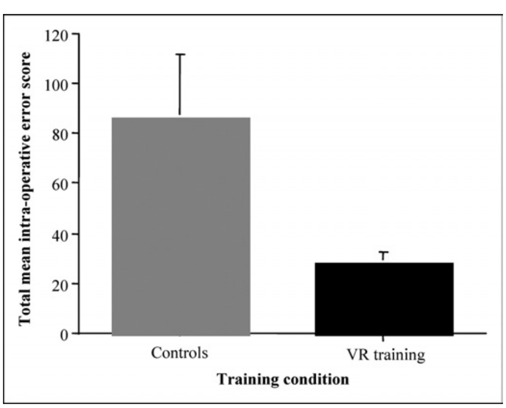

Fig. 5. Total mean and SD scores of the errors enacted during the complete LC.

The most crucial part of an LC is the dissection of the triangle of Calot. Interestingly, in the present study, this part of the procedure was found to show the largest difference in errors between the 2 groups. This finding indicates that the general skills acquired in the LapSim system, together with the procedure-specific skills from the dissection module, are synergistically beneficial. We trained and assessed for the entire LC and we had residents complete their first 10 procedures. The entire learning curve for a new procedure includes both technical and nontechnical skills. In this study we focused only on technical skills trained in the simulator and if these skills transferred to the operating task. Surprisingly, we could not show any reduction in error rate in any of the 2 groups during their 10 LCs. On the other hand, the trained group consistently made fewer errors and performed surgery faster as compared with the control group. If this difference between the groups remains over time will be investigated in a separate, follow-up study after 50 and 100 performed LCs, and a second video review will be performed using the same scoring system.

A potential weakness in our study design was the use of 9 different supervisors in the OR. The reason for this was the strictly chosen test group, with optimal lack of contamination from previous laparoscopic training. To be able to identify these subjects we needed to recruit on a national basis. The importance of the teacher when assessing the learning curve for the trainee has been shown previously [23]. We believe that the differences as a result of multiple supervisors in the present study is of less importance because equal differences between trained subjects and controls were found in centers with multiple participants.

The LapSim is simply a tool, whereas the training paradigm is an integral part of the curriculum. We used a proficiency-based, progression training paradigm for a number of reasons. First, proficiency-based progression sets a very clear benchmark for residents to reach, which they know is reasonable because it was established by experienced laparoscopic surgeons. Second, proficiency-based progression is a much more equitable and fairer training paradigm because it is flexible enough to deal with the truly gifted surgical trainee, as well as those who take slightly longer to acquire the skills. Third, the strength of the approach is that progression does not occur until proficiency has been shown, so one can be assured of at least a defined skill level based on a real-world skills set.

Conclusions

In conclusion, we believe that the results in this study show that skills acquired in the LapSim simulator improve the initial learning curve in LCs, and that the system is clinically validated for this purpose. It also is clear that all new laparoscopists should train on the simulator until they reach the established proficiency level, before performing laparoscopically on patients. Furthermore, our results support the ongoing implementation of simulation technology into the medical training and education system. VR to OR, as a training and scientific validation paradigm, probably will have to continue for the foreseeable future until the surgical establishment is comfortable that this new approach to training is safe and efficient.

Acknowledgment

The authors thank Jakob Bergström at the Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, for all help with the statistical analysis.

References

[1] Darzi A, Munz Y. The impact of minimal invasive surgical techniques. Annu Rev Med 2004;55:223–37.
[2] Moore MJ, Bennet CL. The learning curve for laparoscopic cholecystectomy. Am J Surg 1995;170:55–9.
[3] Deziel DJ, Millikan KW, Economou SG, et al. Complications of laparoscopic cholecystectomy: a national study of 4,292 hospitals and an analysis of 77,604 cases. Am J Surg 1993;165:9–14.
[4] Joice P, Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery—a human reliability analysis. Appl Ergon 1998;29:409–14.
[5] Kohn LT, Corrigan JM, Donaldsson MS. To Err Is Human: Building a Safer Health System. Washington, DC: Institute of Medicine; 1999.
[6] Pfaff H. Surgical safety and overwork. Br J Surg 2004;91:1533–5.
[7] Cuschieri A, Francis N, Crosby J, et al. What do master surgeons think of surgical competence and revalidation? Am J Surg 2000;182:110–6.
[8] Bashook PG, Swing S. Toolbox of assessments methods. Chicago, IL: Accreditation Council for Graduate Medical Education and American Board of Medical Specialities; 2000.
[9] Darzi A, Mackay S. Skills assessment of surgeons. Surgery 2002;131:121–4.
[10] Satava RM, Cushieri A, Hamdorf J. Metrics for objective assessment. Surg Endosc 2003;17:220–6.
[11] Sidhu RS, Grober ED, Musselman LJ, et al. Assessing competency in surgery: where to begin? Surgery 2004;135:6–20.
[12] Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. BMJ 2003;327:1032–7.
[13] Gallagher AG, Satava RM. Objective assessment of experience, junior and novice laparoscopic performance with virtual reality: learning curves and reliability measures. Surg Endosc 2002;16:1746–52.
[14] 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;236:458–64.
[15] Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004;91:146–50.
[16] 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;241:364–72.
[17] Hyltander A, Liljegren E, Rhodin PH, et al. The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surg Endosc 2002;16:1324–8.
[18] Peters M, Laeng B, Latham K, et al. A redrawn Vendenberg and Kuse mental rotation test: different versions and factors that affect performance. Brain Cogn 1995;28:39–58.
[19] Wechsler D. WAIS III manual. San Antonio, TX: The Psychological Corporation; 2003.
[20] Seymour NE, Gallagher AG, Roman SA, et al. Analysis of errors in laparoscopic surgical procedures. A new methodology. Surg Endosc 2004;18:592–5.
[21] Martin P, Bateson P. Measuring behaviour. An introductory guide. Cambridge: Cambridge University Press; 1993.
[22] Akaike H, Nakagawa T. *Statistical Analysis and Control of Dynamic Systems*. Akaike H, Momma MA, trans. Tokyo, Japan: KTK Scientific; Dordrecht; Boston, MA: Kluwer Academic Publishers; 1988.
[23] Ahlberg G, Kruuna O, Leijonmarck C-E, et al. Is the learning curve for laparoscopic fundoplication determined by the teacher or the pupil? Am J Surg 2005;189:184–9.