ASERNIP-S

Australian Safety and Efficacy Register of New Interventional Procedures - Surgical

Systematic review

Surgical simulation for training: skills transfer to the operating room (update)

ASERNIP-S report no. 80

(update of ASERNIP-S report no. 61)

Australian Safety and Efficacy Register of New Interventional Procedures — Surgical

The Royal Australasian College of Surgeons

Surgical simulation for training: skills transfer to the operating room (update)

ISBN 978-09808175-1-5

Published November 2012

This report should be cited in the following manner:

Dawe S, Windsor J, Cregan P, Hewett P, Maddern G. Surgical simulation for training: skills transfer to the operating room (update). ASERNIP-S report no. 80. Adelaide, South Australia: ASERNIP-S, November 2012.

Copies of these reports can be obtained from:

ASERNIP-S
199 Ward Street,
North Adelaide, SA 5006
AUSTRALIA
Ph: 61-8-8219 0900
Fax: 61-8-8219 0999
Email: asernips@surgeons.org
http://www.surgeons.org/audit & surgical research/asernip-s

The systematic review of surgical simulation for training: skills transfer to the operating room (update) was ratified by:

The ASERNIP-S Advisory Committee on September 2012

and approved by the

Research, Audit and Academic Surgery Board

and

The Council of the Royal Australasian College of Surgeons on October 2012

Table of contents

Tables

Figures

Executive summary

Objective

The objective of this systematic review update was to assess the evidence published since 2006 to determine whether skills acquired through simulation-based training transfer to the operative setting.

Methods

This report updates the research on 'surgical simulation for training: skills transfer to the operating room' published subsequent to the ASERNIP-S report no. 61 (Sturm et al 2007).

Search strategy – Studies were identified by searching MEDLINE, EMBASE, CINAHL, The Cochrane Library and Current Contents from the period January 2007 to September 2011. The Clinical Trials Database (US), NHS Centre for Research and Dissemination Databases (UK), National Research Register (UK), Meta Register of Controlled Trials, and the Australian Clinical Trials Registry were also searched in September 2011.

Study selection – Only studies that reported the use of simulation for surgical skills training, and reporting on the transferability of these skills to the patient care setting, were included for review. To be eligible for inclusion, studies must have contained data on training and/or measures of performance in the simulated setting, and measures of performance in the operative setting. Identified measures of surgical task performance included accuracy of skills, error rates, time to complete the task, and achievement of performance to criterion levels. Outcomes of interest included performance (measured by various validated and non-validated global rating scales and/or task-specific checklists), patient comfort/discomfort scores, and intra- and postoperative complications.

Data collection and analysis – Data from the included studies were extracted by one researcher using standardised data extraction tables developed a priori and were checked by a second researcher. Statistical pooling was not appropriate due to the heterogeneity of the included studies.

Results

A total of 20 randomised controlled trials and three non-randomised comparative studies were included in this review. The review examined surgical simulation, and included studies with various training techniques in the surgical setting. The studies reported on different indications, simulation-based training methods, training times, and the amount of guidance and feedback provided to trainees. Simulation-based training was compared to no simulation training in 20 studies. Of the remaining three studies, two compared simulation-based training with patient-based training, and one used interactive seminar-based education as the comparator. Where simulation-based training was compared to no simulation training, it was usually an adjunct to normal surgical training programs. However, one of the 20 studies compared two different

simulation-based training methods with two comparators (no simulation-based training [control] and didactic lecture-based training).

For laparoscopic cholecystectomy, bilateral tubal ligation, salpingectomy, Nissen fundoplication, diagnostic arthroscopy of the knee, and totally extraperitoneal inguinal hernia repair, camera navigation, participants who received simulation-based training prior to patient-based assessments performed better (higher global assessment score and/or shorter time to complete task) than participants who did not have this training. Simulator-trained groups generally made fewer errors than control groups in subsequent patient-based assessments.

For colonoscopy, cystourethroscopy, endoscopic sinus surgery, and transurethral resection of the prostate, participants who received simulation-based training appeared to perform better (higher global assessment score and/or shorter time to complete task) than controls in subsequent patient-based assessments.

There were no differences in time to complete task between simulator-trained participants compared with controls when performing oesophagogastroduodenoscopy or nasolaryngoscopy. However, the simulator-trained group required significantly less assistance from the supervisor to complete the task during oesophagogastroduodenoscopy than the control group. There was no significant difference between the simulator-trained participants and the controls for the flexible laryngoscopy procedure time on the standardised patient or the discomfort assigned by the standardised patient but the authors noted the data was positively skewed by two extremely high values.

For other surgical procedures, abdominal fascial closure, cardiopulmonary bypass weaning following cardiac surgery, phacoemulsification on cataract surgery, and knowledge, attitude and skills in the operating room, participants who received simulation-based training prior to patient-based assessments performed better (higher global assessment score and/or shorter time to complete task) than participants who did not have this training.

One study compared patient-based training with simulation-based training for colonoscopy and found that participants who had trained exclusively on a simulator without any mentoring or supervision performed at an equivalent standard on the assessment procedure to those who had received patient-based training. One study compared patient-based training with simulation-based training for in-surgery laparoscopic camera navigation and found that simulator-based camera navigation training for laparoscopic surgery was as effective as, and more time efficient than, traditional teaching of this task.

Conclusions

The studies included in this update on whether surgical skills acquired through simulation-based training transfer to the operating room were of a higher quality (including considerably more RCTs) than those found in the 2006 systematic review (Sturm et al 2007). These studies have strengthened the evidence base. However, the studies still have variable training and assessment methods, making comparison between studies difficult. Overall the current evidence demonstrates that simulation-based training, as part of a surgical skills training program and

incorporating the achievement of reaching predetermined proficiency levels, results in skills transfer to the operating setting.

Classification and recommendations

On the basis of the evidence presented in this systematic review, the ASERNIP-S Review Group agreed on the following classifications and recommendations concerning the transferability of skills acquired through simulation-based training to the surgical setting:

Classifications

Evidence rating

The evidence-base in this review is rated as average. The studies included were of variable quality, and did not have comparable simulation-based methods for the same indications, resulting in an inability to draw solid conclusions.

Efficacy rating

Efficacy cannot be determined. The studies did not have comparable simulation-based methods for the same indications, resulting in an inability to draw solid conclusions.

2011 clinical and research recommendations

Sturm et al (2007) recommended that further research be conducted into the transfer of skills acquired via simulation-based training to the patient setting, to strengthen the evidence base. This recommendation is still relevant in 2011 although several well-designed studies have strengthened the evidence base since 2006. Areas still requiring further study include:

- 'the nature and duration of training required to deliver the greatest transfer effect
- the stage of training at which trainees receive maximum skill transfer benefits from different forms of simulation,
- the effect of different levels of mentoring during the training period on transfer rates, and
- changes in staff productivity as a result of simulation-based training' (Sturm et al 2007)

Further research could also explore the way that simulation-based technical skills training environments might be used to train and assess non-technical skills, such as decision-making.

Important note

The information contained in this report is a distillation of the best available evidence located at the time the searches were completed as stated in the protocol. Please consult with your health care professional if you have further questions relating to the information provided, as the clinical context may vary from patient to patient.

The ASERNIP-S Classification System

Evidence rating

The evidence for ASERNIP-S systematic reviews is classified as Good, Average or Poor, based on the quality and availability of this evidence. High quality evidence is defined here as having a low risk of bias and no other significant flaws. While high quality randomised controlled trials are regarded as the best kind of evidence for comparing interventions, it may not be practical or ethical to undertake them for some surgical procedures, or the relevant randomised controlled trials may not yet have been carried out. This means that it may not be possible for the evidence on some procedures to be classified as good.

Good

Most of the evidence is from a high quality systematic review of all relevant randomised trials or from at least one high quality randomised controlled trial of sufficient power. The component studies should show consistent results, the differences between the interventions being compared should be large enough to be important, and the results should be precise with minimal uncertainty.

Average

Most of the evidence is from high quality quasi-randomised controlled trials, or from non-randomised comparative studies without significant flaws, such as large losses to follow-up and obvious baseline differences between the comparison groups. There is a greater risk of bias, confounding and chance relationships compared with high-quality randomised controlled trials, but there is still a moderate probability that the relationships are causal.

An inconclusive systematic review based on small randomised controlled trials that lack the power to detect a difference between interventions and randomised controlled trials of moderate or uncertain quality may attract a rating of average.

Poor

Most of the evidence is from case series, or studies of the above designs with significant flaws or a high risk of bias. A poor rating may also be given if there is insufficient evidence.

Efficacy Classification

At least as efficacious compared to comparator procedure(s)

This grading is based on the systematic review showing that the new instrument is at least as efficacious as the comparator.

Efficacy cannot be determined

This grading is given if evidence is insufficient to determine the efficacy of the new intervention.

Less efficacious compared to comparator procedure(s)

This grading is based on the systematic review showing that the new intervention is not as efficacious as the comparator.

Research recommendations

It may be recommended that an audit or a controlled (ideally randomised) clinical trial be undertaken in order to strengthen the evidence base.

Clinical recommendations

Additional recommendations for use of the training techniques in clinical practice may be provided to ensure appropriate use by sufficiently qualified/experienced centres and/or individuals.

The ASERNIP-S Review Group

ASERNIP-S Director

Professor Guy Maddern

ASERNIP-S

Royal Australasian College of Surgeons

199 Ward Street

North Adelaide SA 5006

Protocol Surgeon

Professor John Windsor

Department of Surgery

University of Auckland

PO Box 92019

Auckland New Zealand

Advisory Surgeon

Mr Patrick Cregan

PO Box 1124

Penrith NSW 2751

Advisory Surgeon

Mr Peter Hewett

142 Ward Street

North Adelaide SA 5006

ASERNIP-S Researcher

Susan Dawe

ASERNIP-S

Royal Australasian College of Surgeons

199 Ward Street

North Adelaide SA 5006

Conflict of interest

Professor John Windsor, Director and Educational Advisor, Simtics Ltd.

1. Introduction

This report updates the research published since the previous systematic review (Sturm et al 2007; 2008) and covers the period January 2007 to September 2011.

Objective

The objective of this systematic review update was to assess the evidence published since 2006 to determine whether skills acquired through simulation-based training transfer to the operative setting.

Context

This context is a summary of the introduction to be found in Sturm et al (2007).

Surgical training consists of developing cognitive, clinical and technical skills, with the latter traditionally acquired through mentoring in the operating room (OR). Surgical simulation offers the opportunity for surgical trainees to practise surgical skills (mental practice and reinforcement) before entering the OR and allows for detailed feedback (proximal and technical), and objective assessment of performance. To establish whether there is benefit in using simulated environments to teach surgical skills, it must be shown that the skills acquired through simulation-based training can positively transfer to clinical practice.

Surgical training

Surgical skills training to date has largely been conducted via the mentored or 'apprenticeship' approach. The mentored approach is dependent on skilled surgeons having the time and resources to train and supervise trainees. The mentoring approach is also dependent on the flow of patients through a hospital and can result in trainees having random ad hoc exposure to less common procedures.

The movement toward increased specialisation in academic teaching hospitals has resulted in more complex and challenging surgical problems (Grober et al 2004), greater volumes of cases, increased service responsibilities, and a need for surgeons to work at maximum efficiency with minimal interruption. In addition, legislated restrictions on surgeons' working hours have reduced the number of hours during which experienced surgeons are available to observe and assist trainees.

The move away from open surgery to less invasive techniques has meant that trainee surgeons now have less opportunity to learn an open procedure prior to learning the minimally invasive technique. Minimally invasive procedures using a laparoscope as well as other endoscopic techniques differ from open surgery in terms of direct tactile contact and visual feedback, and an increased need for hand-eye coordination (Gallagher et al 1998). Simulation allows trainees to practise these skills before entering the OR environment.

Surgical simulation

Simulation is an instructional strategy used to teach technical skills, procedures, and operations, by presenting learners with situations that resemble reality (Krummel 1998). Surgical simulation involves the use of objects, devices, electronic and/or mechanical surgical simulators, cadavers, animals and animal organs to reproduce or represent, under test conditions, situations that are likely to occur in the actual operative setting (Krummel 1998).

Surgical competence encompasses a combination of requisite knowledge, technical skills, cognitive skills, decision-making ability, communication skills, leadership skills and professional ethics (Moorthy et al 2003). Simulated training allows trainees to practise the cognitive and technical skills of a procedure under various conditions without the pressures of the OR, and allows for the teaching of rare or unusual cases. The trainees' actions can be analysed, errors identified and corrected, and performance scored under standardised, though not real, conditions. Simulation-based skill training allows an individual to acquire skills to the point where many psychomotor skills and spatial judgments have become automated (Gallagher et al 2005).

Types of surgical simulation

The different forms of simulation are summarised below:

Synthetic (inanimate) models and box trainers

Physical simulators, such as box trainers, do not directly measure movements or skills, and require a trained observer to determine performance (Fried 2006). Their relatively low acquisition cost, high availability and easy portability make this type of simulator the most widely available and validated surgical training system (Roberts et al 2006).

Live animal models

Anaesthetised, live animals provide a high-fidelity, non-patient environment that allows trainees to develop the psychomotor and cognitive skills required for the operative setting (Wagh and Waxman 2006). The numbers of animals needed as well as cultural, financial and ethical issues limit their use.

Cadaveric models

The limited supply of cadavers in Australia, coupled with concerns regarding disease transmission from human tissues and fluids, and ethical and cultural issues, limit this mode of training.

Ex vivo animal tissue models

Using anatomic sections or tissues from euthanased animals (ex vivo) is another form of simulation in surgical skills training. Dedicated 'wet rooms' within skills centres are mandatory if this training model is to be employed.

Virtual reality (computer-based) models

Virtual reality (VR) surgical simulators use computer-generated instruments through specially designed interfaces to manipulate computer-generated objects. An attractive feature of VR surgical simulators is that they can provide objective and repeated measurements, such as the

time taken to complete a task, the errors made in the process and also the efficiency with which the movements were made in the accomplishment of the task (Haque and Srinivasan 2006). These metrics present the opportunity for the assessment of competency without the need for an observer to be present.

Most surgical VR systems function as part-task trainers that aim to increase surgeon skill by shaping behaviours required for performing surgery (Seymour and Rotnes 2006). Realistic tactile sensations (i.e. haptics) in the use of surgical instruments is imperfect in some VR simulators (Seymour and Rotnes 2006). Although work is progressing to improve realistic haptics in VR trainers (Roberts et al 2006), this development is expensive.

VR technology has developed software that attempts to replicate skills required for entire minimally invasive surgery procedures, but this provides limited practice in decision-making and relatively poor haptic feedback (van Empel et al 2012).

Augmented reality simulators

Augmented reality (AR) combines physical reality (such as a box trainer) and VR into one system (Botden and Jakimowicz 2009; van Empel et al 2012). Haptic feedback is maintained, using original laparoscopic instruments and tactile tasks, and objective measures of performance are generated.

Universal simulators

The universal simulators (frequently a mannequin linked to a sophisticated computer program) go beyond basic skills training and are designed to recreate specific anatomy and physiology, allowing trainees to practise all the skills necessary to perform a particular operation (Roberts et al 2006). They allow the production of realistic clinical environments where teams work within simulated scenarios to practise crisis management, team response, communication, and other complex tasks. These simulators will not replace basic skills training, but may help to bridge the gap to the OR (Roberts et al 2006). However, hybrid simulators are expensive (Roberts et al 2006).

Skills transfer to the operating room

Training in the OR is expensive in terms of staffing and time (Roberts et al 2006). Surgical skills training courses delivered through dedicated skill laboratories, and using surgical simulation, are becoming accepted as standards of training. In recent years, VR trainers have become increasingly popular but are expensive due to their low production volume and high development cost. However, the fundamental assumption of surgical simulation-based training is that the skills acquired in simulated environments are directly transferable to the clinical setting.

Systematic review update

Moher and Tsertsvadz (2006) defined an update of a systematic review as 'a discrete event with the aim to search for and identify new evidence to incorporate into the previously completed systematic review'. The emergence of new information may undermine the validity of systematic reviews and therefore they must be kept up-to-date (Moher and Tsertsvadz 2006).

In preparing this update, 42 related systematic and literature reviews that were published since 2006 were pearled for new studies. Besides Sturm et al (2007), seven other systematic reviews were found but these did not evaluate skills transfer to the operative setting. However, it reflects the growing interest in simulation-based training in many surgical specialities including arthroscopy (Modi et al 2010), cardiac surgery (Lodge and Grantcharov 2011), central venous catheterisation (Ma et al 2011), colorectal surgery (Miskovic et al 2010), gynaecological endoscopy (Mettler and Dewan 2009), laparoscopic surgery (Gurusamy et al 2008; Thijssen and Schijven 2010), neurosurgery (Alaraj et al 2011; Malone et al 2010), urology (Ahmed et al 2011; Autorio et al 2010; Lendvay 2011), surgical simulators in general (Schout et al 2010) and simulation-based surgical education and training (Scott et al 2011).

Summary

Surgical simulation-based training does not require the use of patients for skills practice, and is less reliant on supervising surgeons' time. Simulation-based training also ensures that trainees have opportunities to practise under various conditions, facilitates the teaching of rare or unusual cases, and provides opportunity for objective standard assessment. However, it is essential to firstly demonstrate that the skills acquired can be transferred to real patients.

The evidence presented in the first systematic review of simulation-trained surgical skills transfer to the OR (Sturm et al 2007; 2008) indicated that there was limited evidence only and more quality research, particularly involving randomised controlled trials, was needed. This systematic review update examined the new evidence which has become available since 2006 in order to enhance the validity of the systematic review.

2. Methodology

Literature search protocol

Inclusion criteria

Articles were selected for inclusion in this systematic review on the basis of the following criteria.

Types of studies

Systematic reviews of randomised controlled trials (RCTs), RCTs, and non-randomised comparative studies were included for review. Where appropriate, additional relevant published material in the form of literature review articles, letters, conference material, commentary, editorials and abstracts were included as background information.

Participants

Surgeons, surgical trainees (residents), medical students or other people involved in human patient care were included.

Only studies that reported on the use of surgical simulation-based training for surgical skills training, and reporting on the transferability of these skills to the patient care setting, were included.

New intervention

Surgical skills training with surgical simulation.

Comparative intervention

Surgical skills training without surgical simulation.

Outcomes

Studies that reported at least one of the following outcomes were included:

- **measures of surgical task performance** in the simulated setting and the clinical setting could include, but not be limited to:
  - accuracy of skill/technique
  - time to complete skill/technique
  - efficiency of movement
  - error rates
  - achievement of performance to criterion level.
- **mortality, morbidity and discomfort** of patients.

Language restriction

Searches were conducted without language restriction. Foreign language papers were excluded unless the findings provided additional information over that reported in well-designed studies published in the English language.

Literature search strategies

Databases searched and search terms used

Searches are shown in Table 1 for the original systematic review (Sturm et al 2007) and the current update.

Table 1 Databases searched

Database	Platform	Edition (Sturm *et al* 2007)	Edition (update 2011)
Cochrane Library		Issue 2, 2006	2010–11
Current Contents	Ovid	Searched 14/12/2006	15/12/2006 to 23/09/2011
EMBASE	Ovid	Week 1 1980 to 14/12/2006	15/12/2006 to 23/09/2011
CINAHL	Webspirs	1982 to 14/12/2006	15/12/2006 to 23/09/2011
PubMed	Entrez	1953 to 14/12/2006	15/12/2007 to 23/09/2011
Clinical Trials Database (USA)		Searched 14/12/2006	Searched 23/09/2011
NHS CRD (UK); NHS HTA (UK)		Searched 15/12/2006	Searched 23/09/2011
National Research Register (UK)		Issue 2, 2006	2011
Current Controlled Trials (mRCT)		Searched 14/12/2006	Searched 23/09/2011

NHS: National Health Service; CRD: Centre for Reviews and Dissemination; HTA: Health Technology Assessment programme; mRCT: metaRegister of Controlled Trials; UK: the United Kingdom; USA: the United States of America.

Search terms

In the Cochrane Library the search term used was 'surgical simulation'.

For EMBASE, CINAHL, Current Contents Connect and PubMed the following search terms were used:

surg* AND simulat* AND (skill* OR train*)

The NHS CRD databases were searched using the same terms as EMBASE and PubMed. The National Research Register, Clinicaltrials.gov, Meta-Register and the Australian Clinical Trials Registry were also searched using the same search terms for RCTs in progress.

Note: * is a truncation character that retrieves all possible suffix variations of the root word, e.g. surg* retrieves surgery, surgical, surgeon, etc.

Literature database and exclusions

Articles were retrieved if they were judged to possibly meet the inclusion criteria based on their abstracts. Two researchers independently applied the selection criteria and any differences were resolved through discussion. The number of articles retrieved is shown in Figure 1 (Section 3). In some cases, when the full text of the article was retrieved, closer examination revealed that it did not meet the inclusion criteria specified by the review protocol. Consequently, these papers were not used to formulate the evidence base for the systematic review (see Appendix A: Excluded studies). However, relevant information contained in these excluded papers was used to inform and expand the review discussion. The bibliographies of all publications retrieved were manually searched for relevant references that may have been missed in the database search (pearling).

Data extraction and assessment of study quality

Data from all included studies were extracted by one researcher and checked by a second using standardised data extraction tables that were developed a priori. Data were only reported if stated in the text, tables, graphs or figures of the article, or if they could be accurately extrapolated from the data presented. If no data were reported for a particular outcome, in particular adverse outcomes, then no value was tabulated. This was done to avoid the bias caused by incorrectly assigning a value of zero to an outcome measurement on the basis of an unverified assumption by the reviewer.

Data analysis

The included studies were categorised initially by the non-simulation-based training method (i.e. simulation-based training versus no simulation training, didactic-lecture-based education, interactive seminar-based education, or patient-based training). Studies were then categorised by intervention and then by the level of evidence. It was judged that no data were suitable for statistical pooling due to the variability in simulation devices and training methods. Where data could not be grouped, the main outcomes have been reported narratively.

Ongoing and unpublished trials

Searches of the Clinical Trials Database, NHS CRD, NHS HTA, Current Controlled Trials and the National Research Register identified a number of studies. The details for each are provided below:

1. Creating a proficiency-based virtual reality simulation training programme for laparoscopic assisted colectomy (LAC). Royal College of Surgeons, Ireland. (ClinicalTrials.gov identifier NCT00752817). September 2008–September 2009. Principal investigator: Paul Neary. Expected enrolment n=16. Design: prospective randomised, controlled, single blinded, multicentre intervention study. Objective: to prove that participants randomised to train under a proficiency-based progression simulation curriculum (using the PROMIS-LAC simulator from Haptica, Ireland) will learn to perform laparoscopic assisted colectomy faster, complete more surgical steps and commit fewer operative errors compared to participants randomised to the current surgical training curriculum.
2. Laparoscopic simulator training and its impact on surgical education. University of Texas Southwestern Medical Center, USA (ClinicalTrials.gov identifier NCT00555243). January 2005–November 2006. Study completed: January 2011. Expected enrolment n=10. Design: prospective randomised, controlled, single blinded, multicentre intervention study. Objective: to determine whether laparoscopic simulators truly affect real time performance in the operating room among gynaecology residents (over the course of one resident rotation of 4 to 6 weeks) and if the level of improvement is inversely related to resident level of training.
3. Trial of proficiency-based simulation training for general surgical trainees. (ClinicalTrials.gov identifier NCT00712387). Royal College of Surgeons, Ireland. July 2008–July 2009. Principal investigator: Professor A. Gallagher. Expected enrolment n=24.

Design: prospective randomised, controlled, single-blinded intervention study. Objective: to determine if junior surgeons trained to predetermined proficiency level on the laparoscopic virtual reality simulator LapSim for a laparoscopic cholecystectomy will make less critical intraoperative errors while performing a supervised laparoscopic cholecystectomy and will perform faster than their traditionally trained colleagues. *See publication:* Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA 3rd, Ramel S, Smith CD, Arvidsson D 2007, 'Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies', *The American Journal of Surgery*, vol. 193(6), pp.797-804.

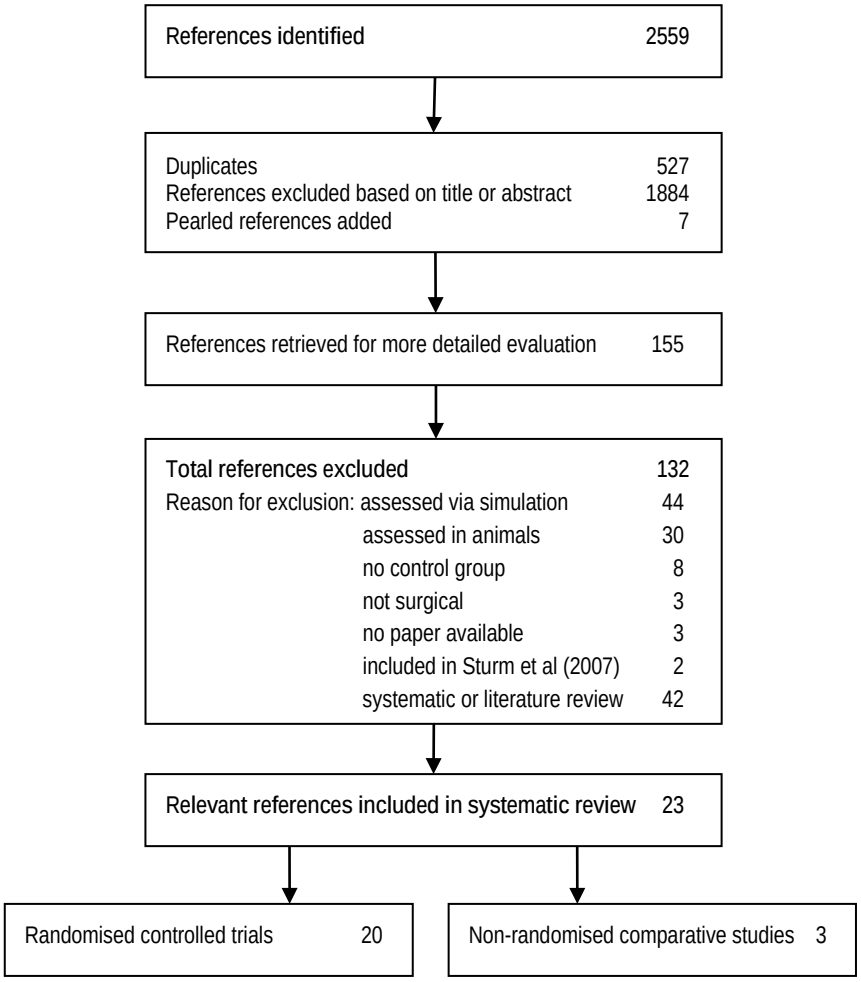
4. Do motion metrics lead to improved skill acquisition on simulators? (ClinicalTrials.gov identifier NCT01052168). Carolinas Healthcare System, USA. November 2009–December 2011. Responsible party: Dr Dimitrios Stefanidis. Expected enrolment n=60. Design: randomised single-blinded intervention study. Objective: to determine if proficiency-based simulator training in laparoscopic suturing to expert-derived levels of speed and motion will result in better operative performance compared to participants training to levels of speed or motion alone.
5. Developing a curriculum to teach laparoscopic colorectal surgery (ClinicalTrials.gov identifier NCT01371136), St Michael's Hospital Toronto, Canada. June 2010–May 2011. Principal investigator: Dr TP Grantcharov. Expected enrolment n=25. Design: randomised single-blinded controlled trial. Objective: to develop and validate a comprehensive ex-vivo curriculum for laparoscopic colorectal surgery by comparing the technical performance of curricular trained and non-trained residents in the operating room, during a procedure on a patient.
6. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy (Protocol). Walsh CM, Sherlock ME, Ling SC, Carnahan H. Cochrane Database of Systematic Reviews 2010, Issue 1. Art. No.: CD008237. DOI: 10.1002/14651858.CD008237. Hospital for Sick Children, Toronto, Ontario, Canada. Responsible party: Dr C M Walsh. Objective: to determine whether virtual reality simulation training can supplement and/or replace early conventional endoscopy training (apprenticeship model) in diagnostic oesophagogastroduodenoscopy, colonoscopy and/or sigmoidoscopy for health professions trainees with limited or no prior endoscopic experience.

3. Studies included in the review

Literature search results

Details of the searching and retrieval process are shown in Figure 1.

Figure 1 Process for selection of studies retrieved from the literature database



Designation of levels of evidence and critical appraisal

The evidence presented in the included studies was classified according to the National Health and Medical Research Council (NHMRC) Hierarchy of Evidence (see Appendix B) (NHMRC 2010). Study quality was assessed according to the methods given in Section 6 of the Cochrane Reviewers' Handbook (Higgins and Green 2009) on a number of parameters such as the quality of the reporting of study methodology, methods of randomisation and allocation concealment, blinding of patients or outcomes assessors, attempts made to minimise bias, sample sizes and their ability to measure 'true effect', applicability of results outside of the study sample as well as examining the statistical methods used to describe and evaluate the study data.

The included studies are shown in Table 2. Details of the methodology assessment and study design of the included studies appear in Appendix C, and the summary of the critical appraisal is in Appendix D. Acronyms and abbreviations used in this report are listed in Appendix G.

Description of studies

A total of 20 RCTs and three non-randomised comparative studies were included in this review (Table 2). Twenty studies compared simulation-based training with no simulation-based training. Of the remaining three studies, two used patient-based training, and another used interactive seminar-based education, as the comparator. Of the 20 studies comparing simulation-based training with no simulation training, Patel et al (2012, electronic publication was available in September 2011) compared two different simulation-based training methods with two comparators (no simulation-based training [control] and didactic lecture-based training).

There were a total of 629 participants in the included studies. Table 3 summarises the training program used with the simulator or simulation in each study and Table 4 describes the statistical analyses used within each study. Details of the methodology assessment and study design of the included studies appear in Appendix C. A summary of the critical appraisal is given in Appendix D. The simulator device specifications and descriptions of non-simulation-based training are given in Appendix F.

Laparoscopic procedures

Nine RCTs and one comparative study examined performance during laparoscopic procedures in participants who had trained using simulation with those who had not received simulation-based training. These included laparoscopic cholecystectomy (Ahlberg et al 2007; Cosman et al 2007; Hogle et al 2009; Sroka et al 2010), bilateral tubal ligation (Banks et al 2007), salpingectomy (Larsen et al 2009), Nissen fundoplication (Van Sickle et al 2008), diagnostic arthroscopy of the knee (Howells et al 2008), and totally extraperitoneal inguinal hernia repair (Zendejas et al 2011). Beyer et al (2011) reported on a comparative prospective study of laparoscopic cholecystectomy performance.

Franzeck et al (2012, electronic publication was available in September 2011) reported on a RCT which compared laparoscopic camera navigation in participants who had trained using simulation with those who had received patient-based training. Studies with this comparator were reported separately in this report.

Endoscopic procedures

Six RCTs and one comparative study examined performance during endoscopic procedures in participants who had trained using simulation with those who had not received simulation-based training. These included colonoscopy (Park et al 2007), cystourethroscopy (Schout et al 2009), oesophagogastroduodenoscopy (Shirai et al 2008), nasolaryngoscopy (Ossowski et al 2008), endoscopic sinus surgery (Fried et al 2010), and transurethral resection of the prostate (Kälström et al 2010). Yi et al (2008) reported on a comparative study of colonoscopy performance.

Haycock et al (2010) reported on a RCT compared colonoscopy procedures in participants who had trained using simulation with those who had received patient-based training. Studies with this comparator were reported separately in this report.

Other surgical procedures

Two RCTs and one comparative study examined performance of other surgical procedures in participants who had trained using simulation with those who had not received simulation-based training. The RCTs included abdominal fascial closure (Palter et al 2011), and knowledge, attitude and skills in the operating room (Patel et al 2012, electronic publication was available in September 2011). Belyea et al (2011) reported on a retrospective comparative series of phacoemulsification performance for cataract surgery.

Bruppacher et al (2010) also conducted a RCT but compared performance during cardiac surgery weaning patients from cardiopulmonary bypass in participants who had trained using high-fidelity simulation with those who had received interactive seminar-based education. Studies with this comparator were reported separately in this report.

Table 2 Summary of included studies

Study	Study type	Level	Training method	Comparator	Indication
Simulation-based training versus no simulation-based training					
Ahlberg et al 2007 (Sweden)	RCT	II	LapSim (without force feedback) v 2.0 simulator	No simulator training	Laparoscopic cholecyst-ectomy
N=13			7	6	
Banks et al 2007 (USA)	RCT	II	Limbs and Things Ltd. laparoscopic simulator	No simulator training	Laparoscopic bilateral tubal ligation
N=20			10	10	
Belyea et al 2011 (USA)	Retrospective comparative study	III-3	Eyesi VR opthalmosurgical simulator	No simulator training	Phacoemulsification for cataract surgery
N=42			17	25	
Beyer et al 2011 (France)	Comparative study	III-3	MISTELS simulator or LAP Mentor VR simulator	No simulator training	Laparoscopic cholecystectomy
N=19			6/6	7	
Cosman et al 2007 (Australia)	RCT	II	LapSim (without haptic feedback) v 1.5 simulator	No simulator training	Laparoscopic cholecystectomy
N=10			5	5	
Fried et al 2010 (UK)	RCT	II	Endoscopic Sinus Surgery Simulator (ES3)	No simulator training	Endoscopic sinus surgery
N=25			12	13	
Hogle et al 2009 (USA)	RCT	II	LapSim simulator	No simulator training	Laparoscopic cholecystectomy
N=12			6	6	
Howells et al 2008 (UK)	RCT	II	Arthroscopy knee bench-top simulator	No simulator training	Arthroscopy
N=20			10	10	

Study	Study type	Level	Training method	Comparator	Indication
Kälström et al 2010 (Sweden)	RCT	II	PelvicVision simulator	No simulator training	Transurethral resection of prostate
N=23			11	12	
Larsen et al 2009 (Denmark)	RCT	II	LapSim v 3.0 simulator	No simulator training	Laparoscopic salpingectomy
N=24			13	11	
Ossowski et al 2008 (USA)	RCT	II	Nasal model for endoscopic simulation	No simulator training	Endoscopy – nasolaryngoscopy
N=20			10	10	
Palter et al 2011 (Canada)	RCT	II	synthetic abdominal wall simulator	No simulator training	Abdominal fascial closure
N=18			9	9	
Park et al 2007 (Canada)	RCT	II	AccuTouch (v 1.2 with haptic feedback from simulated patient) colonoscopy simulator	No simulator training	Colonoscopy
N=24			12	12	
Patel et al 2012* (UK)	RCT	II	Second Life operating theatre Imperial College simulated operating suite	[Didactic lecture**] No training (Control)	Knowledge, attitude and skills in the OR
N=60			15/15	15/15	
Schout et al 2009 (Netherlands)	RCT	II	URO Mentor virtual reality simulator	No simulator training	Cystourethroscopy
N=100			50	50	
Shirai et al 2008 (Japan)	RCT	II	GI Mentor II simulator	No simulator training	Endoscopy – oeosophago-gastroduodenoscopy
N=20			10	10	
Sroka et al 2010 (Canada)	RCT	II	FLS simulator	No simulator training	Laparoscopic cholecystectomy
N=16			8	8	
Van Sickle et al 2008 (USA)	RCT	II	MIST-VR task 3 (immersion jig) and standard box trainer	No simulator training	Laparoscopic Nissen fundoplication
N=22			11	11	
Yi et al 2008 (South Korea)	Non-randomised comparative study	III-2	KAIST-Ewha Colonoscopy Simulator II	No simulator training	Colonoscopy
N=11			5	6	
Zendejas et al 2011 (USA)	RCT	II	Guildford MATTU TEP hernia repair simulator	No simulator training	Totally extraperitoneal inguinal hernia repair
N =50			26	24	

Study	Study type	Level	Training method	Comparator	Indication
*Simulation-based training versus didactic lecture-based education***					
Patel et al 2012* (UK)	RCT	II	Second Life operating theatre Imperial College simulated operating suite	Didactic lecture** [no training (Control)]	Knowledge, attitude and skills in the OR
N=60			15/15	15/15	
Simulation-based training versus interactive seminar-based education					
Bruppacher et al 2010 (Canada)	RCT	II	SimMan Universal patient simulator	Interactive seminar-based education	Cardiopulmonary bypass weaning following cardiac surgery (anaesthesiology)
N=20			10	10	
Simulation-based training versus patient-based training					
Haycock et al 2010 (UK)	RCT	II	Olympus (Endo TS-1) colonoscopy simulator	Patient-based training	Colonoscopy
N=36			19	18	
Franzeck et al 2012* (Switzerland)	RCT	II	LAP Mentor VR simulator or haptic PROMIS hybrid simulator	Patient-based training (OR training)	Camera navigation
N=24			12	12	

* included as electronic publication available in September 2011.
** two comparators used in this study.
OR: operating room; RCT: randomised controlled trial; UK: United Kingdom; USA: United States of America; VR: virtual reality.

Table 3 Description of training in included studies

Study	Training method	Description of simulation training	n/N
Simulation-based training versus no simulation-based training			
Ahlberg et al 2007 RCT, Level II	LapSim Basic Skills package version 2.0 (virtual reality without force feedback)	LapSim basics skills training to proficiency. One-hour training session under supervision of a laparoscopic expert followed by 15 minutes of recovery with a maximum of eight sessions per day during one week until proficiency was reached. Proficiency on simulator was defined by calculating median value for every parameter in each of six tasks from all five laparoscopic surgeons (experts).The participants assigned to simulator training then practised under supervision and received feedback given by the simulator as well as oral feedback given by the supervisor after each completed task until they showed proficiency on each of the six examination tasks at least twice.	7/13
Banks et al 2007 RCT, Level II	Limbs and Things Ltd. laparoscopic simulator	One-hour focused didactic instruction followed by two hours of hands-on teaching in skills laboratory with three stations: suturing pigs' feet, knot-tying boards, and a simulator station with a laparoscopic simulator and an operative laparoscopic tower.	10/20
Belyea et al 2011 Retrospective comparative study, Level III-3	Eyesi virtual reality opthalmosurgical simulator	Post 2006 all ophthalmology residents were expected to spend a minimum of two hours per year using the simulator as verified by a login record.	17/42
Beyer et al 2011 Comparative study, Level III-3	MISTELS simulator (McGill Inanimate System of Training and Evaluation of Laparoscopic Skills) or LAP Mentor VR simulator	*Group 1:* MISTELS simple simulator training – five individual 60-minute sessions over one month. *Group 2*: LAP Mentor virtual reality simulator training – five sessions in one month organised in pairs with each pair undergoing five 120-minute sessions attempting all nine exercises, the intracorporeal knot suture exercise, and one cholecystectomy in each session. All sessions were supervised by the same teacher.	6/6/19
Cosman et al 2007 RCT, Level II	LapSim simulator system with basic skills package version 1.5.	Practised the clipping task on the LapSim virtual reality simulator, following a distributed training protocol including access for a maximum of one hour per day, until participants satisfied the performance criteria on two successive repetitions of the tasks.	5/10
Fried et al 2010 RCT, Level II	Endoscopic Sinus Surgery Simulator (ES3)	Participants were fully trained to proficiency with ES3 in addition to receiving conventional textbook-based and video recorded educational material.	12/25
Hogle et al 2009 RCT, Level II	LapSim simulator	The training curriculum was fully completed when level 3 was passed for each module. Tasks included camera navigation, instrument navigation, coordination, grasping, lifting and grasping, cutting and clip applying. The participants were asked to independently complete two simulation training sessions per week. No other simulation training practice was allowed outside the training curriculum.	19/38
Howells et al 2008 RCT, Level II	Arthroscopy knee bench-top simulator, and a standard 30˚ arthroscope	Three sessions of six simulated arthroscopies in one week, 18 simulated arthroscopies were supervised by Howells and followed a fixed protocol for diagnostic arthroscopy of the knee agreed by two surgeons experienced in this area. Participants also had traditional training.	10/20

Study	Training method	Description of simulation training	n/N
Kälström et al 2010 RCT, Level II	PelvicVision simulator	Proficiency-based virtual reality simulator with haptics, training in transurethral resection of prostate (TURP) procedure training. Five-day course on treatment of benign enlargement of the prostate included theory, diagnostic methods, the instrumentation used in TURP procedure, and risk factors.	11/24
Larsen et al 2009 RCT, Level II	LapSim Gyn v 3.01	Proficiency-based virtual reality simulator training in laparoscopic salpingectomy (for ectopic pregnancy). The programme comprised two parts: training in the two basic skills of 'lifting and grasping' and 'cutting'; one procedure-specific task. The basic skills training was done in each training cycle of 45-60 minutes and the salpingectomy repeated continually during the remainder of the cycle. Training sessions were repeated until expert criterion level was reached in two consecutive and independent simulations. Participants also had standard clinical training.	13/24
Ossowski et al 2008 RCT, Level II	Nasal Model for endoscopic simulation	15-minute video instruction on endoscopy followed by 15 minutes practice on simulator.	10/20
Palter et al 2011 RCT, Level II	Abdominal wall simulator	Participants practised abdominal wall closure on the model to a predefined level of proficiency. Each session was a maximum of 1.5 hours in length and occurred no longer than three weeks apart. Individual feedback was limited to technical performance only. Proficiency was reached when the trainee could independently perform a square knot, demonstrated knowledge of instruments required, and placed sutures correctly 1 cm apart, and 1 cm from the wound edge.	9/18
Park et al 2007 RCT, Level II	AccuTouch colonoscopy simulator (version 1.2)	Two to three hours of practice independently on simulator with access to the range of cases on simulator.	12/24
Patel et al 2012* RCT, Level II	Second Life operating theatre	Second Life Virtual World operating theatre introduction using avatars in groups of five students per one-hour session with additional information supplied by instructor through text or voice chat.	15/60
	Imperial College simulated operating suite	Simulated operating room curriculum-based instruction delivered to five students at a time for one hour.	15/60
Note: This study also included below	Didactic lecture**	Lecture with PowerPoint presentation lasting one hour; included instructional videos regarding gowning and gloving and delivered to five students per session with the opportunity to ask questions at completion of the lecture.	15/60
Schout et al 2009 RCT, Level II	URO Mentor virtual reality simulator	The training protocol consisted of seven flexible cystourethroscopy tasks which included stone manipulation tasks numbers three and eight, and basic tasks numbers four, five and nine of the URO Mentor.	50/100
Shirai et al 2008 RCT, Level II	GI Mentor II	Five one-hour training sessions on simulator within two weeks, i.e. two psychomotor tasks (level one EndoBubble and EndoBasket) were performed three times each and then oesophagogastroduodenoscopy (OGD) training modules completed. Patient case 1-I (lower difficulty level) was performed in each session and remaining time used for other cases of OGD module. Participants were not supervised or instructed during the simulator training. Participants also had 15 hours of bedside training.	10/20

Study	Training method	Description of simulation training	n/N
Sroka et al 2010 RCT, Level II	Fundamentals of Laparoscopic Surgery (FLS) simulator	Regular residency training plus FLS simulator proficiency-based training through McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) Program. FLS simulator training included CD-ROM of didactic material and five MISTELS tasks: peg transfer, circle cut, placement of ligating loop, and simple suture tied with extra- and intracorporeal techniques.	8/16
Van Sickle et al 2008 RCT, Level II	MIST-VR task 3 (immersion jig) and standard box trainer, the foam Nissen suturing model and the intracorporeal slip-square knot.	Supervised simulation-based laparoscopic suturing curriculum with specific training tasks and proficiency levels. Each resident was required to reach training performance goals for each task on two consecutive trials before being allowed to progress on to the next task.	11/22
Yi et al 2008 Comparative study, Level III-2	KAIST-Ewha Colonoscopy Simulator II	Simulation-based training included practising the targeted skills of colonoscopy using two training scenarios that have different colon flexures and degrees of difficulty. Training scenario A is designed to teach practical skills to navigate the colon applying torque and up-down angulations. Scenario B is designed to teach skills to manage a loop formed in the sigmoid colon.	5/11
Zendejas et al 2011 RCT, Level II	Guildford MATTU TEP hernia task trainer	A simulation-based mastery learning curriculum for total extraperitoneal inguinal hernia repair using online course (nine web-based modules) followed by skills training on a simulator consisting of one-on-one practice sessions until expert performance was achieved. Mastery was defined as successful repair of both hernias in less than two minutes on two consecutive attempts.	26/50
Simulation-based training versus didactic lecture-based education			
Patel et al 2012* RCT, Level II *Note: This study also included above*	Second Life operating theatre Imperial College simulated operating suite Didactic lecture**	Second Life Virtual World operating theatre introduction using avatars in groups of five students per one-hour session with additional information supplied by instructor through text or voice chat. Simulated operating theatre curriculum-based instruction delivered to five students at a time for one hour. Lecture with PowerPoint presentation lasting one hour; included instructional videos regarding gowning and gloving and delivered to five students per session with opportunity to ask questions at the completion of the lecture.	15/60 15/60 15/60
Simulation-based training versus interactive seminar-based education			
Bruppacher et al 2010 RCT, Level II	SimMan Universal Simulator	Individual two-hour simulation session videotaped to facilitate debriefing and reinforce learning objectives. High-fidelity simulation which closely mimics a cardiac operating room with medications and cardiopulmonary bypass (CPB) machine that was handled by an actor perfusionist and loop video recordings of real patients in different phases of CPB removal.	10/20
	Interactive seminar-based training	Two-hour seminar by anaesthesiologist using PowerPoint slides, hand-outs and face to face discussion of four paper-based scenarios similar to those in the simulation training.	18/36

Study	Training method	Description of simulation training	n/N
Simulation-based training versus patient-based training			
Haycock et al 2010 RCT, Level II	Olympus Endo TS-1 colonoscopy simulator. It also provides a simulated, 3-D endoscope imager view identical to that provided by ScopeGuide (Olympus, Keymed, Essex, UK)	16 hours of standardised simulator-training program with minimal tutoring and feedback.	18/36
	Patient-based training	16 hours (four half-day sessions) by expert trainer using a ScopeGuide imager; including performing a minimum of eight colonoscopies under one-to-one supervision.	18/36
Franzeck et al 2012* RCT, Level II	LAP Mentor VR simulator system and haptic PROMIS hybrid simulator	Students followed a standardised protocol, performing 40 minutes of camera navigation-specific tasks on the different simulators (25 minutes on the basic task modules of the LAP Mentor camera manipulation and 15 minutes on the laparoscope orientation core modules on the PROMIS) and 20 minutes of training on non-camera or camera-specific simulator exercises of free choice. Participants trained twice a week for one hour for three weeks (total six hours of training) using two Xitact IHP instrument haptic ports for the interfaces for laparoscopic instruments, and a third unidirectional two Xitact IHP instrument tracking port for the camera navigation.	12/24
	Patient-based training	Six laparoscopic interventions in OR including hemicolectomy, rectum resection, gastric bypass, and cholecystectomy-camera navigation at the surgeon's direction. Total time spent in the OR and actual time navigating the camera were documented by the participant immediately after each operation using standard forms. There was no significant difference between the actual camera training time compared with the simulator-trained group, but participants who underwent patient-based training spent four times longer in the OR.	

*electronic publication available in September 2011.
** two comparators used in this study.
3-D: three-dimensional; AG: abbreviation of Aktiengesellschaft, a German name for a type of company; MISTELS: McGill Inanimate System for Training and Evaluation of Laparoscopic Skills; OR: operating room; RCT: randomised controlled trial; USA: United States of America; VR: virtual reality.

Table 4 Description of statistical analyses used in included studies

Study	Training method	Statistical analysis
Simulation-based training versus no simulation-based training		
Ahlberg et al 2007 RCT, Level II N=13	LapSim simulator (n=7) No simulator training (n=6)	The primary (total number of errors) and secondary (total time) outcomes, linear mixed-effects models were used to evaluate the expected mean difference between the simulator-trained and traditionally-trained. The participants' first 10 individual full laparoscopic cholecystectomies in the OR were recorded on videotape. Surgery numbers one, five and 10 were assessed by two observers (laparoscopy experts) who were blinded to training status of participants. Procedures were scored on a minute-by-minute basis for 28 surgical errors. Inter-rater reliability between the two observers was 0.98 throughout the assessment, calculated as concordance or in median 0.80 (95% confidence interval), when Cohen's κ coefficient was calculated. A covariance structure of compound symmetry was found to be the most appropriate for taking into account correlation within a particular student on different laparoscopies performed. The choice of covariance structure was evaluated by Akaike's information criteria. Because of heterogeneous variances for the training groups, different variance parameters were estimated. Because of the small number of subjects, univariate models were applied to make inferences about the possible confounders regarding the different psychological and background variables. Converted procedures were treated as if values were missing completely at random. The Cook's distance and scatter plots of the predicted values versus the raw residuals were used as methods of detecting outliers. Data analysis used SAS software (version 9.1.3); graphs and descriptive statistics used Statistica software (version 7).
Banks et al 2007 RCT, Level II N=20	Limbs and Things Ltd. laparoscopic simulator (n=10) No simulator training (n=10)	Wilcoxon rank-sum test was used to compare differences in pre- and post-test scores, the global rating scale, and the task-specific checklist between the surgical simulator laboratory and control groups. Comparison of pass/fail ratings between the two groups was performed with Fisher's exact text. Inter-rater reliability was calculated with Kappa (or proportion of agreement) for the checklist and intraclass correlation (ICC) coefficient for the global rating scale. Spearman correlation coefficient was used to measure the association of the global rating scale and checklist scores in the laboratory with those in the operating room. Data analysis used SAS software (version 9.1). Results were considered statistically significant at $p<0.05$.
Belyea et al 2011 Comparative study, Level III-3 N=42	Eyesi simulator (n=17) No simulator (n=25)	Results of each outcome measure in the two groups were compared using Student's *t* tests. Regression analyses of phacoemulsification time, phacoemulsification power, and adjusted phacoemulsification time were performed to compare the rates of progression in surgical skill over the course of third year of ophthalmology residency training in the two groups. Statistical analysis was performed using SPSS software version 19.
Beyer et al 2011 Comparative study, Level III-3 N=19	MISTELS (n=6) LAP Mentor (n=6) No simulator training (n=7)	The GOALS score obtained were compared among the three groups by means of a nonparametric test (Mann-Whitney *U* test). Links between quantifiable variables were measured by the Spearman rank correlation coefficient (*r*). All the data were analysed by SPSS version 13.0. For all the bilateral tests, the threshold of significance was fixed at 5%.
Cosman et al 2007 RCT, Level II N=10	LapSim simulator (n=5) No simulator training (n=5)	The data were interrogated using Mann-Whitney *U* test to determine whether there were any differences between the performances of the experimental and control groups for the pre-test, post-test and assessment task, or whether there were differences between them in level of training or extent of laparoscopic experience. Inter-rater reliability for assessment scales was calculated using the ICC method.

Study	Training method	Statistical analysis
Simulation-based training versus no simulation-based training		
Fried et al 2010 RCT, Level II N=25	Endoscopic Sinus Surgery Simulator (ES3) (n=12) No simulator training (n=13)	The inter-rater reliability analysis was performed with the Fleiss kappa statistics, which demonstrated a kappa value of 0.87 *(p*=0.21; 95% CI 0.78–0.96). Results were analysed with the Mann-Whitney *U* test, with statistical significance accepted at the 95% level. Internal rate consistency was verified with the Pearson moment correlation and SPSS software version 17.
Hogle et al 2009 RCT, Level II N=12	LapSim simulator (n=6) No simulator training (n=6)	Within each group, the mean post-training scores from the supervising attending surgeon and the video reviewers for each domain were compared for each domain using nonparamentric Wilcoxon rank sum test. Scores from all reviewers were combined based on work published previously by the group. Tests were considered significant at $p<0.05$.
Howells et al 2008 RCT, Level II N=20	Arthroscopy knee bench-top model (n=10) No simulator training (n=10)	The d'Agostini and Pearson omnibus normality test was used to check for normal distribution of data. Data from the non simulator-trained group were skewed, and so nonparametric tests were used. Data analysis was performed using Mann-Whitney *U* tests and SPSS software version 12.0.
Kälström et al 2010 RCT, Level II N=23	Pelvic Vision simulator (n=11) No simulator training (n=12)	Statistics used were paired samples *t*-test and sign test (for nonparametric data) and the statistical analysis was performed using SPSS version 16.02.
Larsen et al 2009 RCT, Level II N=24	LapSim Gyn v3.0.1 (n= 3) No simulator training (n=11)	Cumulated scores presented as medians (average score of two observers) were compared using non-parametrical analysis (Mann-Whitney *U* test). A two tailed *p*-value of 0.05 or less was considered to be statistically significant and an inter-rater agreement and ɣ coefficient of 0.8 or more for each to be acceptable. Analysis was performed using SPSS 13.0 for Windows; graphics using GraphPad Prism.
Ossowski et al 2008 RCT, Level II N=20	Nasal model endoscopic simulator (n=10) No simulator training (n=10)	All scores were compared using the Mann-Whitney *U* statistical test, a nonparametric test that does not assume data are normally distributed. This test was chosen because the data showed considerable positive skewness due to extremely high values for one or two students.
Palter et al 2011 RCT, Level II N=18	Abdominal wall simulator (n=9) No simulator (n=9)	Descriptive statistics including means and standard deviations were calculated for all variables. Variables were normally distributed and therefore the differences in the primary (multiple choice test scores) and secondary outcome measures (technical skills assessment scores) between the two groups were calculated using Mann-Whitney *U* nonparametric test. An effect size for the primary outcome measure was calculated using Cohen's *d* test. Data were reported as median (interquartile range). All statistical analysis was performed using SPSS software version 16.0.
Park et al 2007 RCT, Level II N=24	AccuTouch colonoscopy simulator (n=12) No simulator training (n=12)	The treatment and control group performances were compared using chi-square tests for outcomes on nominal scales (reaching caecum and penalty points for critical flaws) and independent groups *t* tests for measures of interval scales (total global rating scores). The correlations were tested with either Pearson or Spearman correlation coefficients, depending on whether the data from the individual parameters satisfied tests of normality and homogeneity of variance. An alpha level of $p<0.05$ was considered statistically significant.

Study	Training method	Statistical analysis
Simulation-based training versus no simulation-based training		
Patel et al 2012* RCT, Level II N=60	Second Life operating theatre (n=15) Imperial College simulated operating suite (n=15) Didactic lecture** (n=15) No training (n=15)	The data were analysed using SPSS version 16 for Mac (SPSS, Inc., Chicago, Illinois, USA). The internal reliability of the knowledge, skills and attitudes items on the observation and self-report scales was assessed using Cronbach's α coefficient. In view of the nonparametric nature of data, Wilcoxon's signed-rank test was used to assess any difference between preintervention and postintervention outcomes. The Kruskal-Wallis test was used to assess any differences among all four groups for all preintervention and postintervention outcomes. Further group comparative analysis was performed using the Mann-Whitney *U* test to detect where the outcome differences were among the specific groups.
Schout et al 2009 RCT, level II N=100	URO Mentor virtual reality simulator (n=50) No simulator training (n=50)	Differences in participants' demographics and opinions were examined using chi-square and independent *t*-tests. Stepwise multiple linear regression was used to analyse the data. To interpret the magnitude of the standard regression coefficient (β), the effect size (ES) indication for correlations was used, with 0.10, 0.30 and 0.50 considered as small, moderate and large ES, respectively. The threshold for statistical significance was $p<0.05$.
Shirai et al 2008 RCT, Level II N=20	GI Mentor II simulator (n=10) No simulator training (n=10)	The Mann-Whitney *U* test was used to determine the significance of differences for each item and the ratio of scores of one point and two points between the two groups. Spearman's rank correlation analysis and calculation of kappa values (as an indicator of inter-rater agreement) were used for the assessments of the evaluations by supervising physicians. In all analyses, $p<0.05$ was considered statistically significant. All analyses were carried out using Microsoft Excel for Windows (version 2003) and StatView for Windows (StatView-J5.0).
Sroka et al 2010 RCT, Level II N=16	FLS simulator (n=8) No simulator training (n=8)	Student's *t* test was used to compare mean FLS and GOALS scores in the training and nontraining groups. Paired *t* tests were used to compare baseline and the final performance within each group. Multivariate analysis was used to assess the effect of simulator training on GOALS score after adjusting for the baseline GOALS score. Data expressed as medians (interquartile range). Statistical analysis was conducted using SPSS version 11.0. $p<0.05$ was considered statistically significant.
Van Sickle et al 2008 RCT, Level II N=22	MIST-VR task 3 and standard box trainer, foam Nissen suturing model and the intracorporeal slip-square knot (n=11) No simulator training (n=11)	Interobserver agreement between the rating surgeon investigators was calculated by dividing the number of agreements between the rates by the sum of the number of agreements and disagreements, and the interobserver agreement >0.8 was used as the cut-off for high-stakes assessment. Student's *t* test with unequal variance was used to compare simulation-based laparoscopic suturing training group using MIST-VR task 3 and the standard curriculum group scores.
Yi et al 2008 Comparative study, Level III-2 N=11	KAIST-Ewha colonoscopy simulator II (n=5) No simulator training (n=6)	$p<0.05$ was considered statistically significant.

Study	Training method	Statistical analysis
Simulation-based training versus no simulation-based training		
Zendejas et al 2011 RCT, Level II N=50	Guildford MATTU TEP hernia task trainer (n=26) No simulator training (n=24)	Continuous outcome variables were evaluated first without adjustment using Student's *t* test, and then using generalised estimating equations (GEE) to adjust for repeated measures on trainees, supervising surgeon, case difficulty, baseline score and (for time outcomes and for bilateral procedures) side (first repair versus second repair). Dichotomous outcome variables were evaluated first without adjustment using χ^2, and then using GEE to adjust for repeated measures on trainees, patient-related factors (age, gender, history of prostate disease, ASA, BMI) and case-related factors (difficulty, recurrent hernia, staff-surgeon). A GEE model with an exchangeable correlation structure was assumed, and the identity link function for a random variable was specified. Inter-rater and test-retest reliability were calculated using ICC coefficient. Associations between intraoperative and postoperative complications were evaluated with multivariate logistic regression. Analyses were performed using SAS version 9.1.3 and JMP version 8.0. All hypothesis testing was two-sided using an alpha level of 0.05.
Simulation-based training versus didactic lecture-based education		
see Patel et al 2012* above		
Simulation-based training versus interactive seminar-based education		
Bruppacher et al 2010 RCT, Level II N=20	SimMan Universal Simulator (n=10) Interactive seminar-based education (n=10)	Separate mixed-design ANOVA, with the group (simulator or seminar) as the between-trainee variable and test phase (baseline [pre-test], post-test and retention test) as the within subject variable, were conducted on the ANTS and checklist scores. Post hoc analyses were conducted using independent samples *t* tests for each time, with Bonferroni corrections for multiple comparisons. Independent *t* tests were conducted to measure the differences between the two groups in trainees' ages, years of clinical training ion anaesthesiology, number of previous simulation sessions, and amount of clinical experience in CPB weaning; and in patients' ages, left ventricular function, and number of grafts performed during surgery. For all statistical tests (run using SPSS version 17), a two-tailed *p*-value of <0.05 was considered significant.
Simulation-based training versus patient-based training		
Haycock et al 2010 RCT, Level II N=36	Olympus Endo TS-1 colonoscopy simulator (n=17) Patient-based training (n=18)	Quantitative data were analysed using the paired Student's *t* test. Categorical data analysis was conducted using Fisher's exact test. For ordinal data, a two-sample *t* test was used to compare variables that were normally distributed, and the Mann-Whitney *U* and Kruskal-Wallis tests were used for variables that were not normally distributed. For both the simulator assessments and the patient assessments, each case was analysed separately and then as a combined score from all three cases. Logistic regression was used to investigate the change in scores from pre-test to post-test for the simulator assessments. A *p*-value of <0.05 was considered significant.
Franzeck et al 2012* RCT, Level II N=24	LAP Mentor and PROMIS simulators (n=12) Patient-based training (n=12)	A sample size of 12 participants in each group will have 80% power to detect a difference in means of 100 minutes (assuming a mean of 300 ± 15 minutes camera training for the simulator group and a mean of 400 ± 115 minutes for the OR group) using a two-group Satterthwaite *t* test with a 0.050 two-sided significance level. Statistical analysis was performed using standard software SPSS version 16 for Windows. To compare continuous variables between two groups the Mann-Whitney *U* test was used. Categorical variables were compared using the χ^2 test or, when appropriate, Fisher's exact test. A *p*-value <0.05 was considered to indicate statistical significance. An inter-rater reliability analysis calculating the single-measure ICC was performed to determine consistency among raters of the videotapes.

* electronic publication available in September 2011.
** two comparators used in this study.
ANOVA: analysis of variance; ANTS: Anaesthetists/Anaesthesiologists non-technical skills rating scale; ASA: America Society of Anaesthesiologists physical status levels 1 to VI; BMI: Body Mass Index; CPB: cardiopulmonary bypass; ES: effect size; GEE: generalised estimating equations; ICC: intraclass correlation; JMP®: statistical discovery software; OR: operating room; RCT: randomised controlled trial; SAS: statistical analysis system computer software/company; SPSS: statistical package for social sciences computer software/company; USA: United States of America.

4. Results

The evidence pertaining to skills transfer was obtained from participants' performance in the OR. The different parameters measured during OR performance have been grouped together as closely as possible into tables; however, there were many differences in assessment tools and techniques.

Twenty studies compared simulation-based training with no simulation-based training. Of the remaining three studies, two used patient-based training, and another used interactive seminar-based education, as the comparator. Of the 20 studies comparing simulation-based training with no simulation training, one study compared two different simulation-based training methods with two comparators (no training [control] and didactic lecture-based training).

Studies were categorised initially by the non simulation-based training method (i.e. simulation-based training versus no simulation-based training; simulation-based training versus didactic lecture-based education; simulation-based training versus interactive seminar-based education; and simulation-based training versus patient-based training). Studies were then categorised according to the intervention and the level of evidence.

In addition to participants' performance in the OR, Sturm et al (2007) presented results of participants' assessment on the simulators. In this update, only Ossowski et al (2008) and Haycock et al (2010) reported performance data from the simulators.

Skills transfer outcomes

There was considerable variation between studies in the reporting of performance data (metrics) during patient-based assessment procedures. These included using a mean or median global rating score, mean errors for the entire procedure, mean task-specific checklist score, mean injection/dissection errors, performance accuracy, and patient comfort/discomfort scores. Performance measures used in the 23 studies are summarised in Table 5.

Table 5 Description of performance measures used in included studies

Study	Training method	Performance measures in the operating room
Simulation-based training versus no simulation-based training		
Ahlberg et al 2007 RCT, Level II N=13	LapSim simulator (n=7) No simulator training (n=6)	For full laparoscopic cholecystectomy (recorded on videotape and assessed by two observers). The outcome measures were: 1. mean total number of errors (data for 28 defined surgical errors) 2. total surgical time for full laparoscopic cholecystectomy
Belyea et al 2011 Comparative study, Level III-3 N=42	Eyesi simulator (n=17) No simulator training (n=25)	This cataract removal (phacoemulsification) required simultaneous use of surgeon's two hands and two feet while the surgeon viewed through a light microscope. The outcome measures were 1. phacoemulsification (phaco) time 2. percentage phacoemulsification power used 3. adjusted phacoemulsification time calculated by multiplying the phacoemulsification time by the phacoemulsification power 4. intraoperative complications.
Beyer et al 2011 Comparative study, Level III-3 N=19	LAP Mentor simulator (n=6) PROMIS simulator (n=6) No simulator training (n=7)	For full laparoscopic cholecystectomy (recorded on videotape and assessed independently by two expert surgeons, who were properly trained in GOALS assessment and blinded to training status). GOALS validated by Vassiliou et al (2005). Each domain is scored from 1 (worst) to 5 (best) and the results were summed to get a total score. GOALS scores on 5-point scale assessed: 1. perception of depth 2. bimanual dexterity 3. efficiency, 4. tissue handling (Beyer et al 2011 modified total score to maximum 20) 5. autonomy was excluded from statistical analysis as it was too difficult to evaluate degree of assistance and counselling by the senior surgeon from the videotape.
Banks et al 2007 RCT, Level II N=20	Limbs and Things Ltd. laparoscopic simulator (n=10) No simulator training (n=10)	For laparoscopic bilateral tubal ligation, one of two attending observers rated performance using three tools: 1. a task-specific checklist score 2. a global rating scale (GRS) score (validated by Reznick et al 1997, a five-point Likert scale that assessed seven aspects of surgical skill: respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants and knowledge of the specific procedure) 3. pass-fail grade.
Cosman et al 2007 RCT, Level II N=10	LapSim simulator (n=5) No simulator training (n=5)	All participants applied clips to and divided either the cystic duct or the cystic artery during laparoscopic cholecystectomy. The videotape recording was assessed independently by five experienced laparoscopic surgeons using: 1. validated error assessment scale (Eubanks et al 1999) 2. median value of errors 3. median value of bimanual coordination Index (five-point Likert scale) 4. median global score (five-point Likert scale; validation not stated.) 5. median value of time to task completion.

Study	Training method	Performance measures in the operating room
Fried et al 2010 RCT, Level II N=25	Endoscopic Sinus Surgery Simulator (n=12) No simulator training (n=13)	The first in vivo ESS procedure performed by all participants, standardised around the completion of five basic tasks, was video recorded and independently assessed by three expert otolaryngologists. The measured variables included case difficulty, task completion rate and six significant results as reported: 1. injection time (minutes) 2. dissection time (minutes) 3. injection errors (total number) 4. surgical confidence (1–10 scale) 5. instrument manipulation dexterity (1–10 scale) 6. navigation errors (average number per minute).
Hogle et al 2009 RCT, Level II N=12	LapSim simulator (n=6) No simulator training (n=6)	During two laparoscopic cholecystectomies, the supervising attending surgeon evaluated participants' performance using the validated GOALS. The videotapes were then used for subsequent blinded evaluation and scoring with GOALS (validated by Vassiliou et al 2005). 1. Mean GOALS score for five domains: depth perception, bimanual dexterity, efficiency, tissue handling and autonomy (Maximum = 25).
Howells et al 2008 RCT, Level II N=20	Arthroscopy knee bench-top model (n=10) No simulator training (n=10)	The same consultant knee surgeon who was blinded to their training status assessed all participants' performance using the intraoperative technique section of the Orthopaedic Competence Assessment Project (OCAP) procedure-based assessment for diagnostic arthroscopy. Nine of the 14 OCAP criteria were relevant but the OCAP rating two-point scale (satisfactory/unsatisfactory) was modified to a five-point scale for the global rating scale (GRS) score (as devised and validated by Reznick et al 1997), and a score of 1 (unsatisfactory) to 5 (excellent) was assigned for each. 1. OCAP checklist. 2. Total GRS score for nine relevant skills of OCAP: follows protocol, handles tissue well, appropriate and safe use of instruments, appropriate pace with economy of movement, acts calmly and effectively with untoward events, appropriate use of assistant, communicates with scrub nurse, clearly identifies common abnormalities, and protecting the articular surface.
Kälström et al 2010 RCT, Level II N=23	PelvicVision (n=11) No simulator training (n=12)	1. Task-specific checklist with 12 items. 2. Global five-point ratings scale including 13 items: respect for tissue, time and movements, eye-hand coordination, foot pedals, videoscope, resection, strategy, tempo, use of assistance, stress level, supervision, communication with supervisor, and knowledge about procedure, plus final result and pass/fail score. 3. Pass/fail score. 4. Self-evaluation. 5. Patient follow-up 6–12 months postoperation using the International Prostate Symptom Score, the bother question maximum urinary flow rate and incontinence score (Linköping incontinence questionnaire).
Larsen et al 2009 RCT, Level II N=24	LapSim Gyn v3.0.1 (n=13) No simulator training (n=11)	Technical performance was measured as total score (10–50 points) using the objective structured assessment of laparoscopic salpingectomy, which comprises a five item general rating scale and five item task specific rating scale as previously validated (Larsen et al 2008). Two independent observers who were blinded to trainee and allocated group assessed the recorded operations. 1. Median total score (10 items). 2. Median total time (minutes) to complete laparoscopic salpingectomy procedure.

Study	Training method	Performance measures in the operating room
Ossowski et al 2008 RCT, Level II N=20	Nasal model endoscopic simulator (n=10) No simulator training (n=10)	All students were timed performing flexible laryngoscopy on a single standardised patient who was blinded to which group the student belonged. The standardised patient filled out a comfort/discomfort score for each student using a visual analogue scale from 0–10. 1. Mean time to complete task (seconds). 2. Comfort/discomfort score by standardised patient.
Palter et al 2011 RCT, Level II N=18	Abdominal wall simulator (n=9) No simulator training (n=9)	At the completion of the fascial closure in a real patient whose surgery required an abdominal incision: 1. each participant completed a multiple-choice test of 22 questions designed to assess how much information they retained from the 10-minute script read to them during the procedure 2. the surgical supervisor assessed the resident's technical performance using Objective Structured Assessment of Technical Skills (OSATS) global rating scale (validated Reznick et al 1997; Martin et al 1997).
Park et al 2007 RCT, Level II N=24	AccuTouch colonoscopy simulator (n=12) No simulator training (n=12)	During the clinical colonoscopy, the resident acted as the primary endoscopist under the supervision of one of three faculty endoscopist evaluators who were blinded to the resident's training group. The maximum time allowed for the resident portion of the case was 30 minutes. The test was considered complete and the recording of time stopped when the caecum was reached. A score of 1 to 5 was assigned for each of seven items (validated by Reznick et al 1997) and summed to generate a total global rating score (maximum score is 35). 1. Mean total GRS score – seven items included atraumatic technique, colonoscope use/advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure and overall performance. 2. Outcomes on nominal scales (number reaching caecum and penalty points for critical flaws, i.e. perforation or significant bleeding during the procedure)
Patel et al 2012* RCT, level II N=60	Second Life operating theatre (n=15) Imperial College simulated operating suite (n=15) Didactic lecture** (n=15) No training (n=15)	The primary outcomes assessed in this study were knowledge, skills and attitudes in the OR. These were assessed using a behavioural scale covering predefined elements for preoperative, intraoperative and postoperative components, and a self-report scale detailing participants' perceived knowledge, skills, and attitudes. Both consisted of a Likert-type checklist rating scale. The content domain for all scales was developed on the basis of interviews with trainers and trainees to determine appropriate content. An emergent theme analysis was performed to highlight integral components for the rating scales. Knowledge was further assessed using multiple-choice questions to establish information deemed necessary after the interviews.
Schout et al 2009 RCT, level II N=100	URO Mentor simulator (n=50) No simulator training (n=50)	The supervisors of the real-time cystourethroscopy (CUS), who were unaware of the participants' training status, scored performance using a five-point global rating scale (GRS), a modification of an assessment method that was validated by Matsumato et al (2001). Mean score and standard deviation for each item were reported. This scale evaluates five domains (three of which are concerned with technical aspects of the performance while the remaining two relate more to non-technical performance such as cognition and knowledge): 1. respect for tissue 2. time and motion 3. handling of endoscope 4. flow of procedure and forward planning 5. knowledge of procedure.

Study	Training method	Performance measures in the operating room
Sroka et al 2010 RCT, Level II N=16	FLS simulator (n=8) No simulator training (n=8)	The participants' performance of dissection of the gall bladder from the liver bed during laparoscopic cholecystectomy, was evaluated in the OR room at baseline and at the end of the study by the attending surgeon and/or an evaluator blinded to their training status. The evaluators used the validated GOALS which measures performance in five domains: three are specific for laparoscopic surgery (depth perception, bimanual dexterity, tissue handling) while two are more generic (efficiency and autonomy). Each domain was scored from 1 (worst) to 5 (best) and the results are summed to get a total score maximum=25. 1. Mean total GOALS score (sum of score for five domains).
Shirai et al 2008 RCT, Level II N=20	GI Mentor II simulator (n=10) No simulator training (n=10)	Each subject performed oesophagogastroduodenoscopy (OGD) twice on volunteers without sedation for assessment. Supervisors who were blinded to training group of participants evaluated performance using a five-point scale for a total of 11 items which was modified for real life from that validated by Moorthy et al (2004). This included establishing objective criteria with a time limit for each item and up to three attempts being allowed for four items (^). The eleven items are: insertion into the oesophagus^, crossing oesophagogastric junction (OGJ)^, passing from OGJ into the gastric antrum, passing through the pyloric ring^, examination of the duodenum bulb, insertion in to the third part of the duodenum^, examination of the gastric antrum, examination of the gastric angle, manipulation for retroflexion, looking down the gastric body, and viewing the fornix. The evaluation was performed simultaneously by two supervisors, but evaluation forms were filled out independently and assigned scores were different. The mean score was used for the analysis. 1. Median scores for 11 items and two procedures. 2. Median total procedure time.
Van Sickle et al 2008 RCT, Level II N=22	MIST-VR task 3 and standard box trainer, foam Nissen suturing model and the intracorporeal slip-square knot (n=11) No simulator training (n=11)	Participants performed two consecutive sutures of a standardised three-suture laparoscopic Nissen fundoplication after the attending surgeon placed the first suture. Recordings of the operative performance for each group were reviewed by two surgeon investigators who were blinded to training group and operative team members. Performance was scored based on a set of tightly predefined errors which had been previously validated (Van Sickle et al 2007). Suturing errors covered the 11 predefined suturing steps. 1. Total time (seconds). 2. Total suturing errors. 3. Excess needle manipulations.
Yi et al 2008 Comparative study, Level III-2 N=11	KAIST-Ewha Colonoscopy Simulator II (n=5) No simulator training (n=6)	Each subject performed colonoscopies on five different patients. The supervising experts evaluated the trainees based on accuracy of the colonoscopy results and the established performance criteria. Questionnaires on the colonoscopy experience were also filled out by the patients. 1. Average total insertion time (minutes). 2. Average overall performance accuracy (scale where 1 = poor and 5 = excellent). 3. Number of red-outs. 4. Number of air inflation. 5. Number of loop formation. 6. Mucosal visualisation (scale where 1 = poor to 5 = excellent). 7. Number of abdominal pressure. 8. Changes in patient's posture. 9. Extent of abdominal pain (scale where 1 = no pain to 5 = worst pain). 10. Extent of abdominal inflation (scale where 1 = no pain to 5 = worst pain). 11. Extent of anus discomfort (scale where 1 = no pain and 5 = worst pain).

Study	Training method	Performance measures in the operating room
Zendejas et al 2011 RCT, Level II N=50	Guildford MATTU TEP hernia task trainer (n=26) No simulator training (n=24)	Each subject performed a laparoscopic totally extraperitoneal inguinal hernia repair on patients deemed suitable by the staff surgeon who was blinded to trainee randomisation. Totally extraperitoneal hernia repairs were standardised to follow the three midline port technique. Trainees performed as much of the repair as possible, as deemed appropriate by the supervising staff surgeon. Cases were video recorded in a de-identified fashion. Operative performance graded by GOALS tool validated by Vassiliou et al (2005), and Sroka et al (2010). Assessment items included: 1. operative time (standardised recording by single, trained investigator). 2. operative performance (GOALS). 3. the percentage of trainee participation (number and duration of staff takeovers recorded by single, trained investigator). 4. intraoperative complications (recorded by observer and supervising surgeon). 5. postoperative complications (abstracted from patient medical records.) 6. need for overnight stay. 7. recurrence of the inguinal hernia repair. 8. groin pain.
Simulation-based training versus *didactic lecture-based education		
Patel et al 2012* RCT, level II N=60	Second Life operating theatre (n=15) Imperial College simulated operating suite (n=15) Didactic lecture** (n=15) No training (n=15)	The primary outcomes assessed in this study were knowledge, skills and attitudes in the OR. These were assessed using a checklist behavioural assessment scale covering predefined elements for preoperative, intraoperative and postoperative components, and a self-report scale detailing participants' perceived knowledge, skills, and attitudes. Both consisted of a Likert-type checklist rating scale. The content domain for all scales was developed on the basis of interviews with trainers and trainees to determine appropriate content. An emergent theme analysis was performed to highlight integral components for the rating scales. Knowledge was further assessed using multiple-choice questions to establish information deemed necessary after the interviews. In addition a multiple-choice questionnaire was used to measure participants' knowledge of OR procedures.
Simulation-based training versus interactive seminar education		
Bruppacher et al 2010 RCT, Level II N=20	SimMan Universal Simulator (n=10) Interactive seminar education (n=10)	For weaning a patient from cardiopulmonary bypass (CPB) following cardiac surgery the validated Anaesthetists' Nontechnical Skills (ANTS) global rating scale (validated by Fletcher et al 2003) was used to assess trainees' cognitive and behavioural performance. A checklist of expected clinical actions for CPB weaning was developed specifically for the purpose of this study to assess technical skills of trainees. 1. global rating score (ANTS) – four domains: task management, team working, situation awareness and decision-making assessed using a four-point rating scale – 4: good, 3: acceptable, 2: marginal, 1: poor. 2. a checklist score for CPB weaning developed specifically for this study using Delphi method with one cardiac surgeon and four cardiac anaesthesiologists with 80% agreement for each task through iterative process.

Study	Training method	Performance measures in the operating room
Simulation-based training versus patient-based training		
Haycock et al 2010 RCT, Level II N=36	Olympus Endo TS-1 colonoscopy simulator (n=17) Patient-based training (n=18)	Assessment of proficiency on real colonoscopy was measured using validated structured assessment tools-the UK Joint Advisory Group (JAG) on Gastrointestinal Endoscopy Direct Observation of Procedural Skills (DOPS) assessment form (Colonoscopy 2008), and the global rating score (as validated by Reznick et al 1997 and Park et al 2010). During three patient-based colonoscopies, an expert assessor who was blinded to the group allocation of the trainee supervised the procedure which was limited to 20 minutes. 1. JAG DOPS assessment score (14 major criteria) 2. global rating score (GRS) score 3. mean time to completion (and completion of case). 4. mean straight depth of insertion.
Franzeck et al 2012* RCT, Level II N=24	LAP Mentor and PROMIS simulators (n=12) Patient-based training (n=12)	All participants performed camera navigation skills assessment test on real patients in OR at the beginning of an actual operation. All patients were placed in supine position. Participants were positioned on the patient's right side and were given 30^0 angled laparoscope introduced into the trocar. They had to centre and hold for five seconds the following positions/organs and had to maintain the correct horizontal alignment during camera movement: left abdominal wall, ascending colon, right lobe of liver, sigmoid colon, caecum, pelvis, trocar entry site in the upper left quadrant (simulated by a finger pressing externally), and descending colon. Maximum duration of the test was set at five minutes. This assessment was videotaped. Criteria assessed included: 1. time to completion 2. organ visualisation 3. horizontal alignment 4. correct scope rotation handling 5. visuospatial tests.

*electronic publication available in September 2011.
** two comparators in this study
ANTS: Anaesthetists'/Anaesthesiologists' Nontechnical Skills; CPB: cardiopulmonary bypass; CUS: cystourethroscopy; DOPS: Direct Observation of Procedural Skills; FLS: Fundamentals of Laparoscopic Surgery; GOALS: Global Operative Asessment of Laparoscopic Skills; GRS: global rating scale; JAG: Joint Advisory Group on gastrointestinal endoscopy; MATTU: Minimal Access Therapy Unit; MIST-VR: Minimally Invasive Surgical Trainer−Virtual Reality; OCAP: Orthopaedic Competence Assessment Project; OGD: oesophagogastroduodenoscopy; OGJ: oesophagogastric junction; OR: operating room; OSATS: Objective Structured Assessment of Technical Skills; RCT: randomised controlled trial; TEP: totally extraperitoneal; UK: United Kingdom.

Overall performance of patient-based procedure

Thirteen of the 23 studies reported an overall performance parameter that was a global summary of all the objective performance parameters measured during patient-based assessment procedures or the assessor's evaluation of overall performance (Table 6; see Table 5 for details of all performance measures in the OR). Two additional studies did not report an overall performance parameter but have been included in Table 6 as they either used the validated Global Operative Assessment of Laparoscopic Skills (GOALS) rating scale (Hogle et al 2009) or a validated five-point competency scale (Shirai et al 2008) for individual tasks.

Eight studies did not report an overall performance parameter using a global rating score (Ahlberg et al 2007; Belyea et al 2011; Franzeck et al 2012; Fried et al 2010; Ossowski et al 2008; Patel et al 2012; Van Sickle et al 2008; Yi et al 2008). However, Yi et al (2008) reported overall performance accuracy and this has been included in Table 6 also. Patel et al (2012) reported a checklist behavioural observation scale assessment of knowledge, attitude and skills in the operating room (OR) of first year medical students who had received simulation-based training or no training before their second attendance in the OR. These results have also been included in Table 6.

Simulation-based training versus no simulation-based training

Fifteen studies reported on the overall performance in the clinical setting between simulation-trained and non simulation-trained participants (Table 6). Beyer et al (2011), Hogle et al (2009), Sroka et al (2010) and Zendejas et al (2011) assessed performance using the validated GOALS (maximum score of 25) and Beyer et al (2011) used a modified the total GOALS score (maximum 20) as explained in Table 5. Various scales were used by Banks et al (2007), Cosman et al (2007), Kälström et al (2010), Howells et al (2008), Palter et al (2011), Park et al (2007) Schout et al (2009), Shirai et al (2008), Patel et al (2012) and Yi et al (2008) (Table 5).

Comparison between studies is made more difficult by the variation in the reporting of overall performance. Beyer et al (2011), Park et al (2007), Schout et al (2009), and Sroka et al (2010) reported mean global assessment scores, and Patel et al (2012) reported mean behavioural observation scale scores. Cosman et al (2007, Larsen et al (2009), Palter et al (2011) and Howells et al (2008) used median global assessment scores. Yi et al (2008) reported average scores for overall performance accuracy. Banks et al (2007) reported mean values of global assessment as a percentage of a maximum score of 35, and Hogle et al (2009) reported mean scores for all five GOALS domains. Shirai et al (2008) reported graphs for all 11 items assessed, indicating the median scores, but not an overall score, and Zendejas et al (2011) reported overall performance ratings (GOALS scale 6–30) as mean difference between the groups. Kälström et al (2010) reported mean values for global assessment as a percentage of maximum score of 35 for operations one and three only, and the difference in performance with simulator practice between the procedures was compared with two procedures without simulator practice.

Table 6 Patient-based assessments: overall performance results

Simulation-based training versus no simulation-based training							
Laparoscopic cholecystectomy							
Cosman et al 2007 RCT Level II	Intervention	N=10	Overall assessment, median global score [five-point Likert scale, maximum is five points]				
			Baseline		After training		
	LapSim VR training	5	NA		3.2		
	No simulator training	5	NA		1.8		
	p-value		NA		0.04		
Hogle et al 2009 RCT Level II	Intervention	N=12	Assessment score for each domain†, GOALS score, mean ± SD [five-point scale, maximum is five points for each domain†]				
			1	2	3	4	5
	LapSim VR training	6	3.6 ± 0.55	3.17 ± 0.42	2.89 ± 0.53	2.96 ± 0.59	3.23 ± 0.44
	No simulator training	6	3.5 ± 0.62	2.90 ± 0.51	2.82 ± 0.62	3.10 ± 0.53	3.11 ± 0.62
	p-value		0.99 NS	0.55 NS	0.93 NS	0.56 NS	0.85 NS
Sroka et al 2010 RCT Level II	Intervention	N=16	Overall assessment, Total GOALS score, mean ± SD [five-point scale, five domains so maximum score is 25 points]				
			Baseline		After training		
	FLS training	8	11.3 ± 2.0		17.4 ± 1.9 (*p*<0.0001 versus baseline)		
	No simulator training	8	12.0 ± 1.8		13.8 ± 2.2 (*p*=0.04 versus baseline)		
	p-value		0.47*		0.0003		
	Intervention	N=16	Difference from baseline score for each domain†, mean difference in GOALS score, ± SD				
			1	2	3	4	5
	FLS training	8	1.25 ± 0.7	1.25 ± 0.6	1.13 ± 1.0	1.13 ± 1.0	0.6 ± 1.1
	No simulator training	8	0.5 ± 0.8	0.5 ± 1.1	0.4 ± 1.1	0.3 ± 0.7	0.3 ± 1.0
	p-value		0.08	0.04	0.24 NS	0.04	0.58 NS
Beyer et al 2010 Comparative study Level III-3	Intervention	N=19	Overall assessment, Total GOALS score , mean [five-point scale, first four domains† only so maximum score is 20 points]				
			Baseline (GOALS 1)		After training (GOALS 2)		
	MISTELS training	6	9.33		12.41 (*p*=0.04 versus baseline)		
	LAP Mentor training	6	9.17		13.17 (*p*=0.03 versus baseline)		
	No simulator training	7	12.21		11.85 (*p*=0.35 NS versus baseline)		
	Progression (GOALS 2 – GOALS 1)				MISTELS versus control (*p*=0.03) LAP Mentor versus control (*p*=0.007) MISTELS versus LAP Mentor (*p*=0.28 NS)		
Laparoscopic bilateral tubal ligation							
Banks et al 2007 RCT Level II	Intervention	N=20	Overall assessment, mean global assessment score (SD) [five-point scale, seven tasks so maximum score is 35 points i.e. 22.3 = 64%]				
			Baseline		After training		
	Limbs and Things laparoscopic simulator training	10	9.1		22.3 (5)		
	No simulator training	10	10.2		15.8 (11)		
	p-value		0.8 (Wilcoxon rank sum test)		0.003 (Wilcoxon rank sum test)		
	Intervention	N=20	Overall assessment, mean task-specific checklist score (SD) [25-point scale, four categories so maximum score is 100 points]				
			Baseline		After training		
	Limbs and Things laparoscopic simulator training	10	25		92 (7)		
	No simulator training	10	26		57 (20)		
	p-value		0.9 (Wilcoxon rank sum test)		0.002 (Wilcoxon rank sum test)		

Simulation-based training versus no simulation-based training				
Laparoscopic salpingectomy				
Larsen et al 2009 RCT Level II	Intervention	N=24	Overall assessment, total score – median (range; IQR range) [five-point scale, 10 items maximum score is 50 points]	
			Baseline	After training
	LapSim Gyn VR training	13	NA	33 (25–39; 32–36)
	No simulator training	11	NA	23 (21–28; 22–27)
	p-value		NA	<0.001
Laparoscopic diagnostic arthroscopy of the knee				
Howells et al 2008 RCT Level II	Intervention	N=20	Overall rating, total GRS score as median (IQR; range) [five-point scale, nine skills so maximum is 45 points]	
			Baseline	After training
	arthroscopy knee bench-top simulator training	10	NA	24 (17–29; 11–38)
	No simulator training	10	NA	10 (9–12; 9–26)†
	p-value		NA	0.0011
Laparoscopic totally extraperitoneal inguinal hernia repair				
Zendejas et al 2011 RCT Level II	Intervention	N=50	Overall assessment, total GOALS score (range 6–30), mean ± SD [five-point scale, five domains so maximum score is 25 points]	
			Baseline first hernia repair (TEP#1)	After training (all totally extraperitoneal inguinal hernia repairs)
	totally extraperitoneal inguinal hernia repair simulator training	26	21.9 ± 2.7	23.3 ± 3.0
	No simulator training	24	18.0 ± 3.8	18.7 ± 3.8
	p-value		0.001	0.0001
Colonoscopy				
Park et al 2007 RCT Level II	Intervention	N=24	Overall GRS score, mean ± SD [five-point scale, seven items so maximum score is 35 points]	
			Baseline	After training
	AccuTouch colonoscopy simulator training	12	NA	17.9 ± 5.2
	No simulator training	12	NA	14.8 ± 2.5
	p-value		NA	0.04
Yi et al 2008 Comparative study Level III-2	Intervention	N=11	Overall performance accuracy, average (SD) [five-point scale (1 poor, 5 excellent) maximum score is five points]	
			Baseline	After training
	KAIST-Ewha Colonoscopy Simulator II training	5	NA	3.6 (0.8)
	No simulator training	6	NA	2.7 (0.8)
	p-value		NA	<0.001

Simulation-based training versus no simulation-based training							
Endoscopy/Cystourethroscopy							
Schout et al 2009 RCT Level II	Intervention	N=100	Overall GRS score, mean (SD) [five-point scale]				
			Baseline		After training		
	URO Mentor VR cystourethroscopy simulator training	50	NA		3.8 (1.2)		
	No simulator training	50	NA		3.0 (1.0)		
	p-value		NA		<0.001		
	Intervention	N=100	Assessment score for each of five domains^, mean (SD) [five-point scale]				
			1	2	3	4	5
	URO Mentor VR cystourethroscopy simulator training	50	3.6 (1.0)	3.6 (1.2)	3.7 (1.2)	3.9 (1.0)	4.0 (1.1)
	No simulator training	50	3.0 (1.3)	2.9 (1.2)	2.9 (1.2)	3.1 (1.3)	3.2 (1.3)
	p-value		0.003	0.001	<0.001	0.001	0.001
Endoscopy/ transurethral resection of the prostate (TURP)							
Kälström et al 2010 RCT Level II	Intervention	N=23	Overall global assessment (5-point rating scale; 13 items), mean score as %				
			Baseline TURP #1	After training TURP#2	After training both groups TURP#3		
	PelvicVision simulator training	11	NR	NR	Graph shows improved scores from TURP#1; 58% to 75%		
	No simulator training	12	NR	NR			
	p-value		NR	NR	0.0000		
Endoscopy/oesophagogastroduodenoscopy							
Shirai et al 2008 RCT Level II	Intervention	N=20	Median scores (5-point scale) plotted for all 11 items – no overall score				
			Baseline	After training			
	GI Mentor II VR simulator training	10	NA	Median scores for 5 out of 11 items were significantly higher than no simulator training group scores i.e. insertion of the endoscope into the oesophagus ($p<0.05$), passing the oesophagogastric junction into the antrum ($p<0.01$), passing through the pyloric ring ($p<0.05$), and examination of the duodenal bulb ($p<0.05$) and viewing the fornix ($p<0.05$)		
	No simulator training	10	NA	There was no significant difference between the two groups for 6 out of 11 items i.e. crossing oesophagogastric junction, insertion in to the third part of the duodenum, examination of the gastric antrum, examination of the gastric angle, manipulation for retroflexion, and looking down the gastric body.			
	p-value		NA	NA			

Simulation-based training versus no simulation-based training				
Abdominal fascial closure				
Palter et al 2011 RCT Level II	Intervention	N=18	Overall assessment, OSATS score, median (IQR range) [five-point scale, 7 items; maximum 35]	
			Baseline on model	After training
	Abdominal wall simulator training	9	22.0 (20.5–23.0)	22.0 (20.0–27.0)
	No simulator training	9	21.0 (20.0–21.0)	16.0 (16.0–19.0)
	p-value		0.48	0.04
Knowledge, attitudes and skills in the operation room				
Patel et al 2012** RCT Level II	Intervention	N=60	Behavioural observation assessment score, mean (range) [five-point scale, 30 domains; maximum 150]	
			Baseline	After training
	Second Life VR simulation training	15	69 (51–84)	115 (101–137); *p*=0.001; improvement=25.6%
	Imperial College SOS training	15	76 (56–97)	132 (103–141); p=0.001; improvement=31.1%
	Didactic lecture-based education	15	65 (55–89)	107 (81–122); *p*=0.001; improvement=23.3%
	No training [control]	15	66 (53–82)	69 (56–89); *p*=0.09; improvement=1.7%
	p-value		NR	NR
Simulation-based training versus interactive seminar-based education				
Cardiopulmonary bypass weaning following cardiac surgery				
Bruppacher et al 2010 RCT Level II	Intervention	N=20	Overall rating, total ANTS assessment score, mean ± SD [five-point scale for four categories maximum score is 20]	
			Baseline	After training (post-test i.e. two weeks)
	SimMan Universal Simulator training	10	10.6 ± 0.46	14.3 ± 0.41
	Interactive seminar training	10	10.0 ± 0.46	11.8 ± 0.41
	p-value		0.331	<0.001
	Intervention	N=20	Overall rating, total ANTS assessment score, mean ± SD [five-point scale for four categories maximum score is 20]	
			Baseline	After training (retention-test i.e. four weeks)
	SimMan Universal Simulator training	10	10.6 ± 0.46	14.1 ± 0.41
	Interactive seminar training	10	10.0 ± 0.46	11.7 ± 0.41
	p-value		0.331	<0.001

Simulation-based training versus patient-based training				
Colonoscopy				
Haycock et al 2010 RCT Level II	Intervention	N=36	Overall assessment, global score as median (IQR range) [five-point scale, seven items so maximum score is 35 points]	
			Baseline	After training
	Olympus (Endo TS-1) colonoscopy simulator training	18	NA	16 (14–19)
	Patient-based training	18	NA	17 (14–19)
	p-value		NA	0.35 (NS)
	Intervention	N=36	Overall assessment, UK JAG DOPS score as median (IQR range) [four-point scale, 14 major domains and 6 minor domains, maximum score NR]	
			Baseline	After training
	Olympus (Endo TS-1) colonoscopy simulator training	18	NA	16 (14–22)
	Patient-based training	18	NA	16 (14–21)
	p-value		NA	0.92 (NS)

* No significant difference between the two groups concerning baseline parameters.

** electronic publication available in September 2011.

† 1: depth perception, 2: bimanual dexterity, 3: efficiency, 4: tissue handling, 5: autonomy.

^ 1. Respect for tissue, 2: time and motion, 3: handling of endoscope, 4: flow of procedure and forward planning, 5: knowledge of procedure.

ANTS: Anaesthetists'/Anaesthesiologists' Nontechnical Skills; FLS: Fundamentals of Laparoscopic Surgery; GI: gastrointestinal; GOALS: Global Operative Asessment of Laparoscopic Skills; GRS: global rating scale; IQR: interquartile; JAG DOPS: Joint Advisory Group on gastrointestinal endoscopy Direct Observation of Procedural Skills, UK; KAIST: Korea Advanced Institute of Science and Technology, Seoul, South Korea; MISTELS: McGill Inanimate System for Training and Evaluation of Laparoscopic Skills; NA: not available; NR: not reported; NS: not significant; OSATS: Objective Structured Assessment of Technical Skills; RCT: randomised controlled trial; SD: standard deviation; TEP: totally extraperitoneal; TURP: transurethral resection of the prostate; UK: United Kingdom.

Laparoscopic cholecystectomy

Randomised controlled trials

Cosman et al (2007) reported that those participants who had trained on the LapSim VR simulator received a significantly higher global score than the control group (3.2 versus 1.8; p=0.04) for applying clips during laparoscopic cholecystectomy on a live human patient.

Hogle et al (2009) reported that there was no significant difference between simulator-trained participants and the control group in the GOALS domains of depth perception, bimanual dexterity, efficiency, tissue handling or autonomy during two consecutive elective laparoscopic cholecystectomies performed one month after training (see Table 6).

Sroka et al (2010) reported that the FLS simulator-trained group improved significantly more than the control group (17.4 ± 1.9 points versus 13.8 ± 2.2; p=0.0003). FLS simulator-trained participants significantly improved their total GOALS score for performing laparoscopic cholecystectomies by a mean of 6.1 ± 1.3 points, improving from 11.3 ± 2.0 to 17.4 ± 1.9 (p<0.0001 versus baseline). In contrast the non simulator-trained group only improved by 1.8 ± 2.1 points from baseline (12.0 ± 1.8 to 13.8 ± 2.2; p=0.04). Of the five GOALS domains evaluated, simulator training was associated with greater improvements in the three laparoscopic-specific domains (bimanual dexterity, tissue handling and depth perception) compared with the more generic domains (efficiency and autonomy).

Comparative study

Beyer et al (2011) reported that the group that trained on the simulators improved their mean GOALS score significantly while the control group did not improve their mean GOALS score. MISTELS simulator-trained participants improved their mean GOALS scores significantly from 9.33 to 12.41 (p=0.04), and LAP Mentor simulator-trained participants also significantly improved (from 9.17 to 13.17; p=0.03). Control group participants suffered a decline in mean GOALS score from 12.21 to 11.85 (p=0.35). There was a significant difference in favour of the MISTELS group versus the control group (p=0.03) and in favour of the LAP Mentor group versus the control group (p=0.007). There was no significant difference between the MISTELS and LAP Mentor groups (p=0.28).

Laparoscopic bilateral tubal ligation

Randomised controlled trials

Banks et al (2007) reported on overall performance between Limbs and Things laparoscopic simulator-trained and non simulator-trained participants when they performed their second bilateral tubal ligation. An objective evaluation of resident performance found that simulator-trained participants scored higher overall on the global rating scale than the control group with only surgical teaching in the OR (64% [SD 5] versus 45% [SD 11]; p=0.003). The maximum GRS score was 35; thus the calculated mean global assessment scores are 22.3 for the simulator-trained and 15.8 for the control group as shown in Table 6.

Banks et al (2007) also reported that the simulator-trained group performed significantly higher on their validated task-specific 25-point checklist assessment for four categories than the control group (mean score 92% versus 57%; p=0.002). The maximum score was 100, thus the checklist scores were 92 for the simulator-trained and 57 for the control group.

Laparoscopic salpingectomy

Randomised controlled trials

Larsen et al (2009) reported an overall assessment of technical performance as a total score (10–50 points) between LapSim Gyn VR simulator-trained participants and non simulator-trained participants when they performed laparoscopic salpingectomy. The median total score on the general and task specific rating scale reached 33 points (interquartile range [IQR] 32–36 points) in the simulator-trained group and 23 points (IQR 22–27 points) in the control group (p<0.001). The authors stated that compared with standard clinical education, proficiency-based VR training in laparoscopic salpingectomy was associated with a clinically important improvement of operative skills during the actual procedure. The previous validation of the rating scale (Larsen et al 2006) indicated that a median of 33 points was equivalent to that of an intermediately experienced gynaecologist assessed in their first complex procedure, while novices (fewer than five procedures) scored a median of 24 points.

Laparoscopic diagnostic arthroscopy of the knee

Randomised controlled trials

Howells et al (2008) reported a total global rating score for arthroscopy knee bench-top simulator-trained participants and non simulator-trained participants. Analysis of performance in

the OR showed that the simulator-trained group significantly outscored the control group. The median total score reached by the simulator-trained group was 24 points (range 17–29; IQR 11–38) and 10 points (range 9–12; IQR 9–26) in the control group (p=0.0011).

Laparoscopic totally extraperitoneal inguinal hernia repair

Randomised controlled trial

Zendejas et al (2011) compared totally extraperitoneal inguinal hernia repairs operative performance ratings (GOALS scale 6–30) for simulator-trained participants and controls immediately after training and the mean difference was significantly +3.6 (range 2.1–5.1; p=0.001). The authors also compared operative performance ratings for all totally extraperitoneal inguinal hernia repairs postrandomisation (mean difference +3.7; range 1.8–5.6; p=0.0001). The authors reported that operative performance ratings were better for those trained to mastery on a simulator than for those participants who had no simulation training.

Colonoscopy

Randomised controlled trials

Park et al (2007) reported an overall global rating score for AccuTouch colonoscopy simulator-trained participants and non simulator-trained participants. The mean total global rating score for the simulator-trained group was 17.9 (SD 5.2) out of a maximum possible score of 35, which was significantly higher than the mean score for the control group of 14.8 (SD 2.5; p=0.04).

Comparative study

Yi et al (2008) reported an overall performance accuracy for KAIST-Ewha Colonoscopy II simulator-trained participants and non simulator-trained participants. The average score was significantly higher in the simulator-trained group at (3.6, SD 0.8) than in the control group (2.7, SD 0.8; p<0.001).

Endoscopy/cystourethroscopy

Randomised controlled trials

Schout et al (2009) reported an overall GRS score for URO Mentor VR cystourethroscopy simulator-trained participants and non simulator-trained participants. The mean total GRS score was significantly higher in the simulator-trained group (3.8, SD 1.2) than in the control group (3.0, SD 1.0; p<0.001). The mean GRS scores for the five domains: respect for tissue, time and motion, handling of endoscope, flow of procedure and forward planning, and knowledge of procedure were all significantly higher in the simulator-trained group than in the control group (see Table 6).

Endoscopy/transurethral resection of the prostate

Randomised controlled trial

Kälström et al (2010) measured the transurethral resection of the prostate (TURP) operative performance using a global assessment five-point rating scale of 13 items on day one (TURP#1) of a five-day course. The participants were then randomised into two groups. Group one was trained on the simulator before undertaking a second transurethral resection of the prostate (TURP#2) while group two performed a second TURP without any simulator training. Group

two then trained on the simulator while group one had no further training. All participants performed a third transurethral resection of the prostate (TURP#3) on day five of the course. The authors compared TURP#1 and TURP#3 and stated that the mean value for global assessment score was significantly higher for TURP#3 than at baseline (p=0.000). Data approximated from the authors' graph showed an estimated increase from 58% to 75% on the global assessment score (p=0.000). The maximum score was 65 so this was calculated to be equivalent to increase in global assessment scores from 37.7 to 48.8 after both groups were simulator-trained. The authors did not report the scores for TURP#1 and TURP#2.

Endoscopy/oesophagogastroduodenoscopy

Randomised controlled trials

Shirai et al (2008) did not report an overall score but evaluated performance according to a five-point competency scale. Eleven items and two endoscopic procedures were evaluated for each group. Manipulative skills were mainly evaluated based on the following items: insertion into the oesophagus, crossing the oesophagogastric junction (OGJ), passing from the OGJ into antrum, passing through the pyloric ring, and insertion into the third part of the duodenum. Observation skills were assessed based on the examination of the gastric antrum and the gastric angle. Both manipulation and observation were evaluated based on items such as the examination of the duodenal bulb, retroflexion, looking down the gastric body, and examining the fornix.

The median scores were significantly higher in the simulator-trained group compared with the control group for five items: insertion of the endoscope into the oesophagus (p<0.05), passing the OGJ into the antrum (p<0.01), passing through the pyloric ring (p<0.05), and examination of the duodenal bulb (p<0.05) and fornix (p<0.05). There was no significant difference between the groups for the other six items: crossing oesophagogastric junction (OGJ), insertion in to the third part of the duodenum, examination of the gastric antrum, examination of the gastric angle, manipulation for retroflexion, and looking down the gastric body.

Abdominal fascial closure

Randomised controlled trial

Palter et al (2011) reported that the OSATS global rating scale assessment of technical skills in the OR was significantly higher in the simulator-trained group (22.0; range 20.0–27.0) compared with the control group (16.0; range 16.0–19.0; p=0.04). There was no difference in technical ability between the two groups before the intervention as measured in the skills laboratory as the median baseline global rating score in the control group was 22.0 (range 20.5-23.0), and for the simulator-trained group 21.0 (range 20.0-21.0; p=0.48 NS). The authors noted that technical measures of performance using the OSATS global rating scale remained stable between initial assessment on the model and the OR assessment in the simulator-trained group (OSATS 21.0 in laboratory to 22.0 in OR), whereas they decreased in the control group (OSATS = 22.0 in laboratory to 16.0 in OR).

Knowledge, skills and attitudes in the operating room

Randomised controlled trial

Patel et al (2012) reported that the behavioural observation scale assessment scores of knowledge, skills and attitudes in the OR were significantly higher in the simulator-trained groups i.e. Second Life VR OR (pretest: mean 35; range 14–43 versus 39; 33–45; p=0.12 and the Imperial College SOS (pretest:38; 26–40 versus post-test: 43; 40–44; p=0.001) compared with the control group (pretest: 66; 53–82 versus post-test: 69; 56–89; p=0.09). The control group did not display any significant improvement (1.7%; p=0.09). The Imperial College SOS group demonstrated the largest percentage improvement (31.1%) in the behavioural observation assessment scale measure than the Second Life VR OR (25.6%) and didactic lecture (23.3%).

Simulation-based training versus interactive-seminar education

Brupaccher et al (2010) used the validated ANTS global rating tool to assess trainee's cognitive and behavioural performance in CPB weaning (Table 6).

Cardiopulmonary bypass weaning following cardiac surgery

Randomised controlled trials

Bruppacher et al (2010) compared the total ANTS global assessment scale score of participants who trained using the SimMan Universal simulator with the participants who undertook interactive seminar-based education (Table 6). The authors reported that after training, the simulator-trained group significantly outperformed the seminar-trained group in both the post-test at two weeks (mean 14.3 ± SD 0.41 versus 11.8 ± 0.41; p<0.001) and the retention test at five weeks (14.1 ± 0.41 versus 11.7 ± 0.41; p<0.001). In addition, all four components of the ANTS (task management, team working, situation awareness and decision-making) were significantly higher for the simulation-trained group compared with the seminar group (p<0.01).

Simulation-based training versus patient-based training

Haycock et al (2010) reported the overall colonoscopy performance on simulator-trained versus patient-trained participants (Table 6).

Colonoscopy

Randomised controlled trials

Haycock et al (2010) reported on the overall performance between Olympic colonoscopy simulator-trained and patient-trained participants for colonoscopy. There was no significant difference between the mean global rating score in patient-trained participants and the simulator-trained participants (17 versus 16; p=0.35). Similarly, there was no significant difference between the mean JAG DOPS scores (16 versus 16; p=0.92). The authors noted that this is a positive result for simulation-based training, since for the equivalent time spent on the simulator, this equal performance in real colonoscopy demonstrates a high degree of skill transfer from the simulator to real colonoscopy.

Performance time

Performance time was reported as the time taken, in minutes or seconds, to conduct the patient-based assessment procedure. Ten studies did not report performance time (Banks et al 2007; Beyer et al 2011; Bruppacher et al 2010; Hogle et al 2009; Howells et al 2008; Palter et al 2011; Park et al 2007; Patel et al 2012; Schout et al 2009; Sroka et al 2010).

Simulation-based training versus no simulation-based training

Eleven studies reported performance times between simulator-trained and non simulator-trained participants for ten different interventions (Ahlberg et al 2007; Beyea et al 2011; Cosman et al 2007; Fried et al 2010; Kälström et al 2010; Larsen et al 2009; Shirai et al 2008; Ossowski et al 2008; Van Sickle et al 2008; Yi et al 2008; Zendejas et al 2011) (Table 7).

Laparoscopic cholecystectomy

Randomised controlled trials

Ahlberg et al (2007) and Cosman et al (2007) reported procedure or dissection times for laparoscopic cholecystectomy for LapSim VR simulator-trained participants and participants who did not have this training. Ahlberg et al (2007) reported that the laparoscopic cholecystectomy procedure surgical time was, on average, 58% longer in the control group when compared with the simulator-trained group. This difference did not reach statistical significance ($p<0.0586$). Cosman et al (2007) reported that the median time taken to task completion in the laparoscopic cholecystectomy procedure was 172 seconds for the control group. Those who trained on the simulator took less time to complete the assessment task (median 94 seconds), but this difference was not statistically significant ($p=0.075$).

Laparoscopic salpingectomy

Randomised controlled trials

Larsen et al (2009) reported that the median total time taken to complete the laparoscopic salpingectomy procedure was 12 minutes (IQR 10–14 minutes) in the LapSim VR simulator-trained group compared with 24 (20–29 minutes) in the control group ($p<0.001$). These data indicate that the time to complete the laparoscopic salpingectomy was reduced by half for those participants who had trained on the simulator.

Laparoscopic Nissen fundoplication

Randomised controlled trials

Van Sickle et al (2008) reported on participants (simulator- and non simulator-trained) who performed the fundal suturing portion of a laparoscopic Nissen fundoplication with an attending surgeon who was blinded to participant training status. A standardised three-suture fundoplication was performed, with the first and most cephalad suture being placed by the attending surgeon. The remaining two sutures were placed by the participant according to the protocol. The MIST-VR and box trainer simulator-trained group completed the suturing task in significantly less time than the control group on both sutures (total time 525.6 ± 189.6 seconds versus 789.5 ± 171.3 seconds; $p<0.003$).

Table 7 Patient-based assessments: performance time results

Simulation-based versus no simulation-based training				
Laparoscopic cholecystectomy				
Ahlberg et al 2007 RCT, Level II	Intervention	N=13	Duration of procedure (% difference)	
			Baseline	Final
	LapSim VR training	7	NA	NR
	No simulator training	6	NA	58% longer than simulator-trained group
	p-value		NA	<0.0586
Cosman et al 2007 RCT, Level II	Intervention	N=10	Duration of procedure (seconds), median	
			Baseline	Final
	LapSim VR training	5	NA	94
	No simulator training	5	NA	172
	p-value		NA	0.075 (NS)
Laparoscopic salpingectomy				
Larsen et al 2009 RCT, Level II	Intervention	N=24	Duration of procedure (minutes), median (range; IQR)	
			Baseline	Final
	LapSim VR training	13	NA	12 (6–24; 10–14)
	No simulator training	11	NA	24 (14–38; 20–29)
	p-value		NA	<0.001
Laparoscopic Nissen fundoplication				
Van Sickle et al 2008 RCT, Level II	Intervention	N=22	Duration of procedure (seconds), median ± SD	
			Baseline	After training
	MIST-VR and box trainer laparoscopic simulator training	11	NA	525.6 ± 189.6
	No simulator training	11	NA	789.5 ± 171.3
	p-value		NA	< 0.003
Laparoscopic totally extraperitoneal inguinal hernia repair				
Zendejas et al 2010 RCT, Level II	Intervention	N=50	Operative time (minutes), mean ± SD	
			Baseline first hernia repair (TEP#1)	Second hernia repair (TEP#2) after randomisation
	Guildford MATTU totally extraperitoneal inguinal hernia simulator training	26	30.9 ± 7.3	NR (on average 6.5 minutes faster than performance time for controls: 95% CI, difference between groups -10.1 to -2.9)
	No simulator training	24	37.4 ± 8.3	NR
	p-value		0.0005	<0.0001
	Intervention	N=50	Operative time (minutes), mean ± SD	
			Baseline first hernia repair (TEP#1)	All totally extraperitoneal inguinal hernia repairs after randomisation
	Guildford MATTU TEP hernia repair simulator training	26	30.9 ± 7.3	29.6 ± 6.7
	No simulator training	24	37.4 ± 8.3	35.7 ± 7.6
	p-value		0.0005	0.0002
Colonoscopy				
Yi et al 2008 Comparative study, Level III-2	Intervention	N=55	Total insertion time (minutes), average (SD)	
			Baseline	After training
	KAIST-Ewha Colonoscopy Simulator II training	25	NA	31.0 (18.7)
	No simulator training	30	NA	41.5 (21.2)
	p-value		NA	0.028

Simulation-based versus no simulation-based training					
Endoscopy/oesophagogastroduodenoscopy					
Shirai et al 2008 RCT, Level II	Intervention	N=20	Duration of procedure (minutes), mean (range)		
			Baseline	After training	
	GI Mentor II simulator training	10	NA	14:40 (12:15 – 16:07)	
	No simulator training	10	NA	14:05 (13:30 – 16:00)	
	p-value		NA	NR (NS)	
Endoscopy/nasolaryngoscopy					
Ossowski et al 2008 RCT, Level II	Intervention	N=20	Duration of procedure (seconds), mean ± SD (range)		
			Baseline	After training	
	Nasal model for endoscopic simulator training	10	NA	56.60 ± 36.97 (19–138)	
	No simulator training	10	NA	50.20 ± 35.49 (16–124)	
	p-value		NA	0.315 (NS)	
Endoscopic sinus surgery					
Fried et al 2010 RCT, Level II	Intervention	N=25	Completion of mucosal injection task (minutes), mean (SD)		
			Baseline	After training	
	ES3 simulator training	12	NA	1.75 (1.04)	
	No simulator training	13	NA	4.67 (2.09)	
	p-value		NA	0.003	
Fried et al 2010 RCT, Level II	Intervention	N=25	Dissection time (minutes), mean (SD)		
			Baseline	After training	
	ES3 simulator training	12	NA	15.44 (6.46)	
	No simulator training	13	NA	7.37 (3.36)	
	p-value		NA	<0.001	
Endoscopy/transurethral resection of the prostate					
Kälström et al 2010 RCT Level II	Intervention	N=23	Procedure time (minutes), mean score as %		
			Baseline TURP #1	After training TURP#2	After training both groups TURP#3
	PelvicVision simulator training	11	NR	NR	Graph shows decreased scores from TURP#1; 42% to 37%
	No simulator training	12	NR	NR	
	p-value		NR	NR	NR
	Intervention	N=23	mean (range)		
			Baseline	After training	
	Pelvic Vision VR simulator training	11	NR	Graph shows VR-training better 26% versus No training better 74%.	
	No simulator training	12	NR		
	p-value		NR	0.025	
Phacoemulsification cataract surgery					
Belyea et al 2011 Comparative study, Level III-3	Intervention	N=42	Phacoemulsification time (minutes), mean (range)		
			Baseline	After training	
	Eyesi simulator training	17		1.88 (0.11–7.20)	
	No simulator training	25		2.41 (0.04–8.33)	
	p-value			0.002	

Simulation-based training versus patient-based training				
Colonoscopy				
Haycock et al 2010 RCT, Level II	Intervention	N=36	Duration of procedure (minutes), median (IQR)	
			Baseline	After training
	Olympus (Endo TS-1) colonoscopy simulator training	17	NA	20 (19–20)
	Patient-based training	18	NA	20 (20–20)
	p-value		NA	0.11
Laparoscopic camera navigation				
Franzeck et al 2012* RCT, Level II	Intervention	N=24	Time to completion (seconds), median ± SD	
			Baseline	After training
	LAP Mentor and PROMIS simulator training	12	179 ± 64	133 ± 35 (*p*=0.05 versus baseline)
	Patient-based training in OR	12	163 ± 67	111 ± 30 (*p*=0.02 versus baseline)
	p-value		0.554	0.12

*electronic publication was available in September 2011
ES3: endoscopic sinus surgery; IQR: interquartile range; KAIST: Korea Advanced Institute of Science and Technology; MATTU: Minimal Access Therapy Unit; MIST-VR: Minimally Invasive Surgical Trainer−Virtual Reality; NA: not applicable; NR: not reported; NS: not significant; OR: operating room; RCT: randomised controlled trial; SD: standard deviation; TEP: totally extraperitoneal; VR: virtual reality.

Laparoscopic totally extraperitoneal inguinal hernia repair

Randomised controlled trial

Zendejas et al (2011) reported on the first totally extraperitoneal inguinal hernia repair after randomisation (TEP#2), and found that simulator-trained participants performed the inguinal hernia repair on average 6.5 minutes faster (raw time, mean difference: -6.5 (-10.1, -2.9), 95% CI; $p<0.0001$) than control participants. Resident participation was also significantly different in favour of those who were simulator-trained (mean participation 88.4 versus 73.7; $p<0.0001$). After correcting time to account for varying participation rates, the difference between the groups was even greater (participation-corrected time 13.1 minutes faster for simulator-trained (-18.4, -7.8), 95% CI; $p<0.0001$). When evaluating subsequent first totally extraperitoneal inguinal hernia repairs, simulator-trained participants remained faster than their control counterparts (raw time, mean difference: -6.6 minutes (-10.1, -3.2), 95% CI: $p<0.0001$). At the third totally extraperitoneal inguinal hernia repair (TEP#3), crossover simulator-trained participants were also faster than their control counterparts. Similar results were found when adjusting operative time for supervising surgeons, case difficulty, side (first repair versus second repair if bilateral) and baseline operative time.

Colonoscopy

Comparative study

Yi et al (2008) reported that the simulator-trained group significantly outperformed the control group in terms of colonoscopy insertion time. Participants trained on the KAIST-Ewha colonoscopy simulator II had an average total insertion time of 31.8 (SD 18.7) minutes compared with control participants 41.5 (SD 21.2; $p=0.028$).

Endoscopy/oesophagogastroduodenoscopy

Randomised controlled trials

Shirai et al (2008) reported that there was no significant difference in the total oesophagogastroduodenoscopy procedure time between the two groups. The mean total oesophagogastroduodenoscopy procedure time was 14:40 minutes (IQR 12:15–16:07) for the GI Mentor II simulator-trained group compared with 14:05 minutes (13:30–16:00) minutes in the control group (p not reported). The authors also reported that the number of one-point scores (i.e. direct assistance required from the supervisor) in the simulator group was significantly lower than that in the control group (8.6% compared with 25.9%; p=0.0017).

Endoscopy/nasolaryngoscopy

Randomised controlled trials

Ossowski et al (2008) reported that there was no significant difference between the two groups for the flexible laryngoscopy procedure time. The mean procedure time was 56.60 seconds (range 19–138) for the simulator-trained group and 50.20 (16–124) seconds for the control group (p=0.315).

Endoscopic sinus surgery

Randomised controlled trials

Fried et al (2010) reported that the completion time of the mucosal injection task was significantly shorter in the simulator-trained group, and with a narrower variability, than in the control group (mean 1.75 minutes, SD 1.04 versus 4.67 minutes, SD 2.09; p=0.003). Dissection time yielded similar observations (7.4 minutes, SD 3.6 versus 15.44 minutes, SD 6.546; p<0.001).

Endoscopy/transurethral resection of the prostate

Randomised controlled trials

Kälström et al (2010) compared transurethral resection of the prostate (TURP) procedures preceded by and not preceded by simulation practice. The change in scores for two TURP operations with simulation practice in between was compared with scores for two operations without added simulation practice. When the number of participants who improved or showed no change in skills after simulation practice was compared with the results of procedures without simulation practice, there was a significant difference (p=0.021) indicating that simulation practice resulted in increased skills. The authors stated that sixteen participants showed greater improvements after simulation practice compared with seven participants who showed greater improvement without simulation practice (one participant could not be evaluated due to exclusion of patient). Although it was not possible to measure any significant difference in single parameters, there was a significant increase in participant performance time which increased without added simulation practice (p=0.025).

Kälström et al (2010) also reported that there was a significant increase in the amount of autonomous procedure time (p=0.000), resection time (p=0.029), and a tendency to decrease haemostasis time (p=0.073) and increased successful orientation (p=0.078) when first TURP and

third TURP procedures (after both groups had received simulator-based training) were compared.

Phacoemulsification for cataract surgery

Retrospective comparative study

Belyea et al (2011) reported that the simulator-trained group had a significantly lower mean phacoemulsification time ($p<0.002$), and adjusted phacoemulsification time ($p<0.0001$) for cataract surgery performance compared with those who had not received simulation-based training.

Simulation-based training versus patient-based training

Two studies reported performance time outcomes for participants who had undergone simulator-based training in comparison to participants who had received patient-based training (Haycock et al 2010; Franzeck et al 2012) (Table 8).

Colonoscopy

Randomised controlled trials

Haycock et al (2010) reported the time taken to perform a real colonoscopy after participants had been trained either using an Olympus colonoscopy II simulator or by patient-based training. There was no significant difference between the two groups (20 minutes versus 20 minutes; $p=0.11$). The authors stated that the simulator training produced equal performance outcomes to patient-based training on real-life cases, demonstrating a high degree of skills transfer from the simulator to the real colonoscopy.

Laparoscopic camera navigation

Randomised controlled trials

Franzeck et al (2012) reported no significant difference between simulator-trained and OR-trained groups in the time taken to complete the post-training camera navigation evaluation (simulator group 133 ± 35 seconds versus OR group 111 ± 30 seconds; $p=0.12$). However, participants in the OR group spent significantly more overall time in the OR than the simulator-trained group spent in the skills laboratory. Thus, the authors concluded that VR simulator based-training is more time efficient than OR training for laparoscopic camera navigation training.

Performance errors

Performance errors were described as movements or events outside the normal procedure.

Simulation-based training versus no simulation-based training

Error data for simulator-trained and non simulator-trained participants were reported for laparoscopic cholecystectomy (Ahlberg et al 2007; Cosman et al 2007), laparoscopic Nissen fundoplication (Van Sickle et al 2008) and endoscopic sinus surgery (Fried et al 2010) (Table 8).

Laparoscopic cholecystectomy

Randomised controlled trials

Ahlberg et al (2007) reported that the mean number of total errors for the entire laparoscopic cholecystectomy procedure was 28.4 in the LapSim VR simulator-trained group compared with 86.2 in the control group (p=0.0037). The simulator-trained group made significantly fewer objectively assessed intraoperative errors compared with the control group in each part of the entire procedure (exposure errors 15.0 versus 53.4; p=0.0402; clipping and tissue division errors 1.9 versus 7.1; p=0.008 and dissection errors 11.5 versus 29.5; p=0.031). The control group made considerably more errors (on average three times more than the simulator-trained group) and showed greater variability in performance, as evident from the larger SDs.

Cosman et al (2007) reported that for each participant the median value of errors for the laparoscopic cholecystectomy procedure was significantly lower for the LapSim VR simulator-trained group (median 10) compared with the control group (median 18; p=0.05).

Laparoscopic Nissen fundoplication

Randomised controlled trials

Van Sickle et al (2008) reported that participants who had completed the simulation-based laparoscopic suturing curriculum made significantly fewer errors than the standard clinically-trained group (total errors 25.6 ± 9.3 versus 37.1 ± 10.2; p<0.01).

Endoscopic sinus surgery

Randomised controlled trials

Fried et al (2010) reported that the ES3 simulator-trained participants made fewer mucosal injection errors than the control group (3.53 errors, SD 1.96 versus 6.89 errors, SD 3.30; p=0.048). The authors also measured average error count per minute during the endoscope navigation portion of the procedure (simulator-trained 1.73, SD 1.05 versus control 0.82, SD 0.86; p=0.032). However, individual sample data review found a single outlier in the ES3 simulator-trained group that had skewed the results.

Table 8 Patient-based assessments: errors made during assessment operations

Simulation-based training versus no simulation-based training				
Laparoscopic cholecystectomy				
Ahlberg et al 2007 RCT, Level II	Intervention	N=13	Total errors for entire procedure, mean	
			Baseline	After training
	LapSim VR training	7	NA	28.4 (Variance 118.69)
	No simulator training	6	NA	86.2 (Variance 916.68)
	p-value		NA	0.0037
Cosman et al 2007 RCT, Level II	Intervention	N=10	Total errors for entire procedure, median	
			Baseline	After training
	LapSim VR training	5	NA	10
	No simulator training	5	NA	18
	p-value		NA	0.05
Laparoscopic Nissen fundoplication				
Van Sickle et al 2008 RCT, Level II	Intervention	N=22	Total suturing errors ± SD	
			Baseline	After training
	MIST-VR and box trainer laparoscopic simulator training	11	NA	25.6 ± 9.3
	No simulator training	11	NA	37.1 ± 10.2
	p-value		NA	<0.01
Endoscopic sinus surgery				
Fried et al 2010 RCT, Level II	Intervention	N=25	Total number of injection errors (SD) *Note: six significant criteria*	
			Baseline	After training
	Endoscopic Sinus Surgery (ES3) simulator training	12	NA	3.53 (1.96)
	No simulator training	13	NA	6.89 (3.3)
	p-value		NA	0.048
	Intervention	N=25	Average number navigation errors per minutes (SD)	
			Baseline	After training
	ES3 simulator training	12	NA	1.73* (1.05)
	No simulator training	13	NA	0.82 (0.86)
	p-value		NA	0.032

*Outlier reported to skew data.
MIST-VR: Minimally Invasive Surgical Trainer−Virtual reality; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SD: standard deviation; VR: virtual reality.

Bimanual dexterity

Bimanual dexterity was described as involving two hands.

Simulation-based training versus no simulation-based training

Bimanual dexterity, coordination or manipulation data for simulator-trained and non simulator-trained participants were reported for laparoscopic cholecystectomy (Cosman et al 2007), Hogle et al 2009; Sroka et al 2010), laparoscopic Nissen fundoplication (Van Sickle et al 2008) and endoscopic sinus surgery (Fried et al 2010) (Table 9).

Laparoscopic cholecystectomy

Randomised controlled trials

Cosman et al (2007) specifically measured bimanual coordination, and reported that the LapSim-trained group had better bimanual coordination (median 3) compared with those who had no simulator training (median 1.8; p=0.05) (Table 9). Bimanual dexterity is one of the five domains in the GOALS scale (see Table 5) and Sroka et al (2010) and Hogle et al (2009) reported values for this domain.

Laparoscopic Nissen fundoplication

Randomised controlled trials

Van Sickle et al (2008) reported that the simulator-trained group made significantly fewer excess needle manipulations than the standard clinically-trained group (excess needle manipulations, 18.5 ± 10.5 versus 27.3 ± 8.5; p<0.05) (Table 9).

Endoscopic sinus surgery

Randomised controlled trials

Fried et al (2010) reported that ES3 simulator-trained participants exhibited a significantly higher level of dexterity with instrument manipulation (using a 10-point scale) compared with control participants (ES3-trained 6.75, SD 2.51 versus control 2.78, SD 1.86; p=0.011) (Table 9).

Patient discomfort

Patient discomfort was described as the pain felt by the patient undergoing the procedure.

Simulation-based training versus no simulation-based training

Ossowski et al (2008) and Yi et al (2008) reported patient discomfort outcomes for patients undergoing assessment procedures after participants had received simulation-based training or no simulation-based training (Table 10).

Endoscopy/nasolaryngoscopy

Randomised controlled trial

Ossowski et al (2008) compared standardised patient discomfort scores of simulator-trained participants with those who were not trained on the simulator and found no significant difference (0.89, SD 0.77 versus 1.33, SD 1.70; p=0.448).

Table 9 Patient-based assessments: bimanual dexterity

Simulation-based training versus no simulation-based training				
Laparoscopic cholecystectomy				
Cosman et al 2007 RCT, Level II	Intervention	N=10	Bimanual Coordination Index Score, median	
			Baseline	After training
	LapSim VR training	5	NA	3
	No simulator training	5	NA	1.8
	p-value		NA	0.05
Sroka et al 2010 RCT, Level II	Intervention	N=16	Bimanual dexterity score, mean difference from baseline ± SD	
			Baseline	After training
	FLS simulator training	8	NR	1.25 ± 0.6
	No simulator training	8	NR	0.5 ± 1.1
	p-value		NR	0.04
Hogle et al 2009 RCT, Level II	Intervention	N=12	Bimanual dexterity score , mean ± SD	
			Baseline	After training
	Simbionix GI Mentor simulator training	6	NA	3.17 ± 0.42
	No simulator training	6	NA	2.90 ± 0.51
	p-value		NA	0.55 (NS)
Laparoscopic Nissen fundoplication				
Van Sickle et al 2008 RCT, Level II	Intervention	N=22	Excess needle manipulations ± SD	
			Baseline	After training
	MIST-VR and box trainer laparoscopic simulator training	11	NA	18.5 ± 10.5
	No simulator training	11	NA	27.3 ± 8.5
	p-value		NA	<0.05
Endoscopic sinus surgery				
Fried et al 2010 RCT, Level II	Intervention	N=25	Instrument manipulation dexterity* , mean ± SD	
			Baseline	After training
	ES3 simulator training	12	NA	6.75 ± 2.51
	No simulator training	13	NA	2.78 ± 1.86
	p-value		NA	0.011

*Measured using a 10-point scale.

ES3: endoscopic sinus surgery; FLS: Fundamentals of Laparoscopic Surgery; GI: gastrointestinal; MIST-VR: Minimally Invasive Surgical Trainer−Virtual Reality; NA: not applicable; NS: not significant; RCT: randomised controlled trial; SD: standard deviation; VR: virtual reality.

Colonoscopy

Randomised controlled trial

Yi et al (2008) reported that in the patient survey, the simulator-trained group showed less discomfort during the colonoscopy compared with the control group. However, there was no significant difference between the two groups except in the extent of anus discomfort (simulator-trained 2.7, SD 0.8 versus control 3.4, SD 0.9; p=0.002).

Table 10 Patient-based assessments: patient discomfort

Simulation-based training versus no simulation-based training				
Endoscopy/nasolaryngoscopy				
Ossowski et al (2008) RCT, Level II	Intervention	N=20	standardised patient discomfort score (SD)	
			Baseline	After training
	Nasal Model for endoscopic simulation training	10	NA	0.89* (0.77)
	No simulator training	10	NA	1.33 (1.7)
	p-value		NA	0.448 (NS)
Colonoscopy				
Yii et al (2008) Comparative study, Level III-2	Intervention	N=11	Patient discomfort score – abdominal pain† (SD)	
			Baseline	After training
	KAIST-Ewha Colonoscopy Simulator II training	5	NA	3.1 (0.8)
	No simulator training	6	NA	3.2 (1.1)
	p-value		NA	0.273 (NS)
	Intervention	N=11	Patient discomfort score – abdominal inflation† (SD)	
			Baseline	After training
	KAIST-Ewha Colonoscopy Simulator II training	5	NA	3.0 (0.9)
	No simulator training	6	NA	3.2 (1.3)
	p-value		NA	0.215 (NS)
	Intervention	N=11	Patient discomfort score – anus discomfort† (SD)	
			Baseline	After training
	KAIST-Ewha Colonoscopy Simulator II training	5	NA	2.7 (0.8)
	No simulator training	6	NA	3.4 (0.9)
	p-value		NA	0.002

*Two outliers took an excessively long time to complete procedure so data may be skewed.

† Measured on scale of 1 (no pain) to 5 (worst pain of life).

KAIST: Korea Advanced Institute of Science and Technology; NS: not applicable; NS: not significant; RCT: randomised controlled trial; SD: standard deviation.

Surgical confidence

The level of participants' surgical confidence was evaluated during the patient-based assessment operations by Fried et al (2010) (Table 11). In addition, 'flow of operation' which referred to a participant's ability to move continuously and fluently through the procedure, confident of each step, was included in the total global rating score used by Banks et al (2007) (see Table 5); Park et al (2007) (see Table 5) and Schout et al (2009) (Table 6) but only reported separately by Schout et al (2009) (Table 11).

Simulation-based training versus no simulation-based training

Fried et al (2010) reported outcomes for the surgical confidence of trainees undergoing assessment procedures after participants had received simulation-based training or no simulation-based training. Schout et al (2009) reported outcomes for flow of operation.

Endoscopic sinus surgery

Randomised controlled trial

Fried et al (2010) reported that the ES3 simulator-trained participants demonstrated significantly greater surgical confidence (measured on a 10-point scale) than those who received no simulation-based training (6.55, SD 2.65 versus 2.67, SD 2.00; $p<0.009$).

Table 11 Patient-based assessments: surgical confidence

Simulation-based training versus no simulation-based training				
Endoscopic sinus surgery				
Fried et al 2010 RCT, Level II	Intervention	N=25	Surgical confidence; mean ± SD [ten-point scale, maximum is ten points]	
			Baseline	After training
	ES3 simulator training	12	NA	6.55 ± 2.65
	No simulator training	13	NA	2.67 ± 2.00
	p-value		NA	0.009
Endoscopy/Cystourethroscopy				
Schout et al 2009 RCT Level II	Intervention	N=100	Assessment score for GOALS domain 4: Flow of operation; mean (SD) [five-point scale]	
			Baseline	After training
	URO Mentor VR cystourethroscopy simulator training	50	NA	3.9 (1.0)
	No simulator training	50	NA	3.1 (1.3)
	p-value		NA	0.001

ES3: endoscopic sinus surgery; NA: not applicable; RCT: randomised controlled trial; SD: standard deviation; VR: virtual reality.

Success rate

Success rate was described as either the percentage of participants who were able to complete the patient-based assessment as specified in the study methods, the number of participants who were able to complete the case independently without assistance of a supervising surgeon, or the number of participants who were given a pass grade.

Banks et al (2007) reported a pass-fail assessment score, and Yi et al (2008) reported 'success rate'. Park et al (2007) and Haycock et al (2010) considered reaching the caecum as an indicator of completing the task or the case, respectively. Shirai et al (2008) reported the rate of one-point (direct assistance by the supervisor was required) and two-point scores (instruction by the supervisor was required) received by trainees carrying out oesophagogastroduodenoscopy (Table 12). Kälström et al (2010) reported the amount of autonomous procedure time and perceived participant competence for transurethral resection of the prostate (but after both groups had been simulator-trained).

Other studies used terms which were related to 'success rate' such as 'autonomy', 'attending takeover' or 'independent of a supervisor'. Hogle et al (2009) and Sroka et al (2010) included 'autonomy' in the GOALS scores (see Table 6) but reported no significant difference between the simulator-trained and non simulation-trained groups. The 'suturing operative errors' measured by Van Sickle et al (2008) for laparoscopic Nissen fundoplication included 'attending takeover' (the attending surgeon has to demonstrate or perform any aspect of suturing or tying), but the number of these were not reported. Zendejas et al (2011) reported that the proportion of participants believed to be able to perform the totally extraperitoneal inguinal hernia repair procedure independent of a supervisor increased from <10% to about 75% (p=0.000) from first hernia repair (TEP#1) to third hernia repair (TEP#3) (but after both groups had been simulator-trained). Other studies relied on a global assessment score (Table 6) to indicate the success of the participants in the patient-based assessment (Cosman et al 2007; Howells et al 2008).

Simulation-based training versus no simulation-based training

Banks et al (2007), Kälström et al (2010), Park et al (2007), Shirai et al (2008), and Yi et al (2008) reported outcomes in relation to participants' ability to independently complete the patient-based assessment operation after participants had received simulator-based training compared to no simulation training.(Table 12).

Laparoscopic bilateral tubal ligation

Randomised controlled trials

Banks et al (2007) reported that 100% of the simulator-trained group passed the patient-based assessment procedure compared with only 30% of the control group (p=0.003).

Table 12 Patient-based assessments: success in completing procedure

Simulation-based training versus no simulation-based training				
Laparoscopic bilateral tubal ligation				
Banks et al 2007 RCT, Level II	Intervention	N=20	Pass rate	
			Baseline	After training
	Limbs and Things laparoscopic simulator training	10	NA	100%
	No simulator training	10	NA	30%
	p-value		NA	0.003 (Fisher's exact test)
Colonoscopy				
Park et al 2007 RCT, Level II	Intervention	N=24	Completion – reaching the caecum, number	
			Baseline	After training
	AccuTouch Colonoscopy simulator training	12	NA	1
	No simulator training	12	NA	0
	p-value		NA	NR
Yi et al 2008 Comparative study, Level III-2	Intervention	N=11	Success rate, average of 5 patients (SD)	
			Baseline	After training
	KAIST-Ewha Colonoscopy simulator training	5	NA	0.76 (0.44)
	No simulator training	6	NA	0.43 (0.50)
	p-value		NA	0.006
Endoscopy/oesophagogastroduodenoscopy				
Shirai et al 2008 RCT, Level II	Intervention	N=20	Direct assistance by the supervisor required, %	
			Baseline	After training
	GI Mentor II simulator training	10	NA	19/220 or 8.6%
	No simulator training	10	NA	57/220 or 25.9%
	p-value		NA	0.0017
Endoscopy/transurethral resection of the prostate				
Kälström et al 2010 RCT, Level II	Intervention	N=23	Perceived resident competence (%)	
			Baseline TURP	After training both groups (TURP#3)
	PelvicVision simulator training	11	<10%	75% (*p*=0.000 versus baseline)
	No simulator training	12		
	p-value		NA	NR
Simulation-based training versus patient-based training				
Colonoscopy				
Haycock et al 2010 RCT, Level II	Intervention	N=36	Completion rate, number (%)	
			Baseline	After training
	Olympus (Endo TS-1) colonoscopy simulator training	17	NA	6 (11%)
	No simulator training	18	NA	4 (7%)
	p-value		NA	0.51 NS (indicates TOR)

GI: gastrointestinal; NA: not applicable; NR: not reported; NS: not significant; RCT: randomised controlled trial; SD: standard deviation; TURP: transurethral resection of the prostate; TOR: transfer to operating room.

Colonoscopy

Randomised controlled trials

Park et al (2007) reported that only one participant from the simulator-trained group reached the caecum while none of the control group completed this task (*P* value not reported).

Comparative study

Yi et al (2008) reported that the simulator-trained group significantly outperformed the control group in terms of success rate of reaching the caecum, with an average over five patients of 0.76, SD 0.44 versus 0.43, SD 0.50 (p=0.006).

Endoscopy/oesophagogastroduodenoscopy

Randomised controlled trials

Shirai et al (2008) reported that the simulator-trained group required significantly less assistance from the supervisor to complete the real oesophagogastroduodenoscopy than the control group (19/220 [8.6%] versus 57/220 [25.9%], respectively; p=0.0017).

Simulation-based training versus patient-based training

Colonoscopy

Randomised controlled trials

Haycock et al (2010) reported that six participants in the simulator-trained group and four in patient-based training group reached the caecum (p=0.51). The lack of difference between the two groups indicates a high level of skills transfer from simulator-based training to the real colonoscopy (the operative setting).

Participant satisfaction of training

To determine participants' opinions of the simulation-based training and patient-based training received, Haycock et al (2010 administered questionnaires to participants at the end of the study.

Simulation-based training versus patient-based training

Colonoscopy

Randomised controlled trials

Haycock et al (2010) reported that participants in both groups rated their training experience highly, with a median score of 8.0 out of 10 for both simulator-based and patient-based training.

Training costs

Three studies mentioned training costs in relation to the purchase of surgical simulators. These were procedure-specific and could not be grouped, and therefore have been reported narratively.

Larsen et al (2009) reported that Bridges et al (1999) compared the operation times of experienced surgeons and trainee surgeons and estimated increased operation time during residency training to costs 'about (1997) USD$48,000 (£31 841.00; €35 097.00) per graduate'. Larsen et al (2009) concluded that 'increased use of simulator training could reduce novice operating time substantially'.

Schout et al (2009) warned of the limitations of the benefits of URO Mentor VR CUS simulator training and suggested careful consideration of the appropriate setting to train, before recommending "…potentially expensive simulator training facilities".

Van Sickle et al (2008) calculated the cost of the simulation equipment used for their study of laparoscopic Nissen fundoplication. At the time of the study, the cost of the MIST-VR system was approximately USD$30,000 and the cost of the equipment needed for the video tower box trainers was approximately USD$50,000 (new). The tensiometer device and laptop cost approximately USD$10,000 and the disposable items (suture, foam models, needle drivers, graspers etc.) were donated by various vendors.

Other patient-based assessment-related outcomes

Other patient-based assessment-related outcomes were procedure-specific, and therefore have been reported in Appendix E.

5. Discussion

Limitations of the evidence

The purpose of this systematic review was to assess whether skills acquired using simulation-based training transfer to the operative setting. Studies were included irrespective of the type of patient-based assessment procedure, or the type of simulation. Many factors determine the transfer of skills, including those that relate to the simulator design and functionality (see Appendix F) and the way that simulation is used as a training device, including pre-learning, the nature and type of proximate and summative feedback, and opportunities for reinforcement of learning. In practice, this means that the evidence for skills transfer reported in this review cannot be attributed to the simulator alone.

Simulation-based training was, in most cases, an adjunct to normal surgical training programs or apprenticeship training. In one study (Bruppacher et al 2010), simulation training was compared with interactive seminar-based education and in another study (Patel et al 2012) in addition to the absence of simulator training (no training), there was a second comparator (a didactic lecture). Therefore where the comparator was not simply the absence of simulator-based training but a different comparator, the study was reported separately in this report. In Haycock et al (2010) and Franzeck et al (2012) simulation-based training was compared with structured patient-based training of the same duration and these studies were also reported separately to the other comparators in this report.

Determining the training methods used in some studies was difficult because the terms 'traditional teaching' or 'bedside teaching' were used to represent groups who did not receive simulation-based training but did participate in normal surgical training.

Reporting of methodological detail in the included studies was generally incomplete (Appendix D provides a summary of the critical appraisal). While only five RCTs did not report the method of randomisation, the majority did not report allocation concealment, intention-to-treat, power calculations, losses to assessment, study period or exclusion criteria.

The sample sizes within the included studies were generally small, with 25 or less participants per group. Only four RCTs had samples of more than 25 (Schout et al 2009; Patel et al 2012; Zendejas et al 2011 and Haycock et al 2010 having 100, 60, 50 and 36 participants, respectively).

Simple statistical analyses were generally used within the studies, and there was multiple testing of many variables within each study, often over a number of procedures, or over a number of assessments. This increases the likelihood of type I error (α), that is, a wrong decision is made when a test rejects a true null hypothesis.

There were large variations in the length of time participants were trained. The end-points of training were not consistent between studies, making it difficult to compare the skill level at the end of the training. Only a few studies used a predefined measure of proficiency or mastery level on the simulator to determine the skill level required at the end of participants' training.

Statistical comparison between studies was made difficult because of other factors that were not consistent between the studies. Variables in the OR such as differences in the severity of patient disease, the degree of independence granted by clinicians and various staff assistants, the mentoring given to participants during the training period, and the complexity of the assessment operations, differed between studies. The different parameters measured during the assessment operations were often ill-defined, making direct comparisons between studies difficult.

The adjustments made for any baseline differences were not uniform between studies, making direct comparisons in changes in performance difficult. Some studies did not perform baseline testing of participants, while some others used the training device or the performance of the assessment procedure, or both. Using the simulator or a patient for baseline testing allows a participant to gain familiarity with the procedure or device, and hence can lead to improvements in clinical performance.

Transfer outcomes

Simulation-based training versus no simulation-based training

Laparoscopic cholecystectomy

There were four RCTs for laparoscopic cholecystectomy, and these included three different types of simulators (LapSim, Simbionix GI Mentor and FLS box trainer). There were variations in assessment methods with Ahlberg et al (2007), Hogle et al (2009) and Sroka et al (2010) assessing performance of the entire laparoscopic cholecystectomy procedure, and Cosman et al (2007) assessing the clip and divide part of the procedure (either cystic duct or cystic artery). Hogle et al (2009) and Sroka et al (2010) measured operative performance using GOALS scores while Ahlberg et al (2007) and Cosman et al (2007) measured performance errors on different scales. Despite these variations in assessment, participants who underwent simulation-based training prior to conducting patient-based laparoscopic cholecystectomy performed better than their counterparts who had no contact with the simulators. This improvement was not universal for all the parameters measured, but overall the non simulator-trained group never outperformed the simulator-trained group.

Ahlberg et al (2007) and Cosman et al (2007) reported that procedure or dissection times for laparoscopic cholecystectomy were shorter (but not significantly so: Ahlberg $p<0.0586$, Cosman $p=0.075$) in the LapSim VR-trained group compared with the non simulator-trained group.

Ahlberg et al (2007) and Cosman et al (2007) also reported that the simulator-trained groups made significantly fewer errors than the control groups during laparoscopic cholecystectomy.

In addition, Beyer et al (2011), from the comparative study, reported that the two simulator-trained groups (MISTELS and LAP Mentor) improved their performance significantly while the control did not improve.

Laparoscopic bilateral tubal ligation

Banks et al (2007) reported on overall performance between Limbs and Things laparoscopic simulator-trained and non simulator-trained participants when performing their second bilateral tubal ligation. An objective evaluation of resident performance found that simulator-trained

participants scored higher overall on the global rating scale than the control group with only surgical teaching in the OR. The simulator-trained group also performed significantly better on the validated, task-specific 25-point checklist assessment than the control group.

Laparoscopic salpingectomy

Larsen et al (2009) reported that the median total score on the general and task-specific rating scale was significantly higher in the simulator-trained group than in the control group.

Laparoscopic Nissen fundoplication

Van Sickle et al (2008) reported that the simulator-trained group (using Mist-VR and box trainer) performed the fundal suturing portion of the laparoscopic Nissen fundoplication in significantly less time than the non simulator-trained group.

Laparoscopic diagnostic arthroscopy of the knee

Howells et al (2008) reported that the arthroscopy knee bench-top simulator-trained group significantly outscored the non simulator-trained group in the operative assessment total global rating score.

Laparoscopic totally extraperitoneal inguinal hernia repair

Zendejas et al (2011) reported that operative performance ratings were better for those trained to mastery on a simulator than for those who had no simulation training in the totally extraperitoneal inguinal hernia repair procedure.This finding supports that found by Sturm et al. (2007). Simulator-trained participants performed totally extraperitoneal inguinal hernia repair significantly faster than control participants.

Colonoscopy

Park et al (2007) reported that the overall global rating score for AccuTouch colonoscopy simulator-trained participants was significantly higher than that for the control group. However, only one participant from the simulator-trained group reached the caecum without assistance from the supervisor, and none of the control group completed the task.

In addition, Yi et al (2008), from the comparative study, reported that the KAIST-Ewha colonoscopy simulator II-trained group significantly outperformed the control group in terms of overall performance accuracy, insertion time, and in terms of success rate of reaching the caecum In the patient survey, there was less discomfort during colonoscopy performed by the simulator-trained group compared to the control group; there was significantly less anus discomfort but the difference was not significant for abdominal pain and inflation.

Endoscopy/cystourethroscopy

Schout et al (2009) reported that the overall GRS score for URO Mentor VR cystourethroscopy simulator-trained participants was significantly higher than for the non simulator-trained participants).

Endoscopy/oesophagogastroduodenoscopy

Shirai et al (2008) evaluated performance on eleven items and two endoscopic procedures, but did not report an overall score. The scores were significantly higher in the GI Mentor II simulator-trained group compared with the control group for five items but there was no

significant difference between the groups in the other six items reported. The simulator-trained group required significantly less assistance from the supervisor to complete the intubation to the caecum than the control group.

Endoscopy/nasolaryngoscopy

Ossowski et al (2008) reported that there was no significant difference between the simulator-trained group and the control group for the flexible laryngoscopy procedure time or the discomfort assigned by the standardised patient but the authors noted that the data was skewed by two extremely high values.

Endoscopic sinus surgery

Fried et al (2010) reported that the completion time of the mucosal injection task was significantly shorter, and with a narrower variability, in the simulator-trained group compared with the control group for endoscopic sinus surgery. Dissection time yielded similar observations. Simulator-trained participants made significantly fewer mucosal injection errors than the control participants and also exhibited a significantly higher level of dexterity with instrument manipulation. Simulator-trained participants demonstrated significantly greater surgical confidence than control participants. Sturm et al (2007) found that simulator experience for endoscopic sinus surgery could be a predictor of first-time OR performance but the results of the study were not significant and the total number of participants was only four.

Endoscopy/transurethral resection of the prostate

Kälström et al (2010) compared simulator-training with no simulator training for transurethral resection of the prostate (TURP). The authors reported TURP#1 (baseline) and TURP#3 (after simulator-training) outcomes. Global assessment scores were significantly better after simulator-training. Data approximated from the authors' graph shows an estimated increase from 58% to 75% on the global assessment score.

Abdominal fascial closure

Palter et al (2011) reported that the assessment of technical skills in the OR for abdominal fascial closure was significantly higher for individuals in the simulator-trained group compared with the control group.

Phacoemulsification for cataract surgery

Belyea et al (2011), from a retrospective comparative study, reported that the simulator-trained participants had a significantly lower performance time for phacoemulsification compared with those who had not received simulator-based training.

Knowledge, skills and attitudes in the operating room

Patel et al (2012) compared the knowledge, attitudes and skills in the OR of novices who had simulation-based training with those that had no training (control group). Those trained in the Imperial College simulated operating suite displayed higher behavioural observation assessment scores than those trained in the Second Life VR operating theatre and both demonstrated significant improvement after training compared to the control group. The control group displayed almost no improvement and certainly no significant improvement.

Patel et al (2012) also compared the knowledge, attitudes and skills in the OR of novices who had simulation-based training with those that had received a didactic lecture instead, as reported below.

Simulation-based training versus didactic lecture-based

Knowledge, skills and attitudes in the operating room

Patel et al (2012) compared the knowledge, attitudes and skills in the OR of novices who had simulation-based training with those that had received a didactic lecture. Those trained in the Imperial College simulated operating suite displayed significantly higher behavioural observation assessment scores than the didactic lecture group. However, the difference was not significant between those trained in the Second Life VR operating theatre compared to the didactic lecture group. This may indicate that virtual operating theatre is not as realistic as the real operating suite for simulation-based training.

Simulation-based training versus interactive-seminar education

Cardiopulmonary bypass weaning following cardiac surgery

Bruppacher et al (2010) compared simulation-based training with interactive seminar-based education for cardiopulmonary bypass weaning following cardiac surgery. Participants who trained using the SimMan Universal simulator significantly outperformed the inter-active seminar group when assessed two weeks after training (post-test). Similar results were found for the retention test at five weeks. In addition, in all four components of the ANTS; 'task management', 'team working', 'situation awareness' and 'decision-making', the simulator-trained group significantly outperformed the seminar-based education group.

Simulation-based training versus patient-based training

Colonoscopy

Haycock et al (2010) compared structured patient-based training with simulation-based training. There was no significant difference between the groups on the patient-based assessment in terms of case completion, time taken, JAG DOPS score or GRS Score. These authors found that equivalent time spent on the simulator training package without any additional mentoring or supervision produced performance outcomes on real-life cases that were equivalent to those of the participants who received patient-based training in the assessment procedure. This demonstrated a high degree of skills transfer from the simulator to real colonoscopy. However, this result is different to that from the study reported by Sturm et al (2007).

Laparoscopic camera navigation

Franzeck et al (2012) compared simulation-based training with structured OR-based training. There was no significant difference between simulator-trained and OR-trained groups in the time taken to complete the post-training camera navigation evaluation. However, participants in the OR group spent significantly more overall time in the OR than the simulator-trained group spent in the skills laboratory. The authors concluded that VR simulator based-training is more time efficient than OR training for laparoscopic camera navigation training.

The results support the previous findings that "…the results of the studies for laparoscopic cholecystectomy indicate that skills acquired on a simulator can subsequently transfer to a patient in an operating theatre" (Sturm et al 2007). In addition, results indicate that skills acquired on a simulator can subsequently transfer to a patient in an operating theatre for laparoscopic procedures in bilateral tubal ligation, salpingectomy, Nissen fundoplication, diagnostic arthroscopy of the knee and totally extraperitoneal inguinal hernia repair.

Similarly, skills acquired on a simulator for endoscopic procedures in colonoscopy, endoscopic sinus surgery, and transurethral resection of the prostate, appear to subsequently transfer to a patient in an operating theatre. In particular, when simulator-based training was compared to patient-based training of the same duration, there appeared to be a high degree of skills transfer from the simulator to real colonoscopy, which differed from the finding reported by Sturm et al (2007). For oesophagogastroduodenoscopy, the simulator-trained group required significantly less assistance from the supervisor to complete the intubation to the caecum than the control group, although scores were significantly higher in the simulator-trained group compared with the control group for only five of the eleven items reported. The only contrary result was found by Ossowski et al (2008) for nasolaryngoscopy but the authors noted that the data was skewed by two extremely high values.

For other surgical procedures including abdominal fascial closure, phacoemulsification for cataract surgery, and introduction to the operating theatre, the results indicate that skills acquired on a simulator or through simulation can subsequently transfer to a patient in an operating theatre.

Other considerations

One of the benefits of demonstrating successful skills transfer following simulation-based training is the reduced need to use patients for training. This is likely to increase patient safety, address some risk management concerns, and improve OR efficiency.

However, as Schlickum et al (2011) note, it is important to ascertain what simulation metrics actually measure. Park et al (2007) found that these metrics have very limited concurrent validity: a difference between experts and novices was found for only 3/14 and 2/8 analysed parameters. Pugh et al (2010) note that previous simulation studies have largely focused on technical skills and measures rather than on the assessment of decision-making skills that are required for operative performance.

The development of technical skills is only one part of surgical training, and no single parameter measured in a simulator can by itself demonstrate that a trainee has acquired an expert level of proficiency or competence (Ahlberg et al 2005). A good example of this is performance time, which was measured by many of the included studies. Although more rapid task completion is a recognised feature of expert performance, measurement of this variable alone does not give an indication of the quality of the task performed, and caution should be taken when interpreting it without any additional objective quality data.

At the present time simulation-based training programs are often adjuncts to traditional surgical training and are often voluntary, which was reflected in the included studies. Using defined

performance criteria to define specific training goals is required for motivation (Ahlberg et al 2005). If the performance criteria levels required for proficiency were included in the training curriculum and published studies, comparison of research outcomes between studies may be possible. Although recently Stefandidis et al (2012) suggested that simulation training to automaticity leads to improved skills transfer compared to traditional proficiency-based training.

Only Van Sickle et al (2008) mentioned details of costs in relation to the purchase of surgical simulators, although Schout et al (2009) warned of limitations and the need for careful consideration of the appropriate setting to train before recommending 'potentially expensive simulator training facilities'. Larsen et al (2009) mentioned the estimated cost of longer operating times required for patient-based training but data are not yet available to demonstrate any cost savings realised through improved OR efficiency and/or the reduced risk to patients from integrating simulation-based training into the curriculum.

The acquisition of skills in surgery is influenced by many factors including internal factors, e.g. motivation, ability and learning processes, and external factors, e.g. the learning environment. Only a few factors have been looked at todate, e.g. learning styles (Windsor 2008), and the stage of training (Loveday et al 2010). Key factors may also modulate the transfer effect from simulation-based training to the operative setting, e.g. instructor feedback (Oestergaard 2012).

Future research

The quality of study design has improved from those studies reviewed by Sturm et al (2007) but the challenge of adequate reporting of methodological detail still remains.

It is recommended that further research be conducted into the transfer of skills acquired via surgical simulation-based training to the patient-based setting. Consistency in training and assessment methods across studies would help provide further insight into the benefits of surgical simulation-based training. Determining the most beneficial ways in which simulation-based training can be used in the surgical curriculum is important.

Future studies will have the opportunity to explore other important dimensions of skills transfer. As suggested by Sturm et al (2007) these would include:

- 'the nature and duration of training required to deliver the greatest transfer effect
- the stage of training at which trainees receive maximum skill transfer benefits from different forms of simulation
- the effect of different levels of mentoring during the training period on transfer rates
- changes in staff productivity as a result of surgical simulation-based training'.

In addition, future research should include a focus on intraoperative decision-making and whether a simulation-based skills measure/s can be determined to focus on decision-making during operations.

6. Conclusions and recommendations

The aim of this systematic review was to update the evidence generated since 2006 in order to determine whether skills acquired through simulation-based training are transferable to the operative setting.

Sturm et al (2007) concluded that:

"Overall the evidence available demonstrates that simulation-based training results in skills transfer to the operative setting. It would therefore appear that simulation-based training provides a safe, effective and ethical way for trainees to acquire surgical skills before entering the operating room. Higher quality studies are required to confirm these findings, and will need to examine different simulation technologies, clinical procedures, training regimens and assessment techniques, if the place of simulation-based training within surgical training programs is to be determined."

The studies included in this update on whether surgical skills acquired through simulation-based training transfer to the operating room were of a higher quality (including considerably more RCTs) than those found in the 2006 systematic review (Sturm et al 2007). These studies have strengthened the evidence base but they still have variable training and assessment methods, making comparison between studies difficult. Overall the current evidence demonstrates that simulation-based training, as part of a surgical skills training program and incorporating the achievement of reaching predetermined proficiency levels, results in skills transfer to the operating setting.

Classification and recommendations

On the basis of the evidence presented in this systematic review, the ASERNIP-S Review Group agreed on the following classifications and recommendations concerning the transferability of skills acquired through simulation-based training to the surgical setting.

Class ifications

Evidence rating

The evidence-base in this review is rated as average. The studies included were of variable quality, and did not have comparable simulation-based methods for the same indications, resulting in an inability to draw solid conclusions.

Efficacy rating

Efficacy cannot be determined. The studies did not have comparable simulation-based methods for the same indications, resulting in an inability to draw solid conclusions.

Clinical and research recommendations

Sturm et al (2007) recommended that further research be conducted into the transfer of skills acquired via simulation-based training to the patient setting, to strengthen the evidence base.

This recommendation is still relevant in 2011 although several well-designed studies have strengthened the evidence base since 2006. Areas still requiring further study include:

- 'the nature and duration of training required to deliver the greatest transfer effect
- the stage of training at which trainees receive maximum skill transfer benefits from different forms of simulation,
- the effect of different levels of mentoring during the training period on transfer rates, and
- changes in staff productivity as a result of simulation-based training' (Sturm et al 2007)

Further research could also explore the way that simulation-based technical skills training environments might be used to train and assess non-technical skills, such as decision-making.

Acknowledgments

The authors wish to acknowledge Associate Professor Wendy Babidge, Ms Prema Thavaneswaran, Ms Karen Humphreys, Ms Robyn Lambert, and Ms Caryn Butler for their assistance during the preparation of this review. The ASERNIP-S project is funded by the Australian Government Department of Health and Ageing.

References

Aggarwal R, Tully A, Grantcharov T, Larsen CR, Miskry T, Farthing A, Darzi A. 2006, 'Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy', *British Journal of Obstetrics and Gynaecology,* vol. 113, pp. 1382–7.

Aggarwal R, Undre S, Moorthy K, Vincent C, Darzi A 2004, 'The simulating operating theatre: comprehensive training for surgical teams', *Quality & Safety in Health Care,* vol. 13(suppl), pp. i27–32.

Ahlberg G, Hultcrantz R, Jaramillo E, Lindblom A, Arvidsson D 2005, 'Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist?' *Endoscopy,* vol. 37(12), pp. 1198–1204.

Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, Ramel S, Smith CD, Arvidsson D 2007, 'Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies', *The American Journal of Surgery,* vol. 193, pp. 797–804.

Ahmed, K, Jawad M, Abboudi M, Gavazzi A, Darzi A, Athanasiou T, Vale J, Khan MS, Dasgupta P 2011, 'Effectiveness of procedural simulation in urology: a systematic review', *The Journal of Urology*, vol. 186, pp. 26–34.

Alaraj A, Lemole MG, Finkle JH, Yudowsky R, Wallace A, Luciano C, Banerjee PP, Rizzi SH, Charbel FT 2011, 'Virtual reality in neurosurgery: a review of current status and future applications', *Surgical Neurology International*, vol. 2, pp. 52–72.

Andreatta PB, Woodrum DT, Gauger PG, Minter RM 2008, 'LapMentor metrics possess limited construct validity', *Simulation in Healthcare,* vol. 3(1), pp. 16–25.

Autorino R, Haber GP, Stein RJ, Rane A, De Sio M, White MA, Yang B, de la Rosette JJ, Kaouk JH, Laguna MP 2010, 'Laparoscopic training in urology: critical analysis of current evidence', *Journal of Neurology*, vol. 24(9), pp1377–90.

Ayodeji ID, Schijven M, Jakimowicz J, Greve JW 2007, 'Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum', *Surgical Endoscopy,* vol. 21, pp. 1641–1649.

Bar-Meir S 2000, 'A new endoscopic simulator', *Endoscopy* , vol. 32(11), pp. 898–900.

Banks EH, Chudoff S, Karmin I, Wang C, Pardanani S 2007, 'Does a surgical simulator improve resident operative performance of laparoscopic tubal ligation?' *The American Journal of Obstetrics and Gynecology,* vol. 197, pp. 541.e1–541.e5.

Belyea DA, Brown SE, Rajjoub 2011, 'Influence of surgery simulator training on ophthalmology resident phacoemulsification performance', *Journal of Cataract and Refractive Surgery,* vol. 37, pp. 1756-61.

Beyer L, De Troyer J, Mancini J, Bladou, F, Berdah, SV, Karsenty G 2011, 'Impact of laparoscopy simulator training on the technical skills of future surgeons in the operating room: a prospective study' *The American Journal of Surgery,* vol. 202, pp. 265-272.

Botden S & Jakimowicz J 2009, 'What is going on in augmented reality simulation in laparoscopic surgery?' *Surgical Endoscopy,* vol. 23, pp. 1693–1700.

Bridges M, Diamond DL 1999, 'The financial impact of teaching surgical residents in the operating room', *The American Journal of Surgery*, vol. 177, pp. 28–32

Bruppacher HR, Alam SK, LeBlanc VR, Latter D, Naik, VN, Savoldelli GL, Mazer CD, Kurrek MM, Joo HS 2010, 'Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery', *Anesthesiology,* vol. 112(4), pp. 985–992.

Colonoscopy 2008, 'DOPS assessment form', *The Joint Advisory Group on gastrointestinal endoscopy*, Accessed from JAG website http://www.thejag.org.uk/TrainingforEndoscopists/DOPSForms.aspx.

Cosman PH, High TJ, Shearer CJ, Merrett ND, Biankin AV, Cartmill JA 2007, 'Skills acquired on virtual reality laparoscopic simulators transfer into operating room in blinded, randomised controlled trial', *Studies in Health Technology and Informatics,* vol. 125, pp.76–81.

Desilets DJ 2011, 'Endoscopic simulators', *Gastrointestinal Endoscopy,* vol. 73(5), pp. 861–867.

Dolmans VE, Schout BM, de Beer NA, Bemelmans BL, Scherpbier AJ, Hendrikx AJ 2009, 'The virtual reality endourologic simulator is realistic and useful for educational purposes', *Journal of Endourology,* vol. 23(7), pp. 1175–81.

Dunkin B, Andrales GL, Apelgren K, Mellinger JD 2007, 'Surgical simulation: current review', *Surgical Endoscopy,* vol. 21(3), pp. 357–66.

Edmond CV Jr. 2002, 'Impact of endoscopic sinus simulator on operating room performance', *Laryngoscope,* vol. 112(7 Pt 1), pp. 1148–58.

Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S, Pellegrini C 1999, 'An objective scoring system for laparoscopic cholecystectomy', *Journal of the American College of Surgeons,* vol. 189(6), pp. 566–74.

Ferlitsch A, Glauginger P, Gupper A, Schillinger M, Haefner M, Gangl A, Scgoefl R 2002, 'Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy', *Endoscopy,* vol. 34, pp. 689–702.

Fletcher G, Flin R, McGeorge P, Glavin R, Maran N and Patey R 2003, 'Anaesthetist's Nontechnical Skills (ANTS): evaluation of a behavioural marker system', *British Journal of Anaesthesia*, vol. 90(5), pp. 580–8.

Franzeck FM, Rosenthal R, Muller MK, Nocito A, Wittich F, Maurus C, Dindo D, Clavien P, Hahnloser D 2012, 'Prospective randomized controlled trial of simulator-based versus traditional in-surgery laparoscopic camera navigation training' *Surgical Endoscopy,* vol. 26(1), pp. 235-41.

Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM 2003, 'Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system', *Surgical Endoscopy,* vol. 17, pp. 964–7.

Fried GM 2006, 'Lessons from the surgical experience with simulators: incorporation into training and utilisation in determining competency', *Gastrointestinal Endoscopy Clinics of North America,* vol. 16(3), pp. 425–34.

Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Gitulescu G, Andrew CG 2004, 'Proving the value of simulation in laparoscopic surgery', *Annals of Surgery,* vol. 240, pp. 518–25.

Fried MP, Sadoughi B, Gibber MJ, Jacobs JB, Lebowitz RA, Ross DA, Bent III JP, Parikh SR, Sasaki CT, Schaefer SD 2010, 'From virtual reality to the operating room: the endoscopic sinus surgery simulator experiment', *Otolaryngology−Head and Neck Surgery,* vol. 142, pp. 202–7.

Gallagher AG, Cowie R, Crothers I, Jordan-Black JA, Satava RM 2003, 'PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance', *Surgical Endoscopy*, vol. 17(9), pp. 1468–71.

Gallagher AG, McLure N, Ritchie K, Sheehy NP 1998, 'An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills', *Endoscope,* vol. 30(7), pp. 617–20.

Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM 2005, 'Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training', *Annals of Surgery,* vol. 241(2), pp. 364–72.

Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA 2004, 'The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures', *Annals of Surgery,* vol. 240(2), pp. 374–81.

Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR 2008, 'Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopy surgery, *British Journal of Surgery*, vol. 95, pp.1088–97. Full report: *Cochrane Database Systematic Review* 2007; Issue 3: CD006575, **'**Virtual reality training for surgical trainees in laparoscopic surgery'; http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006575.pub2

Haque S and Srinivasan S 2006, 'A meta-analysis of the training effectiveness of virtual reality surgical simulators', *IEEE Transactions on Information Technology in Biomedicine,* vol. 10(1), pp. 51–8.

Haycock AV, Bassett P, Bladen J, Thomas-Gibson S 2009, 'Validation of the second-generation Olympus colonoscopy simulator for skills assessment', *Endoscopy,* vol. 41(11), pp. 952–8.

Haycock A, Koch AD, Familiari P, van Delft F, Dekker E, Petruzziello L, Haringsma J, Thomas-Gibson S 2010, 'Training and transfer of colonoscopy skills: a multinational, randomized, blinded, controlled trial of simulator versus bedside training', *Gastrointestinal Endoscopy,* vol. 71(2), pp. 298–307.

Hesselfeldt R, Kristensen MS, Rasmussen LS 2005, **'**Evaluation of the airway of the SimMan full-scale patient simulator**'**, *Acta Anaesthesiologica Scandinavica,* vol. 49(9), pp. 1339–45.

Higgins JPT, Green S (editors) 2009, *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.2 [updated September 2009], The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Hogle NJ, Chang L, Strong VEM, Welcome AOU, Sinaan M, Bailey R, Fowler DL 2009, 'Validation of laparoscopic surgical skills training outside the operating room: a long road', *Surgical Endoscopy,* vol. 23, pp. 1476–82.

Howells NR, Gill HS, Carr AJ, Price AJ Rees JL 2008, 'Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study', *Journal of Bone and Joint Surgery*, vol. 90-B(4), pp. 494–9.

Jensen AR, Wright AS, McIntyre LK, Levy AE, Foy HM, Anastakis DJ, Pellegrini CA, Horvath KD 2008, 'Laboratory-based instruction for skin closure and bowel anastomosis for surgical residents, *Archives of Surgery*, vol. 143(9), pp. 852−858; discussion pp. 858-9.

Kälström R, Hjertberg H, Kjolhede H, Svanvik J 2005, 'Use of a virtual reality, real-time, simulation model for the training of urologists in transurethral resection of the prostate', *Scandinavian Journal of Urology and Nephrology,* vol. 39, pp. 313–20.

Kälström R, Hjertberg H, Svanvik J 2010a, 'Construct validity of a full procedure, virtual reality, real-time, simulation model for training in transurethral resection of the prostate', *Journal of Endourology*, vol. 24, pp. 109–15.

Kälström R, Hjertberg H, Svanvik J 2010, 'Impact of virtual reality-simulated training on urology residents' performance of transurethral resection of the prostate', *Journal of Endourology,* vol. 24(9), pp. 1521-28.

Koch AD, Haringsma J, Schoon EJ, Man RA, Kuipers EJ 2008, 'A second-generation virtual reality simulator for colonoscopy: validation and initial experience', *Endoscopy,* vol. 40, pp. 735–8.

Krummel TM 1998, 'Surgical simulation and virtual reality: the coming revolution', *Annals of Surgery,* vol. 228(5), pp. 635–7.

Larsen CR, Grantcharov T, Aggarwal T, Tully A, Sorensen JL, Dalsgaard T, Ottesen B 2006, 'Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator', *Surgical Endoscopy,* vol. 20(9), pp. 1460–6.

Larsen CR, Grantcharov T, Schouenborg L, Søerensen JL, Ottesen C, Ottesen BS 2008, 'Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure specific rating scale', *British Journal of Obstetrics and Gynaecology,* vol. 115, pp. 908–16.

Larsen CR, Søerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS 2009, 'Effect of virtual reality training on laparoscopic surgery: randomised controlled trial', *British Medical Journal,* vol. 338, pp. b1802.

Lendvay TS 2011, 'Surgical simulation in pediatric urological education', *Current Urology Reports*, vol 12, pp. 137–43.

Lodge D and Grantcharov T 2011, 'Training and assessment of technical skills and competency in cardiac surgery', *European Journal of Cardiothoracic Surgery*, vol. 39(3), pp. 287–93.

Loveday BPT, Oosthuizen GV, Diener BS, Windsor JA 2010, 'A randomized trial evaluating a cognitive simulator for laparoscopic appendectomy', *Australia and New Zealand Journal of Surgery*, vol. 80, pp.588–94.

Ma IW, Brindle ME, Ronksley PE, Lorenzetti DL, Sauve RS, Ghali WA 2011, ' Use of simulation-based education to improve outcomes of central venous catheterization: as systematic review and meta-analysis', *Academic Medicine,* vol. 86(9), pp. 1–10.

Malone HR, Syed ON, Downes MS, D'Ambrosio AL, Quest DO, Kaiser MG 2010, 'Simulation in neurosurgery: A review of computer-based simulation environments and their surgical applications', *Neurosurgery*, vol. 67(40), pp. 1105–116.

Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M 1997, 'Objective structured assessment of technical skill (OSATS) for surgical residents', *British Journal of Surgery*, vol. 84, pp. 273–8.

Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD 2001, 'A novel approach to endourological training: training at the Surgical Skills Center', *Journal of Urology,* vol. 166, pp. 1261–6.

Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD 2002, 'The effect of bench model fidelity on endourological skills: a randomized controlled study', *Journal of Urology,* vol. 167(3), pp. 1243–7.

Mettler LL, Dewan P 2009, 'Virtual reality simulators in gynecological endoscopy: a surging new wave', *Journal of the Society of Laparoendoscopic Surgeons*, vol. 13, pp. 279–86

Miskovic D, Wyles SM, Ni M, Darzi AW, Hanna GB 2010, 'Systematic review on mentoring and simulation in laparoscopic colorectal surgery', *Annals of Surgery*, vol. 252(6), pp. 943–51.

Modi CS, Morris G, Mukherjee R 2010, 'Computer-simulation training for knee and shoulder arthoscopy surgery', *Arthroscopy: The Journal of Arthroscopic and Related Surgery*, vol. 26(6), pp. 832 40.

Moher D and Tsertsvadze A 2006, 'Systematic reviews: when is an update an update?', Lancet, vol. 367(9514), pp. 881–3.

Moorthy K, Munz Y, Jiwanji M, Bann S, Chang A, Darzi A 2004, 'Validity and reliability of a virtual reality upper gastrointestinal simulator and cross validation using structured assessment of individual performance with video playback', *Surgical Endoscopy,* vol. 18, pp. 328–33.

Moorthy K, Munz Y, Sarker SK, Darzi A 2003, 'Objective assessment of technical skills in surgery', *British Medical Journal,* vol. 327(7422), pp. 1032–7.

NHMRC 2010, 'How to use the evidence: assessment and application of scientific evidence', National Health and Medical Research Council (NHMRC), Canberra. Australia. Available from www.nhmrc.gov.au

OCAP 2006. 'Orthopaedic Competence Assessment Project (OCAP) - List of performance-based assessments in orthopaedic training' used by Howells *et al*. (2008) [The Orthopaedic Competence Assessment Project website. http://www.ocap.org.uk] Accessed January 30, 2012

Oestergaard J, Bjerrum F, Maagaard W, Winkel P, Larsen CR, Ringsted C, Gluud C, Grantcharov T, Ottesen B, Soerensen JL 2012, 'Instructor feedback versus no instructor

feedback on performance in a laparoscopic virtual reality simulator: a randomized educational trial', *BMC Medical Education*, vol. 12(1), p.7.

Ossowski KL, Rhee DC, Rubenstein EN, Ferguson BJ 2008, 'Efficacy of sinonasal simulator in teaching endoscopic nasal skills', *Laryngoscope,* vol. 118, pp. 1482–5.

Palter VN, Grantcharov T, Harvey A, MacRae HM 2011, 'Ex vivo technical skills training transfers to the operating room and enhances cognitive learning: a randomized controlled trial'. *Annals of Surgery*, vol. 253(5), pp. 886-88.

Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK 2007, 'Randomized controlled trial of virtual reality simulator training: transfer to live patients', *The American Journal of Surgery,* vol. 194, pp. 205–11.

Patel V, Aggarwal R, Osinibi E, Taylor D, Arora S, Darzi A 2012, 'Operating room introduction for the novice' *The American Journal of Surgery*, vol. 203(2), pp. 266-75.

Pugh C, Plachta S, Auyang E, Pryor A, Hungness E 2010, 'Outcome measures for surgical simulators: is the focus on technical skills the best approach?', *Surgery*, vol. 147(5), pp. 646-54.

Reznick RK, Regehr G, MacRae H, Martin J, McCulloch W 1997, 'Testing technical skill via an innovative 'bench station' examination', *The American Journal of Surgery,* vol. 173(3), pp. 226–30.

Roberts KE, Bell RL, Duffy AJ 2006, 'Evolution of surgical skills training', *World Journal of Gastroenterology,* vol. 12(20), pp. 3219–24.

Schijven M, Jakimowicz J 2003, 'Virtual reality surgical laparoscopic simulators', *Surgical Endoscopy,* vol. 17, pp. 1943–50.

Schlickum M, Hedman L, Enochsson L, Henningsohn L, Kjellin A, Felländer-Tsai L 2011, ' Surgical Simulation Tasks Challenge Visual Working Memory and Visual-Spatial Ability Differently', *World Journal of Surgery*, vol. 35, pp. 710–715.

Schreuder HW, van Dongen KW, Roeleveld SJ, Schijven MP, Boerders IA 2009, 'Face and construct validity of virtual reality simulation of laparoscopic gynecology surgery', *American Journal of Obstetrics and Gynecology,* vol. 200, pp. 540–8.

Schout BMA, Ananias HJK, Bemelmans BLH, d'Ancona FCH, Muijtjens AMM, Dolamans VEMG, Scherpbier AJJA, Hendrikx AJM 2009, 'Transfer of cysto-urethroscopy skills from a virtual-reality simulator to the operating room: a randomized controlled trial', *BJU International,* vol. 106, pp. 226–31.

Schout BMA, Hendrikx AJM, Scheele F, Bemelmans BLH, Scherpbier AJJA 2010, 'Validation and implementation of simulators: a critical review of present, past and future', *Surgical Endoscopy*, vol. 24, pp. 536–46.

Schout BM, Muijtjens AM, Hendrikx AJ, Ananias HJ, Dolmans VE, Scherpbier AJ, Bemelmans BL 2009, 'Acquisition of flexible cystoscopy skills on a virtual reality simulator by experts and novices', *BJU International,* vol. 105(2), pp. 234–9.

Scott DJ, Pugh CM, Ritter EM, Jacobs LM, Pellegrini CA, Sachdeva AL 2011, ' New directions in simulation-based surgical education and training: validation and transfer skills, use of

nonsurgeons as faculty, use of simulation to screen and select surgery residents, and long-term follow-up of learners', *Surger*y, vol. 49(6), pp. 735–44.

Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM 2002, 'Virtual reality training improves operating room performance: results of a randomised, double-blinded study', *Annals of Surgery,* vol. 236(4), pp. 458–63.

Seymour NE & Rotnes JS 2006, 'Challenges to the development of complex virtual reality surgical simulations', *Surgical Endoscopy,* vol. 20(11), pp. 1774–7.

Shirai Y, Yoshida T, Shiraishi R, Okamoto T, Nakamura H, Harada T, Nishikawa J, Sakaida I 2008, 'Prospective randomized study on the use of a computer-based endoscopic simulator for training in esophagogastroduodenoscopy', *Journal of Gastroenterology and Hepatology,* vol. 23, pp. 1046–50.

Slater GH, Jourdan I, Flöscher DJ, Snook AL, Cooper M, D'Allessandro P, Rangeley C, Bailey ME 2001, 'The Guildford MATTU TEP hernia model', *Surgical Endoscopy,* 14(5), pp. 493–6.

Solverson DJ, Mazzoli RA, Raymond WR, Nelson ML, Hansen EA, Torres MF, Bghandari A, Hartranft CD 2009, 'Virtual reality simulation in acquiring and differentiating basic ophthalmic microsurgical skills', *Simulation in Healthcare,* vol. 4, pp. 98–103.

Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM, 'Fundamentals of Laparoscopic Surgery simulator training to proficiency improves laparoscopic performance in the operating room−a randomized controlled trial', *The American Journal of Surgery*, vol. 199, pp. 115–20.

Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith W D 2012, 'Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial', *Annals of Surgery*, vol. 255(1), pp. 30–7.

Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ 2007, Surgical simulation for training: skills transfer to the operating room, ASERNIP-S Report 61. Adelaide, South Australia.

Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ 2008, 'A systematic review of skills transfer after surgical simulation training', *Annals of Surgery*, vol. 248(2), pp. 166–79.

Thijssen AS and Schuijven MP 2010, 'Contemporary virtual reality laparoscopic simulators: quicksand or solid grounds for assessing surgical trainees?', *The American Journal of Surgery*, vol. 199, pp. 529–41.

Valentine RJ & Rege RV 2004, 'Integrating technical competency into the surgical curriculum: doing more with less', *Surgical Clinics of North America,* vol. 84(6), pp. 1647–66.

van Empel PJ, van der Veer WM, van Rijessen LB, Cuesta MA, Scheele F, Bonjer J, Meijerink WJ 2012, 'Mapping the maze of minimally invasive surgery simulators', *Journal of Laparoendoscopic & Advanced Surgical Techniques,* vol. 22(1), pp. 1–10.

Van Sickle KR, Baghai M, Huang IP, Goldenberg A, Smith CD, Ritter EM 2007, 'Construct validity of an objective assessment method for laparoscopic intracorporeal suturing and knot tying', *The American Journal of Surgery,* vol. 196, pp. 74–80.

Van Sickle KR, McClusky DA, Gallagher AG, Smith CD 2005, 'Construct validation of the ProMIS simulator using a novel laparoscopic suturing task', *Surgical Endoscopy,* vol. 19, pp. 1227–31.

Van Sickle KR, Ritter EM, Baghai M, Goldenberg AE, Huang IP, Gallagher AG, Smith CD 2008, 'Prospective, randomized, double-blind trail of curriculum-based training for intracorporeal suturing and knot tying', *Journal of American College of Surgeons,* vol. 207, pp. 560–8.

Vassiliou MC, Feldman LS, Andrew GG, Bergman S, Leffondré K, Stanbridge D, Fried GM 2005, 'A global assessment tool for evaluation of intraoperative laparoscopic skills', *The American Journal of Surgery,* vol. 190, pp. 107–13.

Wagh MS & Waxman I 2006, 'Animal models for endoscopic simulation', *Gastrointestinal Endoscopy Clinics of North America,* vol. 16(3), pp. 451–6.

Webster R, Sassani J, Shenk R, Harris M, Gerber J, Benson A, Blumenstock J, Billman C, Haluck R 2005, 'Simulating the curvilinear capsulorhexis procedure during cataract surgery on the EYESi system', *Studies in Health Technology and Informatics,* vol. 111, pp. 592–5.

Woo H, Kim W, Ahn W, Lee D, Yi S 2008, 'Haptic interface of the KAIST-Ewha colonoscopy simulator II', *IEEE Transactions on Information Technology in Biomedicine,* vol. 12(6), pp. 746–53.

Windsor JA, Diener S, Zoha F 2008, 'Learning styles and laparoscopic experience in psychomotor skill performance using a virtual reality surgical simulator', *The American Journal of Surgery*, vol. 195, pp.837–42.

Yi SY, Ryu KH, Na YJ, Woo HS, Ahn W, Kim WS, Lee DY 2008, 'Improvement of colonoscopy skills through simulation-based training', In JD Westwood, RS Haluck, HM Hoffman, GT Mogel, R Phillips, RA Robb, KG Vosburgh (eds) *Medicine Meets Virtual Reality* IOS Press Inc., Amsterdam, the Netherlands, vol. 16, pp. 565–8.

Zendejas B, Cook DA, Bingener J, Huebner M, Dunn WF, Sarr MG, Farley DR 2011, 'Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: A randomized controlled trial', *Annals of Surgery*, vol. 254(3), pp. 502-511.

Zhang A, Hünerbein M, Dai Y, Schlang PM, Beller S 2008, 'Construct validity testing of a laproscopic surgery simulator (LAP Mentor®)', *Surgical Endoscopy,* vol. 22 (6), pp. 1440–44.

Appendix A: Excluded studies

The following articles were excluded from the methodological assessment as outlined in the methods section of the review.

Study	Reason for exclusion
Aggarwal R, Black SA, Hance JR, Darzi A, Cheshire NJW 2006, 'Virtual reality simulation training can improve inexperienced surgeons' endovascular skills', *European Journal of Vascular and Endovascular Surgery,* vol. 31(6), pp. 588-593.	Assessed in a simulator/via simulation
Aggarwal R, Tully A, Grantcharov T, Larsen CR, Miskry T, Farthing A, Darzi A 2006, 'Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy', *BJOG: An International Journal of Obstetrics and Gynaecology*, vol. 113(12), pp. 1382-7.	Assessed in a simulator/via simulation
Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A 2007, 'Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery', *Annals of Surgery*, vol. 246(5), pp. 771–9.	Assessment in animals
Andreatta PB, Woodrum DT, Birkmeyer JD, Yellamanchilli RK, Doherty GM, Gauger PG, and Minter RM 2006, 'Laparoscopic skills are improved with LAP Mentor™ Training: results of a randomized, double-blinded study', *Annals of Surgery*, vol. 243(6), pp. 854–863.	Assessment in animals
Arora S, Aggarwal R, Moran A, Sirimanna P, Crochet P, Darzi A, Kneebone R, Sevdalis N 2011, 'Mental practice: effective stress management training for novice surgeons', *Journal of American College of Surgeons,* vol. 212(2), pp. 225–33.	Assessed in a simulator/via simulation
Arora S, Sevdalis N, Aggarwal R, Sirimanna P, Darzi A, Kneebone R 2010, 'Stress impairs psychomotor performance in novice laparoscopic surgeons', *Surgical Endoscopy*, vol. 24(10), pp. 2588–93.	Assessed in a simulator/via simulation
Bennet A, Birch DW, Menzes C, Vizhul A, Karmali S 2011, 'Assessment of medical student laparoscopic camera skills and the impact of formal camera training', *American Journal of Surgery*, vol. 201(5), pp. 655–659.	Assessed in a simulator/via simulation
Burchard ER, Lockrow EG, Zahn CM, Dunlow SG, Satin AJ 2011 'Simulation training improves resident performance in operative hysteroscopic resection techniques', *American Journal of Obstetrics and Gynecology*, vol.197, pp. 542.31–542.e4.	Assessed in a simulator/via simulation
Burkhart HM, Riley JB, Hendrickson SE, Glenn GF, Lynch JJ, Arnold JJ, Dearani JA, Schaff HV, Sundt TM 2010, 'The successful application of simulation-based training in thoracic surgery residency', *The Journal of Thoracic and Cardiovascular Surgery*, vol. 139(3), pp. 707–12.	Assessed in a simulator/via simulation
Burton KS, Pendergrass TL, Byczkowski TL, Taylor RG, Moyer MR, Falcone RA, Geis GL 2011, 'Impact of Simulation-Based Extracorporeal Membrane Oxygenation Training in the Simulation Laboratory and Clinical Environment', *Simulation in Healthcare,* vol. 6(5), pp. 284–291.	Not surgical
Chaer RA, Derubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D, Morrissey NJ, Faries PL, McKinsey JF, Kent KC 2006, 'Simulation improves resident performance in catheter-based intervention: results of a randomized controlled study', *Annals of Surgery*, vol. 244(3), pp. 343–52.	Included in 2006 systematic review
Chandra V, Nehra D, Parent R, Woo R, Reyes R, Hernandez-Boussard T, Dutta S 2010, 'A comparison of laparoscopic and robotic assisted suturing performance by experts and novices', *Surgery,* vol.147(6), pp. 830–9.	Assessed in a simulator/via simulation
Chatterjee S. R 2007, 'Durability of endourologic skills: two-year follow-up study', *Journal of Endourology*, vol. 21(8), pp. 843–6.	Assessed in a simulator/via simulation
Chipman JG, Schmitz CC 2009, 'Using objective structured assessment of technical skills to evaluate a basic skills simulation curriculum for first-year surgical residents', *Journal of American College of Surgeons*, vol. 209(3), pp. 364–370e2.	Assessment in animals

Chou DS, Abdelshehid C, Clayman RV, McDougall EM 2006, 'Comparison of results of virtual-reality simulator and training model for basic ureteroscopy training', *Journal of Endourology,* vol. 20(4), pp. 266–71.	Assessment in animals
Chudnoff SG, Liu CS, Levie MD, Bernstein P and Banks EH 2009, 'Efficacy of a novel educational curriculum using a simulation laboratory on resident performance of hysteroscopic sterilization', *Fertility and Sterility*, Sep 25, doi:10.1016/j.fertnstert.2009.08.008	Assessed in a simulator/via simulation
Clevin L, Grantcharov TP 2008, 'Does box model training improve surgical dexterity and economy of movement during virtual reality laparoscopy? A randomised trial', *Acta Obstetricia et Gynecologica Scandinavica,* vol. 87(1), pp. 99–103.	Assessed in a simulator/via simulation
Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T, Ericsson KA, Darzi A 2011, 'Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills', *Annals of Surgery*, vol. 253(6), pp. 1216–1222.	Assessment in animals
da Cruz JAS, Sandy NS, Passerotti CC, Nguyen H, Antunes AA, Dos Reis ST, Dall'Oglio MF, Duarte RJ, Srougi M 2010, 'Does training laparoscopic skills in a virtual reality simulator improve surgical performance?', *Journal of Endourology*, vol. 24(11), pp. 1845–9.	Assessment in animals
Dănilă R, Gerdes B, Ulrike H, Domínguez Fernández E, Hassan I 2009, 'Objective evaluation of minimally invasive surgical skills for transplantation. Surgeons using a virtual reality simulator', *Chirurgia (Bucur),* vol. 104(2), pp. 181–5.	Assessed in a simulator/via simulation
Dawson DL, Meyer J, Lee ES, Pevec WC 2007, 'Training with simulation improves residents' endovascular procedure skills', *Journal of Vascular Surgery*, vol. 45(1), pp.149–54.	Assessed in a simulator/via simulation
Debes AJ, Aggarwal R, Balasundaram I, Jacobsen MB 2010, 'A tale of two trainers: virtual reality versus a video trainer for acquisition of basic laparoscopic skills', *American Journal of Surgery*, vol. 199(6), pp. 840–5.	Assessed in a simulator/via simulation
Diesen DL, Erhunmwunsee L, Bennett KM, Ben David K, Yurcisin B, Ceppa EP, Omotosho PA, Perez A, Pryor A 2011, 'Effectiveness of laparoscopic computer simulator versus usage of box trainer for endoscopic surgery training of novices', *Journal of Surgical Education*, vol. 68(4), pp. 282–289.	Assessment in animals
Domuracki KJ, Moule CJ, Owen H, Kostandoff G, Plummer JL 2009, 'Learning on a simulator does transfer to clinical practice', *Resuscitation*, vol. 80(3), pp. 346–9.	Not surgical
Edelman DA, Mattos MA, Bouwman DL 2011, 'Impact of fundamentals of laparoscopic surgery training during medical school on performance by first year surgical residents', *Journal of Surgical Research*, vol. 170(1), pp. 6–9.	Assessed in a simulator/via simulation
Essani R, Scriven RJ, McLarty AJ, Merriam LT, Ahn H, Bergamaschi R 2009, 'Simulated laparoscopic sigmoidectomy training: responsiveness of surgery residents', *Diseases of the Colon & Rectum*, vol. 52(12), pp. 1956–61.	Assessed in a simulator/via simulation
Eversbusch A, Grantcharov TP 2004, 'Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer', *Surgical Endoscopy*, vol. 18(10), pp. 1514–8.	Assessed in a simulator/via simulation
Feifer A, Al-Ammari A, Kovac E, Delisle J, Carrier S, Anidjar M 2011, 'Randomized controlled trial of virtual reality and hybrid simulation for robotic surgical training' *British Journal of Urology International,* vol. 108(10), pp. 1652–6.	Assessed in a simulator/via simulation
Feudner EM, Engel C, Neuhann IM, Petermeier K, Bartz-Schmidt KU, Szurman P 2009 ,'Virtual reality training improves wet-lab performance of capsulorhexis: results of a randomized, controlled study', *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 247(7), pp. 955–63.	Assessment in animals
Fisher L, Ormonde DG, Riley RH, Laurence BH 2010, 'Endoscopic skills training in a simulated clinical setting', *Simulation in Healthcare*, vol. 5(4), pp. 232–7.	Assessed in a simulator/via simulation

Ganai S, Donroe JA, St Louis MR, Lewis GM, Seymour NE 2007, 'Virtual-relaity training improves angled telescopioc skills in novice laparoscopists', *American Journal of Surgery*, vol. 193(2), pp. 260–5.	Asessment in animals
Gauger PG, Hauge LS, Andreatta PB, Hamstra SJ, Hillard ML, Arble EP, Kasten SJ, Mullan PB, Cederna PS, Minter RM 2010, 'Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons', *American Journal of Surgery*, vol. 199, pp. 72–89.	No control group as both groups had simulator training.
Ghaderi I, Vaillancourt M, Sroka G, Kaneva PA, Seagull FJ, George I, Sutton E, Park AE, Vassiliou MC, Fried GM, Feldman LS 2011, 'Performance of simulated laparoscopic incisional hernia repair correlates with operating room performance', *American Journal of Surgery*, vol. 201(1), pp. 40–45.	No control group as both groups had simulator training.
Gillen S, Fiolka A, Kranzfelder M, Wolf P, Feith M, Schneider A, Meining A, Friess H, Feussner H 2011, 'Training of a standardized natural orifice transluminal endoscopic surgery cholecystectomy using an ex vivo training unit', *Endoscopy*, vol. 43(10), pp. 876–81.	Assessment in animals
Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch JP. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery* 2004; 91(2): 146–150.	Included in 2006 systematic review
Grierson L, Melnyk M, Jowlett N, Backstein D, Dubrowski A 2011, 'Bench model surgical skill training improves novice ability to multitask: a randomized controlled study', *Studies in Health Technology and Informatics*, vol. 163, pp. 192–8.	Assessed in a simulator/via simulation
Haycock AV, Youd P, Bassett P, Saunders BP, Tekkis P, Thomas-Gibson S 2009, 'Simulator training improves practical skills in therapeutic GI endoscopy: results from a randomized, blinded, controlled study', *Gastrointestinal Endoscopy*, vol. 70(5), pp. 835–845.e2.	Assessed in a simulator/via simulation
Hogle NJ, Widmann WD, Ude AO, Hardy MA, Fowler DL 2008, 'Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games?', *Journal of Surgical Education*, vol. 65(6), pp. 431–5.	Assessment in animals
Jacobs K, Ivy J, Judson P 2010, 'Advancement of surgical skills with implementation of virtual reality simulation training', *Gynecologic Oncology*, Conference Abstract.	Assessed in a simulator/via simulation
Jiang G, Chen H, Wang S, Zhou Q, Li X, Chen K, Sui X 2011, 'Learning curves and long-term outcome of simulation-based thoracentesis training for medical students', *BMC Medical Education*, vol. 11, p. 39.	Not surgical
Joyce DL, Dhillon TS, Caffarelli AD, Joyce DD, Tsirigotis DN, Burdon TA, Fann JI 2011, 'Simulation and skills training in mitral valve surgery', *The Journal of Thoracic and Cardiovascular Surgery*, vol. 141, pp. 107–112.	Assessment in animals
Kahol K, Ashby A, Smith M. Ferrara JJ 2011, 'Quantitative evaluation of retention of surgical skills learned in simulation', *Journal of Surgical Education*, vol. 67(6), pp. 421–6.	Assessed in a simulator/via simulation
Kahol K, Satava RM, Ferrara J, Smith ML 2009, 'Effect of short-term pretrial practice on surgical proficiency in simulated environments: a randomized trial of the 'preoperative warm-up' effect', *Journal of American College of Surgeons*, vol. 208(2), pp. 255–68.	Assessed in a simulator/via simulation
Kanumuri P, Ganai S, Wohaibi EM, Bush RW, Grow DR, Seymour NE 2008, 'Virtual reality and computer-enhanced training devices equally improve laparoscopic surgical skill in novices', *Journal of the Society of Laparoendoscopic Surgeons*, vol. 12(3), pp. 219–26.	Assessment in animals
Kirby TO, Numnum TM, Kilgore LC, Straughn JM 2008, 'A prospective evaluation of a simulator-based laparoscopic training program for gynecology residents', *Journal of American College of Surgeons*, vol. 206(2), pp. 343–8.	Assessed in a simulator/via simulation
Koch J, Clements S, Abbott J 2011, 'Basic surgical skills training: does it work?', *The Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol. 51(1), pp. 57–60.	Assessed in a simulator/via simulation

Kolozski NO, Kaneva P, Brace C, Chartrand G, Vaillancourt M, Cao J, Banaszek D, Demyttenaere S, Vassiliou MC, Fried GM, Feldman LS 2011, 'Mastery versus the standard proficiency target for basic laparoscopic skill training: effect on skill transfer and retention', *Surgical Endosccopy*, vol. 25(7), pp. 2063–7.	Assessed in a simulator/via simulation
Kundhal PS, Grantcharov TP 2009, 'Psychomotor perfromances measured in a virtual environment correlates with technical skills in the operating room', *Surgical Endosccopy*, vol. 23, pp. 645–649.	No control group as all had simulator training.
Lauscher JC, Ritz JP, Stroux A, Buhr HJ, Grone J 2010, 'A new surgical trainer (BOPT) improves skill transfer for anastomotic techniques in gastrointestinal surgery into the operating room: a prospective randomized trial', *World Journal of Surgery*, vol. 34(9), pp. 2017–25.	Assessment in animals
Lee J, Kahol K, Kerbl DC, Alipanah N, Hsueh TY, Mucksavage P, Pick DL, Louie MK, Winfield HN, Mcdougall EM 2010, 'Impact of pre-operative warm-up exercises on surgical performance in urology', *Journal of Endourology*, Conference Abstract.	Oral presentation; no paper available.
Lendvay TS, Casale P, Sweet R, Peters C 2008, 'VR robotic surgery: randomized blinded study of the dV-Trainer robotic simulator', *Studies in Health Technology and Informatics*, vol. 132, pp. 242–4.	Assessed in a simulator/via simulation
Lerner MA, Ayalew M, Peine WJ, Sundaram CP 2010, 'Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system', *Journal of Endourology*, vol. 24(3), pp. 467–72.	Assessed in a simulator/via simulation
Loucas C, Nikiteas N, Kanakis M, Georgiou E 2011, 'The contribution of simulation training in enhancing key components of laparoscopic competence', *American Surgeon*, vol. 77(6), pp.708–15.	Assessed in a simulator/via simulation
Lucas S, Tuncel A, Bensalah K, Zeltser I, Jenkins A, Pearle M, Cadeddu J 2008, 'Virtual reality training improves simulated laparoscopic surgery performance in laparoscopy naive medical students', *Journal of Endourology*, vol. 22(5), pp. 1047–51.	Assessment in animals
Lucas SM, Zeltser IS, Bensalah K, Tuncel A, Jenkins A, Pearle MS, Cadeddu JA 2008, 'Training on a virtual reality laparoscopic simulator improves performance of an unfamiliar live laparoscopic procedure', *Journal of Urology*, vol. 180(6), pp. 2588–91.	Assessment in animals
Maagaard M, Sorensen JL, Oestergaard J, Dalsgaard , Grantcharov TP, Ottesen BS, Larsen CR 2011, 'Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer', *Surgical Endoscopy.,* vol. 25(3), pp. 722–7.	Assessed in a simulator/via simulation
Madan AK, Frantzides CT 2007, 'Prospective randomized controlled trial of laparoscopic trainers for basic skills acquisition', *Surgical Endoscopy*, vol. 21(2), pp. 209–13.	Assessment in animals
Martinek J, Suchanek S, Stefanova M, Rotnaglova B, Zavada , Strosova A, Zavoral M 2011, 'Training on an ex vivo animal model improves endoscopic skills: A randomized, single-blind study', *Gastrointestinal Endoscopy*, vol. 74(2), pp. 367–73.	Assessed in a simulator/via simulation
Matsuda T, Mcdougall EM, Ono Y, Hattori R, Baba S, Iwamura M, Terachi T, Naito S, Clayman RV 2010, 'Education of laparoscopic surgical skill: virtual reality simulator, objective measurement of performance and double blind video assessment', *International Journal of Urology*; Conference Abstract.	Assessed in a simulator/via simulation
McCluney AL, Vassiliou MC, Kaneva PA, Cao J, Stanbridge DD, Feldman LS, Fried GM 2007, 'FLS simulator performance predicts intraoperative laparoscopic skill', *Surgical Endoscopy*, vol. 21, pp. 1991–1995	No control group as all had simulator training.
McDougall EM, Kolla SB, Santos RT, Gan JM, Box GN, Louie MK, Gamboa AJ, Kaplan AG, Moskowitz RM, Andrade LA, Skarecky DW, Osann KE, Clayman RV 2009, 'Preliminary study of virtual reality and model simulation for learning laparoscopic suturing skills', *Journal of Urology*, vol. 182(3), pp. 1018–25.	Assessment in animals

Munz Y, Almoudaris AM, Moorthy K, Dosis A, Liddle AD, Darzi AW 2007, 'Curriculum-based solo virtual reality training for laparoscopic intracorporeal knot tying: objective assessment of the transfer of skill from virtual reality to reality', *American Journal of Surgery*, vol. 93(6), pp .774–83.	Assessed in a simulator/via simulation
Naylor RA, Hollett LA, Valentine RJ, Mitchell IC, Bowling MW, Ma AM, Dineen SP, Bruns BR, Scott DJ 2009, 'Can medical students achieve skills proficiency through simulation training?', *American Journal of Surgery,* vol. 198(2), pp. 277–82.	Assessed in a simulator/via simulation
Orejuela F, Gerten K, Lockrow E, Noel KA, Chohan L, Kilpatrick CC, Vaught J, Brock E, Schaffer J, Gala R 2010, 'An evaluation of validated laparoscopic skills simulators and the impact on operating room performance in obstetrics and gynecology residents', *Journal of Pelvic Medicine and Surgery*, Conference Abstract.	Oral presentation; no paper available
Panait L, Hogle NJ, Fowler DL, Bell RL, Roberts KE, Duffy AJ 2011, 'Completion of a novel, virtual-reality-based, advanced laparoscopic curriculum improves advanced laparoscopic skills in senior residents', Journal of Surgical Education, vol. 68(2), pp. 121–5.	Assessed in a simulator/via simulation
Parent RJ, Plerhoples TA, Long EE, Zimmer DM, Teshome M, Mohr CJ, Ly DP, Hernandez-Boussard T, Curet MJ, Dutta S 2010, 'Early, intermediate, and late effects of a surgical skills "boot camp" on an objective structured assessment of technical skills: a randomized controlled study', Journal of American College of Surgeons, vol. 210(6), pp. 984–9.	Assessed in a simulator/via simulation
Polterauer S, Grimm C, Hanzal E, Wenzl R, Hefler L, Leodolter S, Husslein PW, Reinthaller A 2010, 'Surgical skills training - Results of a prospective trial' (German), Geburtshilfe und Frauenheilkunde, vol. 70(12), pp. 990-3.	Assessed in a simulator/via simulation
Prabhu A, Smith W, Yurko Y, Acker C, Stefanidis D 2010, 'Increased stress levels may explain the incomplete transfer of simulator-acquired skill to the operating room', *Surgery*, vol. 147(5), pp. 640–5.	Assessment in animals
Price J, Naik V, Boodhwani M, Brandys T, Hendry P, Lam BK 2011, 'A randomized evaluation of simulation training on performance of vascular anastomosis on a high-fidelity in vivo model: the role of deliberate practice', *The Journal of Thoracic and Cardiovascular Surgery*, vol. 142(3), pp. 496–503.	Assessment in animals
Pugh C, Plachta S, Auyang E, Pryor A, Hungness E 2010, 'Outcome measures for surgical simulators: is the focus on technical skills the best approach?', *Surgery*, vol. 147(5), pp. 646–54.	Assessed in a simulator/via simulation
Robinson WP, Schanzer A, Cutler BS, Cardin L, Larkin A, Whitten R, Baril D, Eslami MH, Arous E, Messina LM 2011, 'An open vascular simulation course consisting of three one-hour sessions increases the knowledge and technical proficiency of junior surgical residents to that of senior residents', *Journal of Vascular Surgery*, Conference Abstract.	Assessed in a simulator/via simulation
Santos BF, Enter D, Soper NJ, Hungness ES 2010, 'Single-Incision Laparoscopic Surgery (SILS) simulator training improves performance', *Gastroenterology*, Conference Abstract.	Assessed in a simulator/via simulation
Sarker SK, Albrani T, Zaman A, Kumar I 2010, 'Procedural Performance in Gastrointestinal Endoscopy: Live and Simulated', *World Journal of Surgery*, vol. 34(8), pp. 1764–70.	Focus on validation of assessment tools; results for live and simulated performance not differentiated; no control group.
Sarker SK, Macioco M, Zaman A, Kumar I 2010, 'Operative performance in laparoscopic cholecystectomy using the procedural-based assessment tool', *American Journal of Surgery*, vol. 200(3), pp. 334–400.	Focus on validation of assessment tools; no control group.
Snyder CW, Vandromme MJ, Tyra SL, Porterfield JR, Clements RH, Hawn MT 2011, 'Effects of virtual reality simulator training method and observational learning on surgical performance', *World Journal of Surgery*, vol. 35(2), pp. 245–52.	Assessment in animals

Solverson DJ, Mazzoli RA, Raymond WR, Nelson ML, Hansen EA, Torres MF, Bhandari A, Hartranft CD 2009, 'Virtual reality simulation in acquiring and differentiating basic ophthalmic microsurgical skills', *Simulation in Healthcare*, vol. 4(2), pp. 98–103.	Assessed in a simulator/via simulation
Stefanidis D, Acker C, Heniford BT 2008, 'Proficiency-based laparoscopic simulator training leads to improved operating room skill that is resistant to decay', *Surgical Innovation*, vol. 15(1), pp. 69–73.	Assessment in animals and via simulation.
Stefanidis D, Korndorffer JR, Markley S, Sierra R, Heniford BT, Scott DJ 2007, 'Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training?', *Journal of American College of Surgeons*, vol. 205(2), pp. 307–13.	Assessment in animals
Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD 2011, 'Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial', *Annals of Surgery*, Jun 1 Epub DOI: 10.1097/SLA.0b013e318220ef31.	Assessment in animals
Stefanidis D, Wang F, Korndorffer JR, Dunne JB, Scott DJ 2010, 'Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload', *Surgical Endoscopy*, vol. 24, pp. 377–382.	Assessment in animals
Stelzer MK, Abdel MP, Sloan MP, Gould JC 2009, 'Dry lab practice leads to improved laparoscopic performance in the operating room', *Journal of Surgical Research*, vol. 154(1), pp. 163–6.	Assessment in animals
Verdaasdonk E, Dankelman J, Lange JF, Stassen LP 2008, 'Transfer validity of laparoscopic knot-tying training on a VR simulator to a realistic environment: a randomized controlled trial', *Surgical Endoscopy*, vol. 22(7), pp. 1636–1642.	Assessment in animals
Vlaovic PD, Sargent ER, Boker JR, Corica FA, Chou DS, Abdelshehid CS, White SM, Sala LG, Chu F, Le T, Clayman RV, McDougall EM 2008, 'Immediate impact of an intensive one-week laparoscopy training program on laparoscopic skills among postgraduate urologists', *Journal of the Society of Laparoendoscopic Surgeons,* vol. 12(1), pp. 1–8.	Assessment in animals
Von Sternberg N, Bartsch MS, Petersik A, Wiltfang J, Sibbersen W, Grindel T, Tiede U, Warnke PH, Heiland M, Russo PA, Terheyden H, Pohlenz P, Springer IN 2007, 'Learning by doing virtually', *International Journal of Oral & Maxillofacial Surgery,* vol. 36(5), pp. 386–90.	Assessment in animals
Walsh CM, Sherlock ME, Ling SC, Carnahan H 2010, *Cochrane Database of Systematic Reviews: Protocols*, Issue 1, John Wiley & Sons, Ltd.	Protocol only
Wetzel CM, George A, Hanna GB, Athanasiou T, Black SA, Kneebone RL, Nestel D, Woloshynowych M 2011, 'Stress management training for surgeons-a randomized, controlled, intervention study', *Annals of Surgery*, vol. 253(3), pp. 488–494.	Assessed in a simulator/via simulation
Wilasrusmee C 2007, 'Vascular anastomosis model: relation between competency in a laboratory-based model and surgical competency', *European Journal of Vascular and Endovascular Surgery*, vol. 34(4), pp. 405–10.	No control group as all had simulator training.
Wohaibi EM, Bush RW, Earle DB, Seymour NE 2010, 'Surgical resident performance on a virtual reality simulator correlates with operating room performance', *Journal of Surgical Research*, vol. 160(1), pp. 67–72.	No control group as all had simulator training.
Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D 2010, 'Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool', *Simulation in Healthcare*, vol. 5(5), pp. 67–71.	Assessment in animals
Zhao YC, 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', *Otolaryngoly–Head and Neck Surgery*, vol. 144(3), pp. 357–64.	Assessment in cadaver bones

Reviews

The following articles were excluded from the methodological assessment as outlined in the methods section of the review because they were reviews. They were pearled for studies to include in this systematic review. They highlight the increase in interest in simulation training across the surgical specialities.

Review	Conclusion
Aggarwal R, Darzi A, Grantcharov TP 2008, 'A systematic review of skills transfer after surgical simulation', *Annals of Surgery*, vol. 248(4), pp. 690–1.	Letter to Editor concludes that there is a real need for investigators in this field to work together in terms of multicentre studies to make simulation-based training the norm, rather than the 'addition' to current training; for example American College of Surgeons Education Institutes must bring together individuals and departments with a shared interest in simulation science.
Ahmed, K, Jawad M, Abboudi M, Gavazzi A, Darzi A, Athanasiou T, Vale J, Khan MS, Dasgupta P 2011, 'Effectiveness of procedural simulation in urology: a systematic review', *The Journal of Urology*, vol. 186, pp. 26–34.	Systematic review concluded that VR, synthetic and animal models can be effective for objective assessment of junior and intermediate trainees in urology but more research is need to validate simulated environments for senior trainees and specialists.
Alaraj A, Lemole MG, Finkle JH, Yudowsky R, Wallace A, Luciano C, Banerjee PP, Rizzi SH, Charbel FT 2011, 'Virtual reality in neurosurgery: a review of current status and future applications', *Surgical Neurology International*, vol. 2, pp. 52–72.	Literature review concluded that fully immersive technology is starting to be applied to the practice of neurosurgery. In the near future, detailed VR neurosurgical modules will evolve to be an essential part of the curriculum of the training of neurosurgeons.
Autorino R, Haber GP, Stein RJ, Rane A, De Sio M, White MA, Yang B, de la Rosette JJ, Kaouk JH, Laguna MP 2010, 'Laparoscopic training in urology: critical analysis of current evidence', *Journal of Neurology*, vol. 24(9), pp1377–90.	Literature review concluded that urology as a specialty needs to develop collaborations with simulator companies so more specific urological modules will become available in the future.
Botden SM, Buzink SN, Schijven MP, Jakimowicz JJ 2007, 'Augmented versus virtual reality laparoscopic simulation: what is the difference?', *World Journal of Surgery*, vol. 31, pp. 764–72.	Literature review concluded that the ProMIS AR laparoscopic simulator offers better realism, haptic feedback, didactic value, and construct validity than does VR, and it also gives useful feedback to determine trainee skill levels. Surgical residents not only learn basic laparoscopic skills at the cholecystectomy level, but also suturing skills.
Choy I, Okrainec A 2010, 'Simulation in surgery: perfecting the practice', *Surgical Clinics of North America*, vol. 90, pp. 457–73.	Literature review concluded that simulation training and competency-based curricula offer an opportunity to improve current training models to address challenges such as the expanding set of skills and knowledge, increased public scrutiny, limited financial resources, restricted work hours, and limited exposure to rare procedures, and create effective, efficient, and safe residency programs.
Clifton N, Klingmann C, Khalil H 2011, 'Teaching otolaryngology skills through simulation', *European Archives of Otorhinolarngology*, vol. 268(7), pp. 949–53.	Literature review concluded that good outcomes research is lacking and extremely difficult to measure but it is emerging and will probably improve in quality in due course. For otolaryngologists in training, simulation appears to help bridge the gap between theory and daily practice.
Curry JI 2011, ' "See one, practise on a simulator, do one" –the mantra of the modern surgeon', *South African Journal of Surgery*, vol. 49(1), pp. 4–6.	Keynote address concluded that for simulation to fulfil its potential it must be submitted to the rigours of educational theory and adult learning as are the other current components of surgical curricula. The process of curriculum design is a fluid one and requires regular review.
Desender LM, Van H, Aggarwal R, Vermassen FE, Cheshire NJ 2011, 'Training with simulation versus operative room attendance', *Journal of Cardiovascular Surgery (Torino)*, vol. 52(1), pp. 17–3.	Literature review concluded that although a simulator may be valid and reliable, the curriculum will determine how rapidly trainees are progressing. Curricula and best practice guidelines for incorporating simulation into training should be developed. Simulation training should be trainee orientated and proficiency-based so that skills can be practised at the learner's pace to a predefined level of skill.

Dunkin B, Adrales GL, Apelgren K, Mellinger JD 2007, 'Surgical simulation: a current review', *Surgical Endoscopy*, vol. 21, pp. 357–66.	Literature review concluded that the future direction of surgical simulation can be analysed from at least three significant and distinct vantage points. These included the fiscal realities of medical education and practice, adult education theory and its relationship to surgical didactic methodology, and competency/proficiency movement.
Fitzgerald TN, Duffy AJ, Bell RL, Berman L, Longo WE, Roberts KE 2008, 'Computer-based endoscopy simulation: emerging roles in teaching and professional skills assessment', *Journal of Surgical Education*, vol. 65(3), pp. 229–35.	Literature review concluded that simulators cannot replace traditional bedside teaching completely, but they may be used as an adjunct teaching aid. They only have the capacity to test a finite skill set, and they do not assess less-tangible factors such as bedside manner and professionalism.
Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR 2008, 'Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopy surgery', *British Journal of Surgery*, vol. 95, pp.1088–97. See Full report below:	From Cochrane database systematic review 2009: concluded that VR training improves standard surgical training and is at least as effective as video trainer training. Only one of the trials used patient-oriented outcomes (Ahlberg et al 2007).
Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR 2009, 'Virtual reality training for surgical trainees in laparoscopic surgery', The Cochrane Collaboration, John Wiley & Sons, Ltd. . Available online at The Cochrane Library CD006575.	Systematic review concluded that VR training improves standard surgical training and is at least as effective as video trainer training. Trials with low risk of bias that use patient-oriented outcomes after VR training as a part of surgical training curriculum are awaited. These trials should include cost-effectiveness as one of the outcomes.
Hatef DA, Holtier LH 2019, 'Systematic review of skills transfer after surgical simulation training', *The Journal of Craniofacial Surgery*, vol. 20 (2), pp. 577–8.	Review of Sturm et al 2008 concluded that systematic review made it clear that skills acquired in simulated setting transfer to the operative setting and that high quality studies are needed to examine this emerging tool in surgical education.
Kneebone R 2010, 'Simulation, safety and surgery', *Quality and Safety in Health Care*, vol. 19 (Suppl. 3), ppi47–53.	Review article concluded that the requirement is to provide experiences that reflect clinical practice, and allow educational goals and outcomes to be achieved. Simulation should be the stepping stone to gaining mastery within a complex clinical world. This paper outlines the concept of Distributed Simulation, using low-cost, portable yet immersive environments to address limitations of access to dedicated facilities.
Laguna MP, de Reijke TM, de la Rosette JJ, 'How far will simulators be involved into training?', *Current Urology Reports*, vol. 10, pp. 97–105.	Literature review concluded that skills transfer from VR simulation to the OR has been proven only for laparoscopic cholecystectomy and colonoscopy/sigmoidoscopy. The extent to which simulation will substitute real operations will ultimately depend on the development of finer assessment and proficiency-based criteria training programs.
Lendvay TS 2011, 'Surgical simulation in pediatric urological education', *Current Urology Reports*, vol 12, pp. 137–43.	Literature review concluded that it is important to design training curricula that educate holistically and not just focus on technical skills training. Medical decision-making simulation is in its infancy and, arguably, the most important aspect of effective surgical practice.
Lewis TM, Aggarwal R, Rajaretnam N, Grantcharov TP, Darzi A 2011, 'Training in surgical oncology–the role of VR simulation', *Surgical Oncology*, vol. 20, pp. 134–9.	Literature review concluded that there is now considerable evidence demonstrating the increased realism of simulators, but more work is needed to motivate and educate the surgical community in the effectiveness of VR simulation in terms of cost and future OR performance.
Lodge D, Grantcharov T 2011, 'Training and assessment of technical skills and competency in cardiac surgery', *European Journal of Cardiothoracic Surgery*, vol. 39(3), pp. 287–93.	Literature review concluded that simulation models must be developed and formally evaluated for the full range of cardiac surgery procedures to allow transition to competency-based residency training.
Ma IW, Brindle ME, Ronksley PE, Lorenzetti DL, Sauve RS, Ghali WA 2011, ' Use of simulation-based education to improve outcomes of central venous catheterization: as systematic review and meta-analysis', *Academic Medicine,* vol. 86(9), pp. 1–10.	Systematic review concluded that simulation-based education for CVC provides benefits in learner and select clinical outcomes (such as decrease in the number of needle passes and decrease in the risk of pneumothoraces) in nonrandomised, two group studies.

Malone HR, Syed ON, Downes MS, D'Ambrosio AL, Quest DO, Kaiser MG 2010, 'Simulation in neurosurgery: A review of computer-based simulation environments and their surgical applications', *Neurosurgery*, vol. 67(40), pp. 1105–116.	Literature review concluded that computer-based neurosurgery simulators are currently limited by the computational burden of accurate tissue deformation, the arduous process of manually segmenting volume-rendered models, and the great expense of sophisticated haptic interfaces. International collaboration among research groups is the key to the establishment of common simulation platforms and expediting evolution of neurological simulators.
McGaghie WC, Draycott TJ, Dunn WF, Lopez CM, Stefanidis D 2011, 'Evaluating the impact of simulation on translational patient outcomes', *Simulation in Healthcare*, vol. 6, Suppl. S42–7.	Literature review concluded that simulation-based medical education, especially those that feature deliberate practice towards mastery learning goals, can achieve translational science research outcomes to improved downstream patient care practices and improved patient and public health. Such outcomes are more likely when the interventions are embedded in rigorous educational and health services research programs that are thematic, sustained, and cumulative.
Mettler LL, Dewan P 2009, 'Virtual reality simulators in gynecological endoscopy: a surging new wave', *Journal of the Society of Laparoendoscopic Surgeons*, vol. 13, pp. 279–86.	Literature review concluded that VR computer simulation should be encouraged in training curriculum for gynaecological endoscopic surgeons. The designing of software for gynaecological procedural modules is imminent.
Michelson JD, Manning L 2008, 'Competency assessment in simulation-based procedural education', *The American Journal of Surgery*, vol. 196, pp 609–15.	Literature review concluded that although the use of clinical benchmarking as the standards-setting mechanism for procedural simulation-based learning, feedback, and assessment is critical to establishing the clinical relevance of simulation, it should be appreciated that this only heightens the need for more extensive and better-delineated evidence-based clinical benchmarks.
Miskovic D, Wyles SM, Ni M, Darzi AW, Hanna GB 2010, 'Systematic review on mentoring and simulation in laparoscopic colorectal surgery', *Annals of Surgery*, vol. 252(6), pp. 943–51.	Systematic review concluded that trainees can obtain similar clinical results like expert surgeons in laparoscopic colorectal surgery if supervised by an experienced trainer. There remains a need for further research into technical assessment and the educational value of simulated training.
Modi CS, Morris G, Mukherjee R 2010, 'Computer-simulation training for knee and shoulder arthroscopy surgery', *Arthroscopy: The Journal of Arthroscopic and Related Surgery*, vol. 26(6), pp. 832 40.	Systematic review concluded that studies have shown steady learning of basic skills, such as identification of anatomy and triangulation tasks, in inexperienced participants using computer simulators However, further higher-quality studies are required to show transfer and predictive validity of computer simulation within the OR.
Nogueira Junior JF, Cruz DN 2010, 'Real models and virtual simulators in otolaryngology: review of literature', *Brazilian Journal of Otorhinolaryngology*, vol.76(1), pp.	Literature review concluded that studies have evaluated the effectiveness in training of virtual simulators in rhinology, otology and laryngology. In Brazil they still do not have a nasosinusal endoscopic surgery simulator available. A nasosinusal endoscopic dissection real model is available; and some centres have temporal bone dissection virtual software.
Okuda Y, Bryson EO, DeMaria S, Jacobsen L, Quinones J, Shen B, Levione AI 2009, 'The utility of simulation in medical education: what is the evidence?', *Mount Sinai Journal of Medicine*, vol 76, pp. 330–43	Literature review concluded that simulation-based training was demonstrated to lead to clinical improvement in two areas of simulation research: residents trained on laparoscopic surgery simulators showed improvement in procedural performance in the operating room, and residents trained on simulators were more likely to adhere to the advanced cardiac life support protocol than those who received the standard training for cardiac arrest patients.
Oetting TA 2009, 'Surgical competency in residents', *Current Opinion in Ophthalmology*, vol. 20, pp. 56–60.	Literature review concluded that cataract training must attempt to reduce the early complications with increased use of organised wet laboratory and simulator training. An organised curriculum with defined expectations using simulation and assessment tools will help residents develop skills in all of the mandated six ACGME competencies.

Reznick RK, MacRae H 2006, 'Teaching surgical skills–changes in the wind', The *New England Journal of Medicine*, vol. 355, pp. 2664–9.	Literature review concludes that additional research is needed to demonstrate effectiveness of simulation training for senior learners, and to address motor-learning issues, such as whether it is preferable to practice whole operations, or build the whole from segments, what practice schedules are optimal, and how to optimise transfer of skills to the OR.
Satava RM 2010, *Surgical Clinics of North America*, vol.90, pp. 623–33.	Literature review concluded that new methods of surgical education, skills training, and assessment, as well as simulation, must be implemented across all disciplines and incorporated in laboratory training, in situ training, and daily clinical practice to a point at which simulation becomes automatically embedded in the culture of surgery. It is critical that an enormous effort be expended to ensure that a uniformity of approach and quality of training emerges.
Schout BMA, Hendrix AJM, Scherpbier AJJA, Bemelmans BLH 2008, 'Update on training models in Endourology: a qualitative systematic review or the literature between January 1980 and April 2008', *European Urology*, vol. 54, pp. 1247–61.	Systematic review concluded that 30 types of training models were described (nine for ureterorenoscopy) but only three randomised controlled trials (RCTs) were found. More randomised controlled validation studies including larger number of participants are needed to determine which models are most valuable for training postgraduates.
Schout BMA, Hendrikx AJM, Scheele F, Bemelmans BLH, Scherpbier AJJA 2010, 'Validation and implementation of simulators: a critical review of present, past and future', *Surgical Endoscopy*, vol. 24, pp. 536–46.	Literature review concluded that validity research is hampered by a paucity of widely accepted definitions and measurement methods of validity. Before undertaking a study to validate a simulator, researchers would be well advised to conduct a training needs analysis to evaluate the existing need for training and to determine program requirements in a training program design, methods that are also used by designers of military simulation programs.
Scott DJ, Pugh CM, Ritter EM, Jacobs LM, Pellegrini CA, Sachdeva AL 2011, 'New directions in simulation-based surgical education and training: validation and transfer skills, use of nonsurgeons as faculty, use of simulation to screen and select surgery residents, and long-term follow-up of learners', *Surgery*, vol. 49(6), pp. 735–44.	American College of Surgeons-Accredited–Education Institutes Consortium work group concluded that ACS-AEI Consortium possesses the expertise to pursue research that addresses validation and transfer of surgical skills and improves quality and safety of surgical care. Such an approach will need gap analyses as well as development and implementation of curricula to address the gaps identified. Outcome measures will need to be defined and new metrics developed for the assessment of outcomes. Requisite standards and processes will need to be established and best practices in simulation-based surgical education will need to be shared across the ACS–AEI Consortium.
Seymour NE 2008, 'VR to OR: A review of the evidence that virtual reality simulation improves operating room performance', *World Journal of Surgery*, vol. 32, pp. 182–8.	Literature review concluded that of the seven published studies of laparoscopic skills transfer identified (2002–2007), one failed to demonstrate transfer of skills, and results were similar for the seven skill transfer studies of VR flexible endoscopic trainers identified (2001–2005).
Spiteri A, Aggarwal R, Kersey T, Benjamin l, Darzi A, Bloom P 2010, 'Phacoemulsification skills training and assessment', British Journal of Ophthalmology, vol. 94, pp. 536–41.	Literature review concluded that improvements in technology can be utilised in ophthalmology and will help address the increasing limited opportunities for training and assessment during training and throughout a subsequent career.
Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ 2008, 'A systematic review of skills transfer after surgical simulation training', *Annals of Surgery*, vol. 248(2), pp. 166–79. Full report: Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ 2007, 'Surgical simulation for training: skills transfer to the operating room', ASERNIP-S Report 61, Adelaide, South Australia. Available at http://www.surgeons.org	Systematic review concluded that skills acquired by simulation-based training seem to be transferable to the operative setting. The studies included in this review were of variable quality and did not use comparable simulation-based training methodologies, which limited the strength of the conclusions. More studies are required to strengthen the evidence base and to provide the evidence needed to determine the extent to which simulation should become a part of surgical training programs.

Thijssen AS, Schijven MP 2010, 'Contemporary virtual reality laparoscopic simulators: quicksand or solid grounds for assessing surgical trainees?', *The American Journal of Surgery*, vol. 199, pp. 529–541.	Literature review concluded that using the right simulator, tasks, and metrics, trainees' and experts' laparoscopic skills can reliably be compared. However, VR simulators cannot yet predict levels of real life surgical skills.
Tan SS and Sarker SK 2011, 'Simulation in surgery: a review', *Scottish Medical Journal*, vol. 56, pp. 104–9.	Literature review concluded that a rigorous educational curriculum incorporating simulation will augment the operative exposure of surgical trainees, improving current training standards and its research.
Tsuda S, Scott D, Doyle J, Jones DB 2009, 'Surgical skills training and simulation', *Current Problems in Surgery*, vol. 46(4), pp. 271–370.	Literature review concluded that the adage of "see one, do one, teach one' has been replaced with a new educational heuristic of 'perfect practice makes perfect'. The new ethos advances patient safety through simulation and proficiency-based skills training.
Undre S and Darzi A 2007, 'Laparoscopy simulators', *Journal of Endourology*, vol. 21(3), pp. 274–280.	Expert review of literature concludes not enough evidence to show definite transfer of skills to operating room.
Walker K 2009, 'Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery (*Br J Surgery* 2008; 95:1088–1097)', *British Journal of Surgery*, vol. 96, pp. 221–3.	Letter to Editor concludes that we have sought to make our 'hybrid' laparoscopic simulator both highly accessible and highly incentivised through a simple intranet-based booking system, usable at short or immediate notice, availability out of hours, a 'no locked doors' policy, using CCTV instead for security, regular maintenance, and a stipulation that 'benchmarks' must be reached before progressing to equivalent tasks in the operating room. As a profession we set standards in examinations, which require the candidate to study in their own time. Surely simulator practice can be done on a similar understanding.

Appendix B: Hierarchy of evidence

Level of evidence	Study design
I	Evidence obtained from a systematic review of all relevant randomised controlled trials.
II	Evidence obtained from at least one properly designed randomised controlled trial.
III-1	Evidence obtained from well-designed pseudo-randomised controlled trials (alternate allocation or some other method).
III-2	Evidence obtained from comparative studies (including systematic reviews of such studies) with concurrent controls and allocation not randomised, cohort studies, case-control studies, or interrupted time-series with a control group.
III-3	Evidence obtained from comparative studies with historical control, two or more single arm studies, or interrupted time series without a parallel control group.
IV	Evidence obtained from case-series, either post-test or pre-test/post-test.

Source: NHMRC (2010).

.

Appendix C: Methodological assessment and study design tables

Simulation-based training versus no simulation-based training

Table C.01 LapSim VR simulator training versus no simulator training

Author	Intervention	Study design	Study population
Ahlberg et al 2007 Location Department of Surgery, Karolinska University Hospital Sweden	Objective: to investigate whether proficiency-based training on LapSim VR system improves objectively assessed, intraoperative performance during the initial learning of laparoscopic cholecystectomy, and if that improvement is persistent over time. Pre-test: before randomisation, all 13 surgical residents (6 men, 7 women) were tested in a series of psychometric tests and in the simulator. The Vandenburg and Kuse mental rotation test was used as high-level visuospatial function test. Verbal working memory was assessed using Wechsler Adult Intelligence Scale. A series of 5 tasks in the simulator were used for baseline skills assessment: grasping, lift grasp, cutting right, cutting left and clip application. Each subject filled in a questionnaire to control for variables associated with flow (i.e. their attitude toward simulator). Intervention: participants were randomised: • training: LapSim basic skills training (without force feedback) to proficiency. Proficiency on simulator defined by calculating the medium value for every parameter in each of six tasks from all five laparoscopic surgeons (experts). • control: no simulator training. Time to assessment: within 2 weeks either after study commencement (control group) or after reaching proficiency (trained group). Assessment: all participants performed first 10 individual full procedure laparoscopic cholecystectomies supervised by surgeon. 120 laparoscopic cholecystectomies were recorded on videotape. Surgery numbers 1, 5, and 10 were assessed for each subject. Procedure divided into 3 phases: exposure of the cystic duct and artery, clip placement followed by division of the cystic duct and artery, and gall bladder excision. Video assessments were performed by two observers, both experienced in laparoscopic surgery, who were blinded to training status, subjects and locations. Procedures were scored on a minute-by-minute basis for 28	Randomised controlled trial Method of randomisation: sealed envelope. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: not stated. Blinding: assessing surgeon blinded to training status and performed evaluation independently. Outcome measures: criteria assessed included: • frequency of error for each error type o errors for the entire procedure o exposure errors (1. lack of progress, 2. burn nontarget tissue, 3. nontarget structure injury, 4. instrument out of view, 5. attending takeover, 6. gall bladder Injury, 7. cystic duct injury, 8, inappropriate dissection, 9. incorrect angle of gallbladder retraction, 10. dropped retraction) o clipping and tissue division errors (1. attending takeover, 2. clip overlap, 3. clip spacing error, 4. poor clip orientation, 5. partial closure, 6. poor	Sample size: n=13 • training: n=7* • control: n=6. Baseline characteristics of participants: 13 surgical residents (PGY 1 or 2) from 9 different institutions in Sweden. All had laparoscopic assisting experience but no previous experience performing laparoscopic cholecystectomy.

	Simulator-trained	Controls
Variable	Median (min–max)	Median (min–max)
Age, y	32 (30–34)	32.5 (29–45)
Lap assist experience	15 (10–25)	18 (10–30)
Visuospatial assessment	8 (6–13)	13.5 (2–16)
Working memory assessment	26 (17–32)	26 (23–35)
Sex, male/female	3/4	3/3
Simulator test (lift and grasp)		
Total time, s	110.3 (61.9–150.9)	114.0 (73.4–182.4)
Left instrument path length, m	1.3 (1.2–1.6)	1.4 (1.1–1.7)
Left instrument angular path, °	317.8 (253.2–387.9)	314.7 (276.4–470.2
Right instrument path length, m	1.1 (1.0–1.4)	1.2 (1.0–1.4)
Right instrument angular path, °	210.8 (24.0–265.8)	279.9 (238.5–358.1)
Tissue damage, #	2.0 (1.0–3.0)	4.0 (2.0–6.0)

	surgical errors as defined (as used by Seymour et al. 2002). Interrater reliability of error assessment greater than 90% before the series. Device: LapSim software program version 2.0 run on a dual Xeon™ 1.8-GHz processor (Intel™ Corporation, Santa Clara, California, USA) using Windows XP and a virtual laparoscopic interface manufactured by Immersion Inc. (San Jose, California, USA) including 2 laparoscopic instruments.	application, 7. poor visualisation, 8 nontarget tissue clipped 9. clip drop, tissue division errors 10. inappropriate division. 11 clip cutting, 12. nontarget injury) o dissection errors (1. lack of progress, 2. burn nontarget tissue, 3. instrument out of view, 4. attending takeover, 5. gallbladder injury, 6. Liver injury, 7. Incorrect plane of tissue, 8. Tearing tissue) • time to task completion.	maximum damage, mm 4.7 (0.0–7) 5.2 (0.7–15.9) No significant difference between the two groups concerning baseline parameters. Inclusion: 13 surgical residents from nine different institutions in Sweden (PGY 1 to 2 years); all with experience in assisting with laparoscopic procedures but no experience in performing a laparoscopic cholecystectomy. Exclusion: not stated. Details of patients for live assessment: patients with a history of uncomplicated gallstone disease.

* confirmed by Ahlberg 3 July 2012
PGY: postgraduate year; VR: virtual reality; USA: the United States of America.

Table C.02 Limbs and Things laparoscopic simulator training versus no simulator training

Author	Intervention	Study design	Study population
Banks et al 2007 Location Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, New York USA	Objective: to assess whether, compared to traditional apprenticeship training, a surgical skills simulator laboratory improves resident knowledge and operative performance of laparoscopic bilateral tubal ligation (BTL). All participants: two consecutive classes of postgraduate year (PGY) 1 residents. Pretest: knowledge pretest (7 questions) and objective assessment of BTL in the skills laboratory rated with three assessment tools: task-specific checklist, global rating scale and pass/fail assessment. Intervention: participants were randomised: • training: surgical skills laboratory with simulator: 1 hour of focused didactic instruction on basic laparoscopy, entry techniques, suturing techniques and suture material. This was followed by 2 hours of hands-on teaching in skills laboratory with 3 stations: suturing pigs feet, knot tying boards, and a simulator station with a laparoscopic simulator and an operative laparoscopic tower. • control: no simulator training, only surgical teaching in operating room. Time to assessment: not stated. Assessment: both groups of residents were observed and evaluated when they performed their second laparoscopic BTL. They were rated with 3 tools: a task-specific checklist; a global rating scale, and a pass-fail grade. Posttest: knowledge post-test (7 questions in different order). Device: laparoscopic simulator (Limbs and Things Ltd., Bristol, UK) and an operative laparoscopic tower.	Randomised controlled trial Method of randomisation: names drawn from a hat. Allocation concealment: 10 residents randomly assigned to laboratory were told not to reveal their participation to classmates, attending physicians or senior residents. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: surgical skills laboratory was given in July of 2005 and 2006 followed by 4-month observation study period. Blinding: assessing observing surgeon blinded to training status. Outcome measures: 3 validated tools: a task-specific checklist; a global rating scale, and a pass-fail grade. • The task specific checklist was an internally validated 25 point checklist that assessed 4 categories of skills: preoperative skills, surgical technique, laparoscopic technique, and laparoscopic BTL-specific skills. • The global rating scale is a validated 5-point Likert scale that assesses 7 aspects of surgical skills: respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants, and knowledge of the specific procedure.	Sample size: n=20 • training: n=10 • control: n=10 Baseline characteristics of participants: information collected but data not shown. Inclusion: two consecutive classes of PGY1 residents. Exclusion: not stated. Details of patients for live assessment: not stated.

BTL: bilateral tubal ligation; PGY: postgraduate year; UK: United Kingdom; USA: the United States of America.

Table C.03 VR simulator Eyesi versus no simulator training

Author	Intervention	Study design	Study population
Belyea et al 2011 Location Department of Ophthalmology, George Washington University, Washington DC USA	Objective: to determine whether the use of an eye-surgery simulator during ophthalmology residency training improves cataract surgery performance. All participants: surgeries performed by 42 third-year George Washington University ophthalmology residents (22 men, 20 women) were reviewed. Resident was the primary surgeon working under the supervision of the same attending ophthalmologist who used the same technique of instruction and instrumentation throughout the study period. Pre-test: NA. Intervention: Participants were grouped according to pre or post simulation training: • group 1: simulation training group – post 2006 all residents were expected to spend a minimum of 2 hours per year using the simulator as verified by a login record. • group 2: control group – residents who operated with the same attending surgeon before virtual reality training was introduced (before 2006). Time to assessment: NA. Assessment: this cataract removal required the simultaneous use of the surgeon's 2 hands and 2 feet while the surgeon viewed through a light microscope. The primary outcome measures were phacoemulsification time, percentage phacoemulsification power used and intraoperative complications. Device: Virtual reality opthalmosurgical simulator Eyesi (VRMagic AB, Mannheim, Germany).	Retrospective comparative study Method of randomisation: NA. Allocation concealment: NA. Level of evidence: III-3. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: NA. Study period: before and after 2006 when virtual reality training was introduced. Blinding: same attending surgeon for all operations. Outcome measures: criteria assessed included: • phacoemulsification time • percentage phacoemulsification power used • adjusted phacoemulsification time calculated by multiplying the phacoemulsification time by the phacoemulsification power • intraoperative complications – recorded at the time of operation and assigned a grade on a scale from 1 to 4 - complication rate (%) and complication grade.	Sample size: n=42 • training: n=17 (8 men, 9 women) • control: n=25 (14 men, 11 women). Baseline characteristics of participants: all residents had performed a mean of 16 phacoemulsification cases (range 12–20 cases) before the start of their third year. The mean number of cases per resident in the simulator and non-simulator groups was 16.8 (range 5–45) and 12.2 (range 4–28), respectively. Inclusion: a total of 592 consecutive third-year resident cataract surgeries performed with the same attending surgeon using the same technique; 306 before simulator training and 286 after simulator training introduced. Exclusion: operations other than phacoemulsification or cases in which the resident was not the primary surgeon were excluded. Details of patients for live assessment: not stated.

AG: abbreviation for Aktiengesellschaft a German term for a type of company, similar to "Inc." or "LLC (limited liability company)" in the USA, public limited company (plc) in the UK; DC: District of Columbia, capital of USA; NA: Not applicable; USA: the United States of America.

Table C.04 MISTELS (FLS) or LAP Mentor VR simulator versus no simulator training

Author	Intervention	Study design	Study population
Beyer et al 2011 Location Centre d'Enseignement et de Recherche Chirurgicale (CERC) Faculté de Médecine de Marseille Secteur Nord, Université de la Mediterranée and Département de Santé Publique, Assistance Publique-Hôpitaux de Marseille, Hôpital de la Timone, Marseille France	Objective: to evaluate the impact of simulator training on residents' 'surgical technical skills' in the operating room. All participants: general surgery or gynecology-obstetrics residents joining the general and digestive surgery Department at the North Hospital in Marseille for 1 rotation (4 months). Pre-test: each resident was submitted to an initial evaluation during their first 2 months of the rotation in a real situation during a laparoscopic cholecystectomy in which the resident was aided by a teaching surgeon. Intervention: three groups of residents were formed (on a volunteer basis), each corresponding to a different rotation; all of them benefited from surgical mentorship as the principal means of training from a team of 8 teaching surgeons. The groups only differed according to the complementary training they received for 1 month: • group 1: MISTELS training on a simple simulator: 5 individual 60-minute sessions over a period of 1 month. • group 2: LAP Mentor training on a virtual simulator: 5 sessions in 1 month organised in pairs with each pair undergoing five 120-minute sessions over a period of 1 month. Residents attempting all 9 basic exercises, the intracorporeal knot suture exercise and 1 cholecystectomy in each session. • group 3: control: mentorship only (no simulator training). Time to assessment: up to 1 month after training, i.e. the fourth month. Assessment: the validated Global Operative Assessment of Laparoscopic skills (GOALS) score was based on dissection of the vesicular bed. Device: group 1 trained on a McGill Inanimate System of Training and Evaluation of Laparoscopic Skills (MISTELS) Simulator (FLS Trainer Box). Group 2 trained on LAP Mentor VR Simulator (Symbionix, Cleveland, Ohio, USA).	Monocentre comparative prospective study Method of randomisation: not applicable. Allocation concealment: not applicable. Level of evidence: III-3. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: between May 2007 and July 2008 over a period of 3 residency rotations. Blinding: two assessing surgeon (experts in laparoscopy and properly trained in assessment of videos), blinded to training status (and whether first or second evaluation), performed independent assessments watching every video recording of the procedures. A random video montage ensured anonymity of the resident and surgical assistant. Outcome measures: criteria assessed included: • GOALS score composed of 5 items: perception of depth, bimanual dexterity, efficiency, tissue handling, autonomy (item 5 was excluded from the statistical analysis as it was difficult to evaluate the degree of assistance and counselling by the senior surgeon from the video). Total score = 20 using 5-point scale for 4 items. • Visual Analog Scale (VAS) used for difficulty of surgery.	Sample size: n=19 • training group 1: n=6 • training group 2: n=6 • control: n=7. Baseline characteristics of participants: there was no significant difference between the 3 groups in terms of overall laparoscopic tasks (operator or assistant operator). There was a significant difference as operator between the MISTELSs and the control group in favour of the control group (p=0.03) and between the MISTELS and the LAP Mentor groups in favour of the LAP Mentor group (p=0.04) but no significant difference between LAP mentor and control groups. Inclusion: volunteer general surgery or gynecology-obstetrics residents joining the general and digestive surgery Department at the North Hospital in Marseille for 1 rotation. Exclusion: not stated. Details of patients for live assessment: not stated.

CERC: Centre d'Enseignement et de Recherche Chirurgicale; FLS: Fundamentals of Laparoscopic Surgery; GOALS: Global Operative Assessment of Laparoscopic Skills; MISTELS: McGill Inanimate System of Training and Evaluation of Laparoscopic Skills; VAS: Visual Analog Scale; VR: virtual reality; USA: the United States of America.

Table C.05 LapSim VR simulator training versus no simulator training

Author	Intervention	Study design	Study population
Cosman et al 2007 Location University of Sydney, Sydney Australia	Objective: to determine whether laparoscopic skills acquired on a virtual reality simulator (LapSim) would transfer into the operating suite. All participants: volunteer basic surgical trainees. Pretest: measured by simulator but not shown: time to task completion, path lengths, angular path, number of incomplete targets, number of misplaced clips, number of dropped clips, maximum stretch damage (as a proportion of force required to rupture the vessel) and the amount of blood loss. Intervention: participants were randomised: • training: practice the clipping task on the LapSim simulator, following a distributed training protocol including access for a maximum of one hour a day, until they satisfied the performance criteria on two successive repetitions of the task. • control: no simulator training. Time to assessment: not stated. Assessment: under consultant supervision, they were instructed to apply clips to and divide the cystic duct or the cystic artery during laparoscopic cholecystectomy on a live human patient. The error assessment scale was based on a similar instrument designed and validated by others (Eubanks et al 1999) for assessment of performance during laparoscopic cholecystectomy. The scale comprised six parts: application of the grasper to Hartmann's pouch, retraction of Hartmann's pouch, application of patient-side clip, application of specimen-side clip, transection and miscellaneous. Device: LapSim System (no haptic feedback available) Basic Skills package version 1.5 by Surgical Science Ltd, Gothenburg, Sweden.	Randomised controlled trial Method of randomisation: not stated. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: August 2002 to December 2002. Blinding: independent assessment by five surgeons blinded to training status. Outcome measures: criteria assessed included: • error assessment scale comprised six parts: application of grasper to Hartman's pouch, retraction of Hartman's pouch, application of the patient side clip, application of the specimen-side clip, transaction and miscellaneous • bimanual coordination and overall assessment (using 5-point Likert scale) • time to task completion.	Sample size: n=10 • training: n=5 • control: n=5. Baseline characteristics of participants: not stated except that 'No statistically significant differences were recorded between the two groups in terms of the level of training (mean ± SD = 1.3 ± 0.5 years) or number of laparoscopic cases performed (mean ± SD = 7.7 ± 12.8)'. Inclusion: volunteer basic surgical trainees. Exclusion: not stated. Details of patients for live assessment: not stated.

SD: standard deviation; VR: virtual reality.

Table C.06 Endoscopic Sinus Surgery Simulator (ES3) training versus no simulator training

Author	Intervention	Study design	Study population
Fried et al 2010 Location Department of Otorhinolaryngology-Head and Neck Surgery, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York USA	Objective: to demonstrate the predictive validity of the transfer of simulator-acquired skills onto live patient endoscopic sinus surgery (ESS) in the operating suite. All participants: otolaryngological surgical residents postgraduate year (PGY) 1-2. Intervention: participants were randomised: • training: subjects were full trained to proficiency with ES3 in addition to receiving conventional textbook-based and video recorded educational material. • control: no simulator training, but received conventional material. Time to assessment: not stated. Assessment: first in vivo ESS procedure performed by subjects, standardised around the completion of basic tasks, was video recorded. Those tasks included scope navigation, mucosal injection, middle turbinate medialisation, uncinectomy and maxillary antrostomy. The video recordings were then edited and de-identified. Device: Endoscopic Sinus Surgery Simulator (ES3) (Lockheed Martin, Inc, Akron, Ohio, USA).	Randomised controlled trial Method of randomisation: not stated. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: 3 had prior experience in ESS. Study period: not stated. Blinding: expert panel blinded to training status rated the video recordings using previously described custom-made software application, incorporating consensual ESS metrics. Outcome measures: criteria assessed over three main recorded tasks (i.e. navigation, injection and dissection) included: • time to completion of task (minutes) • case difficulty • tool manipulation • tissue respect • task completion rate • surgical confidence (10-point scale) • number of errors as defined.	Sample size: n=28 (25) • training: n=12 • control: n=13. Baseline characteristics of participants: not stated. Inclusion: otolaryngological surgical residents PGY 1–2 with complete lack of previous hands-on experience in ESS. Exclusion: not stated. Details of patients for live assessment: not stated.

ESS: endoscopic sinus surgery; PGY: postgraduate year; USA = the United States of America.

Table C.07 LapSim VR simulator training versus no simulator training

Author	Intervention	Study design	Study population
Hogle et al 2009 Location Department of Surgery, College of Physicians and Surgeons, Columbia University, New York USA	Objective: study 1 – to compare performance in the operating room after training on a laparoscopic simulator and after no training. All participants: first year surgical residents. Prestudy baseline and poststudy simulator testing were completed for all participants. Intervention: participants were randomised: • training: the training curriculum was fully completed when level 3 was passed for each module. Tasks included: camera navigation instrument navigation, coordination, grasping, lifting and grasping, cutting, and clip applying. The participants were asked to independently complete two simulation training sessions per week. No other simulation training or practice was allowed outside the training curriculum. • control: no simulator training. Time to assessment: within 1 month of randomisation. Assessment: at 1 month after baseline testing, the participants were recorded during their next two elective laparoscopic cholecystectomies. The supervising attending surgeon evaluated performance using the Global Operative Assessment of Laparoscopic Skills (GOALS). The video tapes were used for subsequent blinded evaluation and scoring with GOALS. Device: LapSim (Surgical Science Ltd, Gothenburg, Sweden).	Randomised controlled trial Method of randomisation: random number generator. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: 1 failed to return any videotapes (training or control group not stated). Study period: not stated. Blinding: unblinded attending surgeon but two video assessing reviewers blinded to training status. Outcome measures: criteria assessed included: • depth perception • bimanual dexterity • efficiency • tissue handling • autonomy.	Sample size: n = 13 (12) • training: n = 6 • control: n = 6. Baseline characteristics of participants: not stated. Inclusion: first year surgical residents. Exclusion: not stated. Details of patients for live assessment: not stated.

GOALS: Global Operative Assessment of Laparoscopic Skills; USA: the United States of America.

Table C.08 Arthroscopy knee bench-top simulator training versus no simulator training

Author	Intervention	Study design	Study population
Howells et al 2008 Location Nuffield Orthopaedic Centre, Oxford, England United Kingdom	Objective: to investigate the transfer validity of arthroscopic skills of surgical trainees from simulator training to the operating theatre in performing laparoscopic diagnostic arthroscopy of the knee. All participants: junior orthopaedic trainees. Pretest: during simulator training a 3D electromagnetic movement tracking system (Patriot, Polhemus, Colchesterm Vermont) was used to assess surgical performance objectively. Intervention: participants were randomised: • training: simulator training (3 sessions of 6 simulated arthroscopies in 1 week. 18 simulated arthroscopies were supervised by lead author and followed a fixed protocol for a diagnostic arthroscopy of the knee agreed by the two surgeons experienced in this area) plus traditional training • control: no simulator training but traditional training. Time to assessment: not stated. Assessment: diagnostic arthroscopy of knee performed in theatre using a procedure-based assessment from the Orthopaedic Competence Assessment Project (9 of the 14 criteria) and the validated Objective Structured Assessment of Technical Skills (OSATS) five-point global rating scale for the 9 criteria. Device: an arthroscopy knee bench-top simulator (Sawbones, Malmö, Sweden) and a standard 30˚ arthroscope with an arthroscopic camera and display system (Smith & Nephew Endoscopy, Huntingdon, UK).	Randomised controlled trial Method of randomisation: sealed envelopes. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: February 2006 to February 2007. Blinding: assessing surgeon blinded to training status. Outcome measures: OSATS global rating scale criteria assessed on 5-point scale included: • follows protocol • handles tissue well • appropriate and safe use of instruments • appropriate pace with economy of movement • act calmly and effectively with untoward events • appropriate use of assistant • communicates with scrub nurse • clearly identifies common abnormalities • protecting the articular surface.	Sample size: n = 20 • training: n = 10 • control: n = 10. Baseline characteristics of participants: information collected but data not shown; 'The groups were equally matched in terms of demographics and previous surgical experience.' Inclusion: junior orthopaedic trainees with less than two years surgical training and minimal previous experience of arthroscopy, having observed or assisted in fewer than 10 arthroscopies, or other minimal-access procedures. Exclusion: not stated. Details of patients for live assessment: not stated.

3D: three-dimensional; OSATS: Objective Structured Assessment of Technical Skills; UK: the United Kingdom.

Table C.09 VR simulator PelvicVision versus no simulator training

Author	Intervention	Study design	Study population
Källström et al 2010 Location Department of Urology, University Hospital Linköping, Linköping Sweden	Objective: to test if practising the transurethral resection of prostate (TURP) procedure in a virtual reality simulator increases the skills and dexterity of urology residents when performing the procedure on patients. All participants: Swedish urology residents with some, but minor, experience with transurethral procedures, agreed to undertake a 5-day course on treatment of benign enlargements of the prostate. Pretest: all participants performed three supervised TURP procedures on patients with the same supervisor for each Group 1 or 2. Intervention: participants were randomised: • group 1: simulation practice – on day 2 each participant was instructed during their simulated procedures and required to practice until reaching the 'expert' level on a specific 'patient' – test procedure (mean time 198 minutes). The total mean practice time was 254 minutes distributed over 8.3 procedures, of which 3.7 were test procedures. • group 2: control – on day 2 performed second TURP procedure without prior simulator training. All participants were instructed in theory, diagnostic methods, the instrumentation used in TURP procedure, and risk factors. On Day 1 all performed TURP procedure under supervision. On day 4, group 1 performed their second TURP procedure with the same supervisor and group 2 practised on the simulator until they reached 'expert' level. On day 5 all participants performed their third TURP procedure with the same supervisor. All TURP procedures were video recorded for minute-by-minute analysis. Time to assessment: consecutive days or one day between assessments. The first TURP was performed on day 1 and the third TURP was performed on day 5. Assessment: the analyses of the video recordings were done 'blindly' (without the knowledge of the course, participant, supervisor, patient identity or the order of procedure) by two surgically experienced urologists. Device: VR simulator PelvicVision (Melerit Medical AB, Linköping, Sweden).	Randomised controlled trial Method of randomisation: not stated. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: one participant could not be evaluated regarding the effect of simulation practice (group1) because of exclusion of one patient due to an upper respiratory tract infection. Study period: during 2006 and 2007. Blinding: the analyses of the video recordings were done 'blindly' (without the knowledge of the course, participant, supervisor, patient identity or the order of procedure) by two surgically experienced urologists. Outcome measures: criteria assessed included: • task-specific checklist with 21 items • global five-point ratings scale including respect for tissue; time and movements; eye-hand coordination; foot pedals; videoscope; resection; strategy; tempo; use of assistants; stress level; supervision; communication with supervisor; knowledge of the procedure; final result • pass/failure score • self-evaluation • patient follow-up 6–12 months postoperation using the International Prostate Symptom Score, the bother question, maximum urinary flow rate and incontinence score (Linköping incontinence questionnaire).	Sample size: n=24 (23) • training: n=12 (11) • control: n=12. Baseline characteristics of participants: no significant differences between the two groups regarding sex (women:men 1:3), age (mean 33 years) or prior experience with transurethral procedures (mean 14 months of residency, mean 1.2 TURP procedures performed completely, and mean 6.2 incompletely). Inclusion: Urology residents with some, but minor, experience with transurethral procedures e.g. cystoscopy, transurethral resection of bladder tumours and TURP procedures, who agreed to undertake a 5-day course on treatment of benign enlargement of the prostate. Exclusion: not stated. Details of patients for live assessment: the mean prostate volume (as measured by transurethral ultrasound) was 37.7cc (19–67cc) and the mean age was 72.5 years. 71 patients who underwent surgery had a follow-up time of 2.3 to 3.8 years.

AB: abbreviation of Aktiebolag, the Swedish term for 'limited company or corporation; TURP: transurethral resection of prostate; VR: virtual reality.

Table C.10 LapSim Gyn VR simulator training versus no simulator training

Author	Intervention	Study design	Study population
Larsen et al 2009 Location Copenhagen University Hospital Rigshospitalet, Copenhagen Denmark	Objective: to assess the effect of virtual reality training on an actual laparoscopic operation. All participants: eight of the total cohort of 42 trainees (38 women, four men) were ineligible as they were too experienced for the study (advanced laparoscopy defined as all laparoscopic procedures involving coordination of more than one instrument) and four came from the two gynaecology departments in the Zeeland region not participating in the trial. 24 first and second year registrars specialising in gynaecology and obstetrics (postgraduate years 3–8). Intervention: participants were randomised: • training: proficiency-based virtual reality simulator training in laparoscopic salpingectomy (for ectopic pregnancy) and standard clinical education • control: no simulator training, but standard clinical training alone. Time to assessment: not stated. Assessment: one author present as observer to ensure standard procedure followed and recording done correctly. Operation recorded on DVD using camera attached to laparoscope for later blinded evaluation. Technical performance measured as total score (10–50 points) using the validated objective structure assessment of laparoscopic salpingectomy. Device: LapSim Gyn v 3.0.1 (Surgical Science Ltd., Gothenburg, Sweden) was run on IBM T42 computer in a docking station (Pentium M 1.8GHz/512 MB RAM; IBM, Armonk, New York, USA) using an interface with a diathermy pedal (Virtual Laparoscopic Interface; Immersion, San Jose, California, USA).	Randomised controlled trial Method of randomisation: stratified randomisation based on previous experience of simple laparoscopy, independently randomised by computer using identifier numbers only. Allocation concealment: all involved departments, supervisors and staff in operating theatres were blinded to trainee's group and the assessors were blinded to both trainee and their allocated group. Level of evidence: II. Intention to treat: not stated. Power calculation: minimal relevance difference between novice and intermediately experienced laparoscopists was six points. With α of 0.5 (two sided) and a power of 80% (β=0.2 giving Zα=1.96 and Zβ=0.84, largest SD=4/40) 18 or more trainees required (additional third to allow for possible drop outs totalling 24). Lost to assessment: two trainees were excluded from simulator-trained group because they failed to complete the training or the operation was cancelled. One trainee was excluded from the control group due to technical fault in DVD recorder. Study period: September 2006 to June 2007. Blinding: assessed by two independent observers blinded to trainee and training status. Outcome measures: criteria assessed included: • technical performance measured as total score (10–50 points using validated objective structured assessment of laparoscopic salpingectomy comprising five item general rating scale and five item specific rating scale) • operation time in minutes.	Sample size: n=24 • training: n=13 • control: n=11. Baseline characteristics of participants: Variable / Simulator-trained group (n=13) / Control group (n=11) Men: 1 / 1 Women: 12 / 10 Mean (range) age (y): 33.3 (30–42) / 32.4 (26–38) Experience of simple laparoscopy: 6 / 5 No experience of simple laparoscopy: 7 / 6 Inclusion: gynaecological specialty trainees years 1 and 2 (postgraduate years 3–8) with no experience in advanced laparoscopy and from one of the seven gynaecology departments in the Zeeland region of Denmark participating in the study. Details of patients for live assessment: only elective right-sided salpingectomy before treatment for infertility or for prophylactic removal of fallopian tubes and ovaries owing to a positive test result for breast cancer gene 1 (BRCA1); not patients who had undergone previous open or laparoscopic surgery below umbilicus, had possible abdominal malignant disease, had an American Surgical Association score ≥3 (patients with severe systemic disease), had a body mass index less than 18 or more than 27, had haemophilia, or had other factors of potential influence on the surgical procedure.

BRCA1: breast cancer gene 1; DVD: Digital Versatile Disc; IBM: International Business Machines Corp.; MB: megabyte; RAM: random-access memory; SD: standard deviation; USA: the United States of America.

Table C.11 Nasal Model simulator training versus no simulator training

Author	Intervention	Study design	Study population
Ossowski et al 2008 Location Department of Otolaryngology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania USA	Objective: to assess whether training naïve subjects on an anatomically correct nasal model (NM) improves performance based on time to complete the required task and on the comfort level ranked by a standardised patient. All participants: 20 medical students with no prior endoscopic experience Intervention: participants were randomised: • training: 15-minute video instruction on endoscopy followed by 15-minute practice session on simulator • control: 15-minute video instruction on endoscopy but no simulator training. Time to assessment: within 90 minutes of the original demonstration. Assessment: all students were timed performing flexible nasolaryngoscopy and rigid nasal endoscopy of the NM. Each student was timed performing flexible laryngoscopy on a single standardised patient. This patient (human volunteer) was blinded to which group the student belonged. The standardised patient (whose nose was anaesthetised 20 minutes before each testing and again every 20 minutes until finished) filled in a comfort/discomfort score for each student on a scale from 0 to 10. Device: nasal model for endoscopic simulation.	Randomised controlled trial Method of randomisation: not stated but stratified by prior video game experience. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: not stated. Blinding: patient blinded to training status. Outcome measures: criteria assessed included: • time to complete tasks on video rigid nasal endoscopy and video flexible laryngoscopy – tested on NM • time to complete flexible laryngoscopy on standardised patient • visual analogue scale scores of comfort/discomfort by patient.	Sample size: n=20 • training: n=10 • control: n=10. Baseline characteristics of participants: not stated. Inclusion: medical students with no prior endoscopic experience. Exclusion: not stated. Details of patients for live assessment: human volunteer.

NM: nasal model; USA: the United States of America.

Table C.12 Abdominal wall simulator training versus no simulator training

Author	Intervention	Study design	Study population
Palter et al 2011 Location University of Toronto, Department of Surgery, St Michael's Hospital, Toronto Canada	Objective: to assess the effect of ex-vivo technical skills training on cognitive learning in the operating room. All participants: novice surgical residents (all in first 5 months of residency) were taught the basics of fascial closure in an initial group training session that included a demonstration of a technically correct fascial closure on a synthetic abdominal wall model by a staff general surgeon. Pretest: each participant then performed 1 closure on a low fidelity synthetic model which was evaluated by two observers using an Objective Structured Assessment of technical Skills (OSATS) global rating scale. Intervention: participants were randomised: • intervention group – each participant practised on the models until technical proficiency was reached. Each session was a maximum 1.5 hours in length and occurred no longer than 3 weeks apart (proficiency reached by second session). • control group – had no further contact with models. Time to assessment: not stated. Assessment: each participant was instructed to close the fascia of a patient whose surgery required an abdominal incision. While closing the abdomen, the same study team member read a script that contained information relevant to the procedure including wound infections, care of wound infections and hernias. The script ran for 10 minutes. The staff member who supervised the abdominal wall closure assessed the residents' technical performance utilising the OSATS global rating scale. The primary outcome measure in this study was the difference in multiple-choice test scores between the residents in the control and intervention groups. The secondary outcome measure was the difference in technical skill proficiency between the ex vivo trained group compared with the untrained group. Device: synthetic abdominal wall model.	Randomised controlled trial Method of randomisation: sealed envelopes. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: minimum effect size of 1.2, an α of 0.5 and a power of 0.80, the minimum number of participants in each group is 9. Lost to assessment: one participant recruited did not complete due to leave of absence for personal reasons (not stated whether from training or control group). Study period: not stated. Blinding: assessing surgeon was blinded to training status of participants. Outcome measures: criteria assessed included: • at the completion of the case, each participant completed a multiple-choice test of 22 questions designed to assess how much information they retained from the script • OSATS global rating scale.	Sample size: n=18 (19 recruited) • training: n=9 • control: n=9. Baseline characteristics of participants: not stated except no difference in technical ability, i.e. for median baseline global rating score of residents in the control group was 22.0 (20.5–23.0) and for the intervention group 21.0 (20.0–21.0) (p=0.48 NS). Inclusion: University of Toronto residents in their first postgraduate year of training in either surgery or obstetrics and gynaecology who volunteered to participate at a program-specific orientation or by email. Exclusion: not stated. Details of patients for live assessment: a patient whose surgery required an abdominal incision.

OSATS: Objective Structured Assessment of Technical Skills.

Table C.13 AccuTouch colonoscopy simulator training versus no simulator training

Author	Intervention	Study Design	Study population
Park et al 2007 Location Faculty of Medicine, University of Toronto, Toronto Canada	Objective: to assess whether training on a computer-based colonoscopy simulator outside of the colonoscopy suite improves performance on a resident's first patient colonoscopy in a clinical setting. All participants: general surgery and internal medicine residents (PGY 1 to 3). Pretest: all residents watched a video introduction to colonoscopy and had an opportunity to familiarise themselves with components and handling of a colonoscope but no instruction or practice before pretest on the colonoscopy simulator using module 1, with a time limit of 30 minutes. 8 parameters of most clinical relevance were selected as outcome measures for analysis. Expert global ratings of performance was completed by a single faculty endoscopist. Intervention: participants were randomised: • training: 2 to 3 hours (mean 125 minutes SD 37 min) of practice independently on simulator with access to the range of cases on simulator • control: no simulator training. Time to assessment: within 14 days (range 2–14 days) of their simulator testing and training. Assessment: performed patient colonoscopy as primary endoscopists under the supervision of 1 of 3 faculty endoscopist evaluators blinded to residents' training groups using previously validated performance metrics. Device: AccuTouch colonoscopy simulator (Version 1.2 with haptic feedback from simulated patient (recording of vital signs, verbalisation of discomfort); Immersion Medical, Gaithersburg, Maryland, USA). The software module included with the simulator contained 6 different and progressively more difficult cases. For each simulated procedure performed, the simulator generated a record of 14 different performance metrics.	Randomised controlled trial Method of randomisation: not stated. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: to detect a size of 1 SD, using a 1-tailed alpha of 0.05 and a power of 0.8, a minimum of 13 subjects in each group was required. 28 participants were recruited. Lost to assessment: recruited 28 but 4 (2 from each group) were unable to complete clinical phase. Study period: not stated. Blinding: assessing surgeon blinded to training status. Outcome measures: criteria assessed included: • their ability to independently reach the caecum • the absence of critical flaws (perforation or significant bleeding) during procedure • global ratings completed by the endoscopist evaluators. The 5 point global rating scale consisted of 7 items with scores summed to generate a total global performance score (out of 35): • atraumatic technique • colonoscope use/advancement • use of instrument controls • flow of procedure • use of assistants • knowledge of specific procedure • overall performance.	Sample size: n=28 recruited (24 completed) • training: n=12 • control: n=12. Baseline characteristics of participants: similar proportions of gender, age and postgraduate year level in each group. In addition, the ratio of general surgery to internal medicine residents was 2:10 in control group and 3:9 in the treatment group. There were no significant differences between groups on any of the 8 computer-recorded performance metrics on the pretest procedure on the simulator. Similarly the faculty-generated global ratings for the pretest showed no significant differences between groups, with the treatment group scoring 13.7 (SD 4.5) out of a maximum possible score of 35 versus 15.4 (SD 4.8) for the control group. Inclusion: general surgical and internal medicine residents with limited endoscopic experience. Exclusion: residents who were primary endoscopist for greater than 3 procedures of any type were excluded. Details of patients for live assessment: patients scheduled to undergo a clinically indicated colonoscopy. Only patients between ages 40 and 75 years were included, with no previous colon or rectal resection, no history of difficult colonoscopy (secondary to anatomy or patient compliance) and no history of inflammatory bowel disease.

PGY: postgraduate year; SD: standard deviation; USA: the United States of America.

Table C.14 Simulated Operating Suite or Second Life virtual operating theatre versus no simulation training

Author	Intervention	Study design	Study population
Patel et al 2012 Location Division of Surgery, Imperial College London, St Mary's Hospital, London United Kingdom	Objective: to establish whether an introduction training session for novice medical students would increase their knowledge regarding operating theatre personnel, layout, dress, and activity and enhance their appropriate behaviour for the initial attendance within the operating theatre. All participants: first year medical students. Pretest: all subjects attended, either in pairs or independently, first operative case in day surgery unit. Immediately before entry the novices completed a demographic questionnaire and knowledge bases multiple-choice questionnaire. A trained observer rated all subjects according to the checklist observation scale. Following completion of the operative case, subjects completed the self-report questionnaire. Intervention: participants were randomised into 4 groups: • training: didactic lecture with PowerPoint lasted 1 hour and included instructional videos regarding gowning and gloving and delivered to 5 students per session with opportunity to ask questions at the completion of the lecture • VR simulation-based training: Second Life Virtual World operating theatre using avatars in groups of 5 students per 1-hour session with additional information supplied by instructor through text or voice chat • SOS training: simulated operating theatre curriculum-based instruction delivered to 5 students at a time for 1 hour • control: no simulation-based training or lecture Time to assessment: both operating theatre attendances and training intervention were performed within a 7-day period for each subject. Assessment: knowledge, skills and attitudes, measured using observed behaviour (checklist observation Likert-type scale) and a self-report Likert-type scale, with knowledge further assessed using multiple choice questions. Devices: Second Life (Linden Research Inc. San Francisco, California, USA) virtual operating theatre, and Simulated Operating Suite (SOS, Imperial College, London, UK).	Randomised controlled trial and comparative training study Method of randomisation: not stated. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: not stated. Blinding: not stated. Outcome measures: criteria assessed included: • knowledge, skills and attitudes, measured using observed behaviour (checklist observation Likert-type scale) • self-report Likert-type scale, with knowledge further assessed using multiple choice questions.	Sample size: n=60 • training didactic lecture: n=15 • training Second Life: n=15 • training SOS: n=15 • control: n=15. Baseline characteristics of participants: 32 women and 28 men ranging in age from 18 to 21 years participated in this study - demographics collected but not shown. The preintervention exposure to the operating theatre revealed no differences among the groups for knowledge assessment (*p*=0.477), observation score (*p*=0.212), and self-report scores (*p*=0.099). Inclusion: first year medical students; 'novices' with no experience in operating theatre. Exclusion: having previously attended the operating theatre. Details of patients for live assessment: not stated.

SOS: simulated operating suite; UK: the United Kingdom; USA: the United States of America; VR = virtual reality.

Table C.15 URO Mentor VR simulator training versus no simulator training

<table>
<tr><th>Author</th><th>Intervention</th><th>Study design</th><th>Study population</th></tr>
<tr>
<td>Schout et al 2009

Location
Catharina Hospital Eindhoven, Eidhoven

The Netherlands</td>
<td>Objective: this randomised, single-blind, multicentre study investigated whether practical skills training on the UM VR simulator improved performance of cystourethroscopy (CUS) in real patients.

All participants: Interns.

Intervention: participants were randomised:
• training: simulation-based training. The UM simulates flexible and semi-rigid ureterorenoscopy (URS) and CUS. The training protocol consisted of seven flexible CUS tasks, which included stone-manipulation tasks numbers 3 and 8, and basic tasks numbers 4, 5 and 9 of the UM.
• control: no simulator training.

Time to assessment: all participants performed flexible CUS in a patient within 3 days (range 0–3) after the preparation phase (15 minute instruction video and 15 minutes maximum with a real cystoscope in a glass globe representing the bladder), irrespective of whether they trained on UM or not.

Assessment: real-time performance scored on a validated five-point Global Rating Scale by supervisors unaware of training status.

Device: URO Mentor (UM, Simbionix USA Corp., Cleveland, Ohio, USA) is a virtual reality (VR) simulator.</td>
<td>Randomised controlled trial

Method of randomisation: using http://www.randomization.com.

Allocation concealment: not stated.

Level of evidence: II.

Intention to treat: not stated.

Power calculation: a priori showed a sample of 102 was needed to detect a between-group difference with an effect size of 0.50, a power of 0.80, and an α of 0.05 (independent t-test).

Lost to assessment: not stated.

Study period: March 2007 to February 2008.

Blinding: assessing supervisors blinded to training status.

Outcome measures: 5 point Global Rating Scale criteria assessed included:
• respect for tissue
• time and motion
• handling endoscope
• flow of procedure and forward planning
• knowledge of procedure.</td>
<td>Sample size: n=100
• training: n=50
• control: n=50.

Baseline characteristics of participants:
<table>
<tr><th>Variable</th><th>Simulator-trained</th><th>Controls</th></tr>
<tr><td>No. participants</td><td>50</td><td>50</td></tr>
<tr><td>Age, years - mean (SD)</td><td>24.2 (3.0)</td><td>23.7 (3.3)</td></tr>
<tr><td>Gender, M/F</td><td>26/74</td><td>34/66</td></tr>
<tr><td>Dexterity Right/left/ambidextrous</td><td>88/10/2</td><td>90/8/2</td></tr>
<tr><td>No. of CUS ever observed – Mean (SD)</td><td>2.7 (5.4)</td><td>2.9 (7.4)</td></tr>
<tr><td colspan="3">Years of training</td></tr>
<tr><td>4</td><td>72%</td><td>80%</td></tr>
<tr><td>5</td><td>20%</td><td>8%</td></tr>
<tr><td>6</td><td>8%</td><td>12%</td></tr>
<tr><td colspan="3">Future interest/specialism</td></tr>
<tr><td>Nonsurgical/unknown</td><td>78%</td><td>69%</td></tr>
<tr><td>Surgical</td><td>22%</td><td>31%</td></tr>
</table>
Inclusion: Interns.

Exclusion: Trainees who had previous experience with UM, had performed CUS in a patient, or were unable to complete the full study protocol for personal reasons.

Details of patients for live assessment: included men aged 38–84 years who had a medical indication of CUS. Exclusion consisted of previous major urological surgery.</td>
</tr>
</table>

CUS: cystourethroscopy ; SD: standard deviation; UM: URO Mentor™ simulator; URS: ureterorenoscopy; USA: the United States of America; VR: virtual reality.

Table C.16 GI Mentor II VR simulator training versus no simulator training

Author	Intervention	Study design	Study population
Shirai et al 2008 Location Yamaguchi University Graduate School of Medicine, Ube, Yamaguchi Japan	Objective: to evaluate the difference between simulator and bedside training at an early stage of endoscopic training, medical residents with no endoscopic experience were randomised to simulator and non-simulator training groups, and their skills evaluated by two blinded supervisors while they performed oesophagogastroduodenoscopy (OGD) on volunteers. All participants: hospital medical residents (PGY1 or 2). Intervention: after 3 hours of explanation regarding the manipulation of an endoscope, endoscopic observation, and endoscopic diagnosis of common diseases, participants were randomized: • training: 5 x 1-hour training on simulator within 2 weeks i.e. two psychomotor tasks level 1 EndoBubble and EndoBasket were performed three times each and then OGD training modules completed. Case 1-I was performed in each session and remaining time used for other cases of OGD module. The residents were not supervised or instructed during the simulator training. Residents also had 15 hours of bedside training • control: no simulator training; only 15 hours of bedside training. Time to assessment: not stated. Assessment: each subject performed endoscopy twice on volunteers without sedation for assessment. After the first evaluation, the supervisor gave the residents some oral advice to improve their skills. The interval between the first and second evaluations was within 1 week. Performance was evaluated according to a five-grade scale of a total of 11 items. The volunteers consisted of doctors and residents. Device: GI-Mentor II simulator (Simbionix USA Corp, Cleveland, Ohio, USA) with a mannequin with integrated force feedback system and a computer simulation program. The software package allows the user to practice OGD, colonoscopy, endoscopic retrograde cholangiopancreatography or CyberScopy. The software includes various cases that simulate both diagnostic and therapeutic scenarios.	Randomised controlled trial Method of randomisation: envelopes. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: October 2004 to March 2006. Blinding: two assessing physicians blinded to training status. Outcome measures: criteria assessed on 5 point scale included: • insertion into oesophagus • crossing the oesophagogastric junction (OGJ) • passing from OGJ into the gastric antrum • passing through the pyloric ring • examination of the duodenal bulb • insertion into the third part of the duodenum • examination of the gastric antrum • examination of the gastric angle • manipulation for retroflexion • looking down the gastric body • viewing the fornix. 5-point scale included: 5 points for 'resident could perform the maneuver as well as supervising physician'; 4 points for 'skill good, but not as good as supervising physician'; 3 points for 'resident could perform without receiving instructions'; 2 points for 'instructions were required'; 1 point for 'direct assistance by the supervisor was required'.	Sample size: n=20 • training: n=10 • control: n=10. Baseline characteristics of participants: there was no difference in the age and male:female ratio between the simulator and non-simulator groups (26.00 ± 0.77 years versus 27 ± 1.91 years and 5:5 versus 6:4 respectively). Inclusion: medical residents with no endoscopic experience. Exclusion: not stated. Details of patients for live assessment: the volunteers consisted of doctors and residents. There was no significant difference in age or sex between volunteers used within each group (some had duodenal ulcer scars, hiatus hernia or reflux oesophagitis but none of these findings were considered to have an influence on the difficulty of performing OGD).

GI: gastrointestinal; OGD: oesophagogastroduodenoscopy; OGJ: oesophagogastric junction; PGY: postgraduate year; USA; the United States of America; VR: virtual reality.

Table C.17 FLS simulator training versus no simulator training

Author	Intervention	Study design	Study population
Sroka et al 2010 Location Stenberg-Bernstein Centre for minimally Invasive Surgery and Innovation, McGill University, Montreal Quebec Canada	Objective: to assess the transfer of skills acquired by novices trained to proficiency on Fundamentals of Laparoscopic Surgery (FLS) simulator to operating room (OR) performance as measured by Global Operative Assessment of Laparoscopic Skills (GOALS). All participants: 19 junior general surgical residents (PGY years 1–3) underwent baseline FLS testing and were assessed in the OR at baseline using a validated global rating scale (GOALS). Two were excluded as GOALS score >15. 17 residents with GOALS score ≤ 15 were randomly assigned to training (n=9) or control (n=8) groups. 16 residents completed the study and performed elective laparoscopic cholecystectomy on patients under supervision of experienced surgeons. Intervention: participants were randomised: • training: regular residency and FLS simulator proficiency-based training through MISTELS Program (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills) using FLS simulator and includes CD-ROM of didactic material and 5 MISTELS tasks: peg transfer, circle cut, placement of a ligating loop, and simple suture tied with extra- and intracorporeal techniques. • control: no simulator training (regular residency training alone). Time to assessment: mean time between baseline and final evaluation was 145 days. Final evaluation performed after proficiency testing was requested by the subject and confirmed by the proctor for the training group and after at least 6 weeks for the nontraining (control) group. Assessment: intraoperative laparoscopic performance was assessed on elective Laparoscopic cholecystectomy patients using validated global rating scale (GOALS) scores for 5 individual domains. Device: Fundamentals of Laparoscopic Surgery (FLS) Training Box simulator (Venture Technologies Inc. through VTiMedical, North Billerica, Maryland, USA).	Randomised controlled trial Method of randomisation: assignment drawn from a box by investigator who was not involved in training or evaluation. Allocation concealment: participants asked to keep status confidential for the study period. Supervising surgeons blinded during assessment operation. Assessors blinded to training status and performed evaluation independently. Level of evidence: II. Intention to treat: not stated. Power calculation: it was calculated that 7 subjects in each group would give a power of 80% to detect a difference of 5 points in GOALS scores with an α of 0.05. Lost to assessment: n=1 from the original training group 9. Study period: not stated. Blinding: assessing surgeon blinded to training status. Outcome measures: criteria assessed included: • depth perception • bimanual dexterity • efficiency • tissue handling • autonomy • total score.	Sample size: n=17 (16) • training: n=8 • control: n=8. Baseline characteristics of participants: Variable / Simulator trained (n=8) / Control (n=8) / *p*-value PGY 1/2/3 / 5/2/1 / 6/2/0 / 0.58 Age (y) / 27 (26.5–28.5) / 27(27-28) / 0.85 Gender (male/female) / 6/2 / 3/5 / 0.13 Hand dominance (right/left) / 7/1 / 7/1 / 1 Time between baseline and final evaluation (days) / 162 (100–256) / 113 (40–167) / 0.13 Laparoscopic cholecystectomy performed as primary during study (number) / 4.5 (3–7) / 3.5 (2–5) / 0.21 Laparoscopic cholecystectomy participated as assistant during study (number) / 4.5 (3.5–8) / 4.5 (4–6) / 0.92 Other laparoscopic cases performed or participated during study (number) / 2.5 (1–3.5) / 2.5 (2–3.5) / 0.75 Data expressed as median (IQR). Baseline FLS scores were similar in the simulator-trained and control groups (49.1 ± 17 versus 39.5 ± 16; *p*=0.27). Baseline OR GOALS scores were similar in the simulator-trained and control groups (11.3 ± 2.0 versus 12.0 ± 1.8; *p*=0.47). Inclusion: general surgical residents (PGYs 1 to 3) with no prior FLS experience and initially assessed in OR with GOALS scores ≤ 15. Details of patients for live assessment: elective laparoscopic cholecystectomy patients but other details not stated.

CD-ROM: Compact Disc-Read-Only Memory; FLS: the Fundamentals of Laparoscopic Surgery; GOALS: Global Operative Assessment of Laparoscopic Skills; IQR: interquartile range; MISTELS: McGill Inanimate System for Training and Evaluation of Laparoscopic Skills; OR: operating room; PGY: postgraduate year; USA: the United States of America.

Table C.18 MIST-VR task 3 and Box trainer simulator training for suturing and knot-tying versus no simulator training

<table>
<tr><th>Author</th><th>Intervention</th><th>Study design</th><th>Study population</th></tr>
<tr>
<td>Van Sickle et al 2008

Location
Department of Surgery, University of Texas Health Center at San Antonio, Texas

USA</td>
<td>Objective: to demonstrate that a structured stepwise curriculum for minimally invasive surgery (MIS) suturing and knot tying based on the concept of training to expert performance levels results improved operative performance on the fundal suturing portion of a laparoscopic Nissen fundoplication when compared with the current unstructured training method.

All participants: volunteer General Surgery or surgical subspeciality residents were PGY-3, PGY-5 or PGY-6 and consisted of six fellows (6 females) and five residents (3 females).

Baseline assessment: all subjects were shown a 15-minute cognitive training video on laparoscopic suturing and knot tying. A short, seven-question knowledge examination was given after subjects viewed the video. A score of 100% was required to continue participating in the study. A baseline perceptual ability was assessed using a validated test (Pictorial Surface Orientation).

Intervention: participants were randomised:
• training: supervised simulation-based laparoscopic suturing curriculum training (virtual reality training i.e. MIST-VR traversal task to use both hands and video tower box trainer simulator with foam models for suturing and knot tying) with specific training tasks and proficiency levels. Each participant was required to reach training performance goals for each task on two consecutive trials before being allowed to progress on to next task.
• control: no formal simulator training (standard clinical training).

Time to assessment: not stated.

Assessment: during a laparoscopic Nissen fundoplication, placement of two consecutive intracorporeally knotted sutures was video recorded for analysis. Single trial of suturing task on human patient under expert supervision (blinded to training status). Recordings of operative performance for each group were reviewed by two surgeon investigators blinded to training status. Performance was scored based on a set of tightly defined errors and this assessment method has been extensively validated (Van Sickle et al 2007).</td>
<td>Randomised controlled trial

Method of randomisation: not stated.

Allocation concealment: not stated.

Level of evidence: II.

Intention to treat: not stated.

Power calculation: not stated.

Lost to assessment: 24 subjects enrolled but complete data for 22 (not stated whether non-completions were from training or control group).

Study period: January 2003 to July 2005.

Blinding: supervising and assessing surgeons were blinded to training status of subjects.

Outcome measures: criteria assessed included:
• total suturing time (seconds)
• total suturing errors
• excess needle manipulations.

Defined errors included:
• missed grasp
• instrument not assisting
• tear or injure tissue
• incomplete or repeated bite
• needle out of view
• missed loop
• tail looped
• failure to square knot
• attending surgeon takeover
• scissors touch tissue.</td>
<td>Sample size: n=22 completed (24 enrolled)
• training: n=11
• control: n=11.

Baseline characteristics of participants:
<table>
<tr><th>Variable</th><th>Simulation-based training group</th><th>Control group</th></tr>
<tr><td>PGY level</td><td></td><td></td></tr>
<tr><td>6</td><td>2</td><td>2</td></tr>
<tr><td>5</td><td>6</td><td>5</td></tr>
<tr><td>3</td><td>3</td><td>4</td></tr>
<tr><td>Gender</td><td></td><td></td></tr>
<tr><td>Male</td><td>6</td><td>7</td></tr>
<tr><td>Female</td><td>5</td><td>4</td></tr>
<tr><td>Prior laparoscopic experience</td><td></td><td></td></tr>
<tr><td>Yes</td><td>6</td><td>5</td></tr>
<tr><td>No</td><td>5</td><td>6</td></tr>
<tr><td>Age, y</td><td>32</td><td>30</td></tr>
<tr><td>Prior laparoscopic cases</td><td>94</td><td>60</td></tr>
<tr><td>Prior laparoscopic cholecystectomy</td><td>83</td><td>50</td></tr>
<tr><td>PicSOr</td><td>0.92</td><td>0.92</td></tr>
</table>
PicSOr, Pictorial Surface Orientation test.
No statistical differences between the 2 groups.

Inclusion: not stated.

Exclusion: not stated.

Details of patients for live assessment: not stated.</td>
</tr>
</table>

	Device: MIST-VR simulator (Mentice AB, Gothenburg, Sweden) Task 3 (immersion jig) and standard box trainer, the foam Nissen suturing model and the intracorporeal slip-square knot. Knots were tested on a tensiometer (In-Spec 2200; Instron® Corporation, Canton, Maryland, USA) configured with a 45N load cell and mini-capstan cord/yarn grips and analysed using Series IX software on a standard laptop computer.		

AB: abbreviation of Aktiebolag, the Swedish term for 'limited company or corporation; MIS: minimally invasive surgery; MIST-VR: Minimally Invasive Surgery Trainer−Virtual Reality; PGY: postgraduate; PICSOR: Pictorial Surface Orientation test; USA: the United States of America.

Table C.19 KAIST-Ewha Colonoscopy Simulator II training versus no simulator training

Author	Intervention	Study design	Study population
Yi et al 2008 Location Department of Internal Medicine, Ewha Womans University, Seoul South Korea	Objective: to determine whether targeted colonoscopy skills are acquired through simulation-based training using the KAIST-Ewha Colonoscopy Simulator II and the acquired skills can be transferred to colonoscopy to actual patients. All participants: 11 subjects consisted of six fellows (6 females) and five residents (3 females). Intervention: participants were divided into two groups: • training: simulation-based training included practising the targeted skills of colonoscopy using two training scenarios that have different colon flexures and degrees of difficulty. The training scenario A is designed to teach practical skills to navigate the colon applying torque and up-down angulations. The scenario B is designed to teach skills to manage a loop formed in the sigmoid colon. • control: no simulator training. Time to assessment: not stated. Assessment: each subject performed colonoscopies on 5 different patients and evaluated by supervising experts based on accuracy of colonoscopy results and the established performance criteria. Questionnaires on the colonoscopy experience were also filled out by the patients. Device: KAIST-Ewha Colonoscopy Simulator II .	Non-randomised comparative study Level of evidence: III-2. Intention to treat: not stated. Power calculation: not stated. Lost to assessment: not stated. Study period: October 2006 – February 2007. Blinding: not stated. Outcome measures: criteria assessed included: • total insertion time (minutes) • success rate • number of red-outs • number of air inflation • number of loop formation • mucosal visualisation • number of abdominal pressure • changes in patient's posture • overall performance accuracy • extent of abdominal pain • extent of abdominal inflation • extent of anus discomfort.	Sample size: n=11 • training: n=5 • control: n=6. Baseline characteristics of participants: not stated. Inclusion: fellows and residents. Exclusion: not stated. Details of patients for live assessment: average of patients' ages for the simulation-trained group was 49.6 (n=25, range 24–71) and 53.5 (n=30, range 25–79) for the control group.

KAIST: Korea Advanced Institute of Science and Technology, Daejeon, South Korea.

Table C.20 Guildford MATTU totally extraperitoneal inguinal hernia training versus no simulator training

<table>
<tr><th>Author</th><th>Intervention</th><th>Study design</th><th>Study population</th></tr>
<tr>
<td>Zendejas et al 2011

Location

Department of Surgery, Mayo Clinic Multidisciplinary Simulation Centre, Mayo Clinic, Rochester

USA</td>
<td>Objective: to evaluate the impact of a simulation-based mastery learning course in totally extraperitoneal inguinal hernia repairs for general residents using operative time, intraoperative performance, and objective measures of patient outcomes.

All participants: general surgery residents.

Pretest: performed a baseline first totally extraperitoneal inguinal hernia repair in the operating room (OR) under staff supervision (TEP#1). Operative performance was assessed using Global Operative Assessment of Laparoscopic Skills (GOALS) score. Operative time and patient outcomes including intraoperative and postoperative complications and overnight stay were recorded.

Intervention: participants were randomised:
training: a simulation-based mastery learning curriculum for totally extraperitoneal inguinal hernia repair using online course (9 web-based modules) followed by skills training on a totally extraperitoneal inguinal hernia repair simulator consisting of supervised one-on-one practice sessions until expert performance was achieved. Mastery was defined as successful repair of both hernias in less than 2 minutes on 2 consecutive attempts.
control: standard practice: based on self-learning (without a formal structured curriculum for totally extraperitoneal inguinal hernia repair and intraoperative learning.
Time to assessment: time between totally extraperitoneal inguinal hernia first (TEP#1) and second repair (TEP#2) was 9.2 ± 6.0 days for intervention (mastery training) and similarly 11 ± 8.5 days for controls (p=0.57).

Assessment: all re-assessed during second totally extraperitoneal inguinal hernia repair (TEP#2) and many observed for several subsequent repairs.

Device: Guildford MATTU TEP hernia task trainer (Limbs and Things Ltd. Bristol, UK).</td>
<td>Single-blinded randomised controlled trial

Method of randomisation: blocks of 4 using opaque sealed envelopes containing computer-generated randomised sequences prepared by a study-independent statistician.

Allocation concealment: opaque sealed envelopes.

Level of evidence: II.

Intention to treat: not stated.

Power calculation: 25 residents per arm would provide 80% power to detect a 5-minute decrease in operative time from baseline to postintervention with an alpha level of 0.05.

Lost to assessment: not stated.

Study period: January to September 2010.

Blinding: cases were videotaped in a de-identified fashion and assessed by staff blinded to training status of subjects.

Outcome measures: criteria assessed included:
operative time
operative performance - GOALS score immediately after each case
resident participation - proportion of procedure performed by trainee
patient outcomes – intraoperative and postoperative complications and overnight stay, recurrence of the inguinal hernia, and chronic groin pain.</td>
<td>Sample size: n=50
training: n=26
control: n=24.
Baseline characteristics of participants:
<table>
<tr><th>Variable</th><th>Simulator-trained</th><th>Controls</th></tr>
<tr><td>PGY 1, N (%)</td><td>9 (34)</td><td>8 (33)</td></tr>
<tr><td>PGY 2, N (%)</td><td>3 (12)</td><td>3 (12)</td></tr>
<tr><td>PGY 3, N (%)</td><td>3 (12)</td><td>2 (8)</td></tr>
<tr><td>PGY 4, N (%)</td><td>5 (19)</td><td>4 (16)</td></tr>
<tr><td>PGY 5, N (%)</td><td>6 (23)</td><td>7 (29)</td></tr>
<tr><td>Male, N (%)</td><td>17 (65)</td><td>18 (75)</td></tr>
<tr><td>Female, N (%)</td><td>9 (35)</td><td>6 (25)</td></tr>
<tr><td>Age mean ± SD</td><td>30 ± 2</td><td>30 ± 3</td></tr>
<tr><td>Right-handed (%)</td><td>23 (88)</td><td>20 (83)</td></tr>
<tr><td>Left-handed (%)</td><td>3 (12)</td><td>4 (16)</td></tr>
<tr><td>Video game exposure – pas)</td><td>3.4 ±1.0</td><td>3.6 ± 0.9</td></tr>
<tr><td>Video game exposure – present</td><td>2.1 ± 0.9</td><td>2.3 ± 0.8</td></tr>
<tr><td>TEP experience</td><td>3.3 ± 0.9</td><td>3.4 ± 0.9</td></tr>
<tr><td>TEP comfort</td><td>2.4 ± 1.0</td><td>2.5 ± 0.9</td></tr>
<tr><td>No. patients/repairs</td><td>26/40</td><td>24/36</td></tr>
<tr><td>Operative performance</td><td>17.5 ± 3.8</td><td>17.4 ± 4</td></tr>
<tr><td>Intraoperative complications, N (%)</td><td>4 (15)</td><td>5 (20)</td></tr>
<tr><td>Postoperative complications, N (%)</td><td>4 (15)</td><td>7 (29)</td></tr>
<tr><td>Overnight stay, N (%)</td><td>3 (12)</td><td>6 (25)</td></tr>
</table>
Inclusion: general surgical residents whose clinical assignments provided opportunity to perform at least one TEP inguinal hernia repair.

Exclusion: not stated.

Details of patients for live assessment: collected patient's medical record on age, sex, risk classification, recurrent hernia, body mass index and history of prostate disease.</td>
</tr>
</table>

GOALS: Global Operative Assessment of Laparoscopic Skills; MATTU: The Minimal Access Therapy Training Unit; OR: operating room; PGY: postgraduate; SD: standard deviation; TEP: totally extraperitoneal; UK: the United Kingdom; USA: the United States of America.

Simulation-based training versus didactic lecture-based education (see Table C.14-Patel et al 2012)

Simulation-based training versus interactive seminar based education

Table C.21 Cardiopulmonary bypass weaning, high-fidelity simulation-based training versus interactive seminar-based education

Author	Intervention	Study design	Study population
Bruppacher et al 2010 Location Department of Anesthesia, St Michael's Hospital, University of Toronto, Toronto, Ontario Canada	Objective: to compare two cardiopulmonary bypass (CPB) teaching modalities: high fidelity simulation-based training versus a traditional interactive seminar format for learning CPB weaning. All participants: 20 residents and fellows in anesthesiology, postgraduate year 4 or higher. Pretest phase: all trainees received the usual residency program syllabus for cardiac surgery and CPB weaning. One week later, trainees had a pretest session in the operating room. Intervention: participants were randomised into two intervention groups: • simulation-based education: individual simulation 2-hour session including debriefing, orientation to high-fidelity simulation in simulation room which closely mimics a cardiac operating room with medications and CPB machine that was handled by an actor perfusionist. A flat 17 inch LCD was placed on the chest of the simulated patient to loop video recordings of real patient hearts in difference phases of CPB removal. All sessions videotaped to facilitate debriefing and reinforce learning objectives of the session • interactive seminar-based education: 2-hour seminar by anaesthesiologist using PowerPoint slides, handouts and face to face discussion of four paper-based scenarios similar to those in the simulation training. Time to assessment: at 2 and 5 weeks after the intervention. Assessment: participants' performance weaning patients from CPB was measured using a previously validated global assessment tool, Anaesthetists' Non-Technical Skills (ANTS) rating scale, to assess trainees' cognitive and behavioural performance in CPB weaning, and a checklist based on the syllabus to assess technical skills of trainees.	Randomised controlled trial Method of randomisation: using sealed envelopes. Allocation concealment: not stated. Level of evidence: II. Intention to treat: not stated. Power calculation: to calculate sample size 17 trainees per arm were needed, based on a two-tailed alpha of 0.05 and a power of 0.8. An interim analysis was planned for when 10 trainees in each group had completed all clinical assessments. Lost to assessment: not stated. Study period: not stated. Blinding: assessing surgeon blinded to training status. Outcome measures: technical skills or discrete tasks were measured using a checklist developed for this study using the Delphi method. Tasks included: Before weaning • ensures rewarming is adequate • checks adequacy of pH/electrolytes/haematocrit • restores adequate ventilation • checks cardiac rate and rhythm • treats inadequate cardiac rate/rhythm (drugs/pace) • determines patient is ready for weaning • checks venous reservoir volume Weaning	Sample size: n=20 • training: n=10 • control: n=10. Baseline characteristics of participants: not stated. Inclusion: residents and fellows in anaesthesiology, postgraduate year 4 or higher. Exclusion: experience in cardiac anaesthesia, defined as participation in more than 10 operations involving CPB. Details of patients for live assessment: patients scheduled for elective coronary artery bypass grafting. Exclusion criteria were surgery including noncoronary artery bypass grafting or non-CPB. 20 patients were excluded from the study by the operating team because of intraoperative issues and concerns about possible difficulties weaning the patient. These patients were replaced with additional 20 patients for the study and the trainee's test was postponed to the next possible date within 1 week.

	Device: SimMan Universal Simulator (Laerdal, Wappingers Falls, New York, USA).	• enquires about pump flow during weaning • checks preload • treats preload with aortic cannula (PAP within 24% of baseline) • checks afterload • controls after load using drugs as needed (MAP>55 and <100mmHg) • determines when pump flow can be stopped After weaning • assesses RV contractility by visual inspection • performs cardiac output with drugs/volume as required (cardiac index 2 l/min) • maintains appropriate heart rate (60–100 beats per minute) • maintains appropriate preload with drugs/volume as required (MAP>55 and <100mmHg) • determines when patient is ready for protamine. MAP: mean arterial pressure; PAP: pulmonary arterial pressure; RV: right ventricle. ANTS global rating scale: • task management • team working • situation awareness • decision-making. Rating options (descriptor) 4=Good (Performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others) 3=Acceptable (Performance was satisfactory standard but could be improved) 2=Marginal (Performance indicated cause for concern, considerable improvement is needed) 1=Poor (Performance endangered or potentially endangered patient safety, serious remediation is required).	

ANTS: Anaesthetists' Non-Technical Skills; CBP: cardiopulmonary bypass; LCD: liquid crystal display; MAP: mean arterial pressure; PAP: pulmonary arterial pressure; RV: right ventricle.; USA: the United States of America

Simulation-based training versus patient-based training

Table C.22 LAP Mentor and haptic PROMIS simulator training versus patient-based camera navigation training

<table>
<tr><th>Author</th><th>Intervention</th><th>Study design</th><th>Study population</th></tr>
<tr>
<td>Franzeck et al 2012*

Location

Department of Visceral and Transplantation Surgery, University Hospital Zurich, Zurich

Switzerland</td>
<td>Objective: to determine whether focused VR simulator-based laparoscopic camera training of novices could improve camera performance in actual clinical situation in the same manner as traditional training in OR.

All participants: consecutive pregraduation medical students on a surgical clerkship at the department of Visceral and Transplantation Surgery of the University Hospital of Zurich without any experience in camera navigation or simulators.

Pretest: all eligible participants completed the validated visuospatial ‘Stumpf-Fay cube perspective’ test and were given an identical 60 minute introduction on the technical functionality and correct handling of an angled laparoscopic camera. All students performed a baseline camera navigation test in the OR (maximum 5 minutes). Pretest was validated by comparing assessment videotapes of students with camera tests of 14 experts.

Intervention: participants were randomised into two groups:
• simulator training: six 1-hour structured simulator-based training sessions in the skills lab over 3 weeks – twice a week (total 6 hours). Students followed a standardised protocol, performing 40 minutes of camera navigation-specific tasks on the different simulators (25 minutes on the basic task modules of the LAP Mentor camera manipulation and 15 minutes on the laparoscope orientation Core modules on the PROMIS™) and 20 minutes of training on non-camera or camera-specific simulator exercises of free choice.
• control: traditional OR training only (no simulator training) – six laparoscopic interventions in OR including hemicolectomy, rectum resection, gastric bypass, and cholecystectomy - camera navigation at the surgeon’s direction.

Time to assessment: not stated.

Assessment: all participants performed camera navigation skills assessment test on real patients in OR at the beginning of an actual operation. All patients were placed in supine position. Participants were</td>
<td>Prospective randomised controlled trial.

Method of randomisation: sealed, opaque envelopes.

Allocation concealment: not stated.

Level of evidence: II.

Intention to treat: not stated.

Power calculation: a sample size of 12 participants in each group calculated to have 80% power to detect a difference in means of 100 minutes.

Lost to assessment: six trainees dropped out – 2 because they failed to acquire enough training and 4 failed to attend camera assessment tests.

Study period: between May 2007 and July 2008.

Blinding: 5 independent assessing experts were blinded to training status and all experienced surgical attending physicians.

Outcome measures: criteria assessed included:
• time to completion
• organ visualisation
• horizontal alignment
• correct scope rotation handling
• visuospatial tests.</td>
<td>Sample size: n=30 (24)
• training: n=12
• control: n=12.

Baseline characteristics of participants: No difference in baseline characteristics noted.

<table>
<tr><th>Variable</th><th>Simulator-trained group</th><th>Control (OR) group</th><th>p-value</th></tr>
<tr><td>Mean age (years)</td><td>26.2 ± 1.9</td><td>25.8 ± 1.1</td><td>0.610</td></tr>
<tr><td>Gender (M:F)</td><td>3:9</td><td>4:8</td><td>0.9</td></tr>
<tr><td>Righthandedness (n (%))</td><td>11 (92)</td><td>12 (100)</td><td>0.9</td></tr>
<tr><td>Visuospatial test (mean points ± SD</td><td>13.3 ± 3.6</td><td>15.2 ± 3.7</td><td>0.223</td></tr>
<tr><td colspan="4">Pretraining camera test</td></tr>
<tr><td>Organ visualisation (mean points ± SD)*</td><td>22.4 ± 5.0</td><td>25 ± 2.8</td><td>0.132</td></tr>
<tr><td>Horizon alignment (mean point ± SD)#</td><td>20.1 ± 4.1</td><td>22.7 ± 3.9</td><td>0.127</td></tr>
<tr><td>Time to completion(s) (mean ± SD)</td><td>179 ± 64</td><td>162.9 ± 67</td><td>0.554</td></tr>
<tr><td>Correct scope rotation handling (no. participants (%))</td><td>7 (58)</td><td>7 (58)</td><td>1</td></tr>
</table>
*Single Measure Intraclass Correlation (SMIC)=0.65.
SMIC = 0.66.

Inclusion: general surgical residents with no prior</td>
</tr>
</table>

	positioned on the patient's right side and were given 30^0 angled laparoscope introduced into the trocar. They had to centre and hold for 5 seconds the following positions/organs and had to maintain the correct horizontal alignment during camera movement: left abdominal wall, ascending colon, right lobe of liver, sigmoid colon, caecum, pelvis, trocar entry site in the upper left quadrant (simulated by a finger pressing externally), and descending colon. Maximum duration of the test was set at 5 minutes. Videos of all tests (including of 14 experts) were independently reviewed by five independent experts according to a structured protocol (experts blinded to participants training group). Device: two Xitact™ IHP instrument haptic ports as interfaces and a third unidirectional electromechanical interface, the Xitact™ IHP instrument tracking port, for the camera navigation (Mentice AB, Gothenburg, Sweden) with the LAP Mentor™ (Simbionix USA, Cleveland, Ohio, USA) software. As a second simulator, a haptic PROMIS™ surgical hybrid simulator (Haptica Ltd., Dublin, Ireland) was used.		endovascular experience. Exclusion: any previous active experience in laparoscopic camera handling in the OR and/or use of a VR simulator checked by questionnaire (n=2 students). Details of patients for live assessment: not stated.

**Electronic publication available in September 2011

AB: abbreviation of Aktiebolag, the Swedish term for 'limited company' or corporation; IHP: Instrument Haptic Port; OR: operating room; SD: standard deviation; SMIC: Single Measure Intraclass Correlation; USA: the United States of Americas; VR: virtual reality.

Table C.23 Olympus colonoscopy (Endo TS-1) **simulator training versus patient-based colonoscopy training**

Author	Intervention	Study design	Study population
Haycock et al 2010 Location Wolfson Unit of Endoscopy, St Mark's Hospital, Imperial College London, Harrow, Middlesex United Kingdom	Objective: to investigate the efficacy of the simulator in training novices in colonoscopy by comparing training outcomes from simulator training with those of standard patient-based training and the transfer of skills from simulator training to real-life colonoscopy. All participants: physicians, surgeons, nurses or other position recognised by the training institution as appropriate for training colonoscopy. Pre-test: all participants performed 3 previously validated cases on the simulator to provide a measure of their baseline experience. Each case had a time limit of 20 minutes to intubate to the caecum and all procedures were recorded automatically by the simulator and evaluated using computer-generated parameters with demonstrated construct validity and selected to have clinical importance by investigators. Intervention: participants were randomised: • training: 16 hours of a standardised simulator-training program with minimal tutoring and feedback from trainer • control: 16 hours of patient-based training (4 half-day sessions) by expert trainer using a ScopeGuide imager; included performing a minimum of 8 colonoscopies under one-to-one supervision. Time to assessment: not stated. Assessment: automatically recorded performance metrics on 3 simulated cases and blinded expert assessment of 3 live cases using previously validated structured assessment tools: UK Joint Advisory Group (JAG) Direct Observation of Procedure Skills (DOPS) and Global Score (GS) sheets. Device: Olympus colonoscopy simulator (Endo TS-1; Olympus Keymed, Southend, UK). It also provides a simulated, 3-D endoscope imager view identical to that provided by ScopeGuide (Olympus Keymed, Essex, UK).	Randomised controlled trial Method of randomisation: a computer-generated, block randomisation protocol with 8 per block. Allocation concealment: not stated. Level of evidence: II. Intention to treat: yes. Power calculation: the study was powered to detect a 5-unit (1 SD) difference between the 2 groups. With a 5% significance and 80% power, 16 participants were required for each group, 32 in total. Lost to assessment: 40 trainees were randomised into two groups and 36 completed the study: 2 did not start, 1 did not finish training and another 1 did not finish assessments. Simulator-trained group n=19 (18 completed assessments); control group n=18 (18 completed assessment). Study period: not stated. Blinding: expert assessor blinded to training status. Outcome measures: patient-based assessment criteria included: • completion of case • maximum tip position • sigmoid • descending • transverse • ascending • caecum • time taken (minutes) • straight insertion (depth (cm) • JAG DOPS • Global Score.	Sample size: n=40 recruited (36 completed) • training: n=19 (18 completed) • control: n=18. Baseline characteristics of participants: Variable | Simulator-trained (n=19) median (IQR) or number (%) | Controls (n=18) median (IQR) or number (%) | *p*-value Age, years (range) | 28 (26–30) | 31 (26–33) | 0.32 Male | 6 (32%) | 10 (56%) | 0.19 Nurse | 3 (16%) | 3 (17%) | 0.25 General trainee | 10 (52%) | 6 (33%) | Specialist trainee | 3 (16%) | 8 (45%) | Other | 3 (16%) | 1 (5%) | Colonoscopies witnessed | 15 (7.5–125) | 45 (21.25–137.5) | 0.15 Colonoscopies assisted | 0 (0–4) | 1 (0–30) | 0.21 Colonoscopies performed | 0 (0–0) | 0 (0–0) | 0.72 Sigmoidoscopies | 0 (0–0) | 0 (0–1) | 0.16 OGD | 0 (0–0) | 0 (0–2) | 0.09 Inclusion: physicians, surgeons, nurses or other position recognised by the training institution as appropriate for training colonoscopy. Exclusion: if they had experience of more than 25 previous colonoscopies or flexible sigmoidoscopies, had previously attended an intensive colonoscopy training course or were previously a participant in a colonoscopy training or simulator training study. Participants who had performed more than 10 laparoscopic surgical

			procedures were also excluded. Details of patients for live assessment: patients scheduled to undergo clinically indicated colonoscopy, excluding patients over 75 years or with history of pelvic or colonic surgery or difficult colonoscopy.

3-D: three dimensional; DOPS

Appendix D: Summary of critical appraisal

	Randomisation	Allocation concealment	Blinding of assessors	Intention to treat	Power calculation	Losses to assessment	Study period	Validated assessment tools	Inclusion criteria	Exclusion criteria	Baseline characteristics
Randomised controlled trials (NHMRC Level II)											
Ahlberg et al 2007	✓	✓	✓	×	×	×	×	×	✓	×	✓
Banks et al 2007	✓	✓	✓	×	×	×	✓	✓	✓	×	×
Bruppacher et al 2010	✓	×	✓	×	✓	×	×	✓	✓	✓	×
Cosman et al 2007	✓	×	✓	×	×	×	✓	✓	✓	×	×
Franzeck et al 2012*	✓	×	✓	×	✓	✓	✓	✓	✓	✓	✓
Fried et al 2010	×	×	✓	×	×	✓	×	×	✓	×	×
Haycock et al 2010	✓	×	✓	✓	✓	✓	×	✓	✓	✓	✓
Hogle et al 2009	✓	×	✓×	×	×	✓	×	✓	✓	×	×
Howells et al 2008	✓	×	✓	×	×	×	✓	✓	✓	×	×
Kälström et al 2010	×	×	✓	×	×	✓	✓	✓	✓	×	✓
Larsen et al 2009	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓
Ossowski et al 2008	×	×	✓	×	×	×	×	×	✓	×	×
Palter et al 2011	✓	×	✓	×	✓	✓	×	✓	✓	×	✓
Park et al 2007	×	×	✓	×	✓	✓	×	✓	✓	✓	✓
Patel et al 2012*	×	×	×	×	×	×	×	×	✓	✓	✓
Schout et al 2009	✓	×	✓	×	✓	×	✓	✓	✓	✓	✓
Shirai et al 2008	×	✓	✓	×	×	×	✓	×	✓	×	✓
Sroka et al 2010	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓

	Randomisation	Allocation concealment	Blinding of assessors	Intention to treat	Power calculation	Losses to assessment	Study period	Validated assessment tools	Inclusion criteria	Exclusion criteria	Baseline characteristics
Randomised controlled trials (NHMRC Level II)											
Van Sickle et al 2008	✓	×	✓	×	×	✓	✓	✓	✓	×	×
Zendejas et al 2011	✓	✓	✓	×	✓	×	✓	✓	✓	×	✓
Non-randomised comparative study (NHMRC Level III-2)											
Yi et al 2008	NA	NA	×	NA	NA	×	✓	×	✓	×	×
Comparative studies (NHMRC Level III-3)											
Belyea et al 2011	NA	NA	×	NA	NA	NA	✓	×	✓	✓	✓
Beyer et al 2011	NA	NA	✓	NA	NA	NA	✓	✓	✓	×	✓

× Not reported
✓ Reported
✓× Reported as negative
* electronic publication available in September 2011
NHMRC: National Health and Medical Research Council; NA: Not applicable

Appendix E: Results tables

Simulation training versus no simulation training

Randomised controlled trials

Table E.01 Performance of laparoscopic cholecystectomy by LapSim-trained and non LapSim-trained participants

Ahlberg et al 2007 RCT, Level II	Simulator training (n=at least 6)	No simulator training (n=at least 6)	*p*-value~
Performance parameter	Mean (95% confidence interval; variance)	Mean (95% confidence interval; variance)	
Errors for the entire procedure	28.4 (23.51–33.32; 118.69)	86.2 (58.18–114.12; 916.68)	0.0037
Exposure errors	15.0 (11.16–18.79; 68.44)	53.4 (16.70–90.13; 623.31)	0.0402
Clipping and tissue division errors	1.9 (0.93–2.87; 5.57)	7.1 (3.95–10.25; 41.11)	0.0080
Dissection errors	11.5 (8.82–14.08; 28.77)	29.5 (13.99–45.01; 61.50)	0.0310
Time taken to perform operation (minutes)	Not stated	Not stated (on average 58% longer than simulator-trained group but not statistically significant)	0.0586
Error reduction rate	Not stated (no reduction shown during their 10 laparoscopic cholecystectomies)	Not stated (no reduction shown during their 10 laparoscopic cholecystectomies)	

~*p*-value in bold indicates statistical significance ($p<0.05$).

Participants who received virtual reality (VR) simulator training performed laparoscopic cholecystectomy with significantly fewer objectively assessed intraoperative errors for the entire procedure *(p*=0.0037) than those in the control group. The difference remained when the errors during the exposure portion of the procedures *(p*<0.04), clipping and tissue division *(p*<0 .008) and dissection *(p*<0.03) were examined separately. The control group also showed considerable variability in performance compared with the VR-trained group.

The VR-trained group consistently made fewer errors and performed surgery faster compared with the control group over first 10 consecutive laparoscopic cholecystectomies, with no reduction in error rate shown in either group.

There was no significant difference between the two groups concerning baseline parameters (i.e. gender, age, postgraduate year and previous laparoscopic experience [assisting]). Residents randomised to the LapSIM VR simulator training group all reached a predefined expert (proficiency) level of performance. Each resident's first 10 laparoscopic cholecystectomies were videotaped. Surgery numbers one, five and 10 for each subject were reviewed independently in a blinded fashion and scored for predetermined errors.

Table E.02 Performance of laparoscopic cholecystectomy by LapSim-trained and non LapSim-trained participants

Cosman et al 2007 RCT, Level II	Simulator training (n=5)	No simulator training (n=5)	Mann-Whitney U-Test	*p*-value~
Performance parameter (median)				
Errors for the entire procedure	10	18	3.5	0.05
Bimanual coordination	3	1.8	3	0.05
Global score	3.2	1.8	2.5	0.04
Time taken to perform operation (seconds)	94	172	4	0.075

~*p*-value in bold indicates statistical significance (*p*≤0.05).

Participants who had trained on the LapSim VR simulator (without haptics) committed significantly fewer intraoperative errors for the entire procedure than those in the control group (median 10 versus 18; *U*=3.5; *p*=0.05). They also had significantly better bimanual coordination (median 3.0 versus 1.8; *U*=3; *p*=0.05) and a significantly higher global score (median 3.2 versus 1.8; *U*=2.5; *p*=0.04). Those who had trained on the simulator took less time to complete the assessment task than those with no simulator training but this was only of borderline statistical significance (median 94 seconds versus 172 seconds; *U*=4; *p*=0.075).

The intra-class correlation (ICC) score for the error scale was calculated as 0.96, indicating 96% agreement between the five assessing surgeons on the error scale. The bimanual index and global rating each had an ICC score of 0.87, indicating 87% agreement between raters for these components of the assessment. Generally, scores above 0.7 are accepted as being indicative of significant correlation between independent reviewers.

A total of 10 trainees were enrolled in the study; five were randomly allocated to the control group, and five to the experimental group. The latter had access to the LapSim® VR simulator (without haptics), for one hour a day only, to practice the clipping task until they satisfied the predetermined proficiency performance criteria on two successive repetitions of the task. The control group were tested on the simulator at enrolment into the study and again prior to performing the assessment task on a live human patient. The assessment task was videotaped in the operating theatre and analysed independently by five laparoscopic surgeons who were unaware of the nature of the training received by each participant. Based on the Mann-Whitney *U*-test for all variables, there were no measured differences between the two groups on the pretest (baseline). No statistical differences were recorded between the two groups in terms of level of training (mean ± SD = 1.3 ± 0.5 years) or number of laparoscopic cases performed (mean ± SD = 7.7 ± 2.8). Only time to task completion reached a statistically significant difference in the post-test on the simulator (U=1; *p*=0.04), with the LapSim-trained group (median 47.5 seconds) taking almost half the time to complete the task as the control group (median 87.1 seconds).

Table E.03 Performance of laparoscopic cholecystectomy by LapSim-trained and non LapSim-trained participants

Hogle et al 2009 RCT, Level II	Simulator training (n=6)	No simulator training (n=6)	*p*-value
Performance parameter*	Mean ± SD	Mean ± SD	
Depth perception	3.6 ± 0.55	3.35 ± 0.62	0.99
Bimanual dexterity	3.17 ± 0.42	2.90 ± 0.51	0.55
Efficiency	2.89 ± 0.53	2.82 ± 0.62	0.93
Tissue handling	2.96 ± 0.59	3.10 ± 0.53	0.56
Autonomy	3.23 ± 0.44	3.11 ± 0.62	0.85

* Each domain is scored from 1 (worst possible score) to 5 (best possible score).
SD: standard deviation.

No significant difference was found between the simulator-trained group and the control group in the Global Operative Assessment of Laparoscopic Skills (GOALS) domains of depth perception, bimanual dexterity, efficiency, tissue handling, or autonomy during two consecutive elective laparoscopic cholecystectomies performed one month after training.

Thirteen first year surgical residents were enrolled in the study. One person failed to return any videotapes of procedures. Of the 12 residents who completed the study, two residents did not train to full criteria but were included in the analysis of the trained group because they had spent significant time on the simulator and were not seen as 'untrained'.

Prestudy baseline and poststudy simulator testing was completed for all participants. The participants were randomised and the trained group used a simulator curriculum for each module. The LapSim simulator tasks included: camera navigation, instrument navigation, coordination, grasping, lifting and grasping, cutting and clip applying. The training curriculum was fully completed when level three was passed for each module. The participants in the training group independently completed two simulation training sessions per week and no additional simulation training was allowed outside the actual training curriculum. At one month after baseline testing, the participants recorded their performance during their next two elective laparoscopic cholecystectomies. The supervising surgeon evaluating their performance used the GOALS. The videotapes were used for subsequent blinded evaluation and scoring using GOALS.

Table E.04 Performance of laparoscopic cholecystectomy by FLS- trained and non FLS-trained participants

Sroka et al 2010 RCT, Level II	Simulator training (n=8)	No simulator training (n=8)	*p*-value~
Performance parameter (mean ± SD)			
Total GOALS score - sum of 5 domains (for each domain 1 = worst possible score, 5 = best possible score)	17.4 ± 1.9	13.8 ± 2.2	0.0003

~*p*-value in bold indicates statistical significance (*p*≤0.05).
FLS: The Fundamentals of Laparoscopic Surgery; GOALS: Global Operative Assessment of Laparoscopic Skills; SD: standard deviation.

There was no difference in the GOALS scores for simulator-trained and control groups at baseline (11.3 ± 2.0 versus 12.0 ± 1.8; *p*=0.47). Participants in the control group improved their performance by mean of 1.8 ± 2.1 (from 12.0 ± 1.8 to 13.8 ± 2.2; *p*=0.04). The simulator-trained group improved significantly more than the non-simulator trained group by a mean of 6.1 ± 1.3 points (from 11.3 ± 2.0 to 17.4 ± 1.9; *p*=0.0005 versus control; *p*<0.0001 versus baseline). After adjusting for gender using multivariate analysis, group allocation remained significantly associated with the change in GOALS score (*p*=0.001), with gender not significant (*p*=0.16). There was no correlation of change in GOALS score with time between assessments (Pearson correlation = 0.13, *p*=0.63).

Of the 17 subjects who were randomised, 16 completed their final evaluations, 8 in each group. The groups were statistically similar at randomisation. There were more women in the control group (63% versus 25%) while the simulator-trained group had a longer time between operating room (OR) assessments. Baseline FLS scores were similar in the simulator-trained and control groups (49.1 ± 17 versus 39.5 ± 16, *p*=0.27). At the final evaluation, FLS scores increased and the standard variations decreased in the simulator-trained group to 95.1 ± 4, compared with 60.5 ± 23 in the control group (*p*=0.004). At baseline, no participant had a score above the level required for FLS certification.

Table E.05 Performance of laparoscopic cholecystectomy by FLS-trained and non FLS-trained participants

Sroka et al 2010 RCT, Level II	Simulator training (n=8)	No simulator training (n=8)	*p*-value~
GOALS performance parameter (mean ± SD)			
Depth perception	1.25 ± 0.7	0.5 ± 0.8	0.08
Bimanual dexterity	1.25 ± 0.6	0.5 ± 1.1	0.04
Efficiency	1.13 ± 1.0	0.4 ± 1.1	0.24
Tissue handling	1.13 ± 1.0	0.3 ± 0.7	0.04
Autonomy	0.6 ± 1.1	0.3 ± 1.0	0.58
Total score*	6.1 ± 1.3	1.8 ± 2.1	0.0003

~*p*-value in bold indicates statistical significance (*p*<0.05).
*Each domain is scored from 1 (worst) to 5 (best) and the results summed to obtain a total score.
FLS: The Fundamentals of Laparoscopic Surgery; GOALS: Global Operative Assessment of Laparoscopic Skills; SD: standard deviation.

Of the five individual domains evaluated by GOALS, simulator training was associated with greater improvements in the laparoscopy-specific domains (bimanual dexterity, tissue handling and depth perception) compared with the more generic domains (efficiency and autonomy). There was no difference in the attending surgeon's assessment of the difficulty of the dissection for the simulator-trained and control groups at the baseline (2.5 versus 3, *p*=0.65) or final evaluations (4.5 versus 2.5, *p*=0.15).

Table E.06 Performance of laparoscopic cholecystectomy by MISTELS and LAP Mentor simulator-trained and non simulator-trained participants

Beyer et al 2011 Comparative study, Level III-3	MISTELS simulator training (n=6)	LAP Mentor simulator training (n=6)	No simulator training (n=7)	*p*-value¯ for MISTELS versus control	*p*-value¯ for LAP-Mentor versus control	*p*-value¯ for MISTELS versus LAP Mentor
Parameter						
GOALS 1	9.33	9.17	12.21	**0.006**	0.06	0.68
VAS 1	3.63	1.17	3.08	0.87	**0.046**	**0.03**
GOALS 2	12.41	13.17	11.85	NR	NR	NR
VAS 2	2.71	2.34	2.64	0.94	0.57	0.42
Progression (GOALS 2-GOALS 1)	3.08	4	-0.36	**0.03**	**0.007**	0.28
p-value¯ for GOALS 2 versus GOALS 1	**0.04**	**0.03**	0.35	NA	NA	NA

GOALS: Global Operative Assessment of Laparoscopic Skills; MISTELS: McGill Inanimate System for the Training and Evaluation of Laparoscopic Skills; NA: not applicable; NR: not reported; VAS: visual analogue scale used for difficulty of the surgery.
¯*p*-values in bold indicate statistical significance ($p<0.05$).

There was a significant improvement in GOALS scores after training in the MISTELS group ($p=0.04$) with a mean score of 9.3 (range 7–11) for the first evaluation and 12.4 (range 11–14) for the second evaluation. This difference was also significant in the LAP Mentor group ($p=0.03$) with a mean score of 9.2 (range 7–13) and then 13.2 (range 12–15). This difference was not significant in the control group ($p=0.35$); the mean score obtained from the first evaluation of the control group was 12.2 (range 10.5–12.5) versus 11.7 (range 8.5–14.5) for the second evaluation. There was significant difference ($p=0.03$) in favour of the MISTELS groups versus the control group. Likewise, the progression score was significantly higher in the LAP Mentor group ($p=0.007$) versus the control group. There was no significant between the MISTELS and the LAP Mentor groups ($p=0.28$). In the MISTELS and LAP Mentor groups, the difference between initial and final GOALS scores decreased with the number of surgical rotations ($r = -0.58$, $p=0.046$). Progression between GOALS 1 and GOALS 2 was significantly higher in MISTELS ($p=0.03$) and the LAP Mentor group ($p=0.007$) versus the control group. As for the inter-rater reliability analysis, the ICC coefficient was 0, 31, but there was an equal gap between the two raters for each assessment and the hierarchy of the scores was the same.

Nineteen candidates were included in the study and evaluated. The three groups were comparable in seniority and in the number of rotations of laparoscopic surgeries performed. As for the laparoscopic tasks performed by each participant during the study, there was no significant difference between the three groups in terms of overall laparoscopic tasks (operator or assistant operator) but there was a significant difference in procedures as operator between the MISTELS and the control groups in favour of the control group ($p=0.03$) and between the MISTELS and the LAP Mentor groups in favour of the LAP Mentor group ($p=0.03$). However, there was no significant difference between the LAP Mentor and the control groups.

The mean GOALS scores obtained from the participants' first laparoscopic cholecystectomy in the OR were comparable except between the MISTELS and the control groups. There was a high positive correlation between the number of surgical rotations and the initial GOALS score ($r=0.64$, $p=0.003$). There was no significant difference between the difficulty VAS of the first and second evaluation laparoscopies. However, there was a significant difference between the VAS for the first laparoscopies performed by the control group compared with those by the LAP Mentor group ($p=0.046$) and between those by the LAP Mentor group compared with those by the MISTELS group ($p=0.03$).

Table E.07 Operating room performance of laparoscopic bilateral tubal ligation by Limbs and Things laparoscopic simulator-trained and non simulator-trained participants

Banks et al 2007 RCT, Level II	Simulator training (n=10)	No simulator training (n=10)	*p*-value~
Performance parameter			
Global score (Mean ± SD)	22.3/35 (64%) ± 5	15.8/35 (45%) ± 11	0.003*
Task-specific 25-point checklist score (Mean ± SD)	92% ± 7	57% ± 20	0.002*
Pass rate	100%	30%	0.003**
Knowledge post-test	94%	59%	0.003*
How prepared for LSC BTL *(scale 1–5, with 5 most prepared)*	3.4	2.5	0.009**

~*p*-value in bold indicates statistical significance (*p*<0.05).
* Wilcoxon rank sum test.
**Fisher's exact test.
LSC: laparoscopic; BTL: bilateral tubal ligation; SD: standard deviation.

In the operating room (OR), the residents who were assigned randomly to the simulator training group scored significantly higher than the control group on all three surgical evaluation tools: the global rating scale (64% versus 45%; *p*=0.003), the task-specific checklist (92% versus 57%; *p*=0.002) and the pass/fail score (100% versus 30%; *p*=0.003). The simulator-trained group also performed significantly better on the knowledge post-test than the control group (94% versus 59%; *p*=0.003). When the checklist scores were broken down by skills, the simulator-trained group rated significantly higher on preoperative skills (surgical set up), basic surgical skills, basic laparoscopic skills and the laparoscopic BTL-specific skills than the control group (*p*<0.005). When the global rating scale was broken down by aspects of surgical skills, the simulator-trained group scored significantly higher in each of the seven categories (respect of tissue, time and motion, instrument handling, knowledge of instruments, operation flow, use of assistants, knowledge of procedure) than the control group (*p*<0.01, all).

The observers of the laparoscopic BTL were evaluated for interrater reliability. The proportion of agreement for interrater reliability on the task-specific checklist was 84%, and interrater reliability on the global rating scale showed an intraclass correlation (ICC) coefficient of 0.86. The evaluators had 100% agreement on the pass/fail rating.

The objective surgical skills assessments of residents after the completion of the simulator training were correlated significantly with the same resident's performance in the OR, with Spearman correlation coefficient of 0.69 (*p*=0.03) for the global rating scale and 0.75 (*p*=0.01) for the checklist (Table D.08).

At baseline, there were no differences between the surgical simulator group and control group on any of the technical skills measures in the laboratory (the global rating scale, the checklist, the pass/fail test) or on the knowledge pretest. On a scale of 1–5 (with 5 being the most comfortable), the two groups were believed to be prepared equally for their first BTL at the outset, and there were no significant differences in the number of laparoscopic BTLs that they had seen or performed before the observed procedure.

Table E.08 Correlation between objective assessment of surgical skills of the same residents in the simulator laboratory and the operating room

Banks et al 2007 RCT, Level II	Simulator scores (n=10)	Operating room score (n=10)	*p*-value~ (paired t-test)
Performance parameter			
Global score (mean ± SD)	20.3/35 (58%) ± 7	22.3/35 (64%) ± 5	0.01
Task-specific 25-point checklist score (mean ± SD)	86% ± 8	92% ± 7	0.003
Pass rate	100%	30%	

~*p*-value in bold indicates statistical significance (*p*<0.05)
SD: standard deviation.

Table E.09 Performance of laparoscopic salpingectomy by LapSim Gyn simulator-trained and non simulator-trained participants

Larsen et al 2009 RCT, Level II	Simulator training (n=11)	No simulator training (n=10)	*p*-value*¯
Performance parameter, median (range; interquartile ranges)			
Surgical performance: total score (points)	33 (25–39; 32–36)	23 (21–28; 22–27)	<0.001
% reaching ≥30 points	82	0	Not stated
Time taken to perform operation (minutes)	12 (6–24; 10–14)	24 (14–38; 20–29)	<0.001

Inter-rater agreement 0.79, γ-coefficient 0.83 (95% confidence interval 0.68 to 0.98).
* Mann-Whitney U test.
¯*p*-value in bold indicates statistical significance (*p*<0.05).

The median total score on the general and task specific rating scale reached 33 points (interquartile 32–36 points) in the simulator-trained group and 23 (22–27 points) in the control group *(p*<0.001) (Table E.09). Thus, after proficiency-based VR training in laparoscopic salpingectomy there was a significant improvement of operative skills compared with the controls when assessed in their first actual laparoscopic salpingectomy procedure in the OR (*p*<0.001). On the rating scale used in this study, which had been previously validated in a separate investigation, novices (fewer than five procedures) scored a median of 24 points, and intermediately experienced trainees (20–50 procedures) scored a median of 33 points compared with a median 39 of points for experts. Thus, after training in the specific procedure to a predefined, proficiency-based level, inexperienced trainees progressed from the performance level of a novice to that of an intermediately-experienced gynaecologist assessed in their first actual procedure. Therefore the learning curve in the OR was shorter for those trained to proficiency on the simulator compared with the control group.

The time to complete the laparoscopic salpingectomy was reduced by half for the simulator-trained group compared with the control group. The median total time to complete the procedure was 12 minutes (interquartile range 10–14 minutes) in the simulator-trained group compared with 24 (20–29 minutes) in the control group *(p*<0.001;Table E.09). Twenty-one operations were assessed.

The time used by the assessors to fill in the rating chart was the mean total operation time plus five minutes for each DVD recording. The inter-rater agreement was 0.79 (166/210). The γ coefficient used to investigate the strength of correlations among the observers at a single subject level reached 0.83 (95% confidence interval 0.69-0.99).

There were no significant differences between the two groups in baseline characteristics (age or gender; 90% women) or performance on the simulator. The median number of simulated salpingectomies needed to reach proficiency level in the simulator trained group was 28 (24-32 salpingectomies). The control group was offered simulator training after the study operation; nine out of 11 trainees in this group volunteered and a median of 26 (23-32) simulated operations were needed to reach the proficiency level (*p*=0.70). The mean time spent on training using the simulator was 7 hours and 0 minutes (5 hours 30 minutes – 8 hours 0 minutes) in the simulator-trained group and 7 hours and 0 minutes (5 hours 15 minutes – 7 hours 45 minutes) in the control group *(p*=0.65). The baseline score (first attempt) was 8 (5–15) in the simulator-trained group and 9 (7–19) in the control group after training *(p*=0.70). Two trainees were subsequently excluded from the simulator-trained group (one failed to complete training and the other was involved in an operation that was cancelled because of anatomical abnormality). One trainee was excluded from the control group because of a technical fault in the DVD recorder used to record the operation.

Table E.10 LapSim Gyn simulator training programme required to reach proficiency level for simulator-trained participants compared with the control group volunteers who trained after the study operation

Larsen et al 2009 RCT, Level II	Simulator training (n=11)	Control group (n=9)*	*p*-value¯
Number (range) of training sessions	28 (16–39^)	26 (19–43^)	0.76
Duration (range) of training	7 hours 15 minutes (4 hours 15 minutes – 9 hours 30 minutes^)	7 hours 0 minutes (4 hours 0 minutes – 9 hours 15 minutes^)	0.70
Median percentage (range) scores on first attempt	8% (5–15)	9% (7–19)	0.70

* Voluntary simulator training after surgery.
¯ Mann-Whitney U test.
^ The range values in text in original article differed from those in table in same article.

Table F.11 Performance of intracorporeal suturing and knot tying during a laparoscopic Nissen fundoplication by MIST-VR simulator-trained and non simulator-trained participants

Van Sickle et al 2008 RCT, Level II	Simulator training (n=11)	No simulator training^ (n=11)	*p*-value*~
Performance parameter (mean ± SD)			
Total suturing time (seconds)	525 ± 190	790 ± 171	<0.003
Total suturing errors	25.6 ± 9.3	37.1 ± 10.2	<0.01
Excess needle manipulations	18.5 ± 10.5	27.3 ± 8.6	<0.05

^ Standard clinical training only.
* Student's t-test with unequal variance.
~*p* value in bold indicates statistical significance (*p*<0.05).
MIST-VR™: Minimally Invasive Surgical Trainer-Virtual Reality; SD: standard deviation.

With regard to the operative performance, participants in the VR simulator-trained group completed the suturing task in significantly less time on both sutures (525 ± 190 seconds versus 790 ± 171, *p*<0.003) and committed significantly fewer total suturing errors (25.6 ± 9.3 versus 37.1 ± 10.2, *p*<0.01). With regard to excess needle manipulations the simulator-trained group also made significantly fewer excess needle manipulations than the group that received no simulator training (18.5 ± 10.5 versus 27.3 ± 8.6, *p*<0.05).

There were no statistically significant differences in age, gender, earlier suturing experience, earlier laparoscopic experience or perceptual ability (PicSOr test scores) between the simulation-trained and the standard training (clinical training only) groups at baseline.

Subjects in postgraduate years 3, 5 or 6 in the Emory University School of Medicine surgery department were recruited and randomised to two groups, the Curriculum training group: receiving training on the MIST-VR simulator and the standard box trainer or the Control group receiving standard clinical training. All residents were shown a 15 minute training video on laparoscopic suturing and knot tying and demographic information including previous laparoscopic operative experience was recorded. Baseline perceptual ability was assessed using a validated test and sufficient basic knowledge in laparoscopic suturing technique was assessed using a short examination.

For the curriculum trained group both MIST-VR and video tower box trainer simulators with foam models were used. Training performance criteria were established by experienced laparoscopic surgeons and research fellows. The mean scores of attending surgeons and clinical fellows over five repetitions of set tasks determined the performance criteria. The subjects in the trained group received training under supervision on the MIST-VR and trained to expert level before being allowed to progress to training on the standard box trainer, where they received supervised training until proficiency was established. For the MIST-VR the task was a traversal task chosen to teach the trainee to use both hands efficiently, while for the box trainer the tasks were designed to educate participants on proper needle orientation and manipulation and to produce a high-quality laparoscopic intracorporeal knot. Participants in the group receiving no simulator training were allowed access to both the VR simulator and the standard box trainer but received no supervision or formal training.

Subjects in both groups performed the fundal suturing portion of a laparoscopic Nissen fundoplication with an attending surgeon blinded to training status. A standardised three-suture fundoplication was performed with the first and most cephalad suture being placed by the attending surgeon. The portion of the procedure completed by the participants was video recorded. Recordings of the operative performance for each group were reviewed by two surgeon investigators who were blinded to the training status of participants and to the identity of the operative team members. Performance was scored based on a set of tightly predefined errors (construct validity demonstrated in a separate study). Interobserver agreement used >0.8 as the cut-off.

Table E.12 Performance of laparoscopic knee arthroscopy by simulator-trained and non simulator-trained participants

Howells et al 2008 RCT, Level II	Simulator training (n=10)	No simulator training (n=10)	*p*-value*¯
Performance parameter; median (interquartile range; range) ^			
Orthopaedic Competence Assessment Project (OCAP) satisfactory scores	75 (45–90;12–90)	≤15^ (0–15;0–80)	0.0007
Objective Structured Assessment of Technical Skill (OSATS) Global Rating Score (9 items; scale 1–5; maximum 45 points)	24 (17–29; 11–39)	10 (9–12; 9–25)	0.0011

^ estimated from graph.
* Mann-Whitney U test.
¯*p*-value in bold indicates statistical significance (*p*<0.05).

Analysis of the performance in the operating theatre of both simulator-trained and control groups using the Orthopaedic Competence Assessment Project (OCAP) checklist and the global rating scale showed significant differences. The simulator-trained group were seen to significantly outscore the untrained group in terms of both the OCAP project (*p*=0.0007) and the global rating scores (*p*=0.0011).

The groups were equally matched in terms of demographics and previous surgical experience (data not published).

Movement analysis showed that the performance of all those in the simulator-trained group improved objectively, with demonstrable learning curves for all three output parameters (time taken, number of hand movements, total path length of hand movements). The learning curves were statistically significant, clearly showing improvements in simulator performance and learning during training (*p*=0.001).

Table E.13 Performance of laparoscopic totally extraperitoneal inguinal hernia repairs by Guildford MATTU TEP hernia simulator-trained and non simulator-trained participants

Zendejas et al 2011 RCT, Level II	Simulator training (n=26), mean ± SD	No simulator training (n=24), mean ± SD	Unadjusted mean change or OR (95% CI)	*p*-value~	Adjusted* Mean change or OR (95% CI)	*p*-value~
TEP#1 operative time (minutes) ; raw	30.9 ± 7.3	37.4 ± 8.3	-6.5 (-10.1 to -2.9)	0.0005	-8.3 (-11.6 to -4.9)	<0.0001
TEP#1 operative time (minutes); participation-corrected	34.4 ± 8.4	47.5 ± 13.9	-13.1 (-18.4 to -7.8)	<0.0001	-15.0 (-20.7 to -9.3)	<0.0001
Percentage resident participation	88.4 ± 9.4	73.7 ± 16.4	+14.6 (8.4 to 20.7)	<0.0001	+12.43 (5.1 to 19.8)	0.0009
TEP#1 operative performance^	21.9 ± 2.7	18.3 ± 3.8	+3.6 (2.1 to 5.1)	0.001	+3.4 (2.4 to 4.5)	<0.0001
Intraoperative complications; N (%)	2 (5)	13 (35)	OR 0.10 (0.02 to 0.50)	0.001	OR 0.14 (0.02 to 0.84)	0.032
Postoperative complications; N (%)	1 (3)	11 (30)	OR 0.06 (0.01 to 0.53)	0.001	OR 0.04 (0.002 to 0.83)	0.037
Overnight stay; N (%)	0 (0)	9	0	0.001	undefined	-
ALL totally extraperitoneal inguinal hernia repairs post-randomisation (excluding crossovers)		(n=14)				
Operative time (minutes); raw	29.6 ± 6.7	35.7 ± 7.6	-6.6 (-10.1 to -3.2)	0.0002	-7.5 (-10.6 to -4.4)	<0.0001
Operative time (minutes); participation-corrected	32.4 ± 7.7	44.2 ± 12.9	-12.6 (-17.8 to -7.3)	<0.0001	-13.5 (-18.3 to -8.8)	<0.0001
Percentage resident participation	90.8 ± 8.3	77.1 ± 16.2	+13.7 (9.3 to 18)	<0.0001	+13.65 (7.3 to 20)	<0.0001
Operative performance^	23.3 ± 3.0	18.7 ± 3.8	+3.7 (1.8 to 5.6)	0.0001	+3.8 (2.6 to 5.0)	<0.0001
Intraoperative complications; N (%)	5 (7)	17 (29)	OR 0.20 (0.06 to 0.66)	0.009	OR 0.15 (0.04 to 0.59)	0.006
Postoperative complications; N (%)	4 (9)	15 (26)	OR 0.16 (0.04 to 0.69)	0.013	OR 0.17 (0.04 to 0.74)	0.018
Overnight stay; N (%)	5 (7)	12 (21)	OR 0.23 (0.05 to 1.0)	0.050	OR 0.37 (0.08 to 1.67)	0.20

~*p*- values in bold indicate statistical significance (*p*<0.05).
*Outcomes of time were adjusted for supervising surgeons, case difficulty, side (first repair versus second repair) and baseline operative time; outcomes of operative performance were adjusted for supervising surgeon, case difficulty and baseline operative performance; patient outcomes were adjusted for patient-related factors (age, gender, history of prostate disease, ASA, BMI) and case-related factors (difficulty, recurrent hernia, staff-surgeon).
^GOALS score; range 6–30.
ASA: American Society of Anaesthesiologists 'physical status levels I to IV; BMI: Body Mass Index; CI: confidence interval; GOALS: Global Operative Assessment of Laparoscopic Skills; OR (statistics): Odds ratio; SD: standard deviation; TEP: totally extraperitoneal.

Operative time

On the first totally extraperitoneal inguinal hernia repair after randomisation (TEP#2), repairs performed by simulator-trained participants were on average 6.5 minutes faster than those performed by the control group (95% CI -10.1 to -2.9; *p*<0.0001) . Mean resident participation was also different (simulator-trained group 88.4 ± 9.4 versus control group 73.7 ± 16.4; *p*<0.0001). After correcting time to account for varying participation rates, the difference between groups was even greater (participation-corrected time 13.1 minutes faster for simulator-trained group; 95% CI -18.4 to -7.8; *p*<0.0001). When evaluating subsequent totally extraperitoneal inguinal hernia repairs, simulator-trained participants remained faster than control participants (raw time, mean difference: -6.6 minutes, *p*<0.0001). At the third totally extraperitoneal inguinal hernia repair (TEP#3), crossover simulator-trained residents were also faster than their control counterparts (raw time, mean difference: -6 minutes, *p*<0.0001). Similar results were found when adjusting operative time for supervising surgeon, case difficulty, side (first repair versus second repair if bilateral) and baseline operative time.

Operative performance

Operative performance ratings (GOALS scale 6–30) were better for simulator-trained participants compared with control participants immediately after training (mean difference +3.6; 95% CI 2.1–5.1; *p*=0.001) and for all totally extraperitoneal inguinal hernia repairs post-randomisation (mean difference +3.7; 95% CI 1.8–5.6; *p*=0.0001). No differences in performance were found between crossover simulator-trained participants and their control counterparts at the third repair (TEP#3), or between groups at the fourth (TEP#4) or fifth repair (TEP#5) (all *p*>0.05). When adjusting for supervising surgeon case difficulty and baseline operative performance, results remained

unchanged. Inter-rater reliability was near-perfect for both live ratings and delayed video review ratings (ICC 0.96). Similarly, agreement between live and video review ratings (test-retest) was excellent (ICC 0.95).

Patient outcomes

Intraoperative and postoperative complications (of any type), and overnight stay were less likely for the second totally extraperitoneal inguinal hernia repair (TEP#2) in the simulator-trained group (OR 0.1, 0.06, and 0, all $p<0.05$. During subsequent repairs, similar findings were observed except that the difference in overnight stay was of borderline statistical significance. No significant differences in patient outcomes between crossover simulator-trained participants (n=10) and their control counterparts (n=9) were observed in the third totally extraperitoneal inguinal hernia repair (TEP#3). With a median follow-up of 5.2 months (range 0–12), the number of patients who experienced a hernia recurrence (nil in the simulator-trained group and one in the control group) or were evaluated for groin pain at least three months post-repair (one each in the simulator-trained and control groups) were similar between groups ($p>0.05$ for both comparisons).

Associations between intraoperative and postoperative complications

At any point, after adjusting for patient factors (age, sex, history of prostate disease, ASA, BMI) and case-related factors (difficulty, recurrent hernia or staff-surgeon), peritoneal tears were independently associated with an increased risk of developing postoperative urinary retention (adjusted OR 5.84, $p=0.006$) and the subsequent need for an overnight stay (adjusted OR 9.45, $p=0.0004$). A history of prostate disease had a similar effect on postoperative urinary retention (adjusted OR 5.1, $p=0.03$) and the need for overnight stay (adjusted OR 2.9, $p=0.04$). Operative time was not a predictor of urinary retention ($p=0.73$).

Baseline characteristics and performance

Of 99 general surgery residents in the training program during the study period, 50 were exposed to at least one TEP repair. All 50 residents consented to participate in this study and performed a total of 219 totally extraperitoneal inguinal hernia repairs in 146 patients during the study period. Trainee demographics and baseline time, performance and patient outcomes were similar between groups. Staff and resident participation as well as other operative and patient characteristics were distributed equally among the groups.

Training curriculum

All participants randomised to the simulator training curriculum (n=26), and those in the control group (n=24) who subsequently crossed over (n=10), successfully achieved mastery criteria during training. For the online module, a mean ± SD of 45 ± 24 minutes and up to three attempts were required to pass. Participants required a median of 16 (range 7–27) simulated repairs during 2 ± 1 practice sessions and 65 ± 40 minutes to master the skills training component. No significant differences between the simulator-trained and control, crossover residents were observed for any of the training parameters, The time intervals between baseline operative assessments first totally extraperitoneal inguinal hernia repair (TEP#1) and the first repair postintervention or randomisation (TEP#2) were similar between the simulator-trained and control groups (9.2 ± 6.0 days versus 11.1 ± 8.5 days; $p=0.57$).

Table E.14 Performance of colonoscopy by AccuTouch colonoscopy simulator-trained and non simulator-trained participants

Park et al 2007 RCT, Level II	Simulator training (n=12)	No simulator training (n=12)	*p*-value¯
Performance parameter (mean ± SD)			
Global Rating Score (maximum possible global rating score =35)	17.9 ± 5.2	14.8 ± 2.5	0.04
Number reaching caecum	1	0	
Number of critical flaws (e.g. clinical complications)	0	0	

¯*p*-value in bold indicates statistical significance (*p*<0.05).
SD: standard deviation.

The mean total global rating score for the treatment group was 17.9 (SD 5.2) out of a possible score of 35, which was significantly higher than the mean score of the control group of 14.8 (SD 2.5)($t_{(22)}$ = 1.84, *p*=0.04). This amounted to an effect size of 0.8 (Cohen's d). One resident from the control group and none from the treatment groups completed the entire colonoscopy to the caecum on their own in the allotted time. There were no clinical complications or terminations due to resident factors in either group (i.e. no 'critical flaws').

Residents in the treatment group trained on the colonoscopy simulator for a mean time of 125 minutes (SD 37 minutes), during which time they accessed the full range of available cases.

The demographic characteristics were not significantly different between groups, with similar proportions of gender, age and postgraduate year level in each group. In addition, the ratio of general surgery to internal medicine residents was 2:10 in the control group and 3:9 in the simulator-trained groups. There were no significant difference between groups on any of the eight computer-recorded performance metrics on the pre-test procedure ($t_{(22)}$ = .24 to $t_{(22)}$ = 1.15, P > 0.05 for each parameter). Similarly, the faculty-generated global ratings for the pre-test showed no significant differences between the groups, with the treatment group scoring 13.7 (SD 4.5) out of a maximum possible score of 35 versus 15.4 (SD 4.8) for the control group ($t_{(22)}$ = 0.91, *p*=0.37).

Table E.15 Performance of colonoscopy by KAIST-Ewha colonoscopy simulator II-trained and non simulator-trained participants

Yi et al 2008 Comparative study, Level III-2	Simulator training (n=25*) (5 participants)	No simulator training (n=30*) (6 participants)	*p*-value¯
Performance criteria (mean ± SD)			
Total insertion time (min)	31.0 ± 18.7	41.5 ± 21.2	0.028
Success rate	0.76 ± 0.44	0.43 ± 0.50	0.006
Number of red-outs	10.9 ± 7.2	23.5 ± 20.6	0.002
Number of air inflation	9.5 ± 6.4	13.5 ± 10.4	0.043
Number of loop formation	2.1 ± 2.1	2.0 ± 1.7	0.414
Mucosal visualisation^	3.5 ± 0.8	2.9 ± 0.7	0.002
Abdominal pressure application	2.6 ± 2.1	2.3 ± 2.4	0.269
Changes in patient's posture	1.7 ± 1.1	2.0 ± 1.9	0.252
Overall performance accuracy^	3.6 ± 0.8	2.7 ± 0.8	<0.001
Extent of abdominal pain#	3.1 ± 0.8	3.2 ± 1.1	0.273
Extent of abdominal inflation#	3.0 ± 0.9	3.2 ± 1.3	0.215
Extent of anus discomfort#	2.7 ± 0.8	3.4 ± 0.9	0.002

***** Participants performed colonoscopies on five different patients.
^ Mucosal visualisation and overall performance accuracy were graded based on the scale 1–5 (1=poor, 5=excellent).
# Extent of abdominal pain, inflation and anus discomfort were graded on a scale 1–5 (1=no pain, 5=worst pain).
¯*p*-values in bold indicate statistical significance (*p*<0.05).
KAIST: Korea Advanced Institute of Science and Technology SD: standard deviation.

The simulator-trained group significantly outperformed the control group in terms of insertion time, success rate, number of red-outs, number of air inflation, mucosal visualisation, and overall performance accuracy (*p*<0.05). However, numbers of loop formation, abdominal pressure application and changes in patient's posture did not show any meaningful difference (*p*>0.05). In the patient survey, the simulator-trained group also showed less discomfort during the colonoscopy compared to the control group. However, there was no significant difference between the two groups except in the extent of anus discomfort.

Both groups were evaluated during colonoscopies on actual patients, which were performed under close supervision of colonoscopy experts. Each subject performed colonoscopies on five different patients. The supervising experts evaluated the trainees based on accuracy of colonoscopy results and the established performance criteria. Questionnaires on the colonoscopy experience were also filled out by the patients.

Basic instruction on operation of the colonoscope and on colonoscopy was given to both groups. Simulation-based training included practising the targeted skills of colonoscopy using two training scenarios that have different colon flexures and degrees of difficulty. The training scenario A is designed to teach practical skills to navigate the colon, applying torque and up-down angulations. The training scenario B is designed to teach skills to manage a loop formed in the sigmoid colon.

The simulator had a scoring system that was based on the performance criteria derived from experts' profiles. The trainees were requested to practice until they reach all the established training goals. The subjects completed, on average, 53.4 (range 26–100) and 68.2 (range 33–105) procedures to pass the scenarios A and B, respectively. The mean of the total training time per subject was 229.4 (range 82–377) and 232.00 (range 141–414) minutes, respectively.

Table E.16 Performance of cystourethroscopy by URO Mentor VR simulator-trained and non simulator-trained participants

Schout et al 2009 RCT, level II	Simulator training (n=50)	No simulator training (n=50)	*p*-value~
Global rating scale (mean ± SD)			
Mean (overall)	3.8 ± 1.2	3.0 ± 1.0	<0.001
1: respect of tissue	3.6 ± 1.0	3.0 ± 1.3	0.003
2: time and motion	3.6 ± 1.2	2.9 ± 1.2	0.001
3: handling of endoscope	3.7 ± 1.2	2.9 ± 1.2	<0.001
4: flow of procedure	3.9 ± 1.0	3.1 ± 1.3	0.001
5: knowledge of procedure	4.0 ± 1.1	3.2 ± 1.3	0.001

~*p*-values in bold indicate statistical significance (*p*<0.05).
SD: standard deviation; VR: virtual reality.

The mean overall global rating scale (GRS) score for performance of cystourethroscopy (CUS) was significantly higher in the simulator-trained group than the control group (3.8 ± 1.2 versus 3.0 ± 1.0, *p*<0.001). A comparison between simulator-trained and control participants showed a significant, moderate to large effect of training over all domains (*p*≤0.003, standard regression coefficient β=0.30–0.47, regression coefficient B=0.80–1.17). The mean (SD) GRS scores of the trained group ranged from 3.6–4.0 (1.0–2.0), whereas those of the control group were 2.9–3.2 (1.0–1.3).

For participants who preferred a nonsurgical specialty training or who were still undecided about their future speciality, simulator training had a positive effect on all GRS scores, whereas for participants with a preference for a surgical speciality simulator training had no positive effect on mean GRS, GRS-2 (time and motion) and GRS-3 (handling of endoscope) scores. Furthermore, participants from University Medical Centre Groningen had significantly lower scores on GRS-3 than Catharina Hospital Eindhoven participants (*p*=0.007). Apart from these two variables (preferred speciality and interaction training –location), there were no significant differences between the simulator-trained and control groups for the other moderator variables; gender, number of CUS ever observed, age, interaction training-gender, interaction training-number of CUS ever observed, and interaction training-age.

All participants thought that the final task of the simulator training session was significantly easier than the first task (mean difficulty 4.4 and 7.1, respectively, on a scale of 1–10, *p*<0.001). Moreover, according to participants, CUS in a patient was significantly more difficult than the final task of the simulator training (mean difficulty 6.1 and 4.4, respectively, *p*<0.001).

In the simulator-trained group there were no significant differences between Catharina Hospital Eindhoven and University Medical Centre Groningen in the stress experienced. However, of the University Medical Centre Groningen controls, 71% said they experienced stress or tension, whereas only 17% of Catharina Hospital Eindhoven controls indicated experiencing stress or tension (P<0.001).

Participants were randomly assigned to one of two groups: URO Mentor VR simulator-trained or control (no simulator training) group. Before performing CUS in a patient, the participants watched a 15-minute instruction video to ensure standard level of background knowledge of the procedure and instrument handling. Next the participants participated in a maximum 15-minute introduction session on movements of the cystoscope in which a real cystoscope in a glass globe, representing the bladder, was used. A strict training protocol was designed and adhered to, and the two instructors observed each other in the first two cases.

Patients with similar levels of difficulty were presented to the groups of participants, as shown by there being no significant differences in supervisors' ratings of patient difficulty between the simulator-trained and the control group.

The supervisors of the real-time CUS, who were unaware of the participants' training status, scored performance using a five-point GRS, a modification of an assessment method that was validated earlier (5-point scale for 5 performance criteria: respect of tissue; time and motion; handling of endoscope; flow of procedure and forward planning; knowledge of procedure). The main focus of this study was on the effect of 'training' on the dependent variables, and on the mean GRS.

Table E.17 Performance of oesophagogastroduodenoscopy by GI-Mentor II simulator-trained and non simulator-trained participants

Shirai et al 2008 RCT, Level II	Simulator training (n=10) Median (interquartile range, range 10th and 90th percentile)*	No simulator training (n=10) Median (interquartile range, range 10th and 90th percentile)*	*p*-value~
1: insertion into oesophagus	*2.0 (1.5–2.5;0.5–3.0)	*1.0 (0.5–1.5; 0.5–2.5)	<0.05
2: crossing the oesophagogastric junction (OGJ)	*2.5 (2.0–3.5;0.5–4.0)	*2.5 (1.5–2.5; 0.5–3.5)	NS
3: passing from OGJ into gastric antrum	*3.4 (2.8–3.8;2.2–3.8)	*1.7 (1.2–2.7; 0.5–3.8)	<0.01
4: passing through the pyloric ring	*2.5 (1.5–3.5;0.5–3.5)	*1.5 (0.5–2.5; 0.5–2.5)	<0.05
5: examination of duodenal bulb	*3.5 (2.5–3.5;1.5–3.5)	*2.5 (1.5–2.5; 0.5–2.5)	<0.05
6: insertion into third part of the duodenum	*2.0 (1.0–4.0;1.0–4.0)	*2.5 (1.0–2.5; 1.0–3.0)	NS
7: examination of the gastric antrum	*2.5 (2.5–3.5;2.0–4.0)	*2.5 (1.5–3.0; 0.5–3.5)	NS
8: examination of the gastric angle	*3.0 (2.0–3.5;1.5–4.5)	*2.6 (1.6–3.0; 1.0–3.5)	NS
9: manipulation for retroflexion	*2.7 (1.7–2.7;1.7–3.3)	*2.7 (1.7–2.7; 0.5–3.3)	NS
10: looking down the gastric body	*3.8 (2.8–3.8;1.5–3.8)	*2.8 (1.7–2.8; 0.5–3.8)	NS
11: viewing the fornix	*3.8 (2.8–3.8;1.5–3.8)	*2.8 (1.0–2.8; 0.5–2.8)	<0.05
Total procedure time (minutes)	14.40 (12.15–16.07;*10–20)	14.05 (13.30–16:00;* 10–20.0)	NS

* Five-point maximum evaluation grade scale values estimated from graphs in article (values not given by authors).
~*p*-values in bold indicate statistical significance (*p*<0.05).
NS: not significant; OGJ: oesophagogastric junction.

The evaluation scores were significantly higher in the simulator group with respect to insertion of the endoscope into the oesophagus, passing the OGJ into the antrum, passing through the pyloric ring, and examination of the duodenal bulb and fornix. There was no significant difference in the total procedure time between the simulator-trained and the control groups (14:40 [12:15–16:07] minutes versus 14.05 [13:30–16:00] minutes). However, the number of 1-point scores (where direct assistance by supervisor was required) for the simulator group was significantly smaller than that for the control group (19 out of 220, 8.6% in the simulator-trained group versus 57 out 220, 25.9% in the control group; *p*=0.0017). There was no significant difference between the groups for 2-point scores (where instructions were required); simulator-trained group 39/220, 17.7% versus controls 50/220, 22.7%; *p*=0.36. There was no significant difference in total time for each trial between the two groups.

Eleven items and two endoscopic procedures for each member were evaluated for each group (n=220 i.e. 10 x 11 x 2=220). Evaluation was performed simultaneously by two supervisors, who filled in the evaluation form independently of each other. When the scores assigned by each were different, the mean score was used for the analysis.

The skill in carrying out each item was evaluated using a five-point grade scale, which considered the ability of manipulation, time required, success/failure after three trails, grade of difficulty of the case, performance of risky manoeuvres, and discontinuation of the procedure at the request of volunteer. This scale was defined as follows:

5 points: the resident could perform the manoeuvre as well as supervising physician
4 points: skill was good, but not as good as that of supervising physician
3 points: the resident could perform the manoeuvre without receiving instructions from the supervisor
2 points: instructions were required
1 point: direct assistance by the supervisor was required.

A manoeuvre was defined as risky when there was a possibility of mucosal injury or perforation due to insertion of the endoscope without any confirmation of the position of the lumen. The time for each item, apart from insertion into oesophagus and insertion into the third part of the duodenum, was set at two minutes. Up to three attempts were allowed for insertion into the oesophagus, crossing the OGJ, passing through the pyloric ring, and insertion into the third part of duodenum. Instructions were provided when the supervisor considered that the manoeuvre was risky or when the endoscope remained at the same site for two minutes or more. The supervisor provided direct assistance when the response to oral instructions was inadequate, the endoscope did not move to next item within the time limit, the volunteer showed signs of discomfort, or the manoeuvre could not be carried out within three trials. After assuming charge of the procedure, the supervisor continued it up to next item, and then resident resumed from that item. The time when the supervisor took over the scope was included in the total procedure time.

After three hours of explanation regarding manipulation of an endoscope, endoscopic observation, and endoscopic diagnosis of common diseases, 20 residents with no prior experience of performing endoscopy were each randomised to simulator and non-simulator groups. Both groups received approximately 15 hours of bedside training. The simulator group also received five one-hour sessions of simulator training within two weeks. There was no significant difference in age and gender between the two groups or between the volunteers used within each group.

Table E.18 Performance of Sinonasal nasal model simulator-trained and non simulator-trained participants

Ossowski et al 2008 RCT, Level II	Simulator training (n=10)	No simulator training (n=10)	*p*-value*~
Performance parameter (mean ± SD)			
NM rigid nasal endoscopy time (seconds)	61.30 ± 56.05	104.30 ± 78.94	0.025
NM flexible laryngoscopy time (seconds)	23.40 ± 29.30	32.20 ± 27.45	0.085
Standardised patient flexible laryngoscopy time (seconds)	56.60 ± 36.97	50.20 ± 35.49	0.315
Standardised patient comfort/discomfort score (visual analogue score)	0.89 ± 0.77	1.33 ± 1.70	0.448

*Mann-Whitney test.
~*p*-value in bold indicates statistical significance (*p*<0.05).
NM = nasal model; SD: standard deviation.

Participants who trained on the Nasal Model (NM) had a significantly shorter rigid nasal endoscopy procedure time compared with the control group (61 seconds versus 104 seconds, *p*=0.025) when tested on the NM simulator. The procedure time for the flexible laryngoscopy on the NM showed a trend toward shorter procedure time for the simulator-trained group, but this did not reach significance (23 seconds versus 32 seconds; *p*=0.085). There was no significant difference between the two groups for the flexible laryngoscopy procedure time on the standardised patient or the discomfort scale assigned by the standardised patient. However, the authors note that these data were affected by positive skewness to the data in the trained group due to extremely high values for two students. For example, the longest procedure time for flexible laryngoscopy on the NM was 106 seconds while all other values were less than 21 seconds. In another example, the longest two procedure times on the standardised patient were 138 and 97 seconds, respectively, while all other procedure times were under 60 seconds.

There was a strong and significant correlation between rigid nasal endoscopy and flexible laryngoscopy procedure time on the NM (*p*<0.001). However there was not a significant correlation between procedure time with the NM and procedure time on the standardised patient. Therefore, performance on NM did not predict performance on the standardised patient. There was, however, a correlation between longer flexible laryngoscopy time on the standardised patient and higher discomfort scores. Interestingly, the authors reported that the students with video game experience had a significantly faster procedure time for flexible laryngoscopy on the standardised patient *(p*=0.038) but there was no significant difference on the NM simulator.

Table E.19 Correlation between procedures rigid nasal endoscopy and flexible laryngoscopy on nasal model and flexible laryngoscopy on standardised patient

Ossowski et al 2008	**Correlation coefficient ρ**	*p*-value~
Relationship		
NM rigid nasal endoscopy and NM flexible laryngoscopy	0.716	<0.001
NM rigid and standardised patient flexible laryngoscopy	0.278	0.236
NM flexible laryngoscopy and standardised patient flexible laryngoscopy	0.335	0.149

~*p*-value in bold indicates statistical significance (*p*<0.05).
NM: nasal model.

Ossowski et al (2008) developed a life-sized model head with human texture and accurate nasal anatomy (the NM). Twenty participants with no prior endoscopic experience were recruited and all were given a 15-minute didactic video introduction to nasal anatomy and the technical use of flexible and rigid endoscopes. The participants were then randomised and stratified according to video game experience. Participants randomised to the training group were given 15-minute to practise, video camera-assisted flexible laryngoscopy and rigid nasal endoscopy on the NM. The control group did not receive this training. All participants were timed performing flexible laryngoscopy and rigid nasal endoscopy on the NM (simulator-tested) and then performing flexible laryngoscopy on a single standardised patient. The patient was blinded to the training status of the participant and filled out a comfort/discomfort score for each student. All testing occurred within 90 minutes of the original demonstration.

Table E.20 Performance of basic sinus surgery tasks by Endoscopic Sinus Surgery simulator-trained and non simulator-trained participants

Fried et al 2010 RCT, Level II	Simulator training (n=12)	No simulator training (n=13)	Difference between groups (95% CI)	Wilcoxon-Mann-Whitney two-sample rank sum U test	*p*-value¯
Performance parameter	Mean ± SD	Mean ± SD			
Mucosal injection time (minutes)	1.75 ± 1.04	4.67 ± 2.09	0.71–3.53	24.0	0.003
Dissection time (minutes)	7.37 ± 3.36	15.44 ± 6.46	5.00–13.62	14.0	<0.001
Injection errors (total number)	3.35 ± 1.96	6.89 ± 3.30	0.21–5.69	24.0	0.048
Surgical confidence (1–10 scale)	6.55 ± 2.65	2.67 ± 2.00	0.90–5.35	20.5	0.009
Instrument manipulation dexterity (1–10 scale)	6.75 ± 2.51	2.78 ± 1.86	0.96–5.27	21.5	0.011
Navigation errors (average number per minute)	1.73 ± 1.05	0.82 ± 0.86	0.08–1.66	39.0	0.032

¯*p*-value in bold indicates statistical significance (*p*<0.05).
CI: confidence interval; SD: standard deviation.

The mucosal injection task was completed in a significantly shorter time in the simulator-trained group and with a narrower range of variability (simulator-trained group mean 1.7 minutes [SD 1.0 minute] versus control group mean 4.7 minutes [SD 2.1 minutes]; *p*=0.003). Dissection time yielded similar observations (simulator-trained group mean 7.4 minutes [SD 3.4 minutes] versus 15.4 minutes [SD 6.5 minutes]; *p*<0.001).

The simulator-trained participants were found to make fewer injection errors (mean 3.3 errors [SD 2.0]) compared with the control group (mean 6.9 errors [SD 3.3]; *p*=0.048). They also demonstrated greater confidence during the dissection portion (simulator-trained group mean 6.6 [SD 2.7] versus control mean 2.7 [SD 2.0]; *p*=0.009) and exhibited a higher level of dexterity with instrument manipulation (simulator-trained group mean 6.7 [SD 2.5] versus control group mean 2.8 [SD 1.9]; *p*=0.011).

The only statistically significant difference in performance that favoured the control group was in the average error count per minute during the endoscopic navigation portion of the procedure (simulator-trained group average 1.7 [SD 1.1] versus control average 0.8 [SD 0.9]; *p*=0.032). However, individual sample data review linked that result to a single outlier in the simulator-trained group.

The simulator-trained participants were fully trained to proficiency on the simulator in addition to receiving conventional textbook-based and video recorded educational material, while the control group only had access to conventional material.

The subjects were assessed on their performance of basic sinus surgery tasks and their first in vivo procedure was video recorded and submitted to a panel of three senior academic otolaryngologists with expertise in endoscopic sinus surgery minimum, and who were blinded to the training status of participants. The inter-rater reliability analysis was performed with Fleiss kappa statistic, which demonstrated a kappa value of 0.87 (*p*=0.021; 95% CI 0.78–0.96).The authors stated that prior work has demonstrated construct validity of the ES3 as a discriminant of users with various levels of experience.

Table E.21 Performance of TURP by PelvicVision VR simulator-trained and non simulator-trained participants

Kälström et al 2010 RCT, Level II	Preoperative	Reference*	Postoperative	Reference*
IPSS (0–35)	21.1	18–24	7.6	3–9
Bother score (0–5)	3.7	4–5	1.5	1–2
Qmax (mL/minute)	8.2	6–12	20.6	22–25
Incontinence %	41	4–46	1–13	1–16
ReTURP %			5.6	6–10
Mortality rate %			1.4–4.2	2–13

Note: follow-up data only provided for the 23 patients whose TURPs were performed during the study.
*The reference data are those of large postoperative studies.
IPSS: International Prostate Symptoms Score; Qmax: maximum urinary flow rate; ReTURP: repeated TURP; TURP: transurethral resection of prostate.

Patient follow-up

The 71 patients who underwent surgery had a follow-up time of 2.3 to 3.8 years (note: only 23 patients were relevant to this review). The International Prostate Symptom Score, bother score, maximum urinary flow rate, and incontinence were evaluated at six to 12 months postoperatively. Eight patients died (11%); five due to causes not related to the surgery (accidents, cancer etc.); and three (4.2%) may have been associated with the surgery. Two of these patients had a cerebrovascular insult at 26 and 51 days postoperatively, respectively, and died at 26 and 51 days later, respectively, and the other patient had known cardiovascular disease and during the postoperative period had haematuria and angina pectoris but was discharged after eight days. This patient suffered a myocardial infarction at home nine days postoperatively and was hospitalised again for eight days. He was found dead in his home at 78 days postoperatively. This gives a surgery-associated mortality of 1.4% to 4.2% (one to three patients).

Kälström et al (2010) also reported the following (other information was only shown in their graphs):

Comparison of first and last TURP procedures (TURP#1 and TURP#3)

There was no significant increase in the amount of autonomous procedure time, resection time, and a tendency to decreased haemostasis time and increased successful orientation time. The resection effectiveness measure as total resection weight and resection weight per minute was significantly lower (p=0.003 and 0.004). The number of aborted procedures due to poor skills/dexterity/judgement decreased from 30% to 0% (p=0.016, sign test), and the proportion of participants believed to be able to perform the procedure independently of a supervisor increased from <10% to about 75% (p=0.000, sign test). There were better scores on the checklist (p=0.000), global assessment (p=0.000), self-evaluation (p=0.000).

Comparison of TURP procedures preceded by and not preceded by simulation practice.

Analysis of the effect of simulation practice was done by comparing the change in skills for each participant. The change in scores for two operations with simulation practice in between was compared with scores for two operations without added simulation practice. Although it was not possible to measure any significant difference in single parameters (except participant operation time, which increased, (p=0.025), there was a tendency for improved skills with simulation practice. When the number of participants who improved or showed no change in skills after simulation practice was compared with the results for procedures without simulation practice, there was a significant difference (p=0.021) indicating that simulation practice resulted in increased skills. Sixteen participants showed greater improvement after simulation practice compared with seven participants who showed greater improvement without practice. One participant could not be evaluated regarding the effect of simulation practice because of exclusion of one patient.

Patient follow-up

There were no significant differences between the groups regarding sex (women:men 1:3), age (mean 33 years) or prior experience with transurethral procedures (mean 14 months of residency, mean 1.2 TURP procedures performed completely and mean 6.2 incompletely). Two surgically experienced urologists analysed the video recordings. They analysed approximately 10 procedures together and then independently analysed the same recordings until reaching an inter-rater agreement of >90% was achieved (agreements/total number of observations).

Table E.22 Performance of fascial closure of abdominal wall by low-fidelity synthetic model simulator-trained and non simulator-trained participants

Palter et al 2011 RCT, Level II	Simulator training (n=9)	No simulator training (n=9)	*p*-value~
OSATS assessment, mean (interquartile range)			
Baseline assessment on model	22.0 (20.5–23.0)	21.0 (20.0–21.0)	0.48
Final assessment in OR	22.0 (20.0–27.0)	16.0 (16.0–19.0)	0.04

~The *p*-values in bold indicate statistical significance (*p*<0.05).
OR: operating room; OSATS: Objective Structured Assessment of Technical Skills.

The validated Objective Structure Assessment of Technical Skills (OSATS) global rating scale assessment of technical skill in the OR was significantly higher for individuals in the simulator-trained group (22.0 [20.0–27.0]) compared with the control group (16.0 [16.0–19.0]; *p*=0.04). After the intervention, the median multiple-choice test score for the simulator-trained group was significantly higher compared with that of the control group (14.0 [11.0–16.0] and 11.0 [9.0–11.0] respectively, *p*=0.03) with an effect size of 1.01. Finally for simulator-trained participants, technical measures of performance using the OSATS global rating scale remained stable between initial assessment and the OR assessment, whereas they decreased for the participants in the control group.

Participants in both groups attended an initial group training session that included a demonstration of a technically correct fascial closure on a synthetic abdominal wall model by a staff general surgeon. Participants in both groups then performed one closure on the model that was evaluated by two observers using the OSATS global rating scale. The OSATS scale is an assessment measure that has been shown to have acceptable interobserver reliability and validity (Reznick et al 1997; Martin et al 1997). Participants assigned to the control group had no further contact with the models.

There was no difference in technical ability between the two groups before the intervention. The median baseline global rating score of residents in the simulator-trained group was 21.0 (20.0–21.0) and in the control group was 22.0 (20.5–23.0) (*p*=0.48 NS).

The participants randomised to the intervention group practised abdominal wall closure in the model to a predefined level of proficiency. Each session was at maximum 1.5 hours in length and individual feedback, limited to technical performance only, was provided from an individual of the study team. Proficiency was reached when the trainee was capable of independently performing a square knot, demonstrated knowledge of what instruments were required for the procedure, and placed sutures correctly 1 cm apart, and 1 cm from the wound edge.

Table E.23 Differences before and after the intervention in subject performance and percentage improvement

Patel et al 2012 RCT, Level II	Pre-test – median (range)	Post-test - median (range)	*p*-value~	% improvement
Second Life simulation training (n=15)				
Knowledge multiple choice questionnaire	35 (14–43)	39 (33–45)	0.012	8
Observation scale	69 (51–84)	115 (101–137)	0.001	25.6
Self-report scale	81 (48–115)	118 (101–164)	0.001	21.3
Imperial College simulation operating suite training (n=15)				
Knowledge multiple choice questionnaire	38 (26–40)	43 (40–44)	0.001	10
Observation scale	76 (56–97)	132 (103–141)	0.001	31.1
Self-report scale	80 (50–126)	142 (138–160)	0.001	35.7
Didactic Lecture (n=15) (no simulation training)				
Knowledge multiple choice questionnaire	39 (31–42)	42 (38–45)	0.001	6
Observation scale	65 (55–89)	107 (81–122)	0.001	23.3
Self-report scale	65 (39–123)	121 (86–144)	0.001	32.2
Control (n=15) (no training)				
Knowledge multiple choice questionnaire	36 (33–41)	37 (33–40)	0.323	2
Observation scale	66 (53–82)	69 (56–89)	0.09	1.7
Self-report scale	93 (37–141)	98 (61–148)	0.123	2.9

~The *p*-values in bold indicate statistical significance ($p<0.05$).

The lecture, Second Life, and Imperial College simulated operating suite (SOS) groups all demonstrated significant improvement in the outcome measures after training, with SOS group demonstrating larger percentage improvements in knowledge, observation, and self-report scale measures compared with the other intervention groups (Table D.23). The control group did not display any significant improvements. Using the Kruskal-Wallis test, significant differences ($p<0.001$) were found after the intervention for all outcome measures between the groups.

Thirty-two women and 28 men ranging from 18 to 21 years participated in the study. The preintervention exposure to the operating theatre revealed no difference among the groups for knowledge assessment ($p=0.477$), observation score ($p=0.212$) and self-report scores ($p=0.099$). Among those receiving the training intervention, there was no significant difference between their evaluations of the training they received ($p=0.36$). The SOS and Second Life groups did not display any significant difference between their opinions as to whether the simulated environment was a realistic representation of a real operating theatre. The internal consistency of the knowledge, attitude, and skills items was high for the observation scale ($\alpha \geq 0.72$) and for the self-report scale ($\alpha \geq 0.881$). An α value >0.7 is considered adequate in terms of research contexts.

Further analysis using Mann-Whitney U test was to establish whether the lecture, Second Life and SOS groups had significantly higher knowledge, observation and self-report scores compared with the control group. For the intervention group (i.e. the lecture, Second Life, and SOS groups after the intervention; Table D.22), the SOS group demonstrated significantly higher knowledge ($p<0.001$), observation ($p=0.008$) and self-report ($p<0.001$) scores than the Second Life group. The SOS group also displayed higher observation ($p<0.001$) and self-report ($p<0.001$) scores than the lecture groups. The lecture group had a significantly higher knowledge score (median 42 versus 39; $p<0.001$) than the Second Life group.

Table E.24 Comparison of the intervention groups for the outcomes measures

Patel et al 2012 RCT, Level II	Median	*p*-value~
SOS versus Second Life		
Knowledge multiple choice questionnaire	43 versus 39	0.001
Observation scale	132 versus 115	0.008
Self-report scale	142 versus 118	0.001
Inference	SOS better than Second Life for knowledge, observation and self-report	
SOS versus lecture		
Knowledge multiple choice questionnaire	43 versus 42	0.784 (NS)
Observation scale	132 versus 107	<0.001
Self-report scale	142 versus 121	<0.001
Inference	SOS better than lecture for observation and self-report	
Second Life versus lecture		
Knowledge multiple choice questionnaire	39 versus 42	<0.001
Observation scale	115 versus 107	0.098 (NS)
Self-report scale	118 versus 121	0.683 (NS)
Inference	Lecture better than Second Life for knowledge	

~The *p*-values in bold indicate statistical significance ($p<0.05$).
NS: not signficant; SOS: Simulated Operating Suite

Further analysis of the observation and self-report scores was performed for knowledge, skills and attitudes.

Observation scale: knowledge, skills, and attitudes

There were no significant differences among the groups for knowledge, skills, and attitudes in the behavioural observation assessment before intervention. The lecture, Second Life and SOS groups all demonstrated significant improvements in knowledge, skills and attitudes. The control group displayed a significant improvement in the skills assessment only ($p=0.013$).

After the intervention, the lecture, Second Life and SOS groups all achieved significantly higher knowledge, skills and attitudes scores than the control group ($p<0.001$) (Table E.25). The SOS group demonstrated higher knowledge (median 39.5 versus 34; $p=0.033$), skills (median 44.5 versus 38; $p=0.001$) and attitude (median 45 versus 40; $p=0.015$) scores than the Second Life group. The SOS group also displayed higher knowledge (median 39.5 versus 35; $p=0.026$) and skills (median 44.5 versus 31; $p<0.001$) scores than the lecture group. The lecture group had higher attitude scores (median 43 versus 40; $p=0.009$) than the Second Life group, with the Second Life group displaying higher skill scores (median 38 versus 31; $p=0.001$) than the lecture group.

Self-report scale: knowledge, skills, and attitudes

There were significant differences among the groups in the self-report assessment before the intervention for knowledge ($p=0.26$) and attitude ($p=0.05$) but not for skills ($p=0.217$) (Table E.26). These were identified using the Kruskal-Wallis test for multiple group comparisons. For the knowledge section of the self-report scale, differences were identified between the control and the lecture group (median 30 versus 21; P = 0.026) and between the control and Second Life groups (median 30 versus 23; $p=0.041$). For the attitude section of the self-report scale, differences were between the control and the lecture group (median 38 versus 28) and between the control and Second Life groups (median 38 versus 31).

There were similar improvements in the knowledge, skills and attitude scores for the self-report scale for the groups as displayed in the observation score before and after the intervention (Table E.25). Because of the significant differences in knowledge and attitudes before intervention, both parameters were excluded from post-intervention analysis.

After the intervention, the lecture, Second Life and SOS groups reported significantly higher skills scores than the control group. The SOS group reported significantly higher skill scores than the Second Life (median 38 versus 33; $p=<0.001$) and lecture groups (median 38 versus 32; $p=0.001$).

Table E.25 Differences before and after the intervention in knowledge, attitudes, and skills on the observation scale

Patel et al 2012 RCT, Level II	Pretest – median (range)	Post-test - median (range)	*p*-value~
Second Life operating theatre training (n=15)			
Knowledge	21 (16–31)	34 (28–41)	0.001
Attitude	25 (18–36)	40 (32–52)	0.001
Skills	19 (11–26)	38 (30–45)	0.001
Imperial College SOS training (n=15)			
Knowledge	27 (15–35)	39.5 (27–45)	0.001
Attitude	28 (22–37)	45 (33–54)	0.001
Skills	18 (13–28)	44.5 (35–49)	0.001
Didactic lecture (n=15) (no simulation training)			
Knowledge	24 (14–31)	35 (25–42)	0.001
Attitude	34 (16–39)	43 (27–52)	0.001
Skills	18 (15–24)	31 (27–42)	0.001
Control (n=15) (no training)			
Knowledge	22 (13–30)	25 (15–32)	0.344 (NS)
Attitude	28 (10–22)	28 (17–23)	0.484 (NS)
Skills	16 (16–33)	20 (21–35)	0.013

~The *p*-values in bold indicate statistical significance ($p<0.05$).
NS: not signficant; SOS: Simulated Operating Suite

Table E.26 Differences before and after the intervention in knowledge, attitudes, and skills on the self-report scale

Patel et al 2012 RCT, Level II	Pretest – median (range)	Post-test - median (range)	*p*-value~
Second Life operating theatre training (n=15)			
Knowledge	23 (11–34)	42 (38–54)	0.001
Attitude	31 (21–48)	47 (43–64)	0.001
Skills	17 (8–32)	33 (22–42)	0.001
Imperial College SOS training (n=15)			
Knowledge	25 (18–42)	50 (39–56)	0.001
Attitude	33 (15–52)	52 (43–66)	0.001
Skills	22 (12–34)	38 (34–43)	0.001
Didactic lecture (n=15) (no simulation training)			
Knowledge	21 (13–43)	39 (30–50)	0.001
Attitude	28 (13–53)	49 (35–58)	0.002
Skills	18 (10–35)	32 (21–41)	0.002
Control (n=15) (no training)			
Knowledge	30 (17–46)	35 (20–52)	0.069 (NS)
Attitude	38 (14–64)	38 (24–64)	0.975 (NS)
Skills	24 (9–19)	23 (16–40)	0.028

~The *p*-values in bold indicate statistical significance ($p<0.05$).
NS: not signficant; SOS: Simulated Operating Suite

Comparative studies

Table E.27 Performance of phacoemulsification by Eyesi VR simulator- trained and non simulator-trained participants

Belyea et al 2011 Retrospective comparative study, Level III-3	Simulator training (n=17)	No simulator training (n=25)	*p*-value~
Parameter	Mean (range)	Mean (range)	
Phacoemulsification time (minutes)	1.88 (0.11–7.20)	2.41 (0.04–8.33)	0.002
Phacoemulsification power (%)	25.32 (2.2–50.0)	28.19 (8.0–70.0)	0.0001
Adjusted phase time (minutes)	47.58 (0.24–280.80)	71.85 (0.32–583.10)	0.0001
Complication rate (%)	0.04	0.06	0.443
Complication grade	2.33	2.47	0.701

~*p*-values in bold indicate statistical significance ($p<0.05$).

The simulator-trained group had a significant lower mean phacoemulsification time ($p=0.002$), adjusted phacoemulsification time ($p=0.0001$), and percentage phacoemulsification power ($p=0.0001$). There was no statistically significant between-group difference in mean complication rate ($p=0.443$) or mean complication grade ($p=0.701$). Eight (47%) of the 17 participants in the simulator-trained group and 12 (48%) of the 25 participants in the control group had no complications. In the regression analysis results, there were significant differences in the slope values (M) between the two groups for each of the three phacoemulsification parameters.

One hundred and twenty-six cases (44.1%) in the simulator-trained group and 34 cases (43.8%) in the control group were performed in the first half of the year. There was no significant difference in phacoemulsification time, phacoemulsification power, adjusted phacoemulsification time or complication grade between the cases performed in the first half of the year and those performed in the second half of the year in the simulator-trained or in the control group. Overall in both groups there was an increase in complication rate between cases performed in the second half of the year and cases performed in the first half ($p=0.0001$). When cases in the simulator-trained group were isolated, however, cases performed in the second half of the year had a significantly lower complication rate than those performed in the first half of the year ($p=0.0001$). Linear regression showed weak positive correlations (mean R=0.189) of no statistical significance between the number of cases per participant and all outcome parameters (phacoemulsification time, phacoemulsification power, adjusted phacoemulsification time, complication rate and complication grade) in both groups.

The authors performed a retrospective review of 592 consecutive third-year resident cataract surgeries which were performed with the same attending surgeon using the same technique and instrumentation. Surgeries performed by 42 third year George Washington University ophthalmology residents (22 men, 20 women) were reviewed. Within that group, 17 residents (8 men, 9 women) received VR training using a simulator and 25 residents (14 men, 11 women) were not exposed to the simulator during residency. All residents had performed a mean of 16 phacoemulsification cases (range 12–20 cases) before the start of third year. The mean number of cases per resident in the simulator- and non simulator-trained groups was 16.8 (range 5–45) and 12.2 (range 4–28), respectively.

Simulation training versus interactive seminar-based education

Table E.28 Performance of cardiopulmonary bypass weaning between SimMan Universal simulator-trained and interactive seminar- trained participants

Bruppacher et al 2010 RCT, Level II	SimMan simulator training (n=10)	Interactive seminar-based training (n=10)	*p*-value~
	Mean ± SD	Mean ± SD	
Pretest: ANTS assessment score (minimum score was 4 and the maximum 16)	10.6 ± 0.46	10.0 ± 0.46	0.331
Post-test (2 weeks): ANTS Score	14.3 ± 0.41	11.8 ± 0.41	<0.001
Retention (5 weeks) test: ANTS Score	14.1 ± 0.41	11.7 ± 0.41	<0.001
Pre-test: checklist performance score, %	62.6 ± 5.3	58.3 ± 5.3	0.571
Post-test: checklist performance score, %	89.9 ± 3.0	75.4 ± 3.0	0.003
Retention test: checklist performance score, %	93.2 ± 2.4	77.0 ± 2.4	<0.001

~*p*- values in bold indicate statistical significance (*p*<0.05).
ANTS: Anaesthetists' Non Technical Skills; SD: standard deviation.

A previously validated global assessment tool, Anaesthetists' Non Technical Skills (ANTS) global rating scale was used to assess participants' cognitive and behavioural performance in cardiopulmonary bypass (CPB) weaning. A 20-point checklist for CPB weaning was developed specifically for the purpose of this study to assess the technical skills of participants.

For the specific checklist assessment, the simulator-trained group scored significantly higher than the seminar-trained group during both post-test (89.9 ± 3.0% versus 75.4 ± 3%; t(18) = -3.365, *p*=0.003) and the retention test (93.2 ± 2.4% versus 77.0 ± 2.4%; t(18) = -4.836, *p*<0.001) phases (Table D.28). For the ANTS assessment, both groups improved from the pre-test to post-test to retention tests as indicated by the magnitude of the means. The significant group by time interaction effect suggests that the trend for improvement differs in magnitude between the simulation-based training group and the seminar group (F(2,36) = 11.9, mean standard error [MSE]=0.47, *p*<0.001). Overall, the simulation group scored higher than the seminar group as indicated by the main effect of group (F(1,18) = 11.5, MSE=4.55, *p*=0.003). Although both groups improved, the simulator-trained group showed significantly higher improvements compared with the seminar-trained group, as indicated by significant group by time interaction (F(2,36)=11.9, MSE=0.47, *p*<0.001). Post hoc *t* tests revealed similar pre-test ANTS score for the two groups (simulator-trained group 10.6 ± 0.46 versus seminar-trained group 10.0 ± 0.46; t(18) = -0.99, *p*=0.331).

After training, the simulator-trained group significantly outperformed the seminar-trained group in both the post-test (14.3 ± 0.41 versus11.8 ± 0.41; t(18) = -4.280, *p*<0.001) and the retention test (14.1 ± 0.41 versus11.7± 0.41; t(18) = -4.249, *p*<0.001) phases (Table 28). All components of the ANTS, such as situation awareness, team working, decision-making and task management were significantly different between the two groups at the post-test and the retention tests (Table D.29).

Table E.29 ANTS performance of cardiopulmonary bypass weaning between simulator-trained and interactive seminar-trained participants

Bruppacher et al 2010 RCT, Level II	Task management	Team working	Situation awareness	Decision-making
Pre-test	All values are expressed as mean ± standard error.			
Seminar group	2.50 ± 0.13	2.6 ± 0.16	2.45 ± 0.11	2.40 ± 0.12
Simulation group	2.65 ± 0.13	2.6 ± 0.16	2.80 ± 0.11	2.55 ± 0.12
Post-test				
Seminar group	2.85 ± 0.14	2.95 ± 0.12	3.05 ± 0.11	2.95 ± 0.11
Simulation group	3.60 ± 0.14*	3.45 ± 0.12*	3.75 ± 0.11*	3.50 ± 0.11*
Retention test				
Seminar group	2.90 ± 0.11	2.85 ± 0.17	2.90 ± 0.10	3.00 ± 0.11
Simulation group	3.55 ± 0.11*	3.35 ± 0.17	3.55 ± 0.10*	3.65 ± 0.11*

**p*<0.01 compared with seminar group (same time, same Anaesthetists' Nontechnical Skills [ANTS] subscale).

ANTS Global Rating Scale

ANTS were analysed according to the global rating score (the sum of the four rating category subscores). Therefore the maximum score was 16 and the minimum was 4. The ANTS global rating scale consists of four categories: task management, team working, situation awareness and decision-making. The scale was defined as follows:
4 points: good (performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others)
3 points: acceptable (performance was of a satisfactory standard but could be improved)
2 points: marginal (performance indicated cause for concern, considerable improvement is needed)
1 point: poor (performance endangered or potentially endangered patient safety, serious remediation is required).

Checklist for CPB weaning

A 20-point checklist for CPB weaning was developed specifically for the purpose of this study to assess the technical skills of participants. Using a Delphi method, including one cardiac surgeon and four cardiac staff anaesthesiologist from the department, tasks included on the checklist required 80% agreement through an iterative process. To analyse the checklist, tasks performed independently were scored 2 points, task performed after prompting were scored 1 point, and not done tasks received 0 points. For the checklist data, the Analysis of Variance (ANOVA) test revealed a main effect of time ($F[2.36]=33.38$, $MSE=0.011$, $p<0.001$) and group ($F[1.18]=9.94$, $MSE=0.020$, $p=0.005$). Similar patterns of improvements were indicated by a nonsignificant time by group interaction ($F[2.36]=1.86$, $MSE=0.011$, $p=0.170$). Post hoc independent samples *t* test revealed equivalent pre-test checklist scores between the two groups (simulator-trained group 62.6 ± 5.3% versus seminar-trained group 58.3 ± 5.3%; $t(18) = -0.578$, $p=0.571$). Within two weeks (post-test) and five weeks (retention test) of completing training, participants were asked to wean a patient from CPB in a real-life clinical setting. Participants were not exposed to any cardiac anaesthesia in the time period between training and the two tests (post-test and retention tests). The protocol for the post-test and retention test phases was identical to that of the pre-test phase.

No significant differences were observed between the two groups in terms of any of the participant or patient variables. ANTS and checklist scores at pre-test did not correlate with participants' age, training level, previous simulation sessions or previous clinical experience in CPB weaning. The interim analysis after 20 participants had completed their training revealed highly significant differences in performances of the two groups in CPB weaning of real patients. After weighing the benefits of continuing the study, authors decided to stop recruitment after a total of 20 trainees and 60 patients.

Simulation training versus patient-based training

Table E.30 Results of patient-based assessments for Olympus colonoscopy simulator-trained and patient-trained participants

Haycock et al 2010 RCT, Level II	Simulator training (n=18)	Patient-based training controls (n=18)	*p*-value
Completion of case): number (%) Each participant performed three colonoscopies (Total=54 per group).	6 (11%)	4 (7%)	0.51
Maximum tip position: number (%)			
Sigmoid	29 (54%)	28 (52%)	0.73
Descending	8 (15%)	12 (22%)	
Transverse	11 (20%)	8 (15%)	
Ascending	0 (0%)	2 (4%)	
Caecum	6 (11%)	4 (7%)	
Time taken (minutes)*	20 (19–20)	20 (20–20)	0.11
Straight insertion depth (cm): mean (standard deviation)	48 (23)	52 (21)	0.35
JAG DOPS*^	16 (14–22)	18 (14–22)	0.92
Global Score*#	16 (14–19)	17 (14–19)	0.35

* Time taken, DOPS score and Global Score as median (interquartile range).
^ JAG DOPS, UK Joint Advisory Group (JAG) on Gastrointestinal Endoscopy Direct Observation of Procedural Skills (DOPS).
#Global Score, an expert global rating of performance adapted for colonoscopy.

There was no significant differences between the simulator-trained and patient-trained groups in terms of case completion, maximum tip position achieved, time taken, straight insertion depth, UK Joint Advisory Group on Gastrointestinal Endoscopy Direct Observation of Procedural Skills (JAG DOPS) score or Global Score (Table E.30).

Table E.31 Results of post-training simulator assessments (three per participant)

Haycock et al 2010 RCT, Level II	Simulator training (n=19); median (IQR) or number (%)	Patient training (n=18); median (IQR) or number (%)	*p*-value¯
General:			
Intubated to caecum	54 (95%)	38 (70%)	0.001
Maximum tip position	40 (40–40)	40 (35–40)	<0.001
Time taken (seconds)	407 (327–504)	743 (504–1200)	<0.001
Technical:			
Patient pain - maximum	0.24 (0.05–0.43)	0.45 (0.19–0.68)	0.002
Insertion length with embedded tip	0.03 (0.02–0.08)	0.07 (0.05–0.12)	<0.001
Insertion length with obscured lens	0.001 (0.00–0.005)	0.03 (0.00–0.14)	<0.001
Insertion force−maximum (units NS)	11.8 (1.6)	13.1 (2.8)	0.003
Correct use of abdominal pressure	45 (79%)	28 (52%)	0.003
Correct use of variable stiffness	20 (35%)	21 (39%)	0.70
Excessive inflation	7 (12%)	13 (24%)	0.14
Looping:			
Number of sigmoid loops during sigmoid intubation	1 (1–1)	1 (1–1)	0.97
Time with sigmoid loop during sigmoid intubation (seconds)	34 (9–61)	68 (31–104)	0.002
Number of sigmoid loops after sigmoid intubation	1 (1–1)	1 (1–2)	<0.001
Time with sigmoid loops after sigmoid intubations (seconds)	33 (2–68)	113 (45–240)	<0.001
Number of transverse loops	1 (1–1)	1 (1–1)	0.98
Time with transverse loop (seconds)	46 (6–81)	6 (17–168)	0.12
Time to resolve alpha loop (seconds)	176 (106–224)	210 (158–383)	0.11

¯*p* -values in bold indicate statistical significance ($p<0.05$).
IQR: interquartile range; NS: not stated.

For simulator post-test (Table E.31), the participants were significantly more likely to complete intubation to caecum (95% versus 70%; $p=0.001$) and took approximately half as long to do so (407 versus 743 seconds; $p<0.001$). They demonstrated superior technical skill in terms of reduced maximum patient pain scores ($p=0.001$) and shorter distances pushing with either an embedded tip ($p<0.001$) or obscured lens ($p<0.001$). They were more likely to use abdominal pressure correctly to assist intubation (79% versus 52%; $p=0.003$), but there were no differences in the correct use of variable stiffness. The simulator-trained participants straightened sigmoid loops more quickly than controls (34 versus 68 seconds; $p=0.002$) and kept the colonoscope straighter once the sigmoid colon had been passed ($p<0.001$). There were no differences in the management of transverse or alpha loops.

Table E.32 Results of participants' feedback questionnaire on training experience

Haycock *et al.* 2010 RCT, Level II	Simulator training (n=18)	Patient-based training (n=18)	*p*-value
Overall, how good was your training?	8.0 (7.0–8.5)	8.0 (6.0–9.75)	0.93
Overall, how useful was your training?	8.0 (7.0–9.0)	8.5 (8.0–10.0)	0.13
Overall, how enjoyable was your training?	8.0 (7.0–10.0)	9.0 (8.3–10.0)	0.18
How well do you think your training prepared you for the assessments?	6.0 (6.0–8.0)	7.0 (5.3–9.0)	0.40
How difficult did you find the simulator assessments?	5.0 (5.0–6.5)	6.0 (3.3–8.0)	0.48
How stressful did you find the simulator assessments?	4.0 (3.0–7.0)	5.0 (4.3–7.0)	0.25
How difficult did you find the patient assessments?	8.0 (5.5–9.0)	8.0 (7.3–8.0)	0.96
How stressful did you find the patient assessments?	6.0 (5.0–8.0)	7.0 (5.3–8.5)	0.28

Note: Answers are on a 10-point visual analogue scale and expressed as median (interquartile range).

Participants from both the simulator-trained and patient-trained groups rated their training experience highly, with a median score of 8.0 out of 10 for both groups. Simulator-trained participants rated their training on the simulator as useful (8.0/10) and as enjoyable (8/0/10) as the control group rated their real patient-based training. Both groups felt reasonably well-prepared for their assessments *(p*=0.40), although they found patient assessments to be difficult (8.0/10).

Baseline characteristics and performance

The subjects and controls were reasonably well-matched in their demographics and previous endoscopy experience, with no participant having any practical colonoscopy experience. Although there were more specialists in training in the control group than in the simulator-trained group, this was not statistically significant. There were some differences in the performance metrics on the simulator pertaining assessment between groups, with the control group performing significantly better than the simulator-trained group on several measures. All participants completed an initial assessment on the simulator (three cases).

Assessment of proficiency on real colonoscopy was measured using previously validated structure assessment tools: the JAG DOPS assessment form and the Global Score, an expert global rating of performance adapted for colonoscopy.

The simulator-trained participants completed 16 hours of a simulator-based training program while the control group completed 16 hours of a patient-based training program. All participants then completed three simulator cases and three live colonoscopies as final assessments and a feedback questionnaire. Primary outcome measures were the expert assessor's score on the JAG DOPS and the Global Score. Secondary outcomes measures were the time to completion, depth of insertion, and improvement in performance parameters at simulated colonoscopy.

Table E.33 Performance of camera navigation by LAP Mentor and PROMIS simulator-trained and OR-trained participants

Franzeck et al 2012* RCT, Level II	Simulator training (n=12)	OR training (n=12)	*p*-value	SMIC
Post-training camera test:				
Organ visualisation (mean points ± SD)	25.8 ± 3.4	26.8 ± 1.9	0.45	NR
Horizon alignment (mean points ± SD)	21.8 ± 3.7	24.1 ± 1.9	0.08	NR
Time to completion (seconds) (mean ± SD)	133 ± 35	111 ± 30	0.12	NR
Correct scope rotation handling (number of participants (%))	8 (66)	10 (83)	0.60	NR
Pre-training camera test:				
Organ visualization (mean points ± SD)	22.5 ± 5.00	25 ± 2.8	0.132	0.65
Horizon alignment (mean points ± SD	20.1 ± 4	22.7 ± 3.9	0.127	0.66
Time to completion (seconds) (mean ± SD)	179 ± 64	163 ± 67	0.554	NR
Correct scope rotation handling (number of participants (%))	7 (58)	7 (58)	1	NR

*electronic publication available in 2011
OR: operating room; NR: not reported; SD: standard deviation; SMIC: single measure intraclass correlation.

There was no significant difference between the simulator-trained group and the OR-trained group in any of the parameters in the post-training evaluation (Table D.33). Participants in both groups spent equal time actually training on camera navigation (mean 272 ± 28 min versus 217 ± 138 min, p=0.20) (Table D.34). However, participants in the OR-trained group spent significantly more overall time in the OR (1002 ± 40 minutes) than the simulator-trained group spent in the skills laboratory (307 ± 27 minutes; p<0.01). Thus, the authors concluded that traditional training in the OR achieved its goal of improved and correct camera navigation but was not as time-efficient as simulator-based training.

Table E.34 Comparison of baseline characteristics and training time between simulation-trained and OR-trained participants

Franzeck et al 2012* RCT, Level II	Simulator training (n=12)	OR training (n=12)	*p*-value˜
Mean age (years)	26.2 ± 1.9	25.8 ± 1.1	0.610
Gender (male:female)	3:9	4:8	0.9
Right-handedness (n (%))	11 (92)	12 (100)	0.9
Visuospatial test (mean points ± SD)	13.3 ± 3.6	15.2 ± 3.7	0.223
Expenditure of time:			
Time spent in skills laboratory/OR (minutes) (mean ± SD)	307 ± 27	1002 ± 140	<0.01
Actual camera training time (minutes) (mean ± SD)	272 ± 28	217 ± 138	0.20
Actual camera training time (% of total time spent in OR/skills laboratory)	88%	22%	<0.01

OR: operating room; SD: standard deviation.
˜The *p*-values in bold indicate statistical significance (p<0.05).
*electronic publication available in 2011

Baseline demographics and pre-training test results

There was no difference in baseline characteristics of the 24 participants regarding age, gender, right-handedness or visuospatial test (Table D.34). The pre-training camera assessment tests did not reveal any group difference (Table D.33). The 14 experts performed significantly better in the camera assessment than the 24 participants, demonstrating that the test can distinguish between novices and experts (construct validity: organ visualisation score (30.9 ± 1.2 versus 23.7 ± 4.2, p<0.001), horizon alignment score (29.2 ± 1.5 versus 21.4 ± 4.1, p<0.001), time to completion (69 ± 12 versus 171 ± 65, p<0.001), and percentage of correct camera rotation (93% versus 100%, p<0.001). Inter-rater reliability, single measure ICC among the five independent experts grading all camera tests were 0.68 for organ visualisation and 0.66 for horizon alignment.

The differences in the pre-training and post-training camera test results are given in Table D.35. Both groups showed significant progress in the organ visualisation score and a significant decrease in mean time taken to complete the tests. Improvements in horizon alignment and scope rotation handling were not significant. There was no significant progress difference for any of the parameters (see Table D.33).

Table E.35 Difference in pre-training and post-training camera navigation test results for simulation-trained and OR-trained participants

Franzeck et al 2012* RCT, Level II	Pre-training	Post-training	*p*-value~
Simulator training (n=12)			
Organ visualisation (mean points ± SD)	22.4 ± 5.00	25.8 ± 3	**0.04**
Horizon alignment (mean points ± SD	20.1 ± 4	21.8 ± 4	0.20
Time to completion (seconds) (mean ± SD)	179 ± 64	133 ± 35	**0.05**
Correct scope rotation handling (number of participants (%))	7 (58)	8 (66)	0.9
OR training (n=12)			
Organ visualisation (mean points ± SD)	25 ± 2.8	26.7 ± 1.9	**0.03**
Horizon alignment (mean points ± SD)	22.7 ± 3.9	24.1 ± 1.9	0.09
Time to completion (seconds) (mean ± SD)	163 ± 67	111 ± 30	**0.02**
Correct scope rotation handling (number of participants (%))	7 (58)	10 (83)	0.4

~*p*-values in bold indicate statistical significance ($p<0.05$).
*electronic publication available in 2011
OR: operating room; SD: standard deviation.

Appendix F: Summary of simulators, didactic lecture-based, interactive seminar-based and patient-based training

Study	Simulator overview
Ahlberg et al 2007 Cosman et al 2007 Hogle et al 2009	The LapSim® virtual reality laparoscopic surgical simulator (Surgical Science, Gothenburg, Sweden) is a real-time interactive computer simulator that replicates both diagnostic and therapeutic procedures (Schijven & Jakimowicz 2003). Haptic feedback is optional. Utilising advanced 3D technology, a combination of computer hardware and software modules recreate the procedures and environment of abdominal keyhole surgery. Practice sessions can vary in graphic complexity as well as in the level of difficulty. Alberg et al (2007) used version 2.0 of the LapSim system basic skills package without force feedback; Cosman et al (2007) used version 1.5 of the basic LapSim package without haptic feedback and with a moderate level of difficulty; Hogle et al 2009 used the LapSim system with basic skills package (no version reported).
Beyer et al 2011 Franzeck et al 2012*	The LAP Mentor™ virtual reality laparoscopic surgical simulator (Simbionix, Cleveland, Ohio, USA) is a high-fidelity, virtual reality, real-time interactive computer simulator with haptic feedback that replicates both diagnostic and therapeutic procedures (Ayodeji et al 2007; Zhang et al 2008).There appears to be some controversy about the construct validity of LAP Mentor (Andreatta et al 2008).
Franzeck et al 2012*	The PROMIS™ surgical hybrid simulator (Haptica Ltd., Dublin, Ireland) is an augmented reality (AR) simulator that combines physical reality (such as a box trainer) and virtual reality into one system (van Empel et al 2012). Haptic feedback is maintained, using original laparoscopic instruments and tactile tasks. AR devices are equipped with modules that simulate a laparoscopic environment and allow performance of tasks related to the box trainer tasks within the construct of the simulator. The PROMIS™ measures movements, with three separate cameras, of marked instruments by a passive vision-tracking system situated in a large mannequin. Construct and face validation for the PROMIS™ was demonstrated by Van Sickle et al 2005.
Franzeck et al 2012*	Two XiTact™ IHPs (instrument haptic ports) as interfaces for the laparoscopic instruments and a third unidirectional electromechanical interface the Xitact ITP instrument tracking port for camera navigation (Mentice AB, Gothenburg, Sweden) were used with the LAP Mentor™ surgical simulator (Simbionix USA, Cleveland, Ohio, USA software). A second simulator, a haptic PROMIS™ surgical hybrid simulator (Haptica Ltd, Dublin, Ireland) was also used.
Larsen et al 2009	The LapSim® Gyn virtual reality laparoscopic surgical simulator (Surgical Science, Gothenburg, Sweden) is a real-time interactive computer simulator that replicates both diagnostic and therapeutic gynaecological procedures (Larsen et al 2006; Aggarwal et al 2006; Schrueder et al 2009). A combination of computer hardware and software modules recreates the procedures and environment of abdominal keyhole surgery including tubal occlusion, salpingectomy, tubotomy and myoma suturing.
Van Sickle et al 2008	The MIST-VR™ (Minimally Invasive Surgical Trainer-Virtual Reality) laparoscopic simulator (Mentice AB, Gothenburg, Sweden) is a low-fidelity virtual reality laparoscopic simulator with combined metrics systems to provide feedback to novice learners during practice. The system teaches basic dexterity skills as the student uses real laparoscopic handles that activate virtual instrument tips within the computer. The acquisition of skills on this device relates to generalisation and automation of skills in laparoscopic navigation (Ahlberg et al 2005; Gallagher et al 2005).
Belyea et al 2011	The Eyesi® virtual reality ophthalmosurgical simulator (VRMagic Holding AG, Mannheim, Germany) provides a simulation of the curvilinear capsulorhexis technique performed during cataract extraction (Webster et al 2005; Solverson et al 2009).
Bruppacher et al 2010	The SimMan® Universal simulator is computerised patient simulator (Laerdal Medical Corporation, Wappingers Falls, NY, USA). It is a portable, instructor-driven, full body, electromechanical mannequin tethered by two small cords to a control system, which includes a laptop computer, a signal generator and an air compressor. There is also a monitor that displays vital signs. The computer is programmed to make the mannequin's vital signs, peripheral pulses, lung sounds and cardiac rhythm respond to both the medical problem programmed and the learner's response to the problem. For the more advanced user, many problems can be simulated, including swelling in the hypopharynx or tongue, decreased cervical range of motion and trismus. Depending upon the learner's action this simulated patient can get better, stay the same, deteriorate or even die. The computer logs when these signs are checked by trainees, as well as when they begin and end chest compressions and ventilation, or when they administer a shock. Instructor-entered information that the mannequin cannot sense, such as a trainee calling for additional help, is also logged. The log is used for the debriefing and it can be printed out. Hesselfeldt et al (2005) note airway differences to real patients.

Study	Simulator overview
Beyer et al 2011 Sroka et al 2010	The Fundamentals of Laparoscopic Surgery program: In 1997, the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) developed an educational program entitled 'The Fundamentals of Laparoscopic Surgery' (FLS). FLS teaches the fundamental knowledge, judgment and technical skills specific to laparoscopic surgery. FLS is comprised of a cognitive knowledge component, as well as a skills-based portion pertinent to laparoscopic surgery. The FLS program uses the FLS laparoscopic box trainer. FLS is CME accredited, and in 2008, the American Board of Surgery (ABS) required FLS as a prerequisite of the ABS Certifying Examination. Examinations in FLS are conducted by trained examiners using standardised criteria. The FLS program is based on the laparoscopic box trainer and combines five training modules including peg transfer, pattern cutting, ligating loop, and intracorporeal and extracorporeal knot-tying (van Empel et al 2012). The McGill Inanimate System of Training and Evaluation of Laparoscopic Skills (MISTELS) was used with FLS to assess technical skill. MISTELS was designed to objectively assess basic laparoscopic skills through a series of structured tasks performed under video guidance in a box trainer (Fried 2004). All tasks are scored according to pre-established standards using time and error measurements (Fraser et al 2003).
Banks et al 2007	The Limbs and Things laparoscopic simulator (Limbs and Things, Bristol, UK) is a box trainer that provides haptic feedback and allows the use of actual instruments with a simulated laparoscopic camera and video screen.
Howells et al 2008	The Sawbones arthroscopy knee benchtop simulator (Sawbones Europe, Malmö, Sweden) is a synthetic inanimate model that provides haptic feedback and allows the use of actual instruments such as a standard 30° arthroscope with an arthroscopic camera and display (Smith & Nephew Endoscopy, Huntingdon, UK).
Ossowski et al 2008	The nasal model (University of Pittsburgh Medical School, Pittsburgh, Pennsylvania, USA) is a life-sized model head with human texture and accurate nasal anatomy.
Palter et al 2011	The abdominal wall simulator (University of Toronto, Toronto, Canada) is a low-fidelity bench-top synthetic abdominal wall model that allows simulation of abdominal fascial closure (Matsumoto et al 2002; Grober et al 2004).
Zendejas et al 2011	The Guildford MATTU TEP hernia repair simulator (Guildford Minimal Access Therapy Training Unit (MATTU) and Limbs and Things Ltd, Bristol, UK) is constructed from rubber and depicts a human pelvis. Ports are available for placement of laparoscopes and trocars that permit residents to practise mesh fixations over indirect, direct, or femoral defects (Slater et al 2001; Valentine and Rege 2004). The totally extraperitoneal (TEP) technique is a well-established method for repairing inguinal hernias laparoscopically.
Haycock et al 2010	The Olympus (ENDO TS-1) colonoscopy simulator (Olympus KeyMed, Southend, UK) is a real-time interactive computer simulator that replicates both diagnostic and therapeutic gastrointestinal endoscopic procedures. All procedures are recorded automatically by the simulator and evaluated using computer-generated parameters with demonstrated construct validity and selected to have clinical importance by investigators (Koch et al 2008; Haycock et al 2009).
Yi et al 2008	The KAIST-Ewha Colonoscopy Simulator II (Korea Advanced Institute of Science and Technology-Ewha Womans University, Seoul, South Korea) is a virtual reality endoscopy simulator. It incorporates force feedback for realistic haptic feedback and measurement of performance data (Woo et al 2008).
Fried et al 2010	The ES3® endoscopic sinus surgery virtual reality simulator (Lockheed Martin Inc., Akron, Ohio, USA) contains a virtual patient that is responsible for the simulation of the endoscopic image, the surgical interface and the user interface. The system also contains a haptic system, allows voice recognition, and provides virtual instruction while training (Edmond 2002). The virtual instructor points out mistakes, errors and misses (Fried et al. 2010). The system records overall and task-specific scores.
Park et al 2007	The AccuTouch® virtual reality endoscopic simulator system (CAE Heathcare, Montreal, Quebec, Canada, previously Immersion Medical, Gaithersburg, Maryland, USA or Immersion Corporation, San Jose, California, USA) is a virtual reality endoscopy simulator and includes training for flexible sigmoidoscopy, colonoscopy and bronchoscopy. It incorporates a mannequin, force feedback and measurement of performance data. Both diagnostic and therapeutic scenarios are provided, and a number of aids are available (Dunkin et al 2007). The AccuTouch® device also simulates patient vital signs and responses to sedation and to pain (Desilets 2011).
Shirai et al 2008	The GI Mentor™ II virtual reality endoscopic simulator (Simbionix™, Cleveland, Ohio, USA) is a real-time interactive computer simulator that replicates both diagnostic and therapeutic of gastrointestinal endoscopic procedures (Bar-Meir 2000). It includes a life-sized plastic head and torso with apertures for upper and lower endoscopy (Ferlitsch et al 2002; Valentine and Rege 2004). Real-time three-dimensional pictures are generated as an endoscope is passed through the torso body (Valentine and Rege 2004). Program software generates force feedback to simulate resistance from touching bowel wall as the endoscope is passed (Ferlitsch et al 2002). A monitor depicts typical findings seen at endoscopy as well as adverse events that must be treated (Valentine and Rege 2004). The simulated procedures look and feel similar to the actual procedures and train tasks that will directly transfer to the performed procedures (Gallagher et al 2005). The simulator uses a 'fading' training strategy where major guides and clues are given at the beginning of training, and are gradually faded out until the trainee can perform the task without support (Gallagher et al 2005).

Study	Simulator overview
Kälström et al 2010	The PelvicVision® virtual reality simulator (Melerit Medical AB, Linköping, Sweden) is a real-time interactive computer simulator that replicates diagnostic and therapeutic procedures. Kälström et al (2005; 2010a) previously described and validated the simulator in regards to face, content and construct validity. It consists of a standard personal computer, a modified resectoscope connected to a haptic robot (Phantom, SenseAble), a 'pelvic floor', and a pair of 'legs'. The software makes it possible to create different patient cases with respect to anatomy and physiology.
Schout et al 2009	The URO Mentor™ virtual reality endourologic simulator (UM, Simbionix, Cleveland, Ohio, USA) is a real-time interactive computer simulator that replicates both diagnostic and therapeutic endourologic procedures (Dolmans et al 2009; Schout et al 2010).
Patel et al 2012*	The Second Life® (Linden Research Inc. San Francisco, CA, USA) virtual reality operating room session occurred within a computer laboratory at Imperial College London and was accessed through a desktop computer. Following the orientation (30-minute introductory session) and subjects' familiarising themselves with their avatars, the subjects participated in training in the virtual operating room (http://slurl.com/secondlife/medical%20School/178/183/22). This is a three-dimensional representation of St Mary's Hospital operating room within the virtual world of Second Life®. All content delivery was based on information derived from theatre induction curriculum. The session was interactive with subjects receiving addition information from the instructors through either text or voice chat. The Imperial College simulated operating suite (SOS) (Imperial College, London, UK) consisted of a replicated operating theatre with an adjacent control room. Within this replicated theatre there was an operating table, a laparoscopic stack system, diathermy, and suction and trolleys containing equipment such as suture material and dressings (Aggarwal et al 2004).
Study	**Didactic lecture overview**
Patel et al 2012*	The didactic lecture was prepared in PowerPoint 2008 (Microsoft Corporation, Redmond, Washington, USA) and lasted for one hour. It consisted of information deemed essential from the theatre induction curriculum at the Imperial College London. This lecture included instructional videos regarding gowning and gloving, which were obtained via the Imperial College London undergraduate teaching intranet.
Study	**Interactive seminar-based education overview**
Bruppacher et al 2010	Each trainee attended an individual two-hour interactive training seminar. A staff anaesthesiologist, experienced in both cardiac anaesthesiology and resident training, lectured at the seminar. The aim was to provide best-practice teaching. The seminar included audiovisual aids such as PowerPoint slides, handouts, and face-to-face discussion of four paper-based scenarios similar to those described in the simulation training. The learning objectives of the seminar covered the same content domains as the debriefing for the simulation group. Trainees of both groups had the possibility to discuss the syllabus with regard to tasks assessed by the checklist. If no further questions were asked, the trainees were allowed to leave their respective sessions before the full two hours of allocated training.
Study	**Patient-based training overview**
Haycock et al 2010	The patient-based sigmoidoscopic examinations involved an attending physician instructing participants by using his or her own teaching preferences and techniques. Residents were expected to learn how to advance the colonoscope independently by the end of 10 sessions. Participants were trained with a video endoscope.

*electronic publication available in 2011

Appendix G: Shortened forms

3D	three-dimensional
AB	Aktiebolag (literally "share company" or "stock company") is the Swedish term for "limited company" or "corporation". When used in company names, it is abbreviated 'AB' or 'Ab' (roughly equivalent to the abbreviations Ltd or PLC).
AG	Aktiengesellschaft (AG), a German name for a type of company, similar to "Inc." or "LLC (limited liability company)" in the USA, public limited company (plc) in the UK, or S.p.A. in Italy
AccuTouch®	The AccuTouch® simulator system is a virtual reality endoscopy simulator and includes training for flexible sigmoidoscopy, colonoscopy and bronchoscopy (CAE Heathcare Inc., Montreal, Quebec, Canada, previously Immersion Medical, Gaithersburg, Maryland, USA or Immersion Corporation, San Jose, California, USA).
ANOVA	analysis of variance test
ANTS	anaesthetists' (anaesthesiologists') non-technical skills
AR	augmented reality
ASA	American Society of Anaesthesiologists' (ASA) physical status levels I to VI
ASERNIP-S	Australian Safety and Efficacy Register of New Interventional Procedures−Surgical
BLT	bilateral tubal ligation
BMI	body mass index
BRCA1	breast cancer gene 1
CAE	CAE Inc. (Montreal, Quebec, Canada)
CAE Healthcare	CAE Healthcare, Inc. (Montreal, Canada) operates as a subsidiary of CAE Inc.
CD-ROM	compact disc−read-only memory
CI	confidence interval
CPB	cardiopulmonary bypass
CRD	Centre for Reviews and Dissemination, United Kingdom
CUS	cystourethroscopy or cystoscopy
DC	District of Columbia, capital of USA
DOPS	Direct Observation of Procedural Skills
DVD	digital versatile disc

ENDO TS-1	The Olympus (ENDO TS-1) VR colonoscopy simulator (Olympus KeyMed, Southend, UK) is a real-time interactive computer simulator that replicates both diagnostic and therapeutic gastrointestinal endoscopic procedures.
ES	effect size
ES3™	endoscopic sinus surgery simulator (Lockheed Martin, Akron, Ohio, USA)
ESS	endoscopic sinus surgery
Eyesi®	ophthalmic surgical simulator (VR Magic, Manneheim, Germany)
FLS	the Fundamentals of Laparoscopic Surgery (SAGES, USA)
FLS Trainer Box	physical simulator for FLS (VTiMedical™, North Billerca, Massachusetts, USA)
GEE	generalised estimating equations
GI Mentor™	endoscopic medical simulator for the training of gastrointestinal upper and lower endoscopic procedures (Symbionix, Cleveland, Ohio, USA)
GOALS	Global Operative Assessment of Laparoscopic Skills
GRS	global rating scale/score
HTA	Health Technology Assessment programme, United Kingdom
IBM	International Business Machines Corporation and subsidiary companies
ICC	intra-class correlation
IPSS	International Prostate Symptoms Score
IQR	interquartile range
IHP	instrument haptic port
ITT	Intention-to-treat is a method that includes noncompliant patients in the groups to which they were originally randomised.
JAG DOPS	Joint Advisory Group on gastrointestinal endoscopy Direct Observation of Procedure Skills (United Kingdom)
JMP®	statistical discovery software (SAS Institute Inc., USA)
KAIST	Korea Advanced Institute of Science and Technology, Seoul, South Korea
KAIST-Ewha	The KAIST-Ewha Colonoscopy Simulator II (Korea Advanced Institute of Science and Technology-Ewha Womans University, Seoul, South Korea) is a virtual reality endoscopy simulator
LAC	laparoscopy assisted colectomy
LAP	laparoscopy
LAP Mentor™	laparoscopic surgical simulator (Symbionix™, Cleveland, Ohio, USA)
LapSim®	laparoscopic surgical simulator (Surgical Science, Gothenburg, Sweden)
LapSim® Gyn	laparoscopic surgical simulator with software for gynaecological procedures (Surgical Science, Gothenburg, Sweden)

LCD	liquid crystal display or flat panel display
LSC	laparoscopic
Mac™	Mac computer (Apple, USA)
MATTU	the Minimal Access Therapy Unit (Guildford, United Kingdom)
MB	megabyte
MAP	mean arterial pressure
MISTELS	McGill Inanimate System for Training and Evaluation of Laparoscopic Skills
MIST-VR™	Minimally Invasive Surgical Trainer−Virtual reality (Mentice AB, Gothenburg, Sweden)
MSE	mean squared error
NA	not applicable
NHMRC	National Health and Medical Research Council, Australia
NHS	National Health Service, United Kingdom
NM	nasal model (University of Pittsburgh Medical School, Pittsburgh, Pennsylvania, USA)
NR	not reported
NS	not significant
OCAP	Orthopaedic Competence Assessment Project
OGD	oesophagogastroduodenoscopy
OGJ	oesophagogastric junction
Olympus	Olympus Corporation, Japan. Olympus KeyMed, Southend, UK (see ENDO TS-1) is a wholly-owned subsidiary of Olympus Corporation, Japan.
OR	operating room
OR (statistics)	Odds ratio
OSATS	Objective Structured Assessment of Technical Skills
PAP	pulmonary arterial pressure
PelvicVision®	is a VR simulator for training and skills assessment of prostate resection (Melerit Medical AB, Linköping, Sweden)
PGY	postgraduate year
phaco	phacoemulsification
PicSOr	Pictorial Surface Orientation test (Queen's University, Belfast, UK)
PQS	performance quality score
PROMIS™	hybrid laparoscopic simulator (Haptica Ltd, Dublin, Ireland)
Qmax	maximum urinary flow rate
RAM	random-access memory

RCT	randomised controlled trial
ReTURP	repeated TURP (transurethral resection of prostate)
RV	right ventricle
SAGES	Society of American Gastrointestinal and Endoscopic Surgeons
SAS	SAS company founded in 1976 to assist customers with SAS computer software ('statistical analysis system')
Second Life®	Second Life is an online free 3D virtual world (Linden Research, Inc., San Francisco, California, USA)
SD	standard deviation
SimMan®	SimMan® Universal simulator is a computerised patient simulator which is a portable, instructor-driven, full body, electromechanical mannequin (Laerdal Medical Corporation, Wappingers Falls, New York, USA).
SMIC	single measure intraclass correlation
SOS	simulated operating suite (Imperial College, London, UK)
SPPS	statistical package for social sciences computer software (SSSP Inc., Chicago, Illinois, USA)
TEP	totally extraperitoneal
TURP	transurethral resection of prostate
UK	the United Kingdom
UM	URO Mentor™ endourology simulator (Symbionix, Cleveland, Ohio, USA)
URO Mentor™	endourology simulator (Symbionix, Cleveland, Ohio, USA)
URS	uretrorenoscopy
USA	the United States of America
VAS	visual analogue scale
VR	virtual reality
Xitact™	XiTact was founded in April 2000 as a spin-off from the Swiss Federal Institute of Technology in Lausanne (EPFL).
Xitact™ ITP	instrument tracking and haptic hardware that for endoscopic surgery training in the fields of laparoscopy, gynaecology, arthroscopy and cardiac surgery (Mentice AB, Gothenburg, Sweden)