Surgical Simulation

A Systematic Review

Leanne M. Sutherland, PhD, Philippa F. Middleton, MPH,* Adrian Anthony, MBBS, FRACS,‡
Jeffrey Hamdorf, MBBS, PhD, FRACS,§ Patrick Cregan, MBBS, FRACS,‖
David Scott, MD, MS, FRACS,¶ and Guy J. Maddern, MBBS, MS, MD, PhD, FRACS*†*

Objective: To evaluate the effectiveness of surgical simulation compared with other methods of surgical training.

Summary Background Data: Surgical simulation (with or without computers) is attractive because it avoids the use of patients for skills practice and provides relevant technical training for trainees before they operate on humans.

Methods: Studies were identified through searches of MEDLINE, EMBASE, the Cochrane Library, and other databases until April 2005. Included studies must have been randomized controlled trials (RCTs) assessing any training technique using at least some elements of surgical simulation, which reported measures of surgical task performance.

Results: Thirty RCTs with 760 participants were able to be included, although the quality of the RCTs was often poor. Computer simulation generally showed better results than no training at all (and than physical trainer/model training in one RCT), but was not convincingly superior to standard training (such as surgical drills) or video simulation (particularly when assessed by operative performance). Video simulation did not show consistently better results than groups with no training at all, and there were not enough data to determine if video simulation was better than standard training or the use of models. Model simulation may have been better than standard training, and cadaver training may have been better than model training.

Conclusions: While there may be compelling reasons to reduce reliance on patients, cadavers, and animals for surgical training, none of the methods of simulated training has yet been shown to be better than other forms of surgical training.

(*Ann Surg* 2006;243: 291–300)

From *ASERNIP-S, Royal Australasian College of Surgeons, Stepney, South Australia, Australia; †University of Adelaide, Adelaide, South Australia, Australia; ‡Department of Surgery, Queen Elizabeth Hospital, Woodville, South Australia, Australia; §Surgical Education, University of Western Australia, Crawley, Western Australia, Australia; ‖Surgical Services Network, Sydney West Area Health Service and Surgical Services Taskforce, NSW Department of Health, New South Wales, Australia; and ¶Royal Australasian College of Surgeons, Melbourne, Victoria, Australia.
Reprints: Guy Maddern, MD, PhD, ASERNIP-S, PO Box 553, Stepney, South Australia 5069, Australia. E-mail: college.asernip@surgeons.org.

Simulation is particularly attractive in the field of surgery because it avoids the use of patients for skills practice and ensures that trainees have had some practice before treating humans.[1] Surgical simulation may or may not involve the use of computers.

One of the barriers to development of virtual reality surgical simulation has been the large amount of computing capacity that has been required to remove delays in signal processing, but this is being addressed by systems that break down the tasks by concentrating on chains of behavior.[2] Designers of surgical simulators attempt to balance visual fidelity, real-time response and computing power, and cost.[3]

What Currently Happens in Surgical Training?

Surgical training consists of developing cognitive, clinical, and technical skills, the latter being traditionally acquired through mentoring.[4] Fewer mentoring opportunities[5] have led to the use of models, cadavers, and animals to replicate surgical situations and, more recently, to development of surgical skills centers or laboratories. However, the effectiveness of skills laboratories in teaching basic surgical skills (eg, instrument handling, knot tying, and suturing) is not yet proven.[6,7] Reznick[8] maintains that the teaching and testing of technical skills have been the least systematic or standardized component of surgical education.

Is Simulation an Effective Method of Training?

This review attempts to gauge the instructional effectiveness of surgical simulation[9] (that repeated use improves performance), as well as construct validity (that the simulator measures the skill it is designed to measure). Validity studies of surgical simulation (mostly computer simulation) have shown mixed results for construct validity while other important aspects of validity (such as predictive validity) and reliability are frequently not tested.[3,9] The ultimate validation is for simulation training to show a positive influence on patient outcomes (adapted from Berg et al[9]).

This review focuses on the use of surgical simulation for training, but simulation is also being used to assess surgeons. For example, the U.K. General Medical Council is using it to assess poorly performing surgeons referred to the Council.[10]

TABLE 1.   Summary of Included Studies

Study	Intervention	No. of Participants
Ahlberg et al (2002)[13]	Computer (MIST-VR) simulator training versus no training	29
Ali et al (2002)[14]	Computer (MIST-VR) (medium level) versus computer (MIST-VR) simulator training (easy level)	27
Anastakis et al (1999)[15]	Model simulator training versus cadaver training versus standard training	23
Eversbusch and Grantcharov (2004)[16]	Computer (GI Mentor II) versus no psychomotor training	20
Fried et al (1999)[17]	Video box simulator training versus no training	12
Gallagher et al (1999)[2]	Computer (MIST-VR) simulator training versus no training	16
Grantcharov et al (2004)[18]	Computer (MIST-VR) simulator training versus no training	16
Hamilton et al (2001)[19]	Model simulator training versus no training	22
Hamilton et al (2002)[20]	Computer (MIST-VR) simulator training versus video box simulator training	49
Harold et al (2002)[21]	Video box simulator training plus additional instruction versus video box simulator training	17
Hyltander et al (2002)[22]	Computer (LapSim) simulator training versus no training	24
Jordan et al (2000)[23]	Computer (MIST-VR) simulator training versus video box simulator training (randomly alternating image) versus video box simulator training (normal image)	24
Jordan et al (2000)b[24]	Video box simulator training: randomly alternating image versus y-axis inverted image versus normal image; versus simplified simulation (direct vision box training)	32
Jordan et al (2001)[25]	Computer (MIST-VR) simulator training versus Z-shaped or U-shaped maze video box simulator training versus no training	32
Keyser et al (2002)[26]	Video box training versus simplified simulation (indirect vision mirrored box training)	22
Kothari et al (2002)[27]	Computer (MIST-VR) simulator training versus video box simulator training	24
Lehmann et al (2005)[28]	Computer (VEST) simulator training versus video box simulator training	32
Mackay et al (2002)[12]	Massed practice computer (MIST-VR) simulator training versus 2 types of distributed practice	41
Matsumoto et al (2002)[29]	Video box simulator training versus model training versus standard training	40
Munz et al (2004)[30]	Computer (LapSim) simulator training versus video box simulator training versus no training	24
Pearson et al (2002)[3]	Computer (MIST-VR) simulator training versus video box simulator training versus self-practice versus didactic instruction versus standard training	43
Risucci et al (2001)[31]	Video box simulator training plus additional instruction versus video box simulator training	14
Scott et al (2000)[32]	Video box simulator training versus no training	22
Seymour et al (2002)[33]	Computer (MIST-VR) simulator training versus standard training	16
Taffinder et al (1998)[34]	Video box simulator training versus no training	10
Torkington et al (2001)[35]	Computer (MIST-VR) simulator training versus standard training versus no training	30
Traxer et al (2001)[36]	Video box simulator training versus no training	12
Watterson et al (2002)[37]	Computer (URO Mentor) simulator training versus standard training	20
Wilhelm et al (2002)[38]	Computer (URO Mentor) simulator training versus standard training	21
Youngblood et al (2005)[39]	Computer (LapSim) simulator training versus video box simulator training versus no training	46

Costs of Surgical Training

While the costs of simulation systems can be high, ranging from about U.S. $5000 for most laparoscopic simulators to up to U.S. $200,000 for highly sophisticated anesthesia simulators,[1] traditional Halstedian training is not without cost either. Bridges and Diamond[11] calculated that the cost of training a surgical resident in the operating room for 4 years was nearly U.S. $50,000 (measured by the additional time that the resident took to complete procedures). Other providers of surgical training are making substantial investments in surgical skills centers.

MATERIALS AND METHODS

This review is a summary and an update of an assessment carried out for ASERNIP-S (the Australian Safety and Efficacy Register of New Interventional Procedures): www. surgeons.org/asernip-s/.

Inclusion Criteria

Randomized controlled trials (RCTs) of any training technique using at least some elements of surgical simulation compared with any other methods of surgical training, or no surgical training were included for review. Participants in the RCTs could have been surgeons, surgical trainees (residents), medical students, or other groups. Included studies must have contained information on either measures of surgical task performance (whether objective or subjective) or measures of satisfaction with training techniques or both.

Search Strategy

Studies were identified by searching MEDLINE, EMBASE, PreMEDLINE, Current Contents, The Cochrane Library (Issue 2, 2005), scholar.google.com, *meta*Register of Controlled Trials, National Research Register (U.K.), NHS Health Technology Assessment (U.K.), and the NHS Centre for Reviews and Dissemination databases

TABLE 2. Included Studies Grouped According to Training Category

Category	Simulator Type	Studies
Computer simulator training		
Computer simulation versus no training	MIST-VR	Ahlberg et al (2002),[13] Gallagher et al (1999),[2] Grantcharov et al (2004),[18] Jordan et al (2001),[25] Torkington et al (2001)[35]
	LapSim	Hyltander et al (2002),[22] Munz et al (2004),[30] Youngblood et al (2005)[39]
	GI Mentor II	Eversbusch and Grantcharov (2004)[16]
Computer simulation versus "standard" training	MIST-VR	Pearson et al (2002),[3] Seymour et al (2002),[33] Torkington et al (2001)[35]
	URO Mentor	Watterson et al (2002),[37] Wilhelm et al (2002)[38]
Computer simulation versus video simulation	MIST-VR	Hamilton et al (2002),[20] Jordan et al (2001),[25] Jordan et al (2000),[23] Kothari et al (2002),[27] Pearson et al (2002)[3]
	LapSim	Munz et al (2004)[30]
	VEST (Virtual Endoscopic Surgery Trainer)	Lehmann et al (2005)[28]
Computer simulation versus physical trainer or model	LapSim	Youngblood et al (2005)[39]
2 or more types of computer simulation	MIST-VR	Ali et al (2002),[14] Mackay et al (2002)[12]
	LapSim	
Video simulator training		
Video simulation versus no training	Video box	Fried et al (1999),[17] Jordan et al (2001),[23] Munz et al (2004),[30] Scott et al (2000),[32] Taffinder et al (1998),[34] Traxer et al (2001)[36]
Video simulation versus "standard" training	Video box	Matsumoto et al (2002)[29]
Video simulation versus simplified simulation	Video box	Jordan et al (2000),[24] Keyser et al (2000)[26]
Video simulation versus model simulation	Video box	Matsumoto et al (2002)[29]
Video simulation plus additional instruction versus video simulation	Video box	Harold et al (2002),[21] Risucci et al (2001)[31]
Physical or model simulator training		
Model simulation versus no training	Model	Hamilton et al (2001)[19]
	Physical trainer (Tower Trainer)	Youngblood et al (2005)[39]
Model simulation versus cadaver training	Model	Anastakis et al (1999)[15]
Model simulation versus standard training	Model	Anastakis et al (1999),[15] Matsumoto et al (2002)[29]
Cadaver simulator training		
Cadaver training versus standard training	Cadaver	Anastakis et al (1999)[15]

Some studies contain multiple comparisons.

(last searched April 2005). PsycINFO, CINAHL, and Science Citation Index were searched on March 25, 2003. Search terms used were surgical simulation or surg* AND (simulat* OR virtual realit*). Additional articles were identified through the reference sections of the studies retrieved.

Articles were retrieved when they were judged to possibly meet the inclusion criteria. Two reviewers then independently applied the inclusion criteria to these retrieved articles. Any differences were resolved by discussion.

Data Extraction and Analysis

The ASERNIP-S reviewer extracted data on to data extraction sheets designed for this review and a second reviewer checked the data extraction.

It was not considered appropriate to pool results across studies, because outcomes were not comparable. Relative risks (RR), for dichotomous outcome measures or weighted mean differences (WMD), for continuous outcome measures with 95% confidence intervals (CI) were calculated for some outcomes in individual RCTs where it was thought that this would aid in the interpretation of results. Calculations were made using RevMan 4.2 (Cochrane Collaboration, 2003). Some of the calculations of continuous measures showed skew (the standard deviation is twice more than the mean), indicating that some of the results were not normally distributed. This usually indicates that there is a strong "floor" or "ceiling" effect for performance of a particular task.[12] Study quality was assessed on a number of parameters such as the quality of methodologic reporting, methods of randomization and allocation concealment, blinding of outcomes assessors, attempts made to minimize bias, sample sizes, and ability to measure "true effect."

TABLE 3. Computer Simulation Versus No Training

Study	Simulator	Outcome	Results	Interpretation
Ahlberg et al (2002)[13]	MIST-VR	Simulated appendectomy performance (score of 30)	Simulation (n = 14): mean 12.3 (SD, 4.5) No training (n = 15): mean 12.0 (SD, 4.5) WMD: 0.30 points; 95% CI, −3.58 to 2.98	No difference detected
Gallagher et al (1999)[2]	MIST-VR	Correct incisions	WMD: 9 more correct incisions (95% CI, 6.17–11.83) in simulation group (n = 8) compared with no training group (n = 8)	MIST-VR superior to no training
Grantcharov et al (2004)[18]	MIST-VR	Laparoscopic cholecystectomy performance	Simulation group (n = 8) more accurate (P = 0.003) than the no training group (n = 8), but operation times were similar	MIST-VR generally superior to no training
Jordan et al (2001)[25]	MIST-VR	Laparoscopic cutting skills	Simulation group (n = 8) significantly better (P = 0.001) than no training group (n = 8)	MIST-VR superior to no training
Torkington et al (2001)[35]	MIST-VR	Speed	Simulation group (n = 10) faster than no training group (n = 10)	Inconsistent results
		Other laparoscopic tasks	No statistically significant differences	
Hyltander et al (2002)[22]	LapSim	Navigation	Simulation group (n = 12) performed better and faster (WMD: 2 minutes less, 95% CI, 1.3–2.7) than the no training group (n = 12)	LapSim superior to no training
Munz et al (2004)[30]	LapSim	Clip application performance (on a water-filled glove)	Simulation group (n = 8) showed greater economy of movement but no differences seen for time taken and number of errors when compared with the no training group (n = 8)	Inconsistent results
Youngblood et al (2005)[39]	LapSim	Laparoscopic performance on live anesthetized pig	Simulation group (n = 17) had significantly higher scores (P = 0.005) than the no training group (n = 13)	LapSim superior to no training
Eversbusch and Grantcharov (2004)[16]	GI Mentor II	Computer simulated colonoscopy performance	Simulation group (n = 10) were faster and showed significantly better performance than the no training group (n = 10)	GI Mentor superior to no training

WMD indicates weighted mean difference; CI, confidence interval.

Details of RCTs

A total of 30 relevant RCTs with 760 participants overall were included (Tables 1 and 2).[2,3,12–39]

The length of time devoted to training varied greatly between the studies, with the shortest reported time being 10 minutes,[37] and the longest being 10 hours.[22] In the study by Harold et al,[21] both training groups were given 1 week to practice using the video box trainer, with the trained group receiving an additional 60-minute course on laparoscopic suturing and in Lehmann et al[28] each group carried out a defined training program over 4 days and then switched to the other group on the fifth day. Fifteen studies did not report the length of training but rather the number of sessions or the number of tasks that were practiced.[2,3,14,16,17,20,23,25–27,31,34,35] The previous experience of participants varied greatly between the studies, ranging from high school students[14] to third, fourth, and fifth year urology residents[32] to expert surgeons.[16,28]

Quality of Studies

Only 3 studies[16,18,25] are likely to have demonstrated adequate allocation concealment, with use of sealed or closed envelopes. It was not possible to blind participants for any of the interventions in the included studies. Outcome assessors were blinded in 15 RCTs.[13,15,18–22,29–33,36,37,39] One of the 2 outcome assessors was blinded in Wilhelm et al.[38] In 14 RCTs, it was not stated whether the outcome assessors were blinded.[2,3,12,14,16,17,23–28,30,31] Because the study durations were generally short, most studies had no losses to follow-up.

TABLE 4. Computer Simulation Versus Standard Training

Study	Comparison(s)	Outcome	Results	Interpretation
Pearson et al (2002)[3]	MIST-VR versus unstructured group (watched a video) versus 2 structured training groups (didactic; and self-practice)	Knot tying time	Significantly less for simulation group (n = 10) compared with unstructured group (n = 9); 3.1 versus 7.5 minutes ($P < 0.0001$); no significant differences between simulation group and structured groups (both n = 8)	Inconsistent results
Seymour et al (2002)[33]	MIST-VR, video and questionnaire versus video and questionnaire only	Laparoscopic cholecystectomy performance	Significantly less errors in simulation group (1.2 errors; n = 8) than standard training group (7.4 errors; n = 8) ($P = 0.006$)	MIST-VR superior to standard training
Torkington et al (2001)[35]	MIST-VR versus surgical drills	Time taken to perform laparoscopic tasks on a box trainer	No statistically significant difference between simulation group (n = 10) and standard training group (n = 10)	MIST-VR not shown to be superior to surgical drills training
Watterson et al (2002)[37]	URO Mentor versus didactic training and demonstration	Distal calculus performance on the URO Mentor	Simulation group (n = 10) had a mean final score of 23.5/25 and standard training group scored 14.7 ($P < 0.0001$)	URO Mentor superior to standard training
Wilhelm et al (2002)[38]	Intensive URO Mentor training versus demonstration and one access to URO Mentor	Proximal ureteral calculus performance on the URO Mentor	Simulation group (n = 11) had a mean final global score 5.2 points better (of 25); 95% CI, 3.04–7.36, than the standard training group	URO Mentor superior to standard training

CI indicates confidence interval.

RESULTS

The 30 included studies have been categorized into 4 types of simulation (computer, video, model, and cadaver) and compared with no training and standard training, as well as with each other.

Computer Simulation Versus No Training (9 Studies) (Table 3)

As may be expected, those trained on computer simulators performed better than those who received no training at all. However, the improvement was not universal.

Computer Simulation Versus Standard Training (5 Studies) (Table 4)

When computer simulator trained students were compared with students who received standard training, the superiority of computer simulation was less pronounced than for "no training" comparisons. The computer simulation versus "standard" training comparisons varied, potentially confounded by the different components of "standard" training, as well as by the different intensities of time allowed on the simulator in the computer simulation groups.

Computer Simulation Versus Video Simulation (7 Studies) (Table 5)

Computer simulation showed mixed results, superior in some studies, but not others and was inferior to video simulation in one study. This may have depended on types of tasks, with computer simulation producing better results for tasks such as incisions, but not for knot tying times. However, there were too few studies to determine this.

Computer Simulation Versus Physical Trainer or Model (1 Study)[39] (Table 6)

One study showed computer simulation training to be superior to training on a physical trainer.

Two or More Types of Computer Simulation; MIST-VR (2 Studies) (Table 7)

One RCT[14] showed that more demanding training may lead to better performance of surgical tasks on MIST-VR. Another RCT[12] failed to show clear differences between massed and distributed practice on MIST-VR.

Video Simulation Versus No Training (6 Studies) (Table 8)

Video simulation groups did not show consistently better results than groups who did not receive training.

Video Simulation Versus Other Forms of Training (5 Studies) (Table 9)

Generally, no differences were seen between video box training and other forms of training such as bench models or standard training.

Physical or Model Simulation Versus Other Forms of Training, Including No Training (4 Studies) (Table 10)

While these comparisons showed mixed results, model training may be better than no training and standard training such as instruction from mentors or manuals.

TABLE 5. Computer Simulation Versus Video Simulation

Study	Comparison(s)	Outcome	Results	Interpretation
Hamilton et al (2002)[20]	MIST-VR versus video box training	Laparoscopic cholecystectomy performance on MIST-VR and video box	Simulation group (n = 24) performed better than video group (n = 25) for both MIST-VR ($P < 0.01$) and video box performance ($P < 0.001$)	MIST-VR superior to video simulation
Jordan et al (2001)[25]	MIST-VR versus video box training	Incisions	Simulation group (n = 8) made more correct incisions (mean 40) than the video group (n = 16) ($P = 0.0001$)	MIST-VR superior to video simulation
Jordan et al (2000)[23]	MIST-VR versus video box training	Incisions	Simulation group (n = 8) made more correct incisions than the video group (n = 16) ($P = 0.016$)	MIST-VR superior to video simulation
Kothari et al (2002)[27]	MIST-VR versus video box training	Knot tying time	No statistically significant difference between simulation group (n = 13) and video group (n = 11)	No difference detected
Lehmann et al (2005)[28]	VEST versus video box training	Instrument (VEST) and camera (video box) task performance	Video box group (n = 16) took significantly less camera time ($P = 0.02$) and made significantly less errors ($P < 0.00001$) than the computer simulation group (n = 16)	Video simulation superior to computer simulation
Munz et al (2004)[30]	LapSim versus video box training	Clip application (on a water-filled glove)	Simulation group (n = 8) and video group (n = 8) similar for economy of movement, time taken and number of errors	No difference detected
Pearson et al (2002)[3]	MIST-VR versus video box training	Knot tying time	No statistically significant difference between simulation group (n = 10) and video training group (n = 8)	No difference detected

TABLE 6. Computer Simulation Versus Physical Trainer

Study	Comparison(s)	Outcome	Results	Interpretation
Youngblood et al (2005)[39]	LapSim versus Tower Trainer (endo-trainer)	Laparoscopic performance on live anesthetized pig	Simulation group (n = 17) had significantly higher scores ($P = 0.005$) than the physical trainer group (n = 16)	LapSim superior to physical trainer

TABLE 7. Two or More Types of Computer Simulation; MIST-VR

Study	Comparison(s)	Outcome	Results	Interpretation
Ali et al (2002)[14]	Medium versus easy level training on MIST-VR	Surgical performance on MIST-VR	Medium training group (n = 13) received better scores (WMD: 10 fewer points, 95% CI, −19.81 to −0.19) than the easy level training group (n = 14)	More intensive training is more effective than easy level training
Mackay et al (2002)[12]	Massed versus distributed practice on MIST-VR	Surgical performance on MIST-VR	Massed group (n = 14) received better scores than the 20-minute distributed group (n = 14) ($P = 0.023$), but no difference was seen when compared with the 15-minute distributed group (n = 13) ($P = 0.18$)	Inconsistent results

WMD indicates weighted mean difference; CI, confidence interval.

TABLE 8.　Video Simulation Versus No Training

Study	Comparison(s)	Outcome	Results	Interpretation
Fried et al (1999)[17]	Video box trainer versus no training	Laparoscopic performance in a pig model	No significant difference seen in total score between video trained group (n = 6) and no training group (n = 6)	No difference detected
Munz et al (2004)[30]	Video box trainer versus no training	Clip application performance (on a water-filled glove)	Video trained group (n = 8) showed greater economy of movement, but no differences seen for time taken and number of errors when compared with the no training group (n = 8)	Inconsistent results
Scott et al (2000)[32]	Video box trainer versus no training	Laparoscopic cholecystectomy	Statistically significant improvements in operative performance were seen in the video trained group (n = 9) compared with the no training group (n = 13) for respect of tissue ($P = 0.035$), instrument handling ($P = 0.005$), and use of assistants ($P = 0.035$) but not for knowledge of instruments or procedure, time, and motion or flow of the operation	Inconsistent results
Taffinder et al (1998)[34]	Video box trainer versus no training	Laparoscopic tasks on a MIST-VR	No significant differences seen in efficiency, errors, or time taken between the video trained group (n = 5) and the no training group (n = 5)	No differences detected
Traxer et al (2001)[36]	Video box trainer versus no training	Tasks on a laparoscopic bench trainer	No significant differences detected in simulated nephrectomy tasks between the video trained group (n = 6) and the no training group (n = 6)	No differences detected
Jordan et al (2001)[25]	Video box trainer versus no training	Incisions	Z-shaped video training group (n = 8) made significantly more correct incisions than either the U-shaped video training group (n = 8) or the no training group (n = 8)	Inconsistent results

Cadaver Training Versus Standard Training (1 Study)

In Anastakis et al,[15] the cadaver trained group received better scores than the standard training group, which learned independently from the manuals, for the global assessment of operative performance on cadavers (66.5 versus 51.5) and the checklist score (69.5 versus 60.5), although it was not stated whether these differences were statistically significant.

DISCUSSION

Computer simulation generally showed better results than no training at all but was not convincingly superior to standard training (such as surgical drills) or video simulation (particularly when assessed by operative performance). Video simulation did not show consistently better results than groups with no training at all, and there were not enough data to determine if video simulation was better than standard training or the use of models. Model simulation may have been better than standard training, and cadaver training may have been better than model training. Only one RCT[39] made a comparison between computer simulation and model training, which showed LapSim to be superior to training on a physical Tower trainer.

With 30 RCTs testing the effect of surgical simulation for training, a clearer result might have been expected. Possible reasons for failing to see a clear benefit for surgical simulation include the following:

1. Small sample sizes. There were only 760 participants in total across all the RCTs, meaning that the ability to detect differences between different forms of training was quite limited.
2. Multiple and confounding comparisons. The power to detect any differences was further diluted by the large number of different comparisons, both within and between studies, and the comparisons may have been confounded by factors such as mentoring, which could have swamped any effect of simulation.
3. Disparate interventions. The component tasks of surgical simulation varied considerably between studies, which

TABLE 9. Video Simulation Versus Other Forms of Training

Study	Comparison(s)	Outcome	Results	Interpretation
Matsumoto et al (2002)[29]	Video box trainer versus standard training (one hour surgical instruction)	Video box trainer tasks	No significant differences in scores or time taken were seen between the video box training group (n = 17) and the standard training group (n = 7)	No differences detected
Keyser et al (2002)[26]	Video box trainer versus simplified simulation	Crossover trial on both forms of trainer	Overall scores on mirrored box trainer (n = 22) better than video (same 22 students): mean difference of 240; 95% CI, 33–447, largely influenced by better intracorporeal knot tying	Inconsistent results
Jordan et al (2000)[24]	Video box trainer versus simplified simulation	Incisions in a video box trainer	Randomly alternating image (n = 8) and normal view video box trained group (n = 8) made more correct incisions than the direct vision box trainer (n = 8) ($P < 0.001$ for both comparisons)	Randomly alternating and normal view superior to direct vision
Matsumoto et al (2002)[29]	Video box trainer versus model (bench trainer)	Video box trainer tasks	No significant differences in scores or time taken were seen between the video box training group (n = 17) and the model group (n = 16)	No differences detected
Harold et al (2002)[21]	Video box trainer versus video plus additional training	Video box trainer tasks	No significant differences in knot tying time, needle placement accuracy, or suture strength between the video box training group (n = 8) and the video plus additional instruction group (n = 9)	No differences detected
Risucci et al (2001)[31]	Video box trainer versus video plus additional training	Video box trainer tasks	Video box trained plus additional instruction group (n = 7) made fewer object passing errors (mean difference of 1.6; 95% CI, 0.01–3.19) compared with the video box trained group (n = 7), but no significant differences were seen for other tasks	Inconsistent results

may have obscured results for individual tasks such as knot tying. On the other hand, it may be more realistic to test a "package" of surgical skills, but there did not appear to be a consensus as to what might constitute a core set or package of surgical skills.

4. The comparators were not standardized. "Standard" training covered a wide range of activities, some of which were of very minimal intensity (eg, reading a manual, watching a video). Even so, the simulated groups did not always show superiority even over groups with minimal training.

5. The surgical simulations may not have been intensive or long enough to show an effect on training. While some simulator interventions were very brief, a "practice" effect with longer simulator exposure was not generally evident, although such a practice effect may only be apparent for quite high levels of simulator use. However, most simulator interventions (and comparators) showed significant improvements pre and post intervention (or comparator).

6. Ideally, outcome assessors would have been blinded to group allocation because many of the measurements can be quite subjective (eg, global scores). However, such blinding was not routinely used in the studies included in this review.

7. Most studies measured simulator, and not actual, performance. Unless the simulator has previously been validated by assessing operative performance, or the study assesses skills gained on the simulator by assessing operative performance, we cannot be sure that observed improvements on the simulator will translate into improved operative performance. The failure to translate simulator performance into actual operative performance (concurrent validity) was illustrated in a validity study by Paisley et al,[10] which also showed that simulator performance did not discriminate between naive and experienced surgeons (construct validity). The reasons for this apparent lack of concurrent validity cannot be ascertained from this review, but the relatively new nature of simulation (particularly computer simulation and virtual reality) may be significant causes. Perhaps surgical simulation is not yet realistic enough, in regard to representation of anatomy or haptics are simulated inadequately. Perhaps the simulated tasks

TABLE 10. Model Simulation Versus Other Forms of Training

Study	Comparison(s)	Outcome	Results	Interpretation
Hamilton et al (2001)[19]	Model training (video, CD-ROM and moulded rubber hernia simulator) versus no training (outside the operating room)	Operative performance (laparoscopic hernia repair)	Model training group (n = 11) performed significantly better than the no training group (n = 11); mean overall performance difference of 1.25 points; 95% CI, 0.53–1.87	Model training superior to no training
Youngblood et al (2005)[39]	Physical trainer (Tower trainer) versus no training	Laparoscopic performance on live anesthetized pig	Physical trainer group (n = 16) and no training group (n = 13) very similar overall performance	Not likely to be detectable differences
Anastakis et al (1999)[15]	Model training versus cadaver training	Performance of 6 surgical tasks in cadavers	The cadaver trained group received better scores than the model trained group (66.5 vs. 64.5) and the checklist score (69.5 vs. 67.5), although it was not stated whether these differences were statistically significant. (overall n = 23; which included a standard training group)	Not likely to be detectable differences
Anastakis et al (1999)[15]	Model training versus standard training (learning from manual)	Performance of 6 surgical tasks in cadavers	The model trained group received better scores than the standard training group (64.5 vs. 51.5), although it was not stated whether these differences were statistically significant. (overall, n = 23, which included a standard training group)	Not likely to be detectable differences
Matsumoto et al (2002)[29]	Model training versus standard training (1-hour instruction)	Testing in a Limbs and Things video box trainer	Model group (n = 16) performed better than the standard training group (n = 7) for global rating scores ($P = 0.012$), checklist scores ($P = 0.006$), number of participants receiving a pass rating ($P = 0.001$), and time required to complete the task ($P = 0.013$)	Model training superior to standard training

are isolated or not presented in the right order, or perhaps they are not even the most appropriate tasks.

8. Although it is very difficult to show equivalence (particularly with the small sample sizes and disparate comparisons), a more optimistic interpretation of the findings of this review would be that surgical simulation may be just as good as other forms of surgical training. In other words, can surgical simulation provide an alternative to standard surgical training? Most of the RCTs in this review showed that both simulator and standard training groups improved significantly from baseline, but that participants' final scores usually did not show differences between the simulator and other groups. Apart from the obvious avoidance of experimenting on human patients and not using cadavers or animals, surgical simulation would enable surgical experience earlier in training, make more cases available, give exposure to a wider range of pathologies in a compressed period, allowing acceleration of the learning curve, and give exposure to new procedures and technologies.[40] Surgical simulation is claimed to be able to save money, although this does not

appear to have been substantiated yet. Computer simulation and virtual reality have received much attention, which has eclipsed investigation of other forms of simulation such as models, which may be cheaper and as effective. Unfortunately, many of the studies in surgical simulation have been of poor experimental design, leading Champion and Gallagher[41] to state that "bad science in the field of medical simulation has become all too common." Even though this review included only RCTs, potentially the most rigorous study design for intervention studies, generally there were significant problems with small sample sizes and the validity and reliability of outcome measurements in particular. The studies by Grantcharov et al,[18] Scott et al,[32] and Seymour et al[33] stand out as higher-quality studies since they measured outcomes by actual operative performance, although only some elements were measured in Grantcharov et al.[18]

CONCLUSION

While there may be compelling reasons to reduce reliance on patients, cadavers, and animals for surgical train-

ing, none of the methods of simulated training (including computer simulation) has yet been shown to be better than other forms of surgical training. In addition, little is known about the real costs (including adverse outcomes in patients) of either simulated or standard surgical training.

Adequately powered, well-designed, and unconfounded RCTs (preferably multicenter with similar protocols) are needed and outcome assessors need to be blinded. Outcomes need to be tested in actual operative circumstances (or on validated systems). In particular, model simulation needs to be further tested against computer simulation. Studies of cost comparisons also need to be done. The RCTs dealt exclusively with technical skills, although other skills such as cognitive skills and communication skills are clearly integral parts of surgical performance.

REFERENCES

1. Issenberg SB, McGaghie WC, Hart IR, et al. Simulation technology for health care professional skills training and assessment. *JAMA*. 1999;282:861–866.
2. 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.
3. Pearson AM, Gallagher AG, Rosser JC, et al. Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying. *Surg Endosc*. 2002;16:130–137.
4. Satava RM, Gallagher AG, Pellegrini CA. Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. *J Am Coll Surg*. 2003;196:933–937.
5. Anastakis DJ. Wanzel KR, Brown MH, et al. Evaluating the effectiveness of a 2-year curriculum in a surgical skills center. *Am J Surg*. 2003;185:378–375.
6. Kuppersmith RB, Johnston R, Jones SB, et al. Virtual reality surgical simulation and otolaryngology. *Arch Otolaryngol Head Neck Surg*. 1996;122:1297–1298.
7. Rogers DA, Elstein AS, Bordage G. Improving continuing medical education for surgical techniques: applying the lessons learned in the first decade of minimal access surgery. *Ann Surg*. 2001;233:159–166.
8. Reznick RK. Teaching and testing technical skills. *Am J Surg*. 1993;165:358–361.
9. Berg D, Raugi G, Gladstone H, et al. Virtual reality simulators for dermatologic surgery: measuring their validity as a teaching tool. *Dermatol Surg*. 2001;27:370–374.
10. Paisley AM, Baldwin PJ, Paterson-Brown S. Validity of surgical simulation for the assessment of operative skill. *Br J Surg*. 2001;88:1525–1532.
11. Bridges M, Diamond DL. The financial impact of teaching surgical residents in the operating room. *Am J Surg*. 1999;177:28–32.
12. Mackay S, Morgan P, Datta V, et al. Practice distribution in procedural skills training: a randomized controlled trial. *Surg Endosc*. 2002;16:957–961.
13. Ahlberg G, Heikkinen T, Iselius L, et al. Does training in a virtual reality simulator improve surgical performance? *Surg Endosc*. 2002;16:126–129.
14. Ali MR, Mowery Y, Kaplan B, et al. Training the novice in laparoscopy: more challenge is better. *Surg Endosc*. 2002;16:1732–1736.
15. Anastakis DJ, Regehr G, Reznick RK, et al. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg*. 1999;177:67–70.
16. Eversbusch A, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopic trainer. *Surg Endosc*. 2004;18:1514–1518.
17. Fried GM, Derossis AM, Bothwell J, et al. Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. *Surg Endosc*. 1999;13:1077–1081.
18. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg*. 2004;91:146–150.
19. Hamilton EC, Scott DJ, Kapoor A, et al. Improving operative performance using a laparoscopic hernia simulator. *Am J Surg*. 2001;182:725–728.
20. Hamilton EC, Scott DJ, Fleming JB, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surg Endosc*. 2002;16:406–411.
21. Harold KL, Matthews BD, Backus CL, et al. Prospective randomized evaluation of surgical resident proficiency with laparoscopic suturing after course instruction. *Surg Endosc*. 2002;16:1729–1731.
22. 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;16:1324–1328.
23. Jordan JA, Gallagher AG, McGuigan J, et al. A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. *Am J Surg*. 2000;180:208–211.
24. Jordan JA, Gallagher AG, McGuigan J, et al. Randomly alternating image presentation during laparoscopic training leads to faster automation to the 'fulcrum effect'. *Endoscopy*. 2000;32:317–321.
25. Jordan JA, Gallagher AG, McGuigan J, et al. Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. *Surg Endosc*. 2001;15:1080–1084.
26. Keyser EJ, Derossis AM, Antoniuk M, et al. A simplified simulator for the training and evaluation of laparoscopic skills. *Surg Endosc*. 2000;14:149–153.
27. Kothari SN, Kaplan BJ, DeMaria EJ, et al. Training in laparoscopic suturing skills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. *J Laparoendosc Adv Surg Tech*. 2002;12:167–173.
28. Lehmann KS, Ritz JP, Maass H, et al. A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Ann Surg*. 2005;241:442–449.
29. Matsumoto ED, Hamstra SJ, Radomski SB, et al. The effect of bench model fidelity on endourological skills: a randomized controlled study. *J Urol*. 2002;167:1243–1247.
30. Munz Y, Kumar BD, Moorthy K, et al. Laparoscopic virtual reality and box trainers. *Surg Endosc*. 2004;18:485–494.
31. Risucci D, Cohen JA, Garbus JE, et al. The effects of practice and instruction on speed and accuracy during resident acquisition of simulated laparoscopic skills. *Curr Surg*. 2001;58:230–235.
32. Scott DJ, Bergen PC, Rege RV, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? *J Am Coll Surg*. 2000;191:272–283.
33. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance results of a randomized, double-blinded study. *Ann Surg*. 2002;236:458–464.
34. Taffinder N, Sutton C, Fishwick RJ, et al. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform*. 1998;50:124–130.
35. Torkington J, Smith SGT, Rees BI, et al. Skill transfer from virtual reality to a real laparoscopic task. *Surg Endosc*. 2001;15:1076–1079.
36. Traxer O, Gettman MT, Napper CA, et al. The impact of intense laparoscopic skills training on the operative performance of urology residents. *J Urol*. 2001;166:1658–1661.
37. Watterson JD, Beiko DT, Kuan JK, et al. Randomized prospective blinded study validating acquisition of ureteroscopy skills using computer based virtual reality endourological simulator. *J Urol*. 2002;168:1928–1932.
38. Wilhelm DM, Ogan K, Roehrborn CG, et al. Assessment of basic endoscopic performance using a virtual reality simulator. *J Am Coll Surg*. 2002;195:675–681.
39. Youngblood PL, Srivastava S, Curet M, et al. Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. *J Am Coll Surg*. 2005;200:564–551.
40. Edmond CV. Virtual reality and surgical simulation. In: Citardi MJ, eds. *Computer-Aided Otorhinolaryngology-Head and Neck Surgery*, New York: Marcel Dekker, 2002.
41. Champion HR, Gallagher AG. Surgical simulation: a 'good idea whose time has come'. *Br J Surg*. 2003;90:767–768.