

A Systematic Review of the Educational Effectiveness of Simulation Used in Open Surgery

Leonie Heskin, FRCSI;

Ciaran Simms, PhD;

Jane Holland, PhD;

Oscar Traynor, FRCSI;

Rose Galvin, PhD

Summary Statement: The role of simulation to teach and access open surgical skills has become more prevalent in recent years. This systematic review synthesizes the totality of evidence with respect to the educational effectiveness of simulators used in open surgical training. A systematic literature search was conducted in PubMed, Embase, CINAHL, Scopus, and Web of Science. Only randomized controlled trials were included that explored the educational efficacy of these simulators. Six randomized controlled trials were included from the 9934 studies found. The methodological quality of the included studies was variable. Overall, the use of the simulators was more educationally effective compared with standard teaching of the skill without a simulator ($P < 0.05$). Two studies showed that the simulator was as good as an animal model of much higher fidelity. Further studies are needed to secure higher evidence for the educational value, validity, and transferability of the skills to the hospital setting for all simulators. (*Sim Healthcare* 14:51–58, 2019)

Key Words: Simulation, task trainer, bench model, open surgery, surgical education, technical skill, open surgery.

The role of simulation in surgical resident training is expanding to supplement traditional in-hospital training. In surgical training, simulation attempts to reproduce the essential learning points of a real-life experience.^{1–4} With an increase in operating theater demands, decreased working time for surgeons to learn, and concern for patient safety, the use of simulation in surgical training seems to be a desirable solution at face value.^{5,6} Simulation used to teach open surgical skills usually consists of synthetic, organic, or animal models to recreate the reality of the operating theater.^{7,8} Occasionally, models of increasing fidelity are used, such as live animals, fresh cadavers, and, less frequently, virtual reality simulators.^{1,8} The benefits of simulation allow the surgeon to learn a procedure in a risk-free environment with formative feedback from experts in that specialty.⁹ This type of simulation also allows the learner to grasp instrument handling and to familiarize themselves with new technology in a flexible and cost-effective setting.^{10–12} Once proficient with deliberate practice, specific scenarios with challenging pathology can enhance decision-making and error avoidance and thus enhance patient safety when back in the hospital setting.¹³ In competency-based education, mastery learning with its stringent proficiency targets has been shown to lead to

increased maintenance of skills.^{14–16} The constancy of the simulator for open surgery also provides formative and summative feedback and provides further motivation for the surgeon to achieve mastery outcomes.

Key questions are as follows: are simulators effective in allowing the educator to achieve their learning objectives and the appropriate fidelity for the procedure and have they aided proven transferability of the skill to the patient? When a simulation center is deciding to purchase simulators for surgical training, their priorities will be cost, reusability of the simulator, associated consumables, cost of faculty to teach with them, and the demand to learn a certain procedure. The purchaser would be very much reassured if there was research in the literature demonstrating educational effectiveness, high levels of validity such as discriminative and predictive validity, and transferability of the skill.^{17–21}

With the evolution of many formerly open procedures to endoscopic or laparoscopic surgery, much of the literature looks at simulation for minimally invasive surgery (MIS) with an emphasis on virtual reality simulation. There is more evidence in the literature on the study of the effectiveness of simulation used in MIS.²² The challenges faced by residents converting from closed to open surgery have led to an increase in the numbers seeking out fellowships to refine their open surgical skills. There are comprehensive articles on simulation with valuable descriptions of types of simulators used for training and assessment of surgical skills.^{3,9} A recent systematic and current review of simulation concentrates on training techniques using simulation and inevitably concentrates on those used in MIS.^{20,23} In developing simulators to teach open surgical procedures, there is an increasing amount of simulators being developed commercially and by educational institutions; however, it is not clear whether the evidence in the literature supports their educational value. Animal or human

From the Department of Surgical Affairs (L.H., O.T.), Royal College of Surgeons in Ireland; Department of Mechanical and Manufacturing Engineering (C.S.), Trinity College Dublin; Department of Anatomy (J.H.), Royal College of Surgeons in Ireland, Dublin; and School of Allied Health, Faculty of Education and Health Sciences, Health Research Institute (R.G.), University of Limerick, Limerick, Ireland.

Reprints: Leonie Heskin, FRCSI, MSc, Royal College of Surgeons in Ireland (RCSI), Department of Surgical Affairs, 2nd Flr, 121 St Stephens Green, Dublin 2, Ireland (e-mail: Leonieheskin@rcsi.com).

The authors declare no conflict of interest.

Copyright © 2018 Society for Simulation in Healthcare
DOI: 10.1097/SIH.0000000000000349

models used for simulation have the highest fidelity to teach open surgical skills; however, they require an expensive facility and there are ethics issues. We wish to concentrate on synthetic and virtual reality simulators in this review because we have more control over their design to address particular learning outcomes. Educational institutions can feedback to the makers of simulators to increase the fidelity of their devices or request the insertion of patient-specific pathology into them, for example. Although there are two interesting reviews specifically looking at simulation in open surgery, there is no systematic review looking at learning outcomes with the use of the individual simulators themselves.^{24,25} The purpose of this systematic review and narrative synthesis is to examine the totality of evidence relating to the educational effectiveness of open surgical simulators or task trainers among surgical trainees.

METHODS

Study Design

We conducted a systematic review and narrative synthesis of randomized controlled trials (RCTs) that examined the educational impact of surgical simulators and/or task trainers when compared with routine practice among preregistration and postgraduate surgical trainees. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses standardized reporting guidelines were followed to ensure the conduct and reporting of the research.²⁶

Study Identification

A comprehensive search string was developed in consultation with the faculty librarian. Databases including PubMed, Embase, CINAHL, Scopus, and Web of Science were searched using a combination of the following key words and search terms: “undergraduate” OR “postgraduate” OR “resident” OR “surg*” OR “trainee” OR “internship” AND “simulat*” OR “task trainer” OR “part task trainer” OR “bench model” OR “physical model” OR “virtual reality” OR “low fidelity” OR “high fidelity” OR “open simulat*” AND “surgical education*” OR “assessment” OR “proficiency” OR “education*” OR “train*” OR “skill*” OR “competence” OR “technical skill” OR “skill acquisition” OR “educational efficacy” OR “OSATS.”

Study Selection

Only RCTs (including quasi-RCTs or cluster RCTs) were included where the population of interest included either or both undergraduate and postgraduate surgical trainees. In RCTs, the unit of randomization is the individual participant, whereas in cluster RCTs, the unit of randomization is the participating center (eg, medical school, emergency department, etc). In RCTs, randomization to groups (intervention or control) is usually completed using computer-generated random numbers. Some studies use other methods to randomize such as day of the week or birthdate. These are termed quasi-RCTs because true randomization has not occurred. We included both of these types of experimental studies because randomization is the only method to prevent systematic differences between baseline characteristics of participants in the intervention and control groups in terms of both known and unknown confounding effects. Empirical evidence suggests that on average, nonrandomized studies produce effect estimates that indicate more extreme benefits of the effects of interventions

than RCTs.²⁷ All postgraduate surgical specialties were included in the review. The types of intervention included any task trainer or bench top model that is used as simulation for open surgical training. Simulators include physical models made from synthetic or biological material or a mixture of both, including commercially available models or a model developed by the authors. For the purposes of this systematic review, we focused on models that provided haptic feedback while performing the skill. We also included any simulator of both low and high fidelity. We were interested in outcomes that captured a change in proficiency or skill or confidence of the trainee after the intervention. Studies investigating the training of closed or endoscopic skills were excluded. Examples of closed skills included laparoscopic, endoscopic, and arthroscopic skills. In this study, we looked at studies, which referred to learning open surgical procedures. We also excluded all computer-based simulators including virtual reality that did not have a simulated instrument attached, software imaging controlled by touch screen, or hybrid augmented reality trainers. Medical skills training that did not involve an invasive surgical element were excluded, such as the physical examination of a patient or taking a blood pressure. Studies examining skills where a needle or a tube was inserted into a body cavity or a blood vessel were also excluded.

Data Extraction and Synthesis

A data extraction form was developed by the review team and data extraction was independently undertaken for each study by two reviewers (L.H., R.G.). Information was extracted on authors, year of publication, country of origin, study design, population studied, intervention examined, comparison group and outcome measures assessed, as well as the duration of follow-up. The methodological quality of the included studies was assessed using the Cochrane risk of bias tool for RCTs.²⁷ Domains include selection bias, performance bias, detection bias, attrition bias, selective reporting of outcomes bias, and other biases associated with the study. Selection bias arises because of inadequate generation of a randomized sequence or inadequate concealment of allocations before assignment. Performance bias occurs where there is knowledge of the allocated interventions by participants and personnel during the study. Detection bias refers to knowledge of the allocated interventions by outcome assessors. Attrition bias occurs where the amount, nature, or handling of outcome data is incomplete. Reporting bias arises because of selective reporting of study outcomes. As per the Cochrane guidance notes, each domain was rated by two independent reviewers (L.H., R.G.) as low, unclear, or high risk of bias.²⁷ A “low risk” of bias was awarded if each criterion was met; a “high risk” of bias was documented if each criterion was not met and an “unclear risk” of bias was determined if there was insufficient information presented in the article to permit judgment of low risk or high risk. A consensus meeting was held to resolve discrepancies between reviewers. Seven discrepancies were identified from a total of 42 assessments by each reviewer. In four cases, the discrepancies across reviewers related to selection bias [randomization and allocation concealment (AC)] and three cases related to detection bias in the studies. All discrepancies were resolved without the need for a third independent reviewer.

RESULTS

Study Identification and Selection

Figure 1 describes the flow of studies in the review. After the removal of duplicate studies, 9934 studies were screened by title/abstract. The full-text articles of 11 studies were retrieved for review, five studies^{28–32} were subsequently excluded and the remaining six RCTs were included in the review.^{33–38} There were no trials looking at the effectiveness of a virtual reality simulator in open surgery.

Descriptive Characteristics of the Studies

The sum of the participants over all RCTs was 197. Two studies evaluated the effectiveness with undergraduate students evaluating suturing skills of these novice participants. The remaining studies evaluated tasks that are more complex for post-graduate trainees. Only two of the studies reported the cost of the simulators being studied, for example, the anterior cruciate ligament (ACL) model cost US \$20 to make and the temporary vascular shunt placement simulator cost US \$40 to make.^{33,34} Table 1 details the descriptive characteristics of the included RCTs. One study investigated an orthopedic simulator, two investigated a minor suturing procedure, and three investigated a vascular surgery model.^{33–38} Two of the studies tested two simulators against each other (high-fidelity model vs. low-fidelity model) when compared with the control group.^{35,37} In four of the six studies, the control group were taught by a video, lecture, and textbook.^{33,35–37} All studies used a form of global rating to document how the participant in the studies performed. Some

of the global rating tools were validated, and some were bespoke or modified. Four of the studies used specific check lists for assessing steps in the procedure in addition to the global score rate. The maximum score for each study was different. Some studies had additional test parameters included such as self-reported confidence levels of the participants, time to complete the test, and final product scoring.^{34,35,37,38}

Methodological Quality of the Included Studies

Table 2 details the methodological quality of the included studies. Four studies had a low risk of selection bias^{34–37} because participants were randomly selected or block randomization was performed. In two studies,^{33,38} random sequence generation was not described. Two studies showed low risk of AC bias because they used a computer-generated allocation.^{35,36} The other four studies did not describe how they allocated their groups. Performance bias was considered to be high risk across all interventions because it was not possible to blind participants and personnel to group allocation. In one study,³⁸ it is not clear whether the examiners were blinded or if they were involved in the training on the simulator, so this study has a high risk of detection bias. All other studies had a low risk as they blinded their examiners by using video recordings of the examinations and the examiners were independent to the study.

In terms of selective reporting of outcomes, one study³³ reported participant's skills performance scores but failed to give a numerical value for their confidence and knowledge

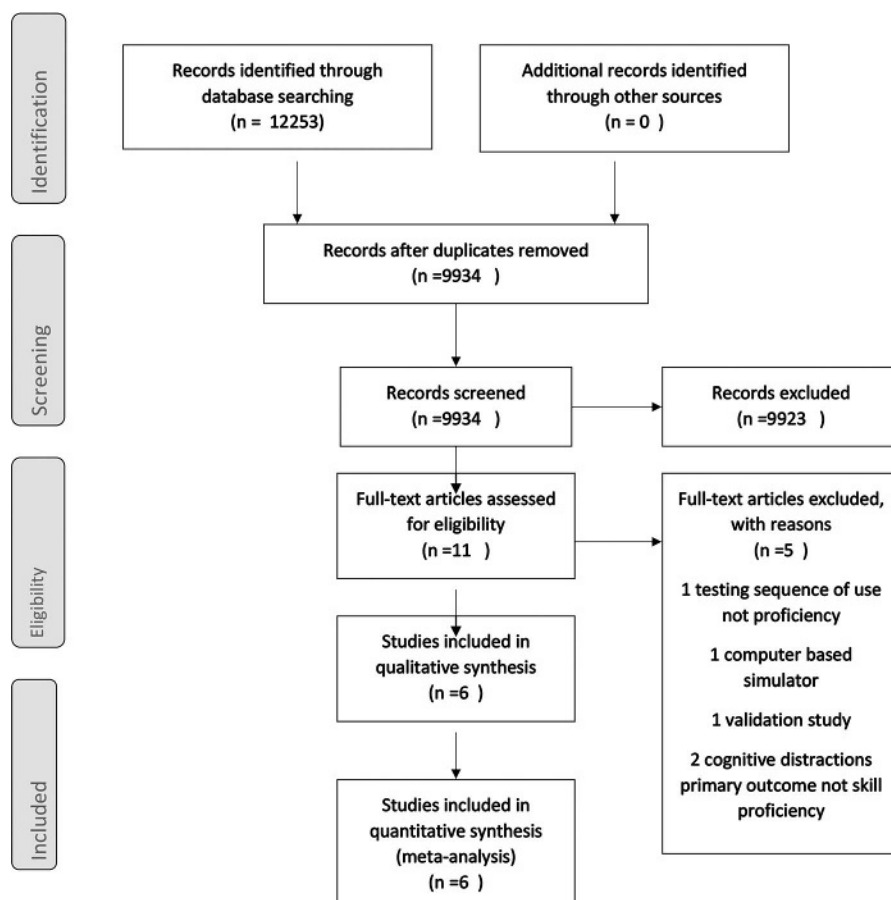


FIGURE 1. The preferred reporting items for systematic reviews and meta-analyses flow diagram.

TABLE 1. Characteristics of the Included Studies

Study	Sample Size (I, C)	Target Population	Intervention	Control/ Comparison	Test Method	Outcome Measure
Denadai et al (2012), ³⁶ Brazil	16 (8, 8)	Med students year 2	Synthetic ethylene-vinyl acetate noncommercial bench model	Video/online/ text book	Pre-post test suturing on ox tongue. Videoed and assessed. Defined time for test = 5-min test.	1. Number of students that completed the task. 2. Number of stitches performed. 3. Global rate score (which included the final product and overall performance) max score 40
Denadai et al (2014), ³⁵ Brazil	40 (8, 8, 8, 8, 8)	Med student year 1, 2.	(n = 8) organic model (n = 8) plastic model (n = 8) chicken (n = 8) pig	(n = 8) Video/ online/text book	Pretest and posttest excision of lesion on ox tongue. Videoed and assessed. Questionnaire.	1. Global rate score. Max score 40 2. Effect size. 3. Self-perceived confidence level.
Brusalis et al (2017), ³³ United States	10 (5, 5)	Postgrad year 1, 3, 4.	Orthopedic noncommercial task trainer to teach ACL graft preparation	Instructional video	8 strand ACL repair on allograft material. Performance videoed and examined blinded. Self-assessment and survey.	1. Checklist 2. Global score. Max points 32 3. Pre and post self-assessment 4. Willingness to use the product again.
Carden et al (2015), ³⁴ United States	54 (26, 28)	Gen surgery residence or equivalent. Postgrad year 1–5.	Dynamic vascular noncommercial vascular simulator	Vessel in a cadaver arm	Performed shunt on a live swine femoral model. Performance videoed. Blindly assessed	1. Checklist, 2. OSATS, 3. Time to hemorrhage control 4. Time to completion of the procedure
Grober et al (2004), ³⁷ Canada	50 (10, 19, 20)	Postgrad year 1–3.	1. Live animal vas deferens 2. Synthetic tubing	Lecture	Pretest = insertion of 2 sutures in a synthetic model. Posttest was like pretest but also all were tested on both the synthetic (1 anastomosis) and live animal vas deferens (2 anastomoses). Videos taken.	1. Check list (30 items). 2. Global rating used. 3. Patent of anastomosis. 4. Suture placement precision. 5. Overall quality of final product. Questionnaire on quality and satisfaction with the model. Delayed outcome 1. Patency 2. Granuloma. Vas deferens working
Sidhu et al (2007), ³⁸ Canada	27 (13 low fidelity, 14 high fidelity)	Postgrad year 1–2, 4, and higher.	Plastic vascular model	Human cadaver arms	1 wk later vascular anastomosis was tested on a live anesthetized animal.	1. Check list. 2. Global score. 3. End product analysis. 4. Time to complete. 5. Hand motion analysis.

scores and could be considered high risk. All articles were considered to have a low risk of attrition bias.

Narrative Synthesis

The variability across the studies in terms of populations studied, simulators examined, comparisons tested, and outcomes explored limited our ability to statistically pool the data. For example, two of the six studies did not perform a pretest of baseline knowledge or skills at the beginning of the RCT,^{34,38} thus limiting our ability to interpret the educational value of the intervention. There was significant heterogeneity across the studies with respect to outcomes reported. Four of the six articles used a check list to explore the impact of the simulator^{33,34,37,38} with scores across the variety of check lists ranging from 10³³ to 30 points.³⁷ Table 3 displays the differences in outcome across the control and intervention groups on the checklist scores. Three of the studies demonstrated a significant educational impact of the simulator postintervention when compared with the comparison group.

There were some checklist components incorporated into the global rating score in one study.³⁵ With regard to the global rating score, only one article used the OSATS (Objective Structured Assessment of Technical Skill)³⁹ in its original format³⁴ with a maxim score of 35. All other studies the authors adapted the global assessment tool, including extra items particular to the procedure, and in some cases, they added a final product score. Therefore, the max score ranged from 35 to 40 and the adaptation made the scores less comparable across the studies. However, all studies that used a pretest showed a significant increase in scores by the participants using the simulators when compared with participants in the controlled groups. Table 4 displays these findings.

The final product analysis, where the appearance of the finished skill is studied, is often excluded from studies on open skills. The quality of the final product analysis was incorporated into the adapted global rating tools used in three studies.^{35,36,38} The study investigating the effectiveness of a silicon tube and live vas deferens and a control go a step further in their posttests

TABLE 2. Methodological Quality of the Included Studies

Author/Year/ Country	Selection Bias – Random Sequence Generation	Selection Bias –AC	Reporting Bias	Performance Bias
Denadai et al (2012), ³⁶ Brazil	16 randomly selected from 60. Risk?	The article does not state whether allocation was concealed. Risk?	No reporting bias Low risk –	In this study, it is not possible to blind the participant from the intervention. High risk +
Denadai et al (2014), ³⁵ Brazil	Assigned to 5 groups by computer-generated process. Low risk –	Informed consent was obtained before participation Low risk –	No reporting bias Low risk –	In this study, it is not possible to blind the participant from the intervention. High risk +
Brusalis et al (2017), ³³ United States	Orthopedic trainees from a single institution recruited. Random sequence generation not described. Risk?	Does not mention allocation. Does mention that the distribution of postgraduate year was equal for both groups but it is not clear whether this is a coincidence or not. Risk?	Check list and global score reported. No numerical value given for contrast in confidence or knowledge. High risk +	In this study, it is not possible to blind the participant from the intervention. High risk +
Carden et al (2015), ³⁴ United States	Scheduling based on resident availability. Block randomization used to ensure equal participation in both groups. The authors admitted that full randomization was not done and stratified randomization on groups due to availability. Not stratified on training level. Low risk –	The article does not state whether allocation was concealed. Risk?	Reporting was clear. Although they did admit that the more senior years did complete the task quicker but this was only significant when comparing 1st yr and 5th yr. Low risk –	In this study, it is not possible to blind the participant from the intervention. High risk +
Grober et al (2004), ³⁷ Canada	Randomized into 3 groups. Unclear how randomization was done. Risk?	Informed consent was obtained before participation Low risk –	No reporting bias Low risk –	In this study, it is not possible to blind the participant from the intervention. High risk +
Sidhu et al (2007), ³⁸ Canada	Randomization done but not shown how. There were even numbers of participant level of experience in the groups. Unclear whether this was a coincidence or arranged. Risk?	The article does not state whether allocation was concealed. Risk?	Results in figures only, not in text. Low risk –	In this study, it is not possible to blind the participant from the intervention. High risk +
Denadai et al (2012), ³⁶ Brazil	Detection bias All recorded performances were evaluated by 2 blinded evaluators Low risk –	Attrition bias No missing data. Low risk –	Other bias The test given to the students only had an allotted time of 5 min. Time to finish and the number of stitches completed in that time was recorded. It is not clear whether the lack of finishing influenced the global score of how they actually performed any stitch?	
Denadai et al (2014), ³⁵ Brazil	Recorded video examined in a blinded fashion by an experienced surgeon who did not take part in the training. Low risk –	No missing data. Low risk –	No teaching of actual operation before the pretest just showed how to use instruments. This may increase the magnitude of the difference in scores between the pre and posttest.	
Brusalis et al (2017), ³³ United States	Assessment of test task videoed without sound to avoid identification of the participant. Examined by expert blinded to group allocation. Low risk –	No missing data. Low risk –	Number of groups small and each group had students from different levels, therefore not necessarily comparative. No pretest to check this skill.	
Carden et al (2015), ³⁴ United States	Edited videos of the task were viewed without sound or identification markers with angles from 2 sides. Two blinded trauma surgeons evaluated the recorded performances. Low risk –	Reported 98% attendance. 26 in 1 group and 28 in the other? why. No missing data to effect the results. Low risk –	Thought the human cadaver model had much higher fidelity, it was not a bleeding model like the synthetic model. It therefore would have not given the same sense of urgency to the participants. It would be a lot more convincing whether the cadaver model was made to bleed too.	
Grober et al (2004), ³⁷ Canada	Video pretest by a blind assessor and same for the matching part of the posttest. Low risk –	Not clear why there were 10 in the control group, 19 in the synthetic model group, and 20 in the high-fidelity group. High risk+	Exclusion criteria were too lax in my opinion. In the posttest, the participants were randomly assigned to which model they get tested on first. One could argue that some got the benefit of extra training on the high-fidelity model? Only a third of the posttest similar to pretest. Uneven numbers in groups.	
Sidhu et al (2007), ³⁸ Canada	The rater was blinded to participant randomization status. Low risk –	No missing data Low risk –	Test took place for a number of days. Some may have had a chance to practice.	

TABLE 3. Check List Scores Across the Intervention and Control Groups

Author/year	Check List Score – Intervention Group Mean $\pm$ SD, <i>P</i> Value (Max Score Available on Checklist)	Check List Score – Control Group Mean $\pm$ SD, <i>P</i> value (Max Score Available on Checklist)
Brusalis et al (2017) ³³ test on allograft	16.4 $\pm$ 5.1 (max 20), <i>P</i> = 0.043	12.6 $\pm$ 3.1 (max 20), <i>P</i> = 0.028
1. Grober et al (2004), ³⁷ anastomosis on silicone tube	Low-fidelity group 24.8 (max 29), <i>P</i> < 0.001 High-fidelity group 25.4 (max 29)	20.5 (max 29), <i>P</i> < 0.001
2. Grober et al (2004), ³⁷ anastomosis on van deferens	Low-fidelity group 25.3 (max 29), <i>P</i> < 0.001 High-fidelity group 26.7 (max 29), <i>P</i> < 0.001	21.9 (max 29), <i>P</i> < 0.001
3. Sidhu et al (2007), ³⁸ test on live animal	Low-fidelity group 16 (junior) (max 24), <i>P</i> = 0.05 18 (senior) (max 24), <i>P</i> = 0.05	High-fidelity group 20 (junior) (max 24), <i>P</i> = 0.05 22 (senior) (max 24), <i>P</i> = 0.05

to examine effectiveness of the simulators.³⁷ They examine the final product just after the initial assessment and then 30 days later to determine whether the structure is still patent, with significant differences still evident across the groups at 30 days.

Although subjective, two of the articles included an additional self-reported confidence or knowledge of skills in the pretest and posttest, which adds to the complexity of the outcomes.^{33,35} In both cases, the use of the ACL simulator and the four models to teach the excision of a lesion significantly increased the confidence of the exposed group when compared with the control group.

DISCUSSION

This systematic review explored the totality of evidence regarding the educational effectiveness of simulators used in open surgical training. We identified six studies that met our inclusion criteria.^{33–38} All simulators examined were noncommercial and made by the authors of the studies. There was significant heterogeneity across the studies in terms of the populations, interventions, comparison groups, and outcomes examined. This limited our ability to pool the data using statistical methods. However, our narrative synthesis demonstrates that overall, the results favor the use of simulators over conventional methods used to teach open surgical skills among

undergraduate and postgraduate trainees. We have to acknowledge in the study of task trainers for an open surgical simulator that the input of the facilitator is important in bringing out the most value of the simulator as a learning tool. One study that developed an in-house aortic aneurysm repair model showed a superior effect when the model was facilitated with a vascular surgeon as compared with a technician.⁴⁰ Most of the included studies addressed this performance bias by ensuring that the control group got similar training.

We used robust and transparent methods to identify, select, appraise, and synthesize the findings from the review. Our comprehensive search string across multiple databases yielded a significant number of studies for consideration. A large number of studies addressed simulation in MIS particularly with the use of virtual reality trainers. The search also yielded studies, which concentrated on different teaching methods, validation studies, and nontechnical skills simulation studies. Some RCTs did not address educational effectiveness directly but looked at issues such as using a carotid artery bench model to assess competency before doing in-hospital surgery or examined the effectiveness of an extended skills course where the simulator only featured in a small part of the entire course.⁴¹ An RCT included in the two previously published open simulation reviews included the transfer of

TABLE 4. Global Rating Scores Across the Intervention and Control Groups

Author/Year	Global Rating Score – Intervention Group Mean $\pm$ SD, <i>P</i> Value (Max Score Available on Checklist)					Global Rating Score – Control Group Mean $\pm$ SD, <i>P</i> value (Max Score Available on Checklist)	
Brusalis et al (2017), ³³ test on allograft	6.2 $\pm$ 1.5 (max 12), <i>P</i> = 0.058					3.4 $\pm$ 2.4 (max 12), <i>P</i> = 0.058	
Carden et al (2015), ³⁴ test on live animal	1.72(max 30 but average for item taken out of 5)					3.44 (max 30 but average for item taken out of 5)	
Denadai et al (2014), ³⁵ test on ox tongue	Group	2	3	4	5	Group	1 = control
	Pre max (40)	10.81 $\pm$ 1.11, <i>P</i> < 0.05	10.81 $\pm$ 1.17, <i>P</i> < 0.05	10.75 $\pm$ 1.39, <i>P</i> < 0.05	10.69 $\pm$ 1.3, <i>P</i> < 0.05	Pretest max (40)	10.7 $\pm$ 1.39, <i>P</i> < 0.05
	Post max (40)	25.38 $\pm$ 3.10, <i>P</i> < 0.05	25.56 $\pm$ 2.63, <i>P</i> < 0.05	25.88 $\pm$ 3.07, <i>P</i> < 0.05	25.69 $\pm$ 2.91, <i>P</i> < 0.05	Post max (40)	14.88 $\pm$ 1.5, <i>P</i> < 0.05
Denadai et al (2012), ³⁶ test on ox tongue	(Interrupted suture) 22.13 $\pm$ 6.52 (max 40), <i>P</i> = 0.0001 (Mattress suture) 20.63 $\pm$ 7.68 (max 40), <i>P</i> = 0.0006					(Interrupted suture) 13.19 $\pm$ 5.48 (max 40), <i>P</i> = 0.0001 (Mattress suture) 12.25 $\pm$ 5.26 (max 40), <i>P</i> = 0.0006	
1. Grober et al (2004), ³⁷ drill anastomosis on silicone tube	Pretest low-fidelity group 16.5 High-fidelity group 16.8 (max 40), <i>P</i> = 0.01 Posttest low-fidelity group 23.5 High-fidelity group 24.8 (max 40), <i>P</i> = 0.01					Pretest 14.0 (max 40) <i>P</i> = 0.01 Posttest 15.2 (max 40) <i>P</i> = 0.01	
2. Grober et al (2004), ³⁷ anastomosis on silicone tube	Low-fidelity group 25.6 (max 40), <i>P</i> = 0.002 High-fidelity group 28 (max 40), <i>P</i> = 0.002					18.9 (max 40), <i>P</i> = 0.002	
3. Grober et al (2004), ³⁷ anastomosis on van deferens	Low-fidelity group 27 (max 40), <i>P</i> = 0.001 High-fidelity group 28 (max 40), <i>P</i> = 0.001					19.2 (max 40), <i>P</i> = 0.001	
4. Sidhu et al (2007), ³⁸ test on live animal	Low-fidelity group 3.1 (junior) 3.5 (senior) (Max 30 but average for item taken out of 5)					High-fidelity group 3.6 (junior) 3.8 (senior) (Max 30 but average for item taken out of 5)	

skills for the Berlin Operation Trainer, which teaches the trainee to stand in the correct position for a bowel anastomosis and creates human anatomical visuals of the inner abdomen. We did not use this study because the trainer necessitates the insertion of animal bowel for the anastomosis.⁴² The non-RCTs relating to open surgical simulators included technical descriptions of materials such as urinary catheters being used in tendon repair or a synthetic bowel being used as the prepuce for a circumcision model.^{43,44} The use of 3D printing was also featured with patient data being used to create a skull for cranial base surgery and a soft kidney phantom with realistic anatomical structures.^{45,46} There are a few virtual reality simulators for teaching open surgical skills, such as Sim-Ortho for open spinal surgery, the virtual reality education surgical tool to teach open hernia surgical repair and the Boston Dynamics INC surgical simulator for teaching suturing.^{47–49} We did not include virtual reality simulators, which do not have physical instrument with haptic feedback but aid the learning of the anatomy, procedural steps, and decision-making. Examples of these include the open inguinal hernia repair simulation model and touch surgery.³¹ We noticed that when using the key word “high fidelity” in our search, it yielded a significant number of studies describing virtual reality simulators as opposed to likeness to real-life models.

It is surprising that so few randomized controlled trials were finally selected. It may be possible that trials studying the educational merit of commercial simulators may have taken place in-house and thus the findings are not published. It may be also difficult to design these studies in the student setting but an element of using cross-over study designs should discourage any ethical issues. Although the literature demonstrating a positive educational contribution of simulators in open surgery is increasing, the evidence for validity and transferability to the patient has not been widely explored. This is predominantly due to the design of the studies where the surgical simulator is often used as a component of a larger training course with lots of other contributing factors to the end of course assessment. Furthermore, it is difficult to prove transferability of specific simulators because there are so many other factors such as in-hospital training that may influence the student experience.

The variability across the studies limits the internal and external validity of the findings of the review. Further methodologically robust longitudinal studies are warranted using standardized methods to conduct and report the findings. There are emerging articles exploring the development of guidelines for specific validation simulators used in surgical practice.⁵⁰ There are some efforts at creating consensus guidelines for the validation of virtual reality simulators used in endoscopic surgical education.⁵¹ It would appear from our review, with the lack of high evidence studies on both commercial and noncommercial simulators, that such guidelines should include educational effectiveness demonstration and evidence of transferability to the hospital setting.^{19,52} In addition, this higher level of evidence would help educators decide on the most effective simulators to add to their curriculum and simulation centers. This type of research is bound to encourage the development of more sophisticated simulators that satisfy the teaching of learning outcomes important for

that open procedure. Demonstration of a reduction in medical error and enhanced recognition of patient safety issues would further enhance the rate of simulator use.

CONCLUSIONS

A small number of studies were found that assessed the educational benefit of open surgical training tools. It was not possible to meta-analyze these studies because of methodological and clinical differences across the studies. Further studies are needed to secure higher evidence for the educational value, validity, and transferability of the skills to the hospital setting for all simulators in use in surgical training. In the interim, this systematic review adds positive encouragement to their use.

REFERENCES

1. Tan SS, Sarker SK. Simulation in surgery: a review. *Scott Med J* 2011; 56(2):104–109.
2. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, et al. Simulation technology for health care professional skills training and assessment. *JAMA* 1999;282(9):861–866.
3. Sarker SK, Patel B. Simulation and surgical training. *Int J Clin Pract* 2007;61(12):2120–2125.
4. Scott DJ, Cendan JC, Pugh CM, Minter RM, Dunnington GL, Kozar RA. The changing face of surgical education: simulation as the new paradigm. *J Surg Res* 2008;147(2):189–193.
5. De Montbrun SL, Macrae H. Simulation in surgical education. *Clin Colon Rectal Surg* 2012;25(3):156–165.
6. Kneebone R. Simulation in surgical training: educational issues and practical implications. *Med Educ* 2003;37(3):267–277.
7. Hamdorf JM, Hall JC. Acquiring surgical skills. *Br J Surg* 2000; 87(1):28–37.
8. Cosman P, Hemli JM, Ellis AM, Hugh TJ. Learning the surgical craft: a review of skills training options. *ANZ J Surg* 2007;77(10):838–845. Review.
9. Hammoud MM, Nuthalapaty FS, Goepfert AR, Casey PM, Emmons S, Espey EL, et al. To the point: medical education review of the role of simulators in surgical training. *Am J Obstet Gynecol* 2008;199(4):338–343.
10. Aucar JA, Groch NR, Troxel SA, Eubanks SW. A review of surgical simulation with attention to validation methodology. *Surg Laparosc Endosc Percutan Tech* 2005;15(2):82–89.
11. Schout BM, Hendriks AJ, Scheele F, Bemelmans BL, Scherpbier AJ. Validation and implementation of surgical simulators: a critical review of present, past, and future. *Surg Endosc* 2010;24(3):536–546.
12. Maran NJ, Glavin RJ. Low- to high-fidelity simulation - a continuum of medical education? *Med Educ* 2003;37(Suppl 1):22–28.
13. Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;46(7):636–647.
14. Wayne DB, Barsuk JH, O'Leary KJ, Fudala MJ, McGaghie WC. Mastery learning of thoracentesis skills by internal medicine residents using simulation technology and deliberate practice. *J Hosp Med* 2008; 3(1):48–54.
15. Siddaiah-Subramanya M, Smith S, Lonie J. Mastery learning: how is it helpful? An analytical review. *Adv Med Educ Pract* 2017;8:269–275.
16. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Medical education featuring mastery learning with deliberate practice can lead to better health for individuals and populations. *Acad Med* 2011; 86(11):e8–e9.
17. Kneebone R, ApSimon D. Surgical skills training: simulation and multimedia combined. *Med Educ* 2001;35(9):909–915.
18. Reznick RK, MacRae H. Teaching surgical skills-changes in the wind. *N Engl J Med* 2006;355(25):2664–2669.
19. Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. *Ann Surg* 2008;248(2):166–179.

20. Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, et al. Surgical simulation: a systematic review. *Ann Surg* 2006; 243(3):291–300. Review.
21. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, et al. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 1999;177(2):167–170.
22. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Ann Surg* 2013;257(4):586–593.
23. Dunkin B, Adrales GL, Apelgren K, Mellinger JD. Surgical simulation: a current review. *Surg Endosc* 2007;21(3):357–366.
24. Davies J, Khatib M, Bello F. Open surgical simulation—a review. *J Surg Educ* 2013;70(5):618–627.
25. Fonseca AL, Evans LV, Gusberg RJ. Open surgical simulation in residency training: a review of its status and a case for its incorporation. *J Surg Educ* 2013;70(1):129–137.
26. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. PRISMA-P Group. *BMJ* 2015;349:g7647.
27. Higgins JP, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available at: handbook.cochrane.org. Accessed November 5, 2018.
28. Dancz CE, Sun V, Moon HB, Chen JH, Özel B. Comparison of 2 simulation models for teaching obstetric anal sphincter repair. *Simul Healthc* 2014;9(5):325–330.
29. Geoffrion R, Suen MW, Koenig NA, Yong P, Brennand E, Mehra N, et al. Teaching vaginal surgery to junior residents: initial validation of 3 novel procedure-specific low-fidelity models. *J Surg Educ* 2016;73(1):157–161.
30. Grierson L, Melnyk M, Jowlett N, Backstein D, Dubrowski A. Bench model surgical skill training improves novice ability to multitask: a randomized controlled study. *Stud Health Technol Inform* 2011; 163:192–198.
31. Khatib M, Hald N, Brenton H, Barakat MF, Sarker SK, Standfield N, et al. Validation of open inguinal hernia repair simulation model: a randomized controlled educational trial. *Am J Surg* 2014;208(2):295–301.
32. Palter VN, Grantcharov T, Harvey A, Macrae HM. Ex vivo technical skills training transfers to the operating room and enhances cognitive learning: a randomized controlled trial. *Ann Surg* 2011;253(5):886–889.
33. Brusalis CM, Lawrence JTR, Ranade SC, Kerr JC, Pulos N, Wells L, et al. Can a novel, low-cost simulation model be used to teach anterior cruciate ligament graft preparation? *J Pediatr Orthop* 2017;37(4):e277–e281.
34. Carden AJ, Salcedo ES, Leshikar DE, Utter GH, Wilson MD, Galante JM. Randomized controlled trial comparing dynamic simulation with static simulation in trauma. *J Trauma Acute Care Surg* 2016;80(5):748–753; discussion 753–4.
35. Denadai R, Oshiiwa M, Saad-Hossne R. Teaching elliptical excision skills to novice medical students: a randomized controlled study comparing low- and high-fidelity bench models. *Indian J Dermatol* 2014; 59(2):169–175.
36. Denadai R, Saad-Hossne R, Oshiiwa M, Bastos EM. Training on synthetic ethylene-vinyl acetate bench model allows novice medical students to acquire suture skills. *Acta Cir Bras* 2012;27(3):271–278.
37. Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, et al. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Ann Surg* 2004;240(2):374–381.
38. Sidhu RS, Park J, Brydges R, MacRae HM, Dubrowski A. Laboratory-based vascular anastomosis training: a randomized controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. *J Vasc Surg* 2007;45(2):343–349.
39. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84(2):273–278.
40. Robinson WP, Baril DT, Taha O, Schanzer A, Larkin AC, Bismuth J, et al. Simulation-based training to teach open abdominal aortic aneurysm repair to surgical residents requires dedicated faculty instruction. *J Vasc Surg* 2013;58(1):247–253.e1–2.
41. Black SA, Harrison RH, Horrocks EJ, Pandey VA, Wolfe JH. Competence assessment of senior vascular trainees using a carotid endarterectomy bench model. *Br J Surg* 2007;94(10):1226–1231.
42. Lauscher JC, Ritz JP, Stroux A, Buhr HJ, Gröne J. A new surgical trainer (BOPT) improves skill transfer for anastomotic techniques in gastrointestinal surgery into the operating room: a prospective randomized trial. *World J Surg* 2010;34(9):2017–2025.
43. Abdulal S, Onyekwelu O. An alternative model for teaching tendon repair and surgical skills in plastic surgery. *JPRAS* 2016;12–15.
44. Abdulmajed MI, Thomas M, Shergill IS. A new training model for adult circumcision. *J Surg Educ* 2012;69(4):447–448.
45. Abe M, Tabuchi K, Goto M, Uchino A. Model-based surgical planning and simulation of cranial base surgery. *Neurol Med Chir (Tokyo)* 1998;38(11): 746–750; discussion 750–1.
46. Adams F, Qiu T, Mark A, Fritz B, Kramer L, Schlager D, et al. Soft 3D-printed phantom of the human kidney with collecting system. *Ann Biomed Eng* 2017;45(4):963–972.
47. Sanders AJ, Warntjes P, Geelkerken RH, Mastboom WJ, Klaase JM, Rödel SG, Luursema JM, Kommers PA, Verwey WB, van Houten FJ, Kunst EE. Open surgery in VR: inguinal hernia repair according to Lichtenstein. *Stud Health Technol Inform* 2006;119:477–9.
48. O'Toole RV, Playter RR, Krummel TM, Blank WC, Cornelius NH, Roberts WR, et al. Measuring and developing suturing technique with a virtual reality surgical simulator. *J Am Coll Surg* 1999;189(1): 114–127.
49. Gladstone HB, Raugi GJ, Berg D, Berkley J, Weghorst S, Ganter M. Virtual reality for dermatologic surgery: virtually a reality in the 21st century. *J Am Acad Dermatol* 2000;42(1 Pt 1):106–112.
50. Agha RA, Fowler AJ. The role and validity of surgical simulation. *Int Surg* 2015;100(2):350–357.
51. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, et al. Consensus guidelines for validation of virtual reality surgical simulators. *Simul Healthc* 2006;1(3):171–179.
52. Schaefer JJ 3rd, Vanderbilt AA, Cason CL, Bauman EB, Glavin RJ, Lee FW, et al. Literature review: instructional design and pedagogy science in healthcare simulation. *Simul Healthc* 2011;(Suppl 6):S30–S41.