Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

Title Page:

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

Mr Mark Higgins[1]

Dr Rakesh Patel[2]

Dr Christopher Madan[3]

1. University of Nottingham – mark.higgins2@nhs.net (corresponding author)
2. University of Nottingham – rakesh.patel@nottingham.ac.uk
3. University of Nottingham – christoper.madan@nottingham.ac.uk

Word Count: 5,528

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

ABSTRACT

Context

Changes to surgical training programmes in the UK has led to a reduction in theatre time for trainees, and an increasing reliance on simulation to provide procedural experience. Whilst simulation offers opportunity for repetitive practice, the effectiveness of simulation as an educational intervention for developing procedural surgical skills is unclear.

Methods

A systematic literature review was undertaken to retrieve all studies describing simulation-based medical education (SBME) interventions for the development of procedural surgical skills using the MEDLINE, PsycINFO, CINAHL, EMBASE and PUBMED databases. Studies measuring skill retention or demonstrating transferability of skills for improving patient outcomes were included in the review.

Results:

SBME is superior to no training and can lead to improvement in procedural surgical skills, such that skills transfer from simulated environments into theatre. SBME results in minimal skill degradation after 2 weeks, although more significant decay results after >90 days. Many studies recruited <10 participants, used a variety of methods and were restricted to endoscopic surgical techniques. All studies did not compare interventions with non-SBME teaching methods for developing procedural surgical skills. No studies compared the curriculum design of different surgical training programmes.

Conclusions

SBME interventions are effective for developing procedural skills in surgery. SBME interventions are also effective for preventing the decay of procedural surgical skills. Although no studies demonstrate non-inferiority of SBME interventions compared to time in theatre developing skills, SBME interventions do enable the transfer of skills into theatre, and the potential for improving patient outcomes.

Keywords: Simulation, Surgery, Surgical simulation, Simulation based medical education (SBME), surgical training, skills training

ACGME Competencies: Practice-based learning and Improvement; Medical Knowledge

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

INTRODUCTION:

The development of expert procedural skills in surgery has traditionally involved significant time spent in theatre watching procedures, alongside practice under the close supervision of experts, and receiving feedback for improvement. Whilst the training of expertise in most domains has popularly, but erroneously, been estimated to be around 10,000 hours [i], the time in surgery has been estimated to require at least double this amount of time.[ii] The particular training challenge in surgery requires both the development of cognitive ability, as well as manual dexterity.

The Halstedian model of surgical education emphasised the exposure of trainees across those training hours to intense and repetitive opportunities for managing patients under the supervision of a skill surgical trainer. However, the reduction of time in the workplace due to employment legislation at the start of 21st century saw a significant change in this delivery model for surgical education. Historical analysis of training log books demonstrated a 50% reduction in recorded events in some cases.[iii] There was also significant variation in the experience of trainees on arguably the most basic of procedures such as appendicectomy, with one study identifying 6 as the mean number of procedures performed whilst under senior supervision, and a range from 0 to 61 procedures attempted by trainees in basic surgical training programmes.[iv]

The changes in working pattern also impacted on the way trainees interact with trainers and their teams. More than a quarter of trainees in their early years of training regularly missed out on training opportunities to cover service commitments.[iv] Almost a half of trainees were unable to attend five or more elective operating sessions on an average week; the minimum level considered necessary for the development of competencies.[iv] The Royal College of Surgeons of England, against this backdrop, suggested simulation-based medical education (SBME) interventions should be a regular and frequent component of skills training.

The benefits of simulation across various systematic reviews of SBME interventions include greater opportunity for focus on specific tasks, rather than whole procedures; providing scope for repetition and feedback; greater control over the case mix for trainees and allowing trainees an environment in which to learn from errors, while protecting the patients from harm.[v] However, outcomes relevant to informing the design of training programmes such as transferability of skills, subsequent decay of skills and impact on patient outcomes remain poorly described. The aim of this research was to undertake a systematic literature review and investigate the effectiveness of SBME interventions for i) increasing transferability of skills acquired from simulated settings to theatre, ii) preventing long-term skills decay or, iii) improve patient-related outcomes.

METHODS:

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

This systematic review was performed using the guidelines set out by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[vi]

Search Methods:

A literature search was performed using the MEDLINE, PsycINFO, CINAHL, EMBASE and PUBMED databases. The search terms "simulation" AND "surgery" AND "validity" OR "transfer" OR "retention" OR "reliability" AND "novice" OR "resident" OR "basic" were used and studies restricted to those published in peer-reviewed journals and in the English language. Variations of terms, such as "surgical" instead of "surgery" and "simulat*" instead of "simulation" were attempted with no additional search results found. This strategy resulted in 1154 citations. which were screened by title to remove duplicates.

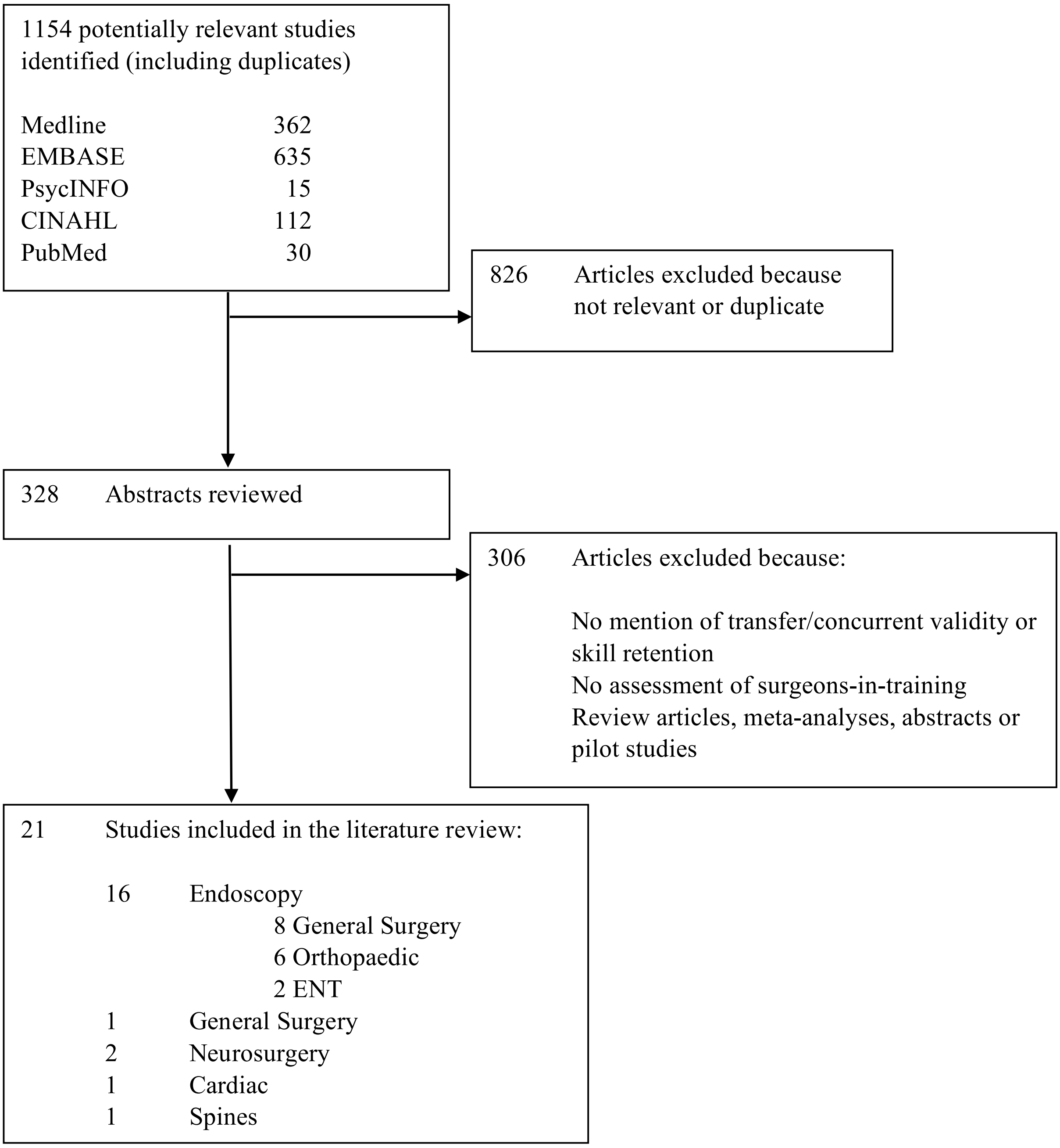
Eligibility criteria: Original research articles that contained a measure of skill retention over time or discussed the use of SBME as a training tool in terms of an impact on surgical skills or outcomes demonstrated in live human patients or animals or on cadavers were included. Articles relating to non-technical skills or non-surgical specialties were excluded, as were review articles and meta-analyses, but a manual search of the references contained within these articles was carried out to highlight additional literature. Pilot studies and published abstracts were discounted and those that focussed on validating a particular simulation model or looked solely at qualified surgeons and did not include surgeons-in-training (residents, trainees, students), were similarly omitted. This strategy resulted in 21 eligible articles. A PRISMA diagram (Figure 1) shows a summary of the process. A list of the studies is included in appendix 1.

Data Analysis

Results were classified into groups depending on whether reported outcomes were a measure of skill improvement or skill retention. A summary of each study can be seen in Table 1, including the demographics, intervention and outcomes, along with a Medical Education Research Study Quality Instrument (MERSQI) score.[vii]

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

Figure 1. PRISMA Diagram



Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

RESULTS:

21 articles met the inclusion criteria: 16 describing transfer of surgical skills from the simulation model to either a cadaveric specimen, live animal model or live patient in the operating room. The remaining 6 articles described skill improvement on a simulation and investigated how these skills decayed over time. Unless otherwise stated, studies did not report the duration or spacing of training, or the interval between training and assessment.

Transfer validity

Live human patients:

Cannon et al. performed a multicentre RCT reviewing competency-based training using a virtual reality (VR) knee arthroscopy simulator. 54 3rd year residents with equivalent baseline experience took part, of which 48 (89%) completed the programme. Training was unsupervised and consisted of 4 rounds on the simulator, requiring a pre-set proficiency to be reached on probing and visualisation tasks. Primary outcome was a global rating score (GRS) and a procedural checklist during a knee arthroscopy on a live patient. All assessments were carried out within 14 days of training. The simulator group achieved higher scores than the control.

Dunn et al. also performed an RCT involving 17 orthopaedic residents, comparing training on VR shoulder arthroscopy simulator to a control group who received no training. Each performed a live shoulder arthroscopy as a baseline. Training consisted of 4 x 15-minute supervised sessions on the simulator over a 90-day period, aiming for 12-15 repetitions per session. A further live shoulder arthroscopy was then performed. Assessments were scored using the ASSET (arthroscopy surgical skill evaluation tool) [viii] and safety score, time and 14-point checklist. Simulation training produced significant improvement in ASSET score and, but no difference in safety score or completion of the checklist.

A similar RCT by Waterman et al. reviewed the impact of training with VR shoulder arthroscopy simulation on surgical skills both on the simulator and on a real patient. 22 orthopaedic trainees performed baseline assessments on shoulder arthroscopy in the operating room and on the simulator. The intervention group were provided 4 x 15-minute supervised training sessions on VR shoulder arthroscopy simulator spaced across a 3-month period, completing a minimum of 50 cases. The control group received no simulator training. Baseline assessment was then repeated. Objective outcomes on the simulator were time and probe/camera distance. Outcomes for the live procedure were ASSET score, time and 14-point checklist. Significant improvements were seen in both groups, however simulator training led to faster completion times of both tasks and higher ASSET safety scores.

Fried et al. randomised 25 naïve trainees to investigate the effectiveness of VR endoscopic sinus surgery simulator training compared to traditional textbook and video-recording instruction. No baseline assessment was performed. Trainees were required to achieve a pre-set proficiency level on the simulator, though this target was not reported. The first endoscopic sinus surgery subsequently carried out by each participant on a patient was assessed based on key steps and scored for efficiency, respect of tissues, confidence, errors, instrument manipulation and task completion. Simulation training led to significant improvements in confidence, efficiency and error rates.

A competency-based approach was also used by Seymour et al. in the setting of laparoscopic gall bladder removal. 16 surgical trainees with equal baseline psychomotor skills were randomised to compare training on a VR laparoscopy simulator to traditional didactic and video instruction. Simulation training consisted of 1-hour supervised sessions directed towards object manipulation and the use of diathermy. Candidates needed to achieve a pre-set proficiency using both hands on consecutive attempts, which required between 3 and 8 sessions. A live gall bladder removal was then performed and assessed for duration and error rates, using a fixed-interval sampling (interval = 1 minute). No difference was found in time taken for completion, but significant reduction in errors, including damage to the gall bladder and non-target tissues.

Zevin et al. randomised 20 inexperienced surgical trainees to compare no training to simulator training using a bench-top model to replicate elements of a basic bariatric surgical procedure. Baseline assessment was performed on the simulator. Training consisted of supervised 1:1 sessions of up to 90 minutes with directed feedback, repeated until a pre-set proficiency was met. Final assessment was performed by all subjects on a live porcine specimen, however only the group who received simulation training subsequently went on to perform the same procedure under supervision on a human patient. Assessments were rated using Bariatric Objective Structured Assessment of Technical Skills (BOSATS), a validated measure of psychomotor skills. The group who experienced simulator training demonstrated significant improvement on the animal model with comparable performance seen subsequently on the live human patient.

Proficiency based simulation training was again employed by Palter et al. in a randomised comparison of bench top simulation or simple technical instruction in learning closure of the abdominal fascia. 18 of 19 junior surgical trainees completed the study (95%). Each received an introduction to the model with a single practice which acted as baseline assessment. Simulation training was provided during 1.5hr sessions with direct supervision, repeated up to 3 times, less than 3 weeks apart, until a set proficiency was reached. Final assessment occurred during supervised abdominal closure on a human patient. Alongside the procedure a novel but clinically relevant script was read out. Technical skills were assessed using a GRS and cognitive learning was assessed by recollection of the script. Simulation training led to significant improvements in cognitive and technical outcomes.

In contrast to supervised, proficiency-based training, Park et al. reviewed the impact of self-directed training with VR simulator on subsequent performance in colonoscopy on a real patient. 24 surgical trainees were randomised and the intervention group allowed access to the simulator for 2-3 hours unsupervised practice. The control group were given no access to the simulator. All participants first performed a simulated procedure to provide a baseline. Within 2 weeks of training each participant performed a supervised colonoscopy on a human patient assessed using a GRS to mark each of 7 key elements of the procedure on a 5-point Likert scale. Both groups demonstrated similar baseline scores, but the group who received simulation training achieved significantly higher outcomes in the final procedure on the patient.

In a non-randomised comparative study, De Oliviera et al. reviewed 3 groups of 3 neurosurgical residents from 3 centres trained using either the established technique of cadaveric dissection, a simulator model using human placentae or video instruction. Baseline ability of the participants was considered similar according to earlier assessment. Each received 2 hours of training per week for 6 consecutive weeks. Each resident then performed a supervised procedure on a human patient to treat a middle cerebral artery aneurysm. The procedure was subjectively broken down into 4 key steps and each assessed using a 5-point Likert scale. The simulation model was considered to be a superior method of training the critical phases of the procedure and led to greater scores in 3 of 4 steps, with equivalent scores in the 4th.

Human cadaveric specimens

Performing procedures on human cadavers has long been considered the gold standard of training, albeit expensive, limited by availability and lacking the vital elements, such as bleeding, healing and outcomes that define success in real surgery. Five studies reviewed the ability for surgical skills learned on the simulator to transfer to a procedure on cadaveric tissue.

Banaszek et al. compared a bench top simulator, a VR simulator and a control group in a randomised study of knee arthroscopy training in 40 medical students with no surgical experience. Each participant underwent baseline assessment on both simulator models after a short orientation to the procedure and simulators. Training was in total 6-8 hours of unsupervised repetitions on the chosen model spaced over 5-weeks. Each student then performed a final knee arthroscopy assessment on both simulator models, as well as human cadaver, and then were asked to perform a surprise task: a medial meniscectomy on the cadaver. Outcome measures were the GRS, 14-point checklist, time and motion analysis of instrument movements provided by the VR simulator. Outcomes were significantly better for the 2 groups receiving simulator training, with VR simulator outperforming the bench top model.

Camp et al. compared training on a VR knee arthroscopy simulator to a cadaveric model, with a non-training control. 57 orthopaedic residents were randomised and 45 (79%) completed the study. Each performed a baseline diagnostic knee arthroscopy on a human cadaver. Training consisted of 4 hours on the assigned simulator, although it was not clear whether this was across one or multiple sessions. VR training was unsupervised with digital feedback given by the simulator. Cadaveric training was supervised with basic feedback provided. A further diagnostic knee arthroscopy on a human cadaver was then performed. Time between baseline and final assessments was 128 days for the VR group, 59 days for the cadaver group and 69 days for the control. Procedures were assessed using ASSET score and time. Both intervention groups achieved significant improvement in outcomes compared to the control, however, improvement was greatest in cadaveric group.

Instead of non-training controls, Rebolledo et al. compared VR simulation training to didactic lectures and instruction using models in the setting of knee and shoulder arthroscopy. 14 orthopaedic residents were randomised. No baseline assessment was performed, but trainees had similar background experience. VR training lasted 2.5 hours and consisted of a basic probe navigation and manipulation programme. This was unsupervised with basic digital feedback provided by the simulator. The duration of the didactic training programme was 2 hours. Training was delivered in a single session. Final assessment consisted of a shoulder and knee arthroscopy on a cadaveric specimen under supervision. Outcome measures included and a subjective “injury grading score” (IGI), designed to be a measure of potential intra-articular injury. VR training led to significantly improved outcomes in shoulder arthroscopy, but no difference seen in knee arthroscopy.

Shi et al. similarly compared training on a VR simulation model to didactic training for pedicle screw insertion, a spinal procedure in which 3-dimensional understanding of screw insertion is essential to prevent neurological injury. 10 orthopaedic trainees were randomised to VR simulator training or didactic teaching with video demonstration. There was no baseline assessment, although none of the participants had relevant prior experience. VR training was provided during a single 30-minute session. Didactic training was provided for 40 minutes with a 10-minute video demonstration. 40 pedicle screws were then inserted by each group in human cadavers. Final assessment of screw position was made, based on degree of penetration and % of screws that were considered acceptable. VR training led to significantly more accurate screw placement with 100% considered acceptable.

Zhao et al. also compared training on a VR simulation model to didactic training in temporal bone dissection. 20 medical students were randomised. No baseline assessment was performed, no participant had relevant prior experience. Simulator training consisted of a 2-hour self-directed curriculum including instructional videos with repetitions on the VR simulator in a single session. Real-time feedback was provided by the simulator with audible alarms to highlight critical mistakes. Didactic training consisted of 2-hour small group teaching using instructional videos and models. Directly after training, each participant was given 1-hour to complete a mastoidectomy on a human cadaver. Outcome measures included a validated overall rating score on a 10-point Likert scale, an end-product score, an injury score and a technique score. VR simulator training performed significantly better in all 4 areas.

Live animal models

In the UK, surgical training on live animal models is strictly controlled by law, but although its use is becoming increasingly rare around the world two studies did assess the transfer of skills gained on a simulator model to a live animal model.

Izawa et al. performed a comparative study of 31 surgical residents in Japan. No baseline assessment was performed, but demographics and prior experience were similar. Initial quasi-randomisation was performed, but alterations were later made due to scheduling. Participants underwent a single training session in the control of bleeding and wound repair following penetrating cardiac trauma. One group trained on a live porcine model and one on an ex-vivo simulation model using cadaveric porcine hearts. Training was supervised and feedback provided. Final assessment was performed 1 week after training. Each participant carried out repair of a standardised penetrating cardiac wound in a live porcine model assessed using the previously validated Objectively Structured Assessment of Technical Skill (OSATS). While each resident was able to achieve satisfactory haemostasis, training on the simulation model led to significantly better outcome scores.

A second study by Stefanidis et al. in USA reviewed the transfer of skills from a bench-top simulator to a live porcine model and the impact that training to automaticity can have on skill transfer. 30 medical students were randomised. Baseline assessment of suturing skills was performed on the simulator. The intervention group received repeated sessions on the simulator of up to 1-hour duration, on different days, allowing repetition with supervision and feedback. Once a pre-set proficiency was reached, participants performed a laparoscopic suturing procedure on a live porcine model. Training then continued with the introduction of a concurrent secondary visuospatial task until proficiency was reached for both tasks. The suturing assessment was then repeated. The control participated in all testing sessions but received no training. Final assessments were carried out within 2 weeks of final training, or else a refresher session was provided. Outcome measures were the Global Objective Assessment of Laparoscopic skill (GOAL) rating scale and any inadvertent injuries to local structures were noted. Simulator training led to better performance compared to baseline and control in both tests, but a higher suturing component of the GOAL and fewer inadvertent injuries in the second test.

Skill retention

Six articles made comment on the retention of skills following a period of simulation training. Dunn et al. studied the impact of VR simulation training on diagnostic shoulder arthroscopy in an RCT of 17 orthopaedic trainees naïve to arthroscopy. Baseline assessment was carried out. 4 x 15-minute training sessions were spaced across 90 days aiming to complete 50 repetitions in total. The control group participated only in assessment stages. Final assessment was made during a further live shoulder arthroscopy. Assessment was repeated after 1 year, during which traditional training continued, but no comment was made on interval surgical experience. Outcomes were scored using ASSET (arthroscopy surgical skill evaluation tool) and safety score, time and a 14-point checklist. While there was a significant immediate positive effect on outcomes, this was lost when retested at 1 year.

Uribe et al. investigated the learning curve for endoscopic sinus surgery using a VR simulation model with benchmark performance set by a group of experts. 26 medical students were enrolled in this observational study but only 5 (19%) completed all stages. Candidates were assessed across a set number of increasingly challenging trials on the simulator. Assessments used a computer-generated score based on time, accuracy and error rate. Due to scheduling conflicts, a number of trainees had unexpected gaps in their training of between 14 and 90 days. 100% of gaps of up to 60 days resulted in no drop in performance, whilst a gap of 90 days did lead to a drop in performance that was recovered by a single further trial on the simulator.

Varley et al. designed their RCT to investigate retention of skills learned on laparoscopic box trainer. 30 medical students with no prior experience of the procedure were enrolled, of which 28 (93%) completed the programme. Training was provided in 2 fundamental tasks (precision cutting and peg transfer) over a single session. Candidates repeated the two tasks at least 5 times each until a pre-set proficiency level was achieved on consecutive attempts. One group repeated the tasks after 4 weeks and the other after 12 weeks. No interval practice or related clinical procedures were permitted. Tasks were assessed by time taken and errors made. 90% of trainees were able to reach proficiency and these skills were retained after 4 weeks. However, there was a clear loss of skills at 12 weeks.

Similarly, Bonrath et al. reviewed the retention of basic laparoscopic skills learned on a bench top simulator. 36 medical students were randomised into two groups and performed 2 repetitions of each task to provide baseline sores. Training was provided over 2 days. Pairs of students carried out 4 repetitions of each task under supervision and immediate feedback was provided. Tasks represented key psychomotor skills, e.g., navigation, cutting etc. Immediately after training, each task was repeated and assessed by time and error rate. Further assessment of group 1 was made after 6 weeks and of group 2 after 11 weeks. No interim exposure to the simulator or similar procedures was permitted. Training produced significant immediate improvements in outcomes. At 6 weeks skill was maintained in all tasks except knot tying, but after 11 weeks there was significant deterioration in scores for the subjectively more challenging tasks.

Two studies looked at short term skills retention in the setting of laparoscopic simulation. Linsk et al. aimed to validate a VR simulation model that replicated an existing physical simulation model. 30

medical students were recruited and 24 (80%) completed the study. After an introductory video participants performed baseline assessments once on each simulator. Supervised training was provided on the chosen simulator model for 30 minutes on 15 consecutive weekdays. Control group only completed the testing phases. Assessments of a basic laparoscopic cutting task were made at the end of training and repeated 2 weeks later. Outcome measures included time, error rate and procedure score. Both training groups improved significantly from baseline and skills were maintained at 2 weeks.

Kolozsvari et al. assessed retention at 1-month and investigated the impact of over-learning to an expert standard compared to training to proficiency. 99 surgical trainees were enrolled and 74 (75%) completed all stages. Baseline assessment was carried out on the simulator for a simple task (peg-transfer) and more complex task (suturing). The simple task was repeated on the simulator until proficiency standards (group 1) or expert standards (group 2) were reached. Trainees then proceeded to the more complex task, which was repeated until proficiency or a maximum of 80 repetitions was reached. An assessment of each task was made 1 month later without any interval practice. Outcome measures were time and errors made. Over-training for the simple task led to a higher starting score and faster learning for the complex task. After 1 month, outcomes remained better than baseline, however there was no beneficial effect of over-training.

DISCUSSION:

This research demonstrates the general effectiveness of simulation-based medical education (SBME) interventions for developing procedural skills which transfer from a simulated educational setting to a surgical theatre context, but the weight of evidence is lacking due to the few number of studies conducted so far, and the low number of participants involved in those studies. The findings from the review demonstrate the definition of ‘simulation’ or ‘simulator’ within the catch all umbrella term ‘SBME intervention’ is variable across all studies, making comparison across the research difficult. Likewise, the review also demonstrates the validity and reliability of assessments evaluating the transferability of skills within and across studies was lacking. Finally, the findings from this research confirm SBME interventions are generally effective for sustaining skills development up to 90 days, however there is evidence from a few studies of a significant and steady decay of skills beyond 90 days.

Even though SBME interventions are commonplace across many surgical education programmes involving many thousands of trainees in some cases, the empirical basis for the widespread use of SBME appears to be drawn from comparative studies which recruited few participants. Whilst there is little role for arguing one way or the other in the case for using SBME as part of the surgical training programme, this research suggests there is need to better understand and develop consensus around the place for SBME in the development of expert surgeons. The case for SBME appears strong in the development of coordination, manipulation and dexterity skills, especially in the context of minimally invasive surgery.[ix] In the US, trainees are now mandated to attend a national programme for the development of basic cognitive and laparoscopic procedural skills, with encouragement for orthopaedic trainees to attend a basic arthroscopy skills course as well. Whilst there are various opportunities for trainees in the UK, there is no nationally mandated programme for skills development in comparison. This review suggests there is a need for much more high-quality outcomes-based research in order to inform the supervision of trainees across the various surgical sub-specialties, as well as policy around the design of national training programmes.

There is now a real need to both achieve consistency and precision around definitions for ‘SBME’ and ‘basic procedural skills’ and raise greater awareness around what these terms actually mean in the development of an expert surgeon. SBME in some respects can be reduced to meaning repeated training on a part-simulator in a laboratory, however in other examples, SBME may involve whole procedures and immersive environments. In medicine SBME can involve lower fidelity training experiences in comparison, yet improved outcomes as seen in Advanced Life Support training.[x] Furthermore, in healthcare more broadly, SBME has extended beyond an educational intervention and can now be viewed as more of a translational science, with evidence of increased workforce and healthcare system level benefits as well as improved patient outcomes.[xi] In order to achieve these changes at the ‘macro-level’, there has also been an implicit realisation along the way in these sub-specialty contexts, that the term ‘basic’ is relative to the competence of the learner, and notions of ‘basic’ should not be confused with ‘simplistic’ given their fundamental role in long-term expertise development.

Basic overhead throwing exercises in sporting contexts lead to “down-stream” improvements in related, but more advanced but similar techniques such as javelin throw and the overhead clear in badminton.[xii] By comparison, psychomotor skills training on a simulator for shoulder arthroscopy is known to improve performance of a subsequent procedural task performed for the first time on a human patient, judged by objective assessment. [xiii] However educationally within sport, there also appears to be equal emphasis given on supporting an individual’s overall development in that domain and ensuring skills are ‘coming together’ when undertaking performance on a given task. Within the education of healthcare professions, the concern remains that there is a greater focus on ensuring trainees have ‘ticked off their competencies’ when making progression decisions, rather than making judgments about expertise development based on whether all these skills are coming together at the right time.[xiv]

Whilst there are a number of reported simulation-related or ‘simulator’ instruments that demonstrate validity or reliability for measuring expertise development,[viii,ix] there is likely to be no one SBME approach that is ‘valid’ or ‘reliable’ in all learner circumstances, given that evidence about performance is drawn from multiple sources when making such a judgment about expertise or competence in practice. For example, the way trainees develop ‘attention skills’ is also critical for demonstrating safe independent practice.[xv] Assessment of visual gaze patterns demonstrate differences between surgeons of varying levels of expertise, with experts exhibiting greater focus on relevant anatomical targets and more rapid visual search patterns alongside exhibiting ‘competence’ or ‘proficiency’ on the task.[xvi,xvii] The actual challenge for surgical education is identifying ways of bringing in the subjective, but cumulatively extremely reliable, judgments made by educators, alongside assessments made using SBME interventions, in a more valid and reliable way.

Finally, this research confirms the anecdotal observation that “unless you use it, you lose it” in terms of the acquisition and retention of procedural surgical skills.[xv] A number of other studies in a medical and healthcare professions context also seem to identify 90 days as a threshold moment for when there is a rapid decay in skills. Furthermore, the phenomenon of decay may not be associated with pure psychomotor or procedural skills, but also cognitive skills as well.[xviii] The reasons for decay are likely related to the way in which skills were acquired in the first place. Traditionally, the origins of much surgical procedural skills training are rooted in behaviourism, where the focus is on repetitive practice and overlearning in order to make unconscious the more routine aspects of some skill. [xix] Whilst this approach is known to improve short-term learning outcomes across cognitive and procedural skills domains,[xx] long term outcomes is poor and in some cases worse in comparison to more evidence-based strategies such as distributed practice,[xxi,xxii] retrieval practice, [xxiii] or deliberate practice[xxiv] for knowledge and skills development. Given there are advantages of overlearning in particular situations,[xxv] there is now perhaps a growing need to re-evaluate the implications of all these findings for curriculum design in surgical education. Given technologies such as augmented and virtual reality also have growing evidence for improved learning outcomes,[xxvi] the need to identify and define effectiveness in terms of relevance for surgical expertise development is even more important due to the associated costs.[xxvii]

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

Table 1. Summary of studies examined.
Studies are listed in the same order they are discussed in the main text.
MERQI = Medical Education Research Study Quality Instrument; De=Design, S=Sampling, Da=Data; SV=Score Validity; O=Outcomes.
GRS = global rating score; ASSET = Arthroscopic Surgery Skill Evaluation Tool.

			MERSQI										
Study	**Participants**	**Design**	**Total**	De	S	Da	SV	A	O	**Task**	**Intervention**	**Outcomes**	**Evidence**
Transfer studies													
Cannon et al. (2014)	48/54 PGY3 surgical trainees (89%)	RCT	17.0	3.0	3.0	3	3	3	2.0	Post-training diagnostic knee arthroscopy performed on live patient	Group 1: VR sim. training; 4 rounds of progressive difficulty with pre-set competency to achieve on final round Group 2: Traditional training in their institution	21-item procedural checklist, Proprietary GRS	Mean of 11 hrs to achieve competence on simulator VR sim. outperformed the control in both outcome measures
Dunn et al. (2015) [transfer]	17 post-graduate orthopaedic surgery residents	RCT	16.0	3.0	2.0	3	3	3	2.0	Pre- and post-training live diagnostic shoulder arthroscopy performed under supervision	Group 1: 4 x 15 min supervised (1:1) training sessions over 90 day period, aiming to achieve 50+ repetitions Group 2: No additional training	ASSET global rating scale, ASSET safety score, Time to completion, 14-point anatomy checklist	Group 1 had improvement in mean ASSET scores and ASSET safety score
Waterman et al. (2016)	22 orthopaedic trainees	RCT	16.0	3.0	2.0	3	3	3	2.0	Task 1: Navigation and manipulation task on VR shoulder arthroscopy sim. Task 2: Diagnostic shoulder arthroscopy on a live human patient in operating room	Group 1: 4 x 15-min sessions on sim across 3-month period Group 2: No additional training	Task 1: Time to completion and distance travelled by instruments, Task 2: ASSET score, 14-point checklist, time to completion	Group 1 had faster completion times of both tasks, more efficient movements on sim, higher ASSET score
Fried et al. (2010)	25/28 PGY 1/2 residents from 4 centres (89%)	RCT	17.0	3.0	3.0	3	3	3	2.0	Assessment of the basic components of the first in-vivo ESS procedure performed by each participant following training	Group 1: Trained to “proficiency” using ES3 VR sim. Group 2: No further training	Case difficulty, Manipulation, Tissue respect, Task completion, Confidence, Number of errors	Group 1 had improvements in Injection time, Dissection time, Injection errors, Surgical confidence, Instrument manipulation, Navigation errors

Seymour et al. (2002)	16 PGY1-4 surgical residents	RCT	16.0	3.0	2.0	3	3	3	2.0	Live lap chole in operating room on human patient under supervision	Group 1: Supervised training on “manipulate and diathermy” task on VR sim. for 1 hr sessions; repeated until pre-set competency achieved (3-8 sessions). Group 2: No sim. training	Time to complete procedure, Errors recorded using fixed-interval time span sampling (an error event irrespective of how many in 1-min period)	Group 1 had fewer errors, were less likely to injure non target tissue, were more likely to make steady progress
Zevin et al. (2017)	20/26 PGY3/4 surgical residents (77%)	RCT	15.5	3.0	1.5	3	3	3	2.0	Laparoscopic Roux-limb and jejunojejunostomy Task 1 (pre-intervention): Box trainer, cadaveric porcine specimen Task 2 (post intervention): Live porcine specimen in OR Task 3 (post-intervention): Live human patient in OR	Group 1: Interactive seminars and trained on benchtop sim. until proficient Group 2: Conventional surgical residency training	Bariatric Objective Structured Assessment of Technical Skills (BOSATS)	Group 1 had improvement in BOSATS that transferred to procedure on human patient
Palter et al. (2011)	18/19 PGY1 general surgical/OBGYN residents (95%)	RCT	16.0	3.0	2.0	3	3	3	2.0	Abdominal fascial closure on live human patient under supervision During procedure a script was read out that contained novel information but relevant to the task	Group 1: 2 training sessions on bench top sim of 1.5hrs duration spaced <3 weeks apart until proficient Group 2: No further training	OSATS GRS, MCQ to assess recollection of script	Sim. training led to improvements in both technical skills and cognitive learning in a clinical setting
Park et al. (2007)	24/28 PGY1-3 residents (86%)	RCT	16.0	3.0	2.0	3	3	3	2.0	Colonoscopy on a live human patient under supervision within 2 weeks of baseline testing	Group 1: 2-3 hours unsupervised sim. practice Group 2: No sim. training	Pre-test colonoscopy on VR was assessed using GRS and computer-generated performance score, Human colonoscopy was assessed using GRS (7 key items)	Group 1 had improvements in GRS

de Oliveira et al. (2018)	9 neurosurgery residents	Quasi-experimental comparitive	15.5	2.0	2.5	3	3	3	2.0	Post-training live open surgery to treat unruptured MCA aneurysm <12mm	Group 1: Training on human cadaver Group 2: Training on placental-model simulator Group 3: Video of training tasks	Sylvian fissure dissection, Use of bipolar cautery, Aneurysm dissection, Aneurysm clipping (all are 5-point Likert scales)	Placental model considered superior in teaching critical phases of surgery with improved or equivalent outcomes in all tasks
Banaszek et al. (2017)	40 medical students	RCT	15.5	3.0	2.0	3	3	3	1.5	Diagnostic knee arthroscopy on human cadaver	Group 1 & 2: Unsupervised training on simulator (VR vs. bench-top) for 6-8 hours over 5 weeks Group 3: No additional training	GRS subjectively collected by expert, 14-point checklist, Procedural Time	Significant improvement in outcomes for both intervention groups VR was superior to bench-top
Camp et al. (2016)	45/57 surgical trainees (79%)	RCT	15.5	3.0	2.0	3	3	3	1.5	Diagnostic knee arthroscopy on cadaveric specimen	Group 1: 4 hours of VR sim training with digital feedback Group 2: 4 hours cadaveric training with basic instruction from senior Group 3: No additional training	ASSET global rating scale	Both training groups outperformed the control, Cadaveric training outperformed the sim. training
Rebolledo et al. (2015)	14 PGY1/2 orthopaedic residents	RCT	15.5	3.0	2.0	3	3	3	1.5	Individual shoulder and knee arthroscopy on a cadaveric specimen	Group 1: 2.5-hours training on VR sim focussed on basic tasks and anatomical landmarks for knee and shoulder arthroscopy Group 2: 2-hours didactic lectures on basic arthroscopy including instruments and models	Time to complete a standardised checklist, "Injury grading index" (IGI) calculated by the senior investigators, NB/ IGI subjective measure of potential intra-articular injury	Group 1 had significantly faster time to checklist completion and improved IGI scores compared to Group 2 for shoulder arthroscopy, The results did not reach significance for knee arthroscopy

Shi et al. (2018)	10 surgical residents	RCT	15.5	3.0	2.0	3	3	3	1.5	40 pedicle screws inserted by each group in human cadaver specimens	Group 1: 30-min VR sim. training of pedicle screw insertion Group 2: 40-min didactic teaching using spine model and 10-min video demo	Assessment of accuracy of pedicle screw placement	Group 1 had 100% acceptable screw position, Group 2 had 85% acceptable screw position, Proportion of screws within pedicle significantly greater in Group 1
Zhao et al. (2011)	20 final year medical students with no previous experience of procedure	RCT	15.0	3.0	1.5	3	3	3	1.5	Post-intervention cortical mastoidectomy on a cadaveric specimen	Group 1: 2-hour self-directed training, including instructional videos & repetition on VR sim. Group 2: 2-hour small-group teaching, including temporal bone models & operative videos	Overall rating on a 10-point Likert scale, End-product score, Injury score, Technique score	Group 1 had higher scores in overall rating, end-product score & technique score, with lower injury score
Izawa et al. (2016)	31 Staff surgeons (14) and residents (17)	Comparative	14.5	2.0	2.0	3	3	3	1.5	Performed procedure on live animal model 1 week after training under observation of 4 blinded assessors	Group 1: Supervised training with live animal model Group 2: Supervised training with ex-vivo simulation model	OSATS which included: GRS, 4-point checklist, Self-reported post-training questionnaire	All participants were able to achieve competence, Ex-vivo model was as effective as the live-animal model as a training tool
Stefanidis et al. (2012)	12/30 medical students (40%)	RCT	14.5	3.0	1.0	3	3	3	1.5	Laparoscopic procedure on live porcine specimen after trainings	Group 1: Trained on FLS suturing model until expert competency achieved, then trained on secondary visual-spatial processing task introduced & training repeated until expert competency achieved on both tasks Group 2: No additional training	Global objective assessment of laparoscopic skill (GOAL) global rating scale Inadvertent injuries to local structures noted	Group 1 had improved scores after second training with fewer inadvertent injuries
Retention studies													

Dunn et al. (2015) [retention]	17 post-graduate orthopaedic surgery residents	RCT	16.0	3.0	2.0	3	3	3	2.0	Pre- and post-training live diagnostic shoulder arthroscopy performed under supervision Assessment repeated after 1 year	Group 1: 4 x 15 min supervised (1:1) training sessions over 90 day period, aiming to achieve 50+ repetitions Group 2: No additional training	ASSET global rating scale, ASSET safety score, Time to completion, 14-point anatomy checklist	Group 1 had improvements in mean ASSET scores and time to completion, but lost on retesting after 1 year
Uribe et al. (2004)	5/26 medical students (19%)	Quasi-experimental comparative	12.5	1.0	1.0	3	3	3	1.5	Novice: abstract environment, no haptics Intermediate: Realistic anatomy (labelled), haptics & simulated bleeding Advanced: pathology & no teaching aids	>1-hour/week supervised training on sim to complete programme of 10 novice, 10 intermediate and 3 advanced trials 13 students experienced a gap between sessions of 14-90 days	Computer generated score based on time taken, accuracy of tasks and errors performed, eg inadvertent injury of local anatomy	100% of students experiencing gap < 60 days had no drop in performance, Gap of 90 days led to drop in performance that was recovered after the first subsequent repetition
Varley et al. (2015)	28/30 medical students and interns (93%)	RCT	15.5	3.0	2.0	3	3	3	1.5	FLS box trainer used (previously validated) Task 1: Peg transfer Task 2: Precision cutting	½ day training on benchtop sim and immediate assessment of all participants Group 1: Repeated testing after 4 weeks Group 2: Repeated testing after 12 weeks	Time taken to complete the procedure, Number of attempts required to reach the proficiency recorded	27/30 participants reached proficiency at end of training, Significant improvements retained at 4 weeks, but not at 12 weeks
Bonrath et al. (2012)	36 medical students	Quasi-experimental pre-/post-test design	12.0	1.5	2.0	3	1	3	1.5	5-day curriculum of complex navigation, manipulation, and cutting tasks on bench-top sim.	Group 1: Repeat testing 6 weeks after training Group 2: Repeat testing 11 weeks after training	Quantitative assessment = time taken, Qualitative assessment = error score	Group 1 had no significant skill deterioration in all but knot tying, Group 2 had significant skill deterioration seen for the "more difficult" tasks

Linsk et al. (2018)	24/30 medical students (80%)	RCT	15.5	3.0	2.0	3	3	3	1.5	Pattern cutting test: a pre-marked circle was cut from a 4x4cm gauze Performed on both physical and VR sim., test repeated after 2 weeks	Group 1: 10+ repetitions of the task on physical sim over 30 min, repeated on 15 consecutive weekdays Group 2: 10+ repetitions of the task on VR sim over 30 min, repeated on 15 consecutive weekdays. Group 3: No further training	Time to complete the procedure, Number of errors, Procedure score	Groups 1 & 2 significantly outperformed Group 3 in post-test, No significant difference between efficacy of physical and virtual training, Score improvements retained at 2 weeks
Kolozsvari et al. (2011)	77/99 surgical trainees (78%)	RCT	15.5	3.0	2.0	3	3	3	1.5	Basic task: Transfer of peg within benchtop simulator Complex Task: Intra-corporeal suturing All participants re-tested on both tasks after a 1 month interval	Group 1: Basic task repeated until proficient and then proceeded to testing of complex task Group 2: Basic task repeated until expert and then proceeded to testing of complex task Group 3: Proceeded to complex task with no further training	Proficiency: 65-s with no errors on 3 consecutive or 5 non-consecutive trials during 2 separate sessions Expert: 48-s with no errors on 3 consecutive or 5 non-consecutive trials during 2 separate sessions	Retention testing demonstrated slight decrease in scores after 1 month but remained higher than baseline for both groups, No benefit found to over-training in skill retention

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

APPENDIX

Studies included in the literature review:

Banaszek, D., You, D., Chang, J., Pickell, M., Hesse, D., Hopman, W.M., Borschneck, D., Bardana, D., 2017. Virtual Reality Compared with Bench-Top Simulation in the Acquisition of Arthroscopic Skill: A Randomized Controlled Trial. J Bone Joint Surg Am 99, e34. https://doi.org/10.2106/JBJS.16.00324

Bonrath, E.M., Weber, B.K., Fritz, M., Mees, S.T., Wolters, H.H., Senninger, N., Rijcken, E., 2012. Laparoscopic simulation training: Testing for skill acquisition and retention. Surgery 152, 12–20. https://doi.org/10.1016/j.surg.2011.12.036

Camp, C.L., Krych, A.J., Stuart, M.J., Regnier, T.D., Mills, K.M., Turner, N.S., 2016. Improving Resident Performance in Knee Arthroscopy: A Prospective Value Assessment of Simulators and Cadaveric Skills Laboratories. J Bone Joint Surg Am 98, 220–225. https://doi.org/10.2106/JBJS.O.00440

Cannon, W.D., Garrett, W.E., Hunter, R.E., Sweeney, H.J., Eckhoff, D.G., Nicandri, G.T., Hutchinson, M.R., Johnson, D.D., Bisson, L.J., Bedi, A., Hill, J.A., Koh, J.L., Reinig, K.D., 2014. Improving residency training in arthroscopic knee surgery with use of a virtual-reality simulator. A randomized blinded study. J Bone Joint Surg Am 96, 1798–1806. https://doi.org/10.2106/JBJS.N.00058

de Oliveira, M.M.R., Ferrarez, C.E., Ramos, T.M., Malheiros, J.A., Nicolato, A., Machado, C.J., Ferreira, M.T., de Oliveira, F.B., de Sousa, C.F.P.M., Costa, P.H.V., Gusmao, S., Lanzino, G., Maestro, R.D., 2018. Learning brain aneurysm microsurgical skills in a human placenta model: predictive validity. J. Neurosurg. 128, 846–852. https://doi.org/10.3171/2016.10.JNS162083

Dunn, J.C., Belmont, P.J., Lanzi, J., Martin, K., Bader, J., Owens, B., Waterman, B.R., 2015. Arthroscopic Shoulder Surgical Simulation Training Curriculum: Transfer Reliability and Maintenance of Skill Over Time. J Surg Educ 72, 1118–1123. https://doi.org/10.1016/j.jsurg.2015.06.021

Fried, M.P., Sadoughi, B., Gibber, M.J., Jacobs, J.B., Lebowitz, R.A., Ross, D.A., Bent, J.P., Parikh, S.R., Sasaki, C.T., Schaefer, S.D., 2010. From virtual reality to the operating room: the endoscopic sinus surgery simulator experiment. Otolaryngol Head Neck Surg 142, 202–207. https://doi.org/10.1016/j.otohns.2009.11.023

Gallagher, A.G., Seymour, N.E., Jordan-Black, J.-A., Bunting, B.P., McGlade, K., Satava, R.M., 2013. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition. Ann. Surg. 257, 1025–1031. https://doi.org/10.1097/SLA.0b013e318284f658

Izawa, Y., Hishikawa, S., Muronoi, T., Yamashita, K., Maruyama, H., Suzukawa, M., Lefor, A.K., 2016. Ex-vivo and live animal models are equally effective training for the management of a penetrating cardiac injury. World J Emerg Surg 11, 45. https://doi.org/10.1186/s13017-016-0104-3

Kolozsvari, N.O., Kaneva, P., Brace, C., Chartrand, G., Vaillancourt, M., Cao, J., Banaszek, D., Demyttenaere, S., Vassiliou, M.C., Fried, G.M., Feldman, L.S., 2011. Mastery versus the standard proficiency target for basic laparoscopic skill training: effect on skill transfer and retention. Surg Endosc 25, 2063–2070. https://doi.org/10.1007/s00464-011-1743-9

Linsk, A.M., Monden, K.R., Sankaranarayanan, G., Ahn, W., Jones, D.B., De, S., Schwaitzberg, S.D., Cao, C.G.L., 2018. Validation of the VBLaST pattern cutting task: a learning curve study. Surg Endosc 32, 1990–2002. https://doi.org/10.1007/s00464-017-5895-0

Palter, V.N., Grantcharov, T., Harvey, A., Macrae, H.M., 2011. Ex vivo technical skills training transfers to the operating room and enhances cognitive learning: a randomized controlled trial. Ann. Surg. 253, 886–889. https://doi.org/10.1097/SLA.0b013e31821263ec

Park, J., MacRae, H., Musselman, L.J., Rossos, P., Hamstra, S.J., Wolman, S., Reznick, R.K., 2007. Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am. J. Surg. 194, 205–211. https://doi.org/10.1016/j.amjsurg.2006.11.032

Rebolledo, B.J., Hammann-Scala, J., Leali, A., Ranawat, A.S., 2015. Arthroscopy skills development with a surgical simulator: a comparative study in orthopaedic surgery residents. Am J Sports Med 43, 1526–1529. https://doi.org/10.1177/0363546515574064

Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Brien, M.K., Bansal, V.K., Andersen, D.K., Satava, R.M., 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann. Surg. 236, 458–463; discussion 463-464. https://doi.org/10.1097/01.SLA.0000028969.51489.B4

Shi, J., Hou, Y., Lin, Y., Chen, H., Yuan, W., 2018. Role of Visuohaptic Surgical Training Simulator in Resident Education of Orthopedic Surgery. World Neurosurg 111, e98–e104. https://doi.org/10.1016/j.wneu.2017.12.015

Stefanidis, D., Scerbo, M.W., Montero, P.N., Acker, C.E., Smith, W.D., 2012. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. Ann. Surg. 255, 30–37. https://doi.org/10.1097/SLA.0b013e318220ef31

Uribe, J.I., Ralph, W.M., Glaser, A.Y., Fried, M.P., 2004. Learning curves, acquisition, and retention of skills trained with the endoscopic sinus surgery simulator. Am J Rhinol 18, 87–92.

Varley, M., Choi, R., Kuan, K., Bhardwaj, N., Trochsler, M., Maddern, G., Hewett, P., Mees, S.T., 2015. Prospective randomized assessment of acquisition and retention of SILS skills after simulation training. Surg Endosc 29, 113–118. https://doi.org/10.1007/s00464-014-3647-y

Waterman, B.R., Martin, K.D., Cameron, K.L., Owens, B.D., Belmont, P.J., 2016. Simulation Training Improves Surgical Proficiency and Safety During Diagnostic Shoulder Arthroscopy Performed by Residents. Orthopedics 39, e479-485. https://doi.org/10.3928/01477447-20160427-02

Zevin, B., Dedy, N.J., Bonrath, E.M., Grantcharov, T.P., 2017. Comprehensive simulation-enhanced training curriculum for an advanced minimally invasive procedure: a randomized controlled trial. Surg Obes Relat Dis 13, 815–824. https://doi.org/10.1016/j.soard.2016.11.019

Zhao, Y.C., Kennedy, G., Yukawa, K., Pyman, B., O'Leary, S., 2011. Improving temporal bone dissection using self-directed virtual reality simulation: results of a randomized blinded control trial. Otolaryngol Head Neck Surg 144, 357–364. https://doi.org/10.1177/0194599810391624

Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulated-based medical education interventions

REFERENCES:

---

[i] Hambrick DZ, Oswald FL, Altmann EM, Meinz EJ, Gobet F, Campitelli G. Deliberate practice: Is that all it takes to become an expert? *Intelligence*. 2014 Jul;45:34-45.

[ii] Jackson GP, Tarpley JL. How long does it take to train a surgeon? *BMJ* 2009;339:b4260

[iii] Greensmith M, Cho J, Hargest R. Changes in surgical training opportunities in Britain and South Africa. *International Journal of Surgery*. 2016 Jan 1;25:76–81.

[iv] Second Annual Report of the JCST trainee survey. https://www.jcst.org/quality-assurance/trainee-survey/ (Last accessed 31/12/2019)

[v] Aggarwal R, Darzi A. Technical-Skills Training in the 21st Century. *New England Journal of Medicine*. 2006 Dec 21;355(25):2695–6.

[vi] Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS ONE*. 2013;8(12):e83138.

[vii] Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA*. 2007 Sep 5;298(9):1002–9.

[viii] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, et al. The Arthroscopic Surgical Skill Evaluation Tool (ASSET). *Am J Sports Med*. 2013 Jun;41(6):1229–37.

[ix] Peters JH, Fried GM, Swanstrom L, Soper NJ, Sillín LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135(1):21–7

[x] Massoth C, Röder H, Ohlenburg H, Hessler M, Zarbock A, Pöpping DM, et al. High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. *BMC Med Educ*. 2019 Dec;19(1):1–8.

[xi] Brazil V. Translational simulation: not ‘where?’ but ‘why?’ A functional view of in situ simulation. *Adv Simul*. 2017 Dec;2(1):1–5.

[xii] O’keeffe SL, Harrison AJ, Smyth PJ. Transfer or specificity? An applied investigation into the relationship between fundamental overarm throwing and related sport skills. *Physical Education and Sport Pedagogy*. 2007 Jun 1;12(2):89–102.

[xiii] Dunn JC, Belmont PJ, Lanzi J, Martin K, Bader J, Owens B, et al. Arthroscopic Shoulder Surgical Simulation Training Curriculum: Transfer Reliability and Maintenance of Skill Over Time. *Journal of Surgical Education*. 2015 Nov 1;72(6):1118–23.

[xiv] Davis MH, Ponnamperuma GG. Education in surgery: competency-based training. *The Bulletin of the Royal College of Surgeons of England*. 2007 Nov;89(10):342-45.

[xv] Perez RS, Skinner A, Weyhrauch P, Niehaus J, Lathan C, Schwaitzberg SD, et al. Prevention of surgical skill decay. *Mil Med*. 2013 Oct;178(10 Suppl):76–86.

[xvi] Tien T, Pucher P, H Sodergren M, Sriskandarajah K, Yang G-Z, Darzi A. Eye tracking for skills assessment and training: A systematic review. *J Surg Res*. 2014 Sep;191(1):169-78.

[xvii] Eivazi S, Hafez A, Fuhl W, Afkari H, Kasneci E, Lehecka M, et al. Optimal eye movement strategies: a comparison of neurosurgeons gaze patterns when using a surgical microscope. *Acta Neurochir (Wien).* 2017;159(6):959–66.

[xviii] Giovanni DD, Roberts T, Norman G. Relative effectiveness of high- versus low- fidelity simulation in learning heart sounds. *Medical Education*. 2009 Jul 1;43(7):661–8.

[xix] Fitts PM, Posner MI. Human performance. Oxford, England: Brooks/Cole; 1967.

[xx] Rohrer D, Taylor K, Pashler H, Wixted J, Wiseheart M. The effect of overlearning on long- term retention. *Applied Cognitive Psychology*. 2005 Apr 1;19:361–74.

[xxi] Rohrer D, Taylor K. The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*. 2006 Dec 1;20:1209–24.

[xxii] Moulton C-AE, Dubrowski A, MacRae H, Graham B, Grober E, Reznick R. Teaching Surgical Skills: What Kind of Practice Makes Perfect?: A Randomized, Controlled Trial. *Annals of Surgery*. 2006 Sep;244(3):400-9.

[xxiii] Dobson J, Linderholm T, Perez J. Retrieval practice enhances the ability to evaluate complex physiology information. *Medical Education.* 2018;52(5):513–25.

[xxiv] Ericsson KA. An expert- performance perspective of research on medical expertise: the study of clinical performance. *Medical Education*. 2007 Dec 1;41(12):1124–30.

[xxv] Marcucci V, Greenawald L, Uribe JL, Shariff FU, Lind DS, Shewokis PA, et al. Overlearning enhances skill retention in a simulated model of laparoscopic cholecystectomy. *Journal of the American College of Surgeons*. 2015 Oct 1;221(4):e74.

[xxvi] Sheik-Ali S, Edgcombe H, Paton C. Next-generation Virtual and Augmented Reality in Surgical Education: A Narrative Review. *Surg Technol Int*. 2019 10;35:27–35.

[xxvii] Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, et al. Surgical Simulation: A Systematic Review. *Annals of Surgery*. 2006 Mar;243(3):291.