

Virtual Reality Training Improves Operating Room Performance

Results of a Randomized, Double-Blinded Study

Neal E. Seymour, MD,* Anthony G. Gallagher, PhD,† Sanziana A. Roman, MD,* Michael K. O'Brien, MD,* Vipin K. Bansal, MD,* Dana K. Andersen, MD,* and Richard M. Satava, MD*

From the *Department of Surgery, Yale University School of Medicine, New Haven, Connecticut, U.S.A., and the †Department of Psychology, Queens University, Belfast, Northern Ireland, U.K.

Objective

To demonstrate that virtual reality (VR) training transfers technical skills to the operating room (OR) environment.

Summary Background Data

The use of VR surgical simulation to train skills and reduce error risk in the OR has never been demonstrated in a prospective, randomized, blinded study.

Methods

Sixteen surgical residents (PGY 1–4) had baseline psychomotor abilities assessed, then were randomized to either VR training (MIST VR simulator diathermy task) until expert criterion levels established by experienced laparoscopists were achieved ($n = 8$), or control non-VR-trained ($n = 8$). All subjects performed laparoscopic cholecystectomy with an attending surgeon blinded to training status. Videotapes of gallbladder dissection were reviewed independently by two investigators blinded to subject identity and training, and

scored for eight predefined errors for each procedure minute (interrater reliability of error assessment $r > 0.80$).

Results

No differences in baseline assessments were found between groups. Gallbladder dissection was 29% faster for VR-trained residents. Non-VR-trained residents were nine times more likely to transiently fail to make progress ($P < .007$, Mann-Whitney test) and five times more likely to injure the gallbladder or burn nontarget tissue (chi-square = 4.27, $P < .04$). Mean errors were six times less likely to occur in the VR-trained group (1.19 vs. 7.38 errors per case; $P < .008$, Mann-Whitney test).

Conclusions

The use of VR surgical simulation to reach specific target criteria significantly improved the OR performance of residents during laparoscopic cholecystectomy. This validation of transfer of training skills from VR to OR sets the stage for more sophisticated uses of VR in assessment, training, error reduction, and certification of surgeons.

The introduction of laparoscopic cholecystectomy and the subsequent rapid growth of minimal access surgery

(MAS) have challenged conventional systems for surgical training and establishment of competency. After 1989, as MAS became more commonly practiced, it became clear that the laparoscopic approach was associated with a significantly higher rate of complications,¹ particularly during surgeons' early experience with these procedures.² The underlying causes of these developments were complex but ultimately related to inadequate training of the skills necessary to overcome the psychomotor hurdles imposed by videoscopic interface. When higher complication rates with MAS were scientifically validated,³ surgeons set about defining more structured training methods, such as the Wolf-

Supported with a grant from the Fulbright Distinguished Scholar Program (A.G.G.).

Presented at the 122nd Annual Meeting of the American Surgical Association, April 24–27, 2002, The Homestead, Hot Springs, Virginia.

Correspondence: Neal E. Seymour, MD, Department of Surgery, Yale University School of Medicine, TMP 202, 330 Cedar Street, New Haven, CT 06520-8062.

E-mail: neal.seymour@yale.edu

Accepted for publication April 24, 2002.

DOI: 10.1097/01.SLA.0000028969.51489.B4

son Minimal Access Training Units in the United Kingdom. However, laparoscopic surgical training has for the most part remained relatively unstructured and patterned on the same mentor–trainee model that served surgical training objectives throughout the last century. At the onset of the 21st century, the surgical education establishment is searching for new and innovative training tools that match the sophistication of the new operative methods.

Concurrent with the growth of MAS, separate developments have brought considerable focus on the issue of errors in medicine. The “Bristol Case”⁴ in the U.K. and the “To Err is Human”⁵ report published by the Institute of Medicine in the United States suggested that better training and objective assessment would be key strategies in attaining the goal of reduced medical errors. Surgeons were already sensitive to these issues and have accepted the idea that new and better evidence-based training is necessary and achievable.

Drawing on the successful paradigm of flight simulation, Satava first proposed training surgical skills in virtual reality (VR) nearly a decade ago.⁶ Since that time, with the advancement of desktop computing power, practical and commercially available VR-based surgical simulators and trainers have been developed. At Queen’s University, Belfast, and at Yale University such systems have been employed for training and assessment of surgical skills.^{7–9} Results from both centers show that VR training results in technical skills acquisition at least as good as, if not better than, programs that employ conventional box trainers.^{10,11}

The most important goal of any training method is to increase the level of skill that can be brought to bear on a clinical situation, but to date no studies have established a clear benefit of VR training that transfers to surgeon skill measured in the operating room (OR). Our current study, a component of the program project “VR to OR,” was undertaken to determine whether training on VR in the skills laboratory generalizes to the clinical OR. A commonly performed laparoscopic procedure was selected for examination, along with a VR trainer task that was felt to most effectively train the desired operative skill.

METHODS

Sixteen surgical residents (11 male, 5 female) in postgraduate year (PGY) 1 to 4 in the Yale University School of Medicine Department of Surgery participated in this study. All study participants were randomly assigned to either a study group that would receive VR training in addition to the standard programmatic training (ST) appropriate for PGY level, or a control group that would receive ST only. Participants were stratified by PGY. All residents in both groups completed a series of previously validated tests to assess fundamental abilities. Visuospatial assessment included the pencil and paper Card Rotation, Cube Comparison, and Map Plan tests.¹² Perceptual ability (reconstruction of 3-D from 2-D images) was assessed on a laptop computer with the Pic-

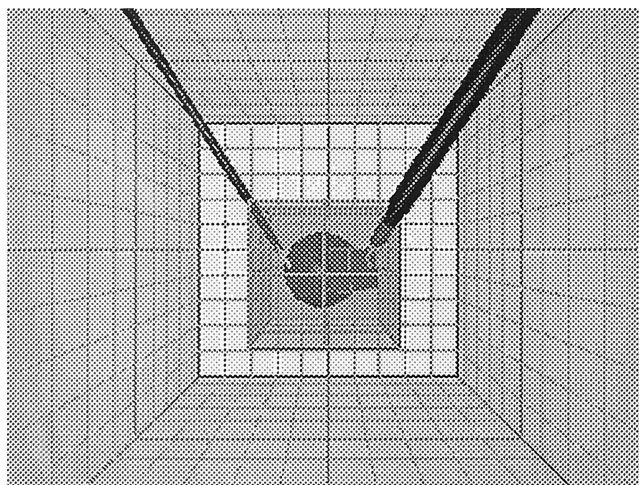


Figure 1. MIST VR screen appearance on “Manipulate and Diathermy” task. The sphere, which must be precisely positioned within a virtual cube, presents a target for the L-hook electrosurgery instrument. Objects may be positioned anywhere within the defined operating space.

torial Surface Orientation test (PicSOR).¹³ Psychomotor ability was assessed with the Minimally Invasive Surgical Trainer-Virtual Reality (MIST VR) system (Mentice AB, Gothenburg, Sweden) with all tasks set at medium level of difficulty.

Apparatus

The MIST VR system (Frameset v. 1.2) was run on a desktop PC (400-MHz Pentium II, 64-Mb RAM) with tasks viewed on a 17-inch CRT monitor positioned at operator eye level. The video subsystem employed (Matrox Mystique, 8-MB SDRAM) delivered a frame rate of approximately 15 frames per second, permitting near-real-time translation of instrument movements to the video screen. The laparoscopic interface input device (Immersion Corporation, San Jose, CA) consisted of two laparoscopic instruments at a comfortable surgical height relative to the operator, mounted in a frame by position-sensing gimbals that provided six degrees of freedom, as well as a foot pedal to activate simulated electrosurgery instruments. With this system, a 3-D “box” on the computer screen represents an accurately scaled operating space. Targets appear within the operating space according to the specific skill task selected and can be grasped and manipulated with virtual instruments (Fig. 1). Each of the different tasks is recorded exactly as performed and can be accurately and reliably assessed.

Training

Four attending surgeons, all with extensive prior experience with laparoscopic procedures, completed 10 trials on the MIST VR “Manipulate and Diathermy” task (see Fig. 1)

at the "Difficult" level to establish the performance criterion levels (mean error score = 50, mean economy of diathermy score = 2). The training goal for residents in the VR group was to perform the same task equally well with both hands on two consecutive trials at the criterion levels set by the experienced surgeons. Training sessions lasted approximately 1 hour. Training was always supervised by one of the authors (A.G.G. or N.E.S.), and explicit attention was paid to error reduction and economy of diathermy.

Operative Procedures

All residents performed laparoscopic cholecystectomy with one of surgeon-investigators, who were blinded to the subject's training status. Before procedures, all were asked to view a short training video demonstrating optimal performance of excision of the gallbladder from the liver using a hook-type monopolar electrosurgical instrument. This video defined specific deviations from optimal performance that would be considered errors. After the viewing, all residents were given an eight-question multiple-choice examination that tested recognition of these errors. During surgery, after division of the carefully identified cystic structures, residents were asked to perform the gallbladder excision using a standardized two-handed method. This phase of the procedure was video-recorded with voice audio by the attending surgeon describing any interventions (attending takeover of one or both instruments). Procedures with attending takeover were flagged for examination of audio.

Error Definition

During unfettered review of archived videotapes of laparoscopic cholecystectomy, potential measures of surgical performance were collated and discussed by the four surgeon-investigators and one behavioral scientist involved in the study. From this list, eight events associated with the excisional phase of the procedure were defined as errors and chosen as the study measurements (Table 1). These measurements excluded any inferences that were not directly observable. All of the events were explicitly defined to facilitate interrater agreement. Clear guidance was given as to when an event was judged to have or have not occurred. The length of time of the gallbladder excision phase was also determined. Timing of length of procedure started with first contact of the electrosurgical instrument with tissue and ended when the last attachment of the gallbladder to liver was divided.

Interrater Reliability Assessment

Each procedural video was viewed without audio by two surgeon-investigators blinded to operating team members. The gallbladder excision phase of the procedure was scored on a minute-by-minute basis using a scoring matrix (see Fig.

Table 1. ASSESSED OPERATIVE ERROR DEFINITIONS

1. LACK OF PROGRESS: No progress made in excising the gallbladder for an entire minute of the dissection. Dealing with the consequences of a predefined error represents lack of progress if no progress is made in excising the gallbladder during this period.
2. GALLBLADDER INJURY: There is gallbladder wall perforation with or without leakage of bile. Injury may be incurred with either hand.
3. LIVER INJURY: There is liver capsule and parenchyma penetration, or capsule stripping with or without associated bleeding.
4. INCORRECT PLANE OF DISSECTION: The dissection is conducted outside the recognized plane between the gallbladder and the liver (i.e., in the submucosal plane on the gallbladder, or subcapsular plane on the liver).
5. BURN NONTARGET TISSUE: Any application of electrocautery to nontarget tissue, with the exception of the final part of the fundic dissection, where some current transmission may occur.
6. TEARING TISSUE: Uncontrolled tearing of tissue with the dissecting or retracting instrument.
7. INSTRUMENT OUT OF VIEW: The dissecting instrument is placed outside the field of view of the telescope such that its tip is unviewable and can potentially be in contact with tissue. No error will be attributed to an incident of a dissecting instrument out of view as the result of a sudden telescope movement.
8. ATTENDING TAKEOVER: The supervising attending surgeon takes the dissecting instrument (right hand) or retracting instrument (left hand) from the resident and performs a component of the procedure.

1) that enabled the observers to record whether an error had or had not occurred during each 60-second period. Errors were recorded using fixed-interval time span sampling (one-zero sampling) described by Martin and Bateson,¹⁴ where a single error event is scored irrespective of how many times during the 1-minute defined period it occurred. Attending takeover events were scored afterward based on review of the flagged videos. Interobserver agreement was determined as described by Kazdin for interval assessments according to the equation: $\text{agreements}/(\text{agreements} + \text{disagreements}) \times 100$.¹⁵

Data are expressed as mean \pm standard error. Statistical comparisons were performed by chi-square analysis, analysis of variance (ANOVA), and Mann-Whitney test (SPSS, Chicago, IL), with statistical significance taken at the $P < .05$ level.

RESULTS

There were no significant differences in any of the initial battery of assessment tests noted between the VR and ST groups (Fig. 2). All residents randomized to the VR group successfully achieved the required criterion levels of performance in three to eight training sessions. All residents in both groups successfully completed the dissection of the gallbladder from the liver bed. The interrater reliability for

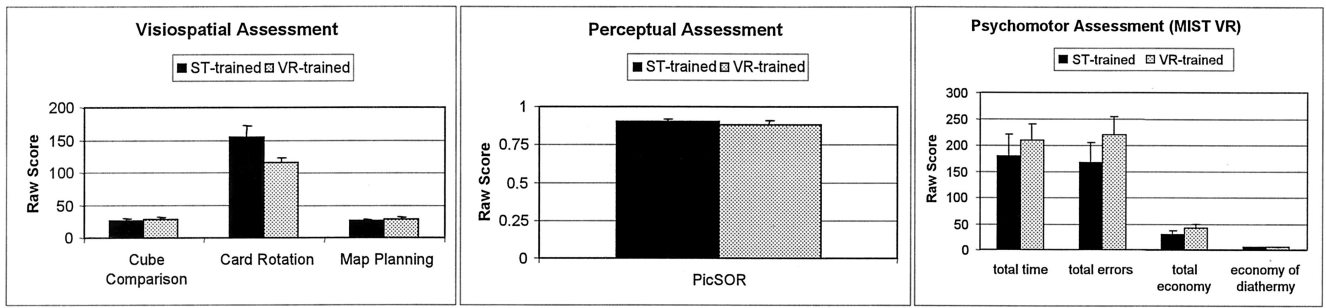


Fig. 2. Results of fundamental abilities assessment. No significant differences were noted in visuospatial, perceptual, or psychomotor abilities between subjects randomized to ST and VR groups when assessed before the training phase of the study.

the assessment of residents’ operative performance during video reviews was $91 \pm 4\%$ (range 84–100%).

The duration of the dissection for the VR-trained group was 29% less than in the ST group, although this difference did not achieve statistical significance (Fig. 3). Gallbladder injury and burn of nontarget tissue errors were five times more likely to occur in the ST group than in the VR group (one of each error in VR residents as compared to five of each error in the ST residents). Separate comparisons between the groups for these errors demonstrated statistical significance in both cases (chi square 4.27, $df = 1$, $P < .039$). ST residents were nine times more likely to be scored as lack of progress, with mean number of lack of progress errors per case of 0.25 versus 2.19 (VR vs. ST groups, respectively; Mann-Whitney, $Z = -2.677$, $P < .008$). There were no tearing tissue errors or noncontact cautery errors in either group. There was one liver injury, three dissection incorrect plane, and six attending surgeon take-over errors scored, all in the ST group. In all error categories except liver injury (one error in VR group) and tearing tissue (no errors either group), more errors were observed in the ST group than in the VR group (Fig. 4). The ST group

made six times as many errors as the VR group (Fig. 5), with four times the variability in the performance of the VR residents as indicated by standard errors. The mean number of scored errors per procedure was significantly greater in the ST than in the VR group (1.19 vs. 7.38, $Z = -2.76$, $P < .006$, Mann-Whitney test).

DISCUSSION

The results of this study demonstrate that it is feasible to train operative skills in virtual reality in surgical trainees without extensive prior MAS experience. Residents who trained on MIST VR made fewer errors, were less likely to injure the gallbladder and burn nontarget tissue, and were more likely to make steady progress throughout the procedure. During VR training it was made clear to residents that speed was not a major training parameter. Instead, training emphasized safe and economical use of electrosurgery instruments and positioning of “tissue” with the nondissecting hand. Completion of the training phase was carefully defined based on objective performance criteria established in

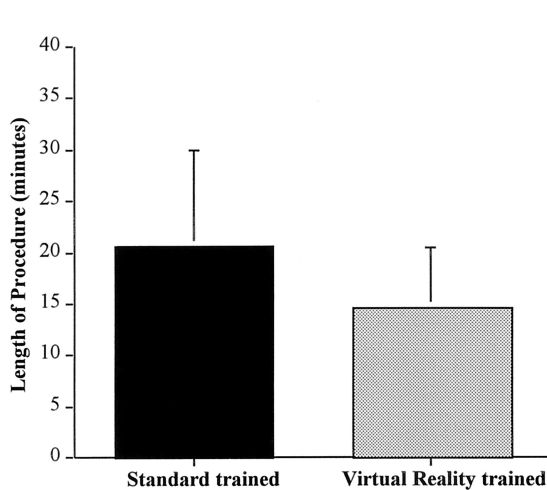


Figure 3. Mean duration of operative procedure for the VR and ST groups.

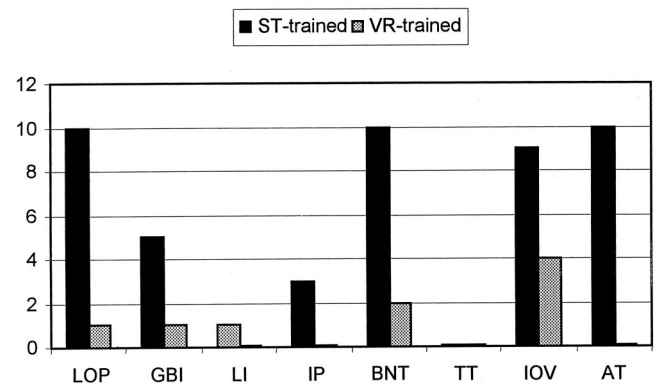


Figure 4. Total error number for each error type. LOP, lack of progress; GBI, gallbladder injury; LI, liver injury; intraperitoneal, incorrect plane of dissection; BNT, burn nontarget tissue; TT, tearing tissue; IOV, instrument out of view; AT, attending takeover. In all error categories except LI and TT, a greater number of errors were observed in the ST group than in the VR group.

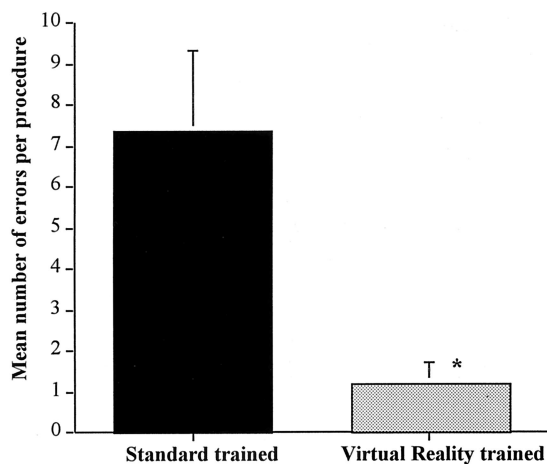


Figure 5. Total number of errors scored per procedure for VR and ST groups. The mean number of errors per procedure was significantly greater in the ST group than in the VR group ($P < .006$).

advance by experienced laparoscopic surgeons. In the planning phase of the study, it was uncertain whether these criterion levels were set too high, but pilot testing demonstrated that these levels could be attained by residents. This is an important point since if the criterion level were set too high, study participants would not be able to reach it. These targeted performance levels may have contributed to the consistent performance demonstrated by the VR-trained group in the OR phase of the study. The requirement that explicit performance criterion levels be reached on two consecutive trials made it unlikely that the resident could achieve it by chance. The performance criterion level established by laparoscopic surgeons at Yale University will need to be validated by the larger surgical community to determine whether they are appropriate measurements and levels to reflect “expert” performance.

No matter how sophisticated laboratory assessment and training methods become, their relationship to OR performance must be established. Our operative assessment methodology was designed to measure observable surgical performance. No inferences were drawn about why the resident performed in a particular way. Global ratings of resident performance and Likert-type scales were avoided in favor of the fixed-interval time span sampling method that identified the presence or absence of predefined error events. This emphasis on observable events resulted in interrater reliability levels that remained above 0.8 throughout the study, with effective blinding of observers to study participant training status. Prior efforts to quantify performance during laparoscopic cholecystectomy have examined similar errors, but with a more global view of the procedure.¹⁶ Although clearly feasible and reliable, this methodology was somewhat time-consuming during both the training and scoring phases.

For the purposes of this investigation, we have chosen a simple operative task that emphasizes technical skills. “Errors” in operative technique were defined as specific events

that represented significant deviations from optimal performance, without linking these events to adverse outcomes or proximate causes. The identification and measurement of these errors permitted assessment of the effectiveness of VR training specifically intended to reduce their incidence. The simplicity of the operative procedure aided in the attainment of this study goal. However, competency comprises disparate cognitive and manual skills elements that do not necessarily lend themselves to unified testing. Anticipating that VR training will rapidly become more varied and realistic, more sophisticated methods of isolating and measuring specific skills in the OR are still needed. We envision the extension of these training tools to other procedures with the aim of eliminating behaviors that lead to adverse clinical outcomes.

The validation of VR training in training operative skills marks a turning point in surgical education. The potential exists to train a resident to a high level of objectively measured skill before he or she is permitted to operate on a patient. VR trainers and simulators offer the advantage of allowing as much training as is required to achieve the training goal. During our study, an investigator observed and instructed the residents during training exercises in order to validate the system. In the immediate future surgical trainees will be able to train whenever they choose, with their performance continuously assessed by the simulator until proficiency in the selected task is attained. With proper software, computer mentoring of the training task is also feasible. The implication is that the surgical education process will soon have the ability to “train out” the learning curve for technical skills on a simulator, rather than on patients; and that a high level of mentoring can be provided without consuming an inordinate amount of a supervising surgeon’s time. VR simulators maintain a log of performance over time, providing an automatic quality assurance tool for objectively assessing the advancement of an individual’s basic technical skills for the program director. It must be emphasized that many more skills are incorporated into the technical training of a surgeon (including the cognitive skills of anatomical recognition, decision making, alternate planning, and so forth), and that the simulators are but one part that can contribute to the overall improvement of performance and assessment of proficiency. Nevertheless, our study validates for the first time the role of VR training on the ability of surgical residents to perform an operative procedure with an improved and, arguably, safer performance. Our findings therefore support the introduction of VR training into surgical education programs.

References

1. Deziel D, Milikan KW, Economou SG, et al. Complications of laparoscopic cholecystectomy: A national survey of 4,292 hospitals and an analysis of 77,604 cases. *Am J Surg* 1993; 165:9–14.
2. The Southern Surgeons Club. The learning curve for laparoscopic cholecystectomy. *Am J Surg* 1995; 170:55–59.

3. Wherry DC, Rob CG, Marohn MR, et al. An external audit of laparoscopic cholecystectomy performed in medical treatment facilities of the Department of Defense. *Ann Surg* 1994; 220:626–634.
4. Senate of Surgery. Response to the General Medical Council Determination of the Bristol Case. London: Senate Paper 5, The Senate of Surgery of Great Britain and Ireland, 1998.
5. Kohn LT, Corrigan JM, Donaldson M. *To Err is Human: Building a Safer Health System*. Washington, DC: Institute of Medicine, 1999.
6. Satava RM. Virtual reality surgical simulator: The first steps. *Surg Endosc* 1993; 7:203–205.
7. Gallagher AG, McClure N, McGuigan J, et al. Virtual reality training in laparoscopic surgery: A preliminary assessment of Minimally Invasive Surgical Trainer Virtual Reality (MIST VR). *Endoscopy* 1999; 31:310–313.
8. Jordan JA, Gallagher AG, McGuigan J, et al. A comparison between random randomly alternating imaging, normal laparoscopic imaging and virtual reality training in laparoscopic psychomotor skill acquisition. *Am J Surg* 2000; 180:208–211.
9. Gallagher AG, McGuigan J, Ritchie K, et al. Objective psychomotor assessment of senior, junior and novice laparoscopists with virtual reality. *World J Surg* 2001; 25:1478–1483.
10. Rosser JC, Rosser LE, Savalgi RS. Skill acquisition and assessment for laparoscopic surgery. *Arch Surg* 1997; 132:200–204.
11. Rosser, JC, Rosser, LE, Savalgi RS. Objective evaluation of laparoscopic surgical skill program for residents and senior surgeons. *Arch Surg* 1998; 133:657–661.
12. Ekstrom RB, French JW, Harman HH, et al. *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service, 1976.
13. Cowie R. Measurement and modelling of perceived slant in surfaces represented by freely viewed line drawings. *Perception* 1998; 27:505–540.
14. Martin P, Bateson P. *Measuring Behaviour: An Introductory Guide*. Cambridge: Cambridge University Press, 1986.
15. Kazdin AE. *Behavior Modification in Applied Settings*. Pacific Grove: Brooks/Cole Publishing Co., 1998.
16. Eubanks TR, Clements RH, Pohl D, et al. An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg* 1991;566–574.

DISCUSSION

DR. CARLOS A. PELLEGRINI (Seattle, WA): The authors of this very important paper tested, in a randomized, double-blind study, the hypothesis that virtual reality simulation training would improve operating room performance. The objective assessment of the laparoscopic cholecystectomy showed that the VR training definitely improved performance when compared to a group of residents trained by traditional means.

A study by our own group in which residents were trained using artificial tissue-like materials shows that these exercises significantly enhanced performance and decreased errors when doing a cholecystectomy in a pig by the residents that were trained this way. Our system allowed us to determine performance objectively at every step of the training phase.

Could you tell us about the individual differences among the residents at the starting time? We found significant differences at the beginning of the training phase, and very little difference, with everybody achieving a performance within 10% of each other, at the end of the training phase.

Was any individual in your study excluded because at the beginning they did not achieve whatever performance levels you have when you take this and other tests? And most importantly, did anyone fail to meet the preset criteria that you had established? Was anybody excluded from this study?

However important these details may be, I would like to make sure that we do not miss the forest for the trees. The real contribution of this presentation, as I see it, is the demonstration that today, using computer simulation and virtual reality environments, we can teach residents skills that in the past we could only do in the operating room or at best in the animal laboratories. Virtual reality simulators allow students of surgery to practice as many times as they need to, which we found to be a very

important element of learning. I think that this would definitely improve the nature of the learning experience, and, most definitely, the quality of life of the resident.

One of the things that we have done is to bring to our laboratories medical students. In fact, we studied fourth-year medical students. I believe that if we want to reverse the trend away from surgery, as President Debas so eloquently described yesterday, we should expose our students to these environments early on in their careers. And herein, I think, lies the strength of the study. Indeed, it is surgeons that have developed and are at the forefront of virtual reality simulation. And since once created, this environment can be modified for other tasks, I believe that surgeons are in a unique position to offer to train medical students in basic skills. Early exposure to these individuals, establishing relationships and friendship at an early stage of medical student careers, I think will have, or may have, a profound effect on their ultimate career choice. Perhaps the authors may wish to comment on this, a less obvious but I think a much more important aspect of their work.

DR. NEAL E. SEYMOUR (New Haven, CT): With regard to your first question, Dr. Pellegrini, none of the subjects were excluded based on the performance criteria levels that we established before they trained. We initially had some concerns that by having experienced laparoscopic surgeons use the MIST-VR to establish a performance benchmark, a new class of difficulty for the task would be created. We did not wish to set a target performance criterion level that could not be achieved by surgical residents at all PGY levels of training. During an early phase of the development of our methods, we tested the ability of residents who were not randomized for the actual study to reach those criteria and levels and found that they could, although there was significant variability in the amount of training that was required to accomplish this. Generally, more senior residents required less training, although the small “n” value does not permit more specific comment on construct validity. There were some differences noted among individuals in the study which I have not presented here today. These pertain to gender-specific performance with regard to the rate and consistency in skills acquisition. Ultimately, the achievement of performance criteria was not a problem for any of the subjects in this study and no one was excluded on that basis.

I absolutely agree with you on the value of VR as a component of a fundamental skills acquisition program, even at the present level of technology of widely available virtual reality simulators and trainers. These devices can, in conjunction with an appropriate curriculum, provide a means of both training and assessing performance. It is not to my mind a replacement for mechanical box trainers and other training techniques, particularly if one examines more complex tasks which currently cannot be achieved in virtual reality. Suturing and knot-tying are VR tasks in development that most readily come to mind in this regard. Simulation of these tasks is a major goal for the engineers and software developers who are currently working in this area. Until this goal is achieved, VR will be extremely valuable for basic skills but somewhat limited for advanced skills acquisition.

However, having said that, direct prospective comparison of box trainers tasks and the MIST-VR in basic skills acquisition anticipatory to the development of laparoscopic suturing and knot-tying abilities have shown that the VR simulator prepares trainees equally well if not better than a box trainer. The training used with the box trainer in that study consisted of the Rosser drills, a well-validated method of preparing trainees and students to do laparoscopic suturing and knot-tying. Preliminary studies like this prompted us to focus on VR as a means of acquiring basic skills in preparation for the OR.

Students are certainly able participants in any VR basic skills acquisition program, and they stand to benefit in a number of ways. Seeing this technology may, in fact, influence a decision to pursue a career in surgery. However, our immediate focus has been on surgical trainees and on giving them the skills they need to perform laparoscopic surgery at a higher level when they enter the operating room.

DR AJIT K. SACHDEVA (Chicago, IL): I believe this a landmark study, as Dr. Pellegrini has mentioned. The study has demonstrated, through a very elegant study design, the transfer of psychomotor skills acquired in a

virtual setting to the real environment. Thus, the predictive validity of the educational intervention has been demonstrated. Also, the measurement approach used by the authors is quite innovative. In the past, evaluations conducted by other investigators have involved the use of global ratings and, prior to that, checklists. The objective evaluation of errors rather than the evaluation of the psychomotor skills using global ratings or checklists is a move in the right direction. I have a number of questions for you, Dr. Seymour.

First, although the cohort of residents was small, did you see any trends by level of learner? Was there a difference in the transfer of skills at the different levels of residents, from years 1 through 4? Second, what was the time interval between the completion of training in the virtual environment and evaluation of skills in the operating room? My third question relates to the second. If the time interval was long, was there any operative experience that might have contributed to the acquisition of the psychomotor skills, and thus have contaminated the results? Finally, what are your plans for further dissemination and validation of your approach? We certainly need larger data sets to further validate your findings, study other types of validity, and assess the generalizability of your approach.

DR. NEAL E. SEYMOUR (New Haven, CT): I appreciate your comment on the predictive validity of the study. We obviously designed our study with predictive validity in mind. We were determined to do our assessment of operative performance without using global ratings and maintaining a focus on errors because we felt that the study strategies would maximize the sensitivity in demonstrating skills transfer. In answer to your first question which pertains to the number of residents and level of training, I believe you were referring to the issue of construct validity. As I pointed out earlier, our study was not designed to test construct validity which has been demonstrated very clearly for the MIST-VR system in the past by a number of investigators including Anthony Gallagher, one of the investigators in the current study. We could see differences between residents at different PGY levels in the number of training sessions it took to achieve performance criterion levels. There were no statistically significant trends in that data, and I have not presented because of, as you rightly pointed out, the small size of the study.

The time interval between completion of the training and performance of the operative procedure was kept to a minimum with some variability due to the vagaries of operative scheduling. Residents were able to do a laparoscopic cholecystectomy with one of the surgeon-investigators within two weeks of the completion of training. Although I think that pushing the operative procedure out to two weeks introduced some question that there might be time-dependent attenuation of the effect of the training, that proved not to be the case. As to our future work in this area, we have shifted our focus to higher end, high fidelity simulators and are examining more sophisticated operative tasks such as clip application and tissue division in laparoscopic cholecystectomy. It is fair to say that these more advanced simulators are much more interesting and exciting for participants to use. This relates to the perception of a more sophisticated task, and the face validity of manipulating recognizable tissue and anatomic structures. The problems of devising the appropriate metrics for the simulators are being solved at present time. As this work evolves, I believe we are going to obtain some very interesting data with these high fidelity simulators as well.

DR. JOHNATHAN L. MEAKINS (Montreal, Quebec, Canada): I second the remarks of Dr. Pellegrini. It is noted that OR time is so precious and expensive that it can't actually be used for practice; whereas, on the other hand, there is an old joke, the punchline of which is, "The way to Carnegie Hall is practice, man, practice." So that is a part of what has been demonstrated.

Cost issues (i.e., OR time, surgeon teaching time, etc.) need to be integrated with the cost of the simulators, how we create the software and how it gets disseminated and need to be integrated into use. These two cost issues need integration with the ways in which we as surgical educators reframe residency programs to deal with modern constraints. I just will mention the 80-hour workweek as an example. There is, however, a second constituency who may well need this kind of training. That is the surgeon in practice who has to deal with a new technique, a new technology, or a

new operation, and needs to learn that in an appropriate way. What is the cost and the time required to create the kind of simulator model that you have got so that it might be applied to the second constituency; that is, those of who need to learn how to do something brand new?

DR. NEAL E. SEYMOUR (New Haven, CT): With regard to the surgeon in practice, I think that we are waiting on the sort of simulations that train very, very specific procedures rather than focus on acquisition of basic skills, although certainly there is still probably plenty of room for basic skills improvement in that target group of surgeons who might want to increase laparoscopic skills. Getting this technology out to surgeons in practice presents an entirely different set of problems that we are currently dealing with. As the machines become more advanced and more available, there will be increased opportunities for community surgeons to have access to simulators to hone skills. We have not worked out the details of how best to achieve that, but it is certainly something to give consideration to.

I am probably not the best person to address the issue of cost of VR training, although I am aware of the considerable cost of the machines that we are using. It is not clear who is best able to pay the development costs of more advanced medical simulation technology or how to generate interest in this sort of endeavor among venture capitalists. I think the fundamental question boils down to who pays for surgical education. In the case of flight simulators, military simulators, and business simulators, it is quite obvious where the benefits lie and intense investment has produced machines that are light years ahead of what I have shown you, and what we currently use today. Again, I think that sorting out how to get money into surgical education is the fundamental problem. Simulators will certainly be a component of the educational process, but producing the best possible simulators is going to take dollars.

DR. LESLIE H. BLUMGART (New York, NY): Three brief questions. Firstly, these operations in your study were performed under the supervision of an attending surgeon. Was it the same attending surgeon throughout? Or could the results have been influenced by different comments from the attending surgeon in the two groups? Secondly, can you use these techniques to teach judgment? For instance, in cholecystectomy, can you, using these virtual reality approaches, teach the surgeon when to convert to an open operation? Finally, can virtual reality be used to train surgeons to do operations which are not necessarily approachable by minimally invasive methods?

DR. NEAL E. SEYMOUR (New Haven, CT): The four attending surgeons who participated in this study did so because of their interest in laparoscopic surgical training. We had a very, very clear protocol which addressed issues of surgeon behavior in the OR, emphasizing what behaviors would and would not be appropriate for the study. Certainly, patient safety was the major concern. In the case of a resident who was not performing the necessary task appropriately, and who was not responding to verbal instructions, would have one or both instruments taken away by the attending. We regarded such an event as an error and as one of the study metrics. The threshold of individuals to intervene is inevitably going to vary variable however, we went to great lengths to try to preserve some uniformity of behavior based on preliminary meetings, discussion, establishment of appropriate OR behaviors, and I think to a very great extent we achieved that. But your question is a difficult one to answer and a difficult problem to test.

Simulators can teach judgment. There are many other simulator paradigms where improved judgment is the major goal of doing simulations. This particular simulator, MIST-VR, is not designed to teach surgical judgment. Another surgical simulator, the sinuscope simulator currently in use, does overlays of anatomic information and shows on a teacher-determined basis what anatomy is revealed to the participant on the simulator, and teaches decision-making based on a continuous, dynamic education process. Rather than doing a terminal phase instruction, this is real-time instruction where anatomic information can be presented to a person so that their judgment will be shown to be adequate or inadequate, as determined by the clinical situation. So, yes, judgment should be topped by simulators and in the future will be taught by simulators with the appropriate curriculum goals in mind.