ASERNIP-S

Australian Safety and Efficacy Register of New Interventional Procedures - Surgical

Systematic Review

Surgical simulation for training: Skills transfer to the operating room

ASERNIP-S REPORT NO. 61

July 2007

Australian Government
Department of Health and Ageing

Australian Safety & Efficacy Register of New Interventional Procedures – Surgical

The Royal Australasian College of Surgeons

Surgical simulation for training: Skills transfer to the operating room

ISBN 0909844 83 6

Published July 2007

This report should be cited in the following manner:

Sturm L, *et al.* Surgical simulation for training: Skills transfer to the operating room. ASERNIP-S Report No. 61. Adelaide, South Australia: ASERNIP-S, July 2007.

Copies of these reports can be obtained from:

ASERNIP-S
PO Box 553,
Stepney, SA 5069
AUSTRALIA
Ph: 61-8-8363 7513
Fax: 61-8-8362 2077
E-Mail: asernips@surgeons.org
http://www.surgeons.org/asernip-s

The Systematic Review of Surgical Simulation for Training: Skills Transfer to the Operating Room
was ratified by:

The ASERNIP-S Management Committee on
June 2007
and
The Council of the Royal Australasian College of Surgeons on
July 2007

Table of Contents

List of Tables

List of Figures

Executive Summary

Objective

To assess whether skills acquired via simulation-based training transfer to the operative setting.

Methods

Search strategy – Studies were identified by searching MEDLINE, EMBASE, CINAHL, The Cochrane Library and Current Contents from inception to December 2006. 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 December 2006.

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. The articles included must have contained training and/or measures of performance in the simulated setting and measures of performance in the operative setting. Measures of surgical task performance included accuracy of skills, time to complete technique, efficiency of movement, error rates and achievement of performance to criterion levels.

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

Results

A total of 12 randomised controlled trials and two non-randomised comparative studies were included in this review. The review looked at simulation as a concept, and as such included studies with various training techniques in the surgical setting. There were differences in indications, simulation-based training methods, training times, and the amount of guidance and feedback provided to trainees. In most cases, simulation-based training was an add-on to normal surgical training programs. Only one study compared simulation-based training with current training methods (patient-based training).

For laparoscopic cholecystectomy, participants who received simulation-based training prior to conducting patient-based assessment generally performed better than their counterparts who did not have this training. This improvement was not universal for all the parameters measured, but the untrained group never outperformed the trained group. Trained groups generally made fewer errors, and

had less instances of supervising surgeon takeover than participants who did not have the training.

For colonoscopy/sigmoidoscopy, simulation-based training prior to patient-based assessment generally appeared to provide participants some advantage over their untrained controls, particularly during the initial stages of learning.

For catheter-based intervention for occlusive vascular disease and TEP hernia repair, simulation-based training appeared to show benefits for participants when later conducting patient-based assessment.

There were no differences in performance between endoscopic sinus surgery simulator-trained residents compared with controls when performing endoscopic sinus surgery.

The study that compared patient-based training with simulation-based training for colonoscopy/sigmoidoscopy found that participants who received training in the assessment procedure exhibited better performance than those who had trained exclusively on a simulator without any mentoring or supervision.

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 via 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.

Clinical and Research Recommendations

It is recommended that further research be done into the transfer of skills acquired via simulation-based training to the patient setting to strengthen the current evidence base. Future studies could explore:

- 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.

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.

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 to 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.

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
PO Box 553
Stepney SA 5069

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

Advisory Surgeon

Mr Peter Cosman
PO Box 196
Brighton Le Sands NSW 2216

ASERNIP-S Researcher

Lana Sturm
ASERNIP-S
Royal Australasian College of Surgeons
PO Box 553
Stepney SA 5069

Conflicts of Interest

- Professor John Windsor
  Director and Educational Advisor, Go Virtual Medical Ltd.
  Founding Director, Advanced Clinical Skills Centre, University of Auckland.
- Mr Patrick Cregan
  Medical Director, Medic Vision Pty. Ltd.

1. Introduction

Objective

The objective of this systematic review is to assess whether skills acquired via simulation-based training transfer to the surgical setting.

Context

Surgical training consists of developing cognitive, clinical and technical skills, with the latter traditionally acquired through mentoring in the operating theatre. Current surgical skills training programs are under pressure due to increased service requirements, new and emerging techniques, a greater focus on surgeons' competence, and concerns regarding patient safety. In order to face these challenges and to improve surgical education, many surgical colleges are re-engineering their training curricula with an increased interest in the use of rapidly developing and emerging educational strategies. Surgical simulation offers the opportunity for surgical trainees to practise surgical skills (mental practice and reinforcement) before entering the operating room 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. It involves surgical trainees learning surgical skills in the operating theatre by first observing, then assisting mentors, before being permitted to operate under supervision. As the trainee's experience grows, he or she is able to take on a more active role, until eventually he or she is able to work unsupervised. In this environment, unskilled trainees learn to perform skills on actual patients. Competency is subjectively assessed by supervising surgeons.

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. 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 mentored approach is dependent on skilled surgeons having the time and resources to train and supervise trainees. A survey in 2005 found that 91% of the current Australian Fellowship are involved in training supervision, providing on average, 10.2 hours of training supervision per week (Royal Australasian College of

Surgeons 2005). This survey also found that the current workforce is ageing and within the next five years, one third of the current active Fellowship is expected to retire from emergency call work and two thirds of this group are also intending to retire from operative practice. This will reduce the number of Fellows available to provide training, supervision and mentoring. In addition to this, legislated restrictions on surgeons' working hours have reduced the number of hours during which experienced surgeons are avaliable to observe and assist trainees. It has also reduced the number of hours trainees can expect to be supervised (Woodrow *et al.* 2006). In the Australian system, as in many other countries, working shifts of 36 hours or more are now largely outlawed and have been replaced by maximum working periods of 13 to 14 hours with mandated rest periods (Maddern 2003). The impact of these changes has resulted in a more consultant-led supervision of care, an increase in the number of surgical trainees per surgical unit, and competition for the surgical cases available from which to gain surgical experience during the course of training (Maddern 2003).

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. For example, cholecystectomy, anti-reflux and bariatric surgery are now being done using a laparoscope (Aggarwal *et al.* 2006) instead of the open surgical approaches used previously. Minimally invasive procedures using a laparoscope as well as other endoscopic techniques differ from open surgery in the way of direct tactile contact and visual feedback, and an increased need for hand-eye coordination (Gallagher *et al.* 1998). Although the skills needed for these surgical modalities, like those needed for open surgery, can be taught in the operating theatre, simulation allows trainees to practise these skills before entering the operating theatre 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 actual performance (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). Of these, technical skills make up the majority of the objective data on surgical training and assessment, although cognitive skills are likely to play a larger part (Satava *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 operating room, 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). This allows the trainee to focus more on learning the steps of the operation and how to handle intraoperative complications, than on the refinement of technical skills (Gallagher *et al.* 2005). Simulation-based training using flight simulators has been mandatory in the United States aviation industry since 1955 (Kaufmann 2001). All commercial and military pilots must train and be certified on a simulator before actual flight (Cohen *et al.* 2006a). Anaesthesiology has applied principles similar to those used in pilot training and now has over 30 years of history in simulation-based training (Reznek *et al.* 2002).

Types of surgical simulation

A range of simulation devices and models now exist, and many hospitals and medical schools have developed skills laboratories where learners from a range of disciplines can practise their skills (Dent 2001). In general, surgical simulation allows for the repetitive performance of a single task to allow the trainee to develop hand-eye co-ordination and motor skills before entering the real-patient setting. It has been shown that training a single task results in better performance than training a variety of skills at once (Kirk 1996). Speed, efficiency and precision are developed with practice and automaticity in the skill is eventually achieved.

The different forms of simulation are described below.

Synthetic (inanimate) models and box trainers

Simulation using physical objects usually involves models of plastic, rubber and latex. These objects are used to render different organs and pathologies and allow a trainee to perform specific tasks and procedures (Cisler and Martin 2006). A box trainer uses the actual instruments and optical system used clinically to manipulate 'synthetic' tissues (Fried 2006). Some physical simulators may also reproduce the haptics (tactile sensations) of the actual surgical environment. Physical simulators are generally part-task trainers and help develop the hand-eye co-ordination and motor skills necessary for specific tasks such as cutting, suturing, grasping or clipping structures. Skills such as suturing and knot-tying can effectively be acquired on these models (Grober *et al.* 2004).

Physical simulators must be re-equipped after each use which requires some time investment. These models have a number of limitations, including limited realism and haptics, and no inherent metrics (performance measurements taken by simulator) for assessment. Physical simulator models do not directly measure movements or skills, and require a trained observer to determine performance (Fried 2006). Assessment occurs when trainees perform a series of standardised tasks, and performance is rated using a task-specific checklist or a global rating form (Reznick and MacRae 2006).

Although some trainees using these types of simulators have found the training unrealistic and boring due to their low fidelity (the extent to which the model imitates reality) (Hyltander 2003), studies show that they are valid and reliable instruments for training specific surgical skills (Grantcharov *et al.* 2004; Seymour *et al.* 2002). Their relatively low acquisition cost, high availability and easy portability, make this type of simulator the most widely available and most 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). Animal models enable trainees to work together as a team on an operation, providing additional insight into setting up an operative case with regard to functional team relationships, hierarchies and gradients of authority (Roberts *et al.* 2006). Animal models have been used extensively in both open (Chapman *et al.* 1996) and laparoscopic (Fried *et al.* 1999; Korndorffer, Jr. *et al.* 2005) surgical training. Anatomical differences between humans and animals make some animal models better than others for training purposes. Examples of animals in use for training include porcine, canine, ovine, baboon and avian models. Training surgical skills using live animals requires access to large animal care facilities and veterinary staff, which is expensive (Cisler and Martin 2006). Live animal model training is more involved than other simulation models because there is a need for sedation, analgesia and peri-operative monitoring (Cisler and Martin 2006), and in some cases require the presence of a training supervisor to train and assess trainees. Animal models provide realistic haptic feedback but the numbers of animals needed as well as cultural, financial and ethical issues limit their use.

Cadaveric models

The use of human cadavers allows surgical trainees to develop a detailed understanding of human anatomy and are a valuable tool for the teaching of whole-body anatomy and the interaction between different body parts when affected by disease processes (Parker 2002). In addition to anatomical dissection courses, surgical trainees have used human cadavers to practise many procedures, including laparoscopy (Levine *et al.* 2006), endoscopy (Rivron and Maran 1991), and saphenous vein cutdown (Wong and Stewart 2004). Preserved human cadavers (including cadavers preserved by plastination) do not bear the same tissue elasticity as live or recently deceased cadavers, thereby losing some fidelity as a surgical instructional model (Herbella and Del Grande 2001). In addition to this, considerable expense is involved in the provision of cadaveric specimens, which are single use and not portable (Wong and Stewart 2004). The limited supply of cadavers in Australia, the decline of anatomical dissection courses in Australian medical schools (Parker 2002), 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 is another form of simulation in surgical skills training. These sections or tissues can be attached to, and supported by, synthetic frames configured to mimic human anatomy and may provide approximate haptic feedback (Hochberger and Maiss 2006). This form of training allows for organs to be prepared and frozen ready for use at a later time. Although individual *ex vivo* models are limited in the variety of procedures they can train, they may, as with some inanimate models, prove beneficial as training tools for skills that require repetitive practice (Hochberger and Maiss 2006). Costs associated with this type of training are relatively low in comparison to live animal training, but still require staff to prepare the organs for simulation, and to supervise and assess trainees (Cisler and Martin 2006). Dedicated 'wet rooms' within skills centres are mandatory if this training model is to be employed. Although data supporting the benefits of *ex vivo* simulation are starting to accrue, more evidence is required to confirm that the use of this simulation model improves outcomes in clinical performance (Hochberger and Maiss 2006).

Virtual reality (computer-based) models

Virtual reality (VR), surgical simulators are the next development in the area of surgical simulation, and are receiving widespread attention. VR systems use computer generated instruments through specially designed interfaces to manipulate computer generated objects. To date, laparoscopic and endoscopic surgery, rather than open surgery have received most of the developmental focus in surgical VR simulators (Carter *et al.* 2005; Seymour 2005). In part, this is the result of efforts to address outcome deficiencies identified early in the clinical experience with laparoscopic cholecystectomy (Deziel *et al.* 1993). It is also attributable to the fact that laparoscopic and endoscopic scenarios are more easily recreated by VR simulators than open surgery because they comprise a two-dimensional visual system with limited haptic interactions (Hyltander *et al.* 2002).

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. VR trainers allow trainees to practise independently, and can be incorporated into a structured curriculum (Roberts *et al.* 2006).

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 (ie haptics) in the use of surgical instruments is imperfect in some VR simulators (Seymour and Rotnes 2006). Although work is progressing to further incorporate haptics into VR trainers (Roberts *et al.* 2006), this is expensive, and studies indicate that a high level of fidelity may not be essential to

train some surgical skills, particularly to junior level trainees (Anastakis *et al.* 1999; Grober *et al.* 2004). Studies comparing VR simulators have demonstrated that they are useful in training students in such procedures as upper endoscopy (Cisler and Martin 2006), haemostasis (Maiss *et al.* 2006), flexible sigmoidoscopy (Datta *et al.* 2002; Tuggy 1998) and colonoscopy (Cohen *et al.* 2006b; Sedlack and Kolars 2003).

Hybrid simulators

Hybrid simulators are a combination of physical simulators and VR simulators. They consist of a physical object (frequently a mannequin) being linked to a sophisticated computer program that provides visual images and/or feedback (Satava 2001). The computer program can simulate patient responses, both physically on the mannequin and physiologically in response to a procedure (Satava 2001). These simulators go beyond basic skills training and are designed to recreate specific anatomy and physiology, and allow 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 operating room (Roberts *et al.* 2006). Hybrid simulators are expensive and are limited in use due to high demands of time and effort needed to prepare and run them (Sewell *et al.* 2004). In anaesthesia and critical care medicine, there is extensive research supporting the use of hybrid simulator models for training (Cooper and Taqueti 2004).

Skills transfer to the operating theatre

Training in the operating room 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 and appear to be gaining acceptance into surgical training programs, although not yet reaching widespread adoption (Satava 2001). These types of simulators are expensive due to their low production volume and high development cost.

The fundamental assumption of surgical simulation-based training is that the skills acquired in simulated environments are directly transferable to the clinical setting. Although the effectiveness of individual simulators to teach surgical skills has not yet been proven (Dutta *et al.* 2006; Sutherland *et al.* 2006) and there is little strong evidence to indicate that they can be used for the determination of competence (Feldman *et al.* 2004), many studies have demonstrated improved simulator performance after simulator training (Haque and Srinivasan 2006; Sutherland *et al.* 2006). Many studies attempting to determine construct validity (that simulator performance correlates with actual technical ability) have focused on being able to discriminate levels of experience, and compare experienced surgeons to those with less or no surgical experience (McDougall *et al.* 2006; Schijven and Jakimowicz 2005;

Srivastava *et al.* 2004). In live animal models, studies have demonstrated improved operative performance in anaesthetised animals after simulation-based training (Hyltander *et al.* 2002; Korndorffer, Jr. *et al.* 2005; Van Sickle *et al.* 2006).

To date, less attention has been focused on correlating simulator performance with operative performance on live human patients. If the positive relationship between technical skills measured in a simulated environment and the technical skills measured in the operating room are robust, simulation-based training could not only predict a trainee's future performance in the operative setting, but could also go some way towards justifying the cost of these training devices. Without evidence that transfer is occurring, simulation-based training will battle to find relevance and acceptance into surgical skill training programs.

Summary

Increasing demands on current surgical training programs has resulted in other approaches to training being investigated and employed. The ability to learn specific technical surgical skills requires deliberate practice, and evidence suggests that many of the skills required for surgery can be acquired away from the operating theatre (Hamdorf and Hall 2000). Surgical simulation-based training is attractive in the field of surgical training because it 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. In order for surgical simulation-based to gain significance in training environments, it is essential to demonstrate that these skills can be transferred to real patients.

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 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.

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

No surgical simulation-based training.

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, which 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 subsequently 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.

Table 1 Databases searched

Database	Platform	Edition
Cochrane Library		Issue 2, 2006
Current Contents	Ovid	Searched 14/12/2006
EMBASE	Ovid	Week 1 1980 to 14/12/2006
MEDLINE	Ovid	1966 to 14/12/2006
CINAHL	Webspirs	1982 to 14/12/2006
PubMed	Entrez	1953 to 14/12/2006
Clinical Trials Database (US)		Searched 14/12/2006
NHS CRD (UK) NHS HTA (UK)		Searched 15/12/2006
National Research Register (UK)		Issue 2, 2006
Current Controlled Trials (mRCT)		Searched 14/12/2006

Search terms

In the Cochrane Library the search term used was:

surgical simulation

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

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

The NHS CRD databases were searched using the above terms. The National Research Register, Clinicaltrials.gov, Meta-Register and the Australian Clinical Trials Registry were also searched using the above 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. In Cochrane the truncation character is *; in Current Contents, EMBASE, CINAHL and MEDLINE (Ovid) it is $. # is a wildcard symbol that substitutes for one required character in Current Contents, EMBASE, CINAHL and MEDLINE (Ovid).

Literature database & exclusions

Articles were retrieved if they were judged to possibly meet the inclusion criteria based on their abstracts. Two ASERNIP-S researchers independently applied the selection criteria and any differences were resolved through discussion. The number of articles retrieved for each search category is listed in Figure 1. 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). 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 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 (ie simulation-based training *vs.* no training, and simulation-based training *vs.* 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 unpublished studies. The details for each are provided below:

1. Evaluation of Surgical Simulator for Practicing a Vascular Anastomosis. University of Western Ontario, Canada (ClinicalTrials.gov identifier NCT00318279). May 2006 – September 2007. Expected enrolment n =40. Design: prospective randomised double-blinded intervention study. To determine if practicing an aorto-saphenous vein anastomosis on a low-fidelity surgical simulator allows trainees to produce a higher quality anastomosis in a shorter period of time, than a group that only learns by watching a video (it is unclear whether this study will assess trainees on a simulator or a human patient).
2. Transfer of Skills from VR-Trainer to Operation Room (ClinicalTrials.gov identifier NCT00311792). Rigshospitalet, Denmark. April 2006 – August 2008. Expected enrolment n =24. Design: prospective randomised double-blinded intervention study. To determine if skills obtained by training in the laparoscopic virtual reality simulator LapSimGyn can be transferred to a real laparoscopic operation measured as improved score in a technical skills assessment.
3. Laparoscopic Simulator Training Study (National Research Register number N0504163844). UK. January 2005 – November 2006. Expected enrolment; n =

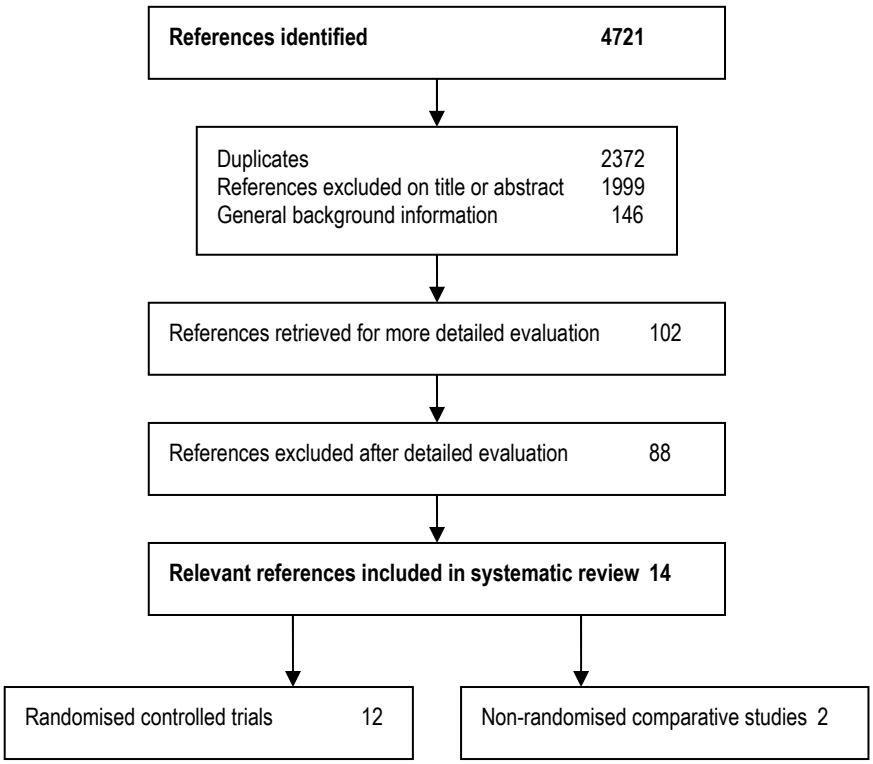
10 surgeons, n = 20 medical students. To determine if a structured training program incorporating simulator technology with in-built measures of performance can improve a surgical trainee's performance on a laparoscopic task/procedure when compared with a control group that does not incorporate simulator technology.

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). Study quality was assessed according to the methods given in Section 6 of the Cochrane Reviewers' Handbook (Higgins and Green 2005) 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. The study profiles for these studies are given in Appendix C.1 and C.2, and the data extraction is given in Appendix D. Table 3 provides a description of the training method used within each study.

Some authors and/or centres have published more than one report on the transfer of skills acquired via surgical simulation to the patient-based setting. As a result, some studies published by the same group may have common pools of patients. These studies have been identified.

Description of studies

A total of 12 randomised controlled trials and two non-randomised comparative studies were included in this review (Table 2). There were a total of 287 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. The simulator device specifications and a description of patient-based training are given in Appendix E. A summary of the critical appraisal is given in Appendix F.

Simulation-based training vs. no simulation-based training

Laparoscopic cholecystectomy

Randomised controlled trials

Four RCTs compared operative laparoscopic cholecystectomy performance in participants that had trained using simulation with those who had not received simulation-based training (Grantcharov *et al.* 2004; Scott *et al.* 2000; Scott *et al.* 1999; Seymour *et al.* 2002) (Table 2). In Grantcharov *et al.* (2004), untrained participants performed a laparoscopic cholecystectomy, and were then given either simulator training or no simulator training before conducting another laparoscopic cholecystectomy. In Scott *et al.* (1999) and Scott *et al.* (2000) untrained participants performed both a laparoscopic cholecystectomy and were tested on a simulator at baseline. They were then given either simulator training or no simulator training, after which they were again assessed on a laparoscopic cholecystectomy procedure and on the simulator. In Seymour *et al.* (2002) participants had pre-study (visuospatial, perceptual and psychomotor) tests to determine baseline performance. Participants were then given either simulator training or no training. All residents then performed a laparoscopic cholecystectomy.

Scott *et al.* (2000) used a random numbers table to randomise participants. The method of randomisation was not stated in the other three studies (Grantcharov *et al.* 2004; Scott *et al.* 1999; Seymour *et al.* 2002). Seymour *et al.* (2002) stratified participants by postgraduate year. Allocation concealment was not described in Scott *et al.* (1999), Scott *et al.* (2000), and Seymour *et al.* (2002). Grantcharov *et al.* (2004) used sealed envelopes to conceal the allocation of participants.

All four studies reported elements of blinding. Grantcharov *et al.* (2004) stated that the supervising surgeon for the assessment procedure was blinded, as well as the surgeons that reviewed and assessed the videotaped procedures. Scott *et al.* (1999) stated that the raters who assessed the operations were blinded to the training status, and Scott *et al.* (2000) stated that both the supervising surgeons for the operative procedures and the evaluators assessing these operations were blinded to the training status of the

participants. Seymour *et al.* (2002) stated that surgeon-investigators present at the assessment procedure were blinded to the participants' training status.

None of the four studies stated whether they employed an intention to treat analysis (Grantcharov *et al.* 2004; Scott *et al.* 1999; Scott *et al.* 2000; Seymour *et al.* 2002).

Three of the four studies did not state if they used a power calculation (Grantcharov *et al.* 2004; Scott *et al.* 1999; Seymour *et al.* 2002). Scott *et al.* (2000) stated that a nonparametric power analysis was performed to determine the sample size needed to detect an effect of training. It was found that twenty seven or more participants were required to provide a power of at least 0.8 with a type I error of 0.05 if the equivalent effect size equaled or exceeded unity.

Losses to assessment were reported by Grantcharov *et al.* (2004) (video recorder malfunction, training: n = 2, control: n = 2) and Scott *et al.* (2000) (scheduling difficulties, training: n = 4, control: n = 1). Scott *et al.* (1999) and Seymour *et al.* (2002) did not report on any losses of participants before assessment.

The study period (the month and/or years during which each study took place) was stated for two studies (Grantcharov *et al.* 2004; Scott *et al.* 2000), but not the other two studies (Scott *et al.* 1999; Seymour *et al.* 2002).

The global rating scale for the assessment of operative performance used by Grantcharov *et al.* (2004), Scott *et al.* (2000) and Scott *et al.* (1999) was originally created and validated by Reznick *et al.* (1997). It was not stated by Seymour *et al.* (2002) whether the assessment methods of the operative procedure had been previously validated, but the test used to assess pre-study abilities had been validated.

For all four studies, inclusion criteria were reported as being surgical residents (at various years of training) (Grantcharov *et al.* 2004; Scott *et al.* 2000; Scott *et al.* 1999; Seymour *et al.* 2002). Exclusion criteria were not reported for any of the four studies (Grantcharov *et al.* 2004; Scott *et al.* 2000; Scott *et al.* 1999; Seymour *et al.* 2002).

The reporting of baseline characteristics of the participants in the studies varied. Grantcharov *et al.* (2004) reported gender, age, time since graduation and the median number of previous cholecystectomies performed. Scott *et al.* (2000) reported the mean number of previous cholecystectomies performed and Seymour *et al.* (2002) reported the gender of the participants. Scott *et al.* (1999) did not report any baseline characteristics. There were no differences in baseline characteristics between the trained and untrained groups in the four studies (Grantcharov *et al.* 2004; Scott *et al.* 2000; Scott *et al.* 1999; Seymour *et al.* 2002).

Non-randomised comparative studies

Schijven *et al.* (2005) compared the operative performance of participants who had undergone a laparoscopic cholecystectomy training course to those who had not done the course. Assessment operations were taped and the videotape assessors were blinded to the participants training status. Losses to assessment were reported as being due to

videotape recording failures (training: n = 2, control: n = 2). Baseline characteristics were reported and included age, gender, years of training, right handedness, the number of previous laparoscopic cholecystectomies performed and the number of participants that had conducted simulator training courses. There was a significant difference for the number of laparoscopic cholecystectomies in favour of the control group (p = 0.008). The study period and inclusion criteria were reported. No exclusion criteria were reported and it was not reported whether the operative assessment methods had been validated.

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Five RCTs investigated transfer of colonoscopy/sigmoidoscopy skills learned on a simulator to a live patient (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Cohen *et al.* 2006b; Tuggy 1998). In all five studies participants either trained on a simulator or received no simulator training before conducting colonoscopy or sigmoidoscopy procedures in patients.

Cohen *et al.* (2006b) used a random number table to randomise participants. The method of randomisation was not stated for the other four studies (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Cohen *et al.* 2006b; Tuggy 1998). Ahlberg *et al.* 2005 reported using sealed envelopes to conceal allocation of the participants into the two arms of the study, but allocation concealment methods were not reported for the four other studies (Sedlack and Kolars 2004; Sedlack *et al.* 2004; Cohen *et al.* 2006b; Tuggy 1998).

In Ahlberg *et al.* (2005), the supervising surgeons were blinded to the training status of the participants, but it was unclear whether these surgeons also conducted the assessment of the procedure. The surgeons supervising and assessing colonoscopies in Cohen *et al.* (2006b) were blinded to participants training status. Evaluating staff in Sedlack and Kolars (2004) and Sedlack *et al.* (2004) were not blinded to participant training status. In Tuggy (1998), patients were blinded to the experience and training status of the participants.

None of the five studies reported an intention to treat analysis (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Cohen *et al.* 2006b; Tuggy 1998).

Cohen *et al.* (2006b) generated Kaplan-Meier curves to determine the number of blocks of 20 cases needed for each group to reach a median of 90% objective and subjective competency. The four other studies did not report on power analyses (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Tuggy 1998).

Losses to assessment were reported by Cohen *et al.* (2006b) as being due to protocol violations (group not specified: n = 4). The other four studies did not report on any losses to assessment (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Tuggy 1998).

Only one out of the five studies reported the study period (Sedlack *et al.* 2004).

It was stated that face and construct validity of the AccuTouch® endoscopy simulator used by Ahlberg *et al.* (2005) was demonstrated by Datta *et al.* (2002) but it was not stated whether the methods used to assess performance had been validated. Sedlack and Kolars (2004) stated that the AccuTouch® colonoscopy device used within the simulation curriculum had been previously validated by Sedlack and Kolars (2003) and Sedlack and Kolars (2002) but it was not reported whether the methods used for assessment had been validated. It was also not stated by Sedlack *et al.* (2004) whether the assessment methods had been validated. Cohen *et al.* (2006b) reported that validation data for the AccuTouch® simulator was presented in Sedlack and Kolars (2004) and that for patient-based assessment, an evaluation form previously used by Cass *et al.* (1996) was used (this study was published as an abstract and does not elucidate the assessment tool). Tuggy (1998) did not report whether the simulator or assessment methods had been previously validated.

The reported inclusion criteria for the studies included the training level of participants (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004), no prior experience in flexible sigmoidoscopy (Tuggy 1998) and the training directors' adherence to the study protocol (Cohen *et al.* 2006b). Exclusion criteria were listed in one study to be prior performance of more than 10 colonoscopies and an inability to adhere to the study protocol (Cohen *et al.* 2006b). None of the other studies listed exclusion criteria (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Tuggy 1998).

Baseline characteristics of participants were reported as gender and pervious experience in colonoscopy for Ahlberg *et al.* (2005) and Sedlack and Kolars (2004), and pre-colonoscopy experience for Cohen *et al.* (2006b). Sedlack *et al.* (2004) stated that no participant had prior endoscopy training. No baseline characteristics were reported for the participants of Tuggy (1998).

Cohen *et al.* (2006b) reported that there were no significant differences between the groups prior to training. Ahlberg *et al.* (2005) reported that participants came from eight different institutions in Sweden and were at different levels of experience (from the 2nd to 5th postgraduate year), but had the common feature that they had no previous experience in colonoscopy and were designated to start colonoscopy training. Sedlack and Kolars (2004) and Sedlack *et al.* (2004) did not state whether there were any significant differences between the two groups prior to training. Tuggy (1998) stated that matched pairs of residents performed examinations sequentially on the same patient to reduce the risk of encountering different colon structures which could affect their performance.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported that their participants were randomised to receive either training on the Procedicus VIST™ simulator, or to no additional training. After the training period, all residents conducted two consecutive mentored catheter-based

interventions for lower extremity occlusive vascular disease in an operating room/angiography suite. In this study, the method of randomisation was not stated. Allocation concealment was through sealed envelopes. The surgeons assessing participants were blinded to the training status of participants. An intention to treat analysis and power calculations were not reported. There were no reported losses to assessment and the study period was not stated. Objectives of the study were to validate the use of the simulator. It was reported that the construct validity of the simulator had been previously established, but it was not reported whether all of the assessment methods had been previously validated. The global rating scale used to assess the endovascular technique and the ability to complete the case was adapted from a previously validated scoring system (Reznick *et al.* 1997). Inclusion criteria included general surgical residents with no endovascular experience. No exclusion criteria were described and although baseline characteristics were collected, they were not reported. The authors reported that there were no pre-study differences between the simulator-trained group and control group prior to training.

Total extraperitoneal (TEP) hernia repair

Randomised controlled trials

Participants of Hamilton *et al.* (2001) performed a laparoscopic TEP hernia repair to determine baseline operative performance. They were then randomised to either receive training using an instructional video, interactive CD ROM and a moulded rubber hernia simulator, or no additional training. After the training period, each participant's operative performance was again measured while performing a laparoscopic TEP hernia repair in the operating room. The method of randomisation and allocation concealment were not reported in this study. Surgeons acted as assistants and evaluators, and were blinded to the residents training status. Intention to treat analyses and power calculations were not reported. Operative performance was evaluated by a global assessment tool previously validated by Reznick *et al.* (1997). Inclusion criteria were 3rd and 4th year surgery residents. No exclusion criteria or baseline characteristics were reported. The study period was given and there were no reported losses to assessment. The authors reported one baseline difference between the groups prior to training.

Endoscopic sinus surgery

Non-randomised comparative studies

Four participants in Edmond (2002) performed endoscopic sinus surgery. Two had trained on the simulator, and two had not. Proctors evaluated patient-based operative performance. The procedures were videotaped and were assessed by videotape assessors who were blinded to the participant's training status. Validation of the simulator was the first phase of this study. It was not stated whether the operative assessment methods had been validated. Inclusion criteria included 1st year ear-nose-throat residents. Any exclusion criteria, the study period and losses to assessment were not reported. The level of experience was the baseline characteristic collected. Some participants had previous

training on the simulator but it was difficult to ascertain if there were baseline differences between the two groups.

Simulation-based training vs. patient-based training

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) randomised participants into two groups that received either sigmoidoscopy simulator training or patient-based training (performance of 10 supervised sigmoidoscopic examinations). All of the other included studies in this review compared simulation-based training with no simulation-based training, where as this study compared simulation-based training with patient-based training. The difference in comparator meant that this study was treated separately. For assessment, all participants performed five patient-based sigmoidoscopies. For randomisation, participants were sequentially allocated by the investigator, but allocation concealment was not reported. Investigators and assessors were not blinded to the training status of participants, although the patients used for the assessments were. An intention to treat analysis was not reported. The authors conducted a power calculation, and found that using a comparison of the means of two independent samples, the sample size required to detect a magnitude difference between arms of 30%, a power of 90% and an alpha value of 0.05, was calculated to be 30 examinations in each arm of the study (assuming 5% of examinations would not be completed, and another 5% due to patient intolerance, 33 patients would need to be recruited per arm). It was stated that there were no losses to assessment, but the study period was not stated. It was not stated whether the assessment tools used for assessment had been validated. Inclusion criteria included internal medicine residents. Exclusion criteria included prior experience in flexible sigmoidoscopy and the prior use of the simulator. Age and the extent of computer usage were the baseline characteristics collected.

Table 2 Summary of included studies

Study	Study type	Level	Training method	Comparator*	Indication
Simulation-based training *vs.* no simulation-based training					
Grantcharov *et al.* 2004 (Denmark)	RCT	II	MIST-VR simulator	no simulator training	laparoscopic cholecystectomy
N			10	10	
Seymour *et al.* 2002 (Northern Ireland)	RCT	II	MIST-VR simulator	no simulator training	laparoscopic cholecystectomy
N			8	8	
Scott *et al.* 2000 (USA)	RCT	II	SCMIS GEM video trainer	no simulator training	laparoscopic cholecystectomy
N			13	14	
Scott *et al.* 1999† (USA)	RCT	II	Video trainer‡	no simulator training	laparoscopic cholecystectomy
N			9	13	
Schijven *et al.* 2005 (Netherlands)	comparative study	III-2	4 day laparoscopic cholecystectomy training course	no training course	laparoscopic cholecystectomy
N			12	12	
Ahlberg *et al.* 2005 (Sweden)	RCT	II	AccuTouch® VR endoscopy§ simulator	no simulator training	colonoscopy
N			6	6	
Sedlack *et al.* 2004 (USA)	RCT	II	AccuTouch® flexible sigmoidoscopy simulator	no simulator training	sigmoidoscopy
N			19	19	
Sedlack and Kolars 2004‖ (USA)	RCT	II	AccuTouch® colonoscopy simulator	no simulator training	colonoscopy
N			4	4	
Cohen *et al.* 2006b (USA)	RCT	II	Simbionix GI Mentor™ simulator	no simulator training	colonoscopy
N			22	23	
Tuggy 1998 (USA)	RCT	II	Gastro-Sim® flexible sigmoidoscopy simulator	no simulator training	sigmoidoscopy
N			5	5	
Chaer *et al.* 2006 (USA)	RCT	II	Procedicus VIST™ simulator	no simulator training	catheter-based intervention for occlusive disease
N			10	10	
Hamilton *et al.* 2001 (USA)	RCT	II	TEP hernia repair rubber model simulator, instructional video and interactive CD ROM	no simulation training	TEP hernia repair
N			10	11	
Edmond 2002 (USA)	comparative study	III-2	ESS simulator	no simulator training	endoscopic sinus surgery
N			2	2	

Table 2. Summary of included studies continued

Study	Study type	Level	Training method	Comparator*	Indication
Simulation-based training *vs.* patient-based training					
Gerson and van Dam 2003 (USA)	RCT	II	VR sigmoidoscopy simulator	supervised patient-based sigmoidoscopies	sigmoidoscopy
N			9	7	

* No simulator training/no training course, refers to control groups who did not receive training on a simulator, or did not participate in a training course, but did continue normal surgical training.
† Potential patient overlap with Scott *et al.* 2000.
‡ Video trainer likely to be SCMIS GEM.
§ Referred to as endoscopy simulator throughout study, but likely to be colonoscopy simulator.
‖ Potential patient overlap with Sedlack *et al.* 2004.
MIST-VR Minimally Invasive Surgical Trainer - Virtual Reality
SCMIS GEM Southwestern Centre for Minimally Invasive Surgery Guided Endoscopic Module
VIST Virtual Interventional Simulator Trainer
GI Gastrointestinal
TEP Total extraperitoneal
ESS Endoscopic sinus surgery

Table 3 Description of training in included studies

Study	Training method	Description of training	n/N
Simulation-based training *vs.* no simulation-based training			
Grantcharov *et al.* 2004 RCT Level II	MIST-VR simulator	10 repetitions of 6 tasks.	10/20
Seymour *et al.* 2002 RCT Level II	MIST-VR simulator	Training until criterion levels reached (3 – 8 hours).	8/16
Scott *et al.* 2000 RCT Level II	SCMIS GEM trainer	Separate 30 minute sessions for 10 days. Tasks practised an average of 138 times (range 94 – 171 times).	13/27
Scott *et al.* 1999 RCT Level II	Video trainer	Separate 30 minute sessions for 10 days.	9/22
Schijven *et al.* 2005 Comparative study Level III-2	Laparoscopic cholecystectomy training course	4 day course including videos, oral presentations, table sessions, instrument displays and repetitive sessions of VR software simulations using the Xitact LS500 laparoscopy simulator platform. Both psychomotor VR simulation (MIST-VR) and procedural laparoscopic cholecystectomy simulation, including the clip-and-cut, navigation and dissection modules (Xitact) were featured.	12/24
Ahlberg *et al.* 2005 RCT Level II	AccuTouch® endoscopy simulator	Median total training time 20 hours (range 15 -25) over at least 4 days.	6/12
Sedlack *et al.* 2004 RCT Level II	AccuTouch® flexible sigmoidoscopy simulator	Independent, supervised 3 hour simulator based training curriculum involving brief multimedia tutorial followed by 8 – 10 simulated scenarios.	19/38
Sedlack and Kolars 2004 RCT Level II	AccuTouch® colonoscopy simulator	6 hours of simulator training over a 2 day period. Performance of 20 – 25 simulated colonoscopies.	4/8
Cohen *et al.* 2006b RCT Level II	Simbionix GI Mentor™	10 hrs over 8 week period.	22/49
Tuggy 1998 RCT Level II	Gastro-Sim® flexible sigmoidoscopy simulator	Initially 5 hours. Then an additional 5 hours.	5/10
Chaer *et al.* 2006 RCT Level II	Procedicus VIST™	A single session of not more than 2 hours duration (mean 90 ± 21 minutes).	10/20
Hamilton *et al.* 2001 RCT Level II	TEP hernia repair rubber model simulator, instructional video and interactive CD ROM	10 separate 30 minute sessions over a 2 week period. Participants asked to alternate daily between simulator and CD ROM.	10/21
Edmond 2002 Comparative study Level III-2	ESS simulator	Not stated.	2/4
Simulation-based training *vs.* patient-based training			
Gerson and van Dam 2003 RCT Level II	VR sigmoidoscopy simulator	Unlimited time on simulator during 2 week period. Overall time on simulator was 138 ± 28 mins (range 66 – 287) per resident.	9/16
	Patient-based training	10 supervised sigmoidoscopic examinations during 2 week period.	7/16

Table 4 Description of statistical analyses used in included studies

Study	Training method	Statistical analysis
Simulation-based training *vs.* no simulation-based training		
Grantcharov *et al.* 2004 RCT Level II	MIST-VR simulator (n = 10) No simulator training (n = 10)	Normal distribution of the data was confirmed using Q-Q plots. The primary outcome measure was the difference in performance scores during laparoscopic cholecystectomy in the operating theatre between the first and second procedures. An independent samples *t*-test was used to examine the difference in improvements demonstrated by the two groups. Result was considered statistically significant at $P \leq 0.05$.
Seymour *et al.* 2002 RCT Level II	MIST-VR simulator (n = 8) No simulator training (n = 8)	Statistical comparisons were performed by chi-square analysis, analysis of variance (ANOVA), and Mann-Whitney test. Result significant at $P < 0.05$.
Scott *et al.* 2000 RCT Level II	SCMIS GEM trainer (n = 13) No simulator training (n = 14)	To determine whether there were differences between the control and trained groups at baseline testing, a two-tailed Wilcoxon rank-sum test was used. To determine if training was beneficial, within-person changes in performance were compared for the control and the trained groups. The amount of improvement varied with baseline performance, so a linear covariance adjustment was used to compensate for differences in baseline scores. The covariance-adjusted improvements for residents in the control and trained groups were compared using a Wilcoxon rank-sum test. To test the hypothesis that the trained group achieved greater adjusted improvement than the control group, a one-tailed test was used. Questionnaire data regarding comfort with laparoscopic surgery were analysed using Fisher's exact test. Tests were considered significant at $P \leq 0.05$.
Scott *et al.* 1999 RCT Level II n = 22	Video trainer (n = 9) No simulator training (n = 13)	To determine differences between control and trained groups after simulator training as well as to determine differences between control and trained groups after patient-based assessment, an exact Wilcoxon test was used. P was considered significant at $P < 0.05$.
Schijven *et al.* 2005 Comparative study Level III-2 n = 24	Laparoscopic cholecystectomy training course (n = 12) No training course (n = 12)	Normal distribution of the primary outcome parameter judgment and the secondary outcome parameters fluency and carefulness was confirmed using Q-Q plots. To determine differences in performance status between the two groups, a Mann-Whitney *U* test was used. *P* was considered significant at $P \leq 0.05$.
Ahlberg *et al.* 2005 RCT Level II n = 12	AccuTouch® endoscopy simulator (n = 6) No simulator training (n = 6)	The number of successful colonoscopies for the two groups was evaluated using a binary logistic regression analysis, controlling for patient gender, order of operation and student background. Confidence intervals were calculated using the profile likelihood estimation. The nine colon segments were categorised into three groups (1 – 4, 5 – 8 and 9), and an ordinal logistic regression analysis was carried out, and controlled for the same factors as mentioned above. The evaluate the training effect on the mean performance time per section, a stratified Mann-Whitney *U* test was performed. The time was ranked within each segment, and gender, gastroscopic experience, and procedure order was used as strata. Multiple regression analysis, controlling for the set of confounders, evaluated the association between training and the total procedure time for the trainees who completed the colonoscopy. The patient's pain score was categorised in three groups and ordinal logistic regression, controlling for different confounders, was used to assess the effect of training. A $P < 0.05$ was used as a criterion for inclusion in the statistics packaged used for the regression models.
Sedlack *et al.* 2004 RCT Level II n = 38	AccuTouch® flexible sigmoidoscopy simulator (n = 19) No simulator training (n = 19)	Median scores for each parameter graded by staff, resident, and patients were analysed by using Wilcoxon rank-sum test. Median differences between paired staff and resident evaluation scores were compared by using a Wilcoxon signed rank test.

Table 4. Description of statistical analyses used in included studies continued

Study	Training method	Statistical analysis
Simulation-based training *vs.* no simulation-based training		
Sedlack and Kolars 2004 RCT Level II n = 8	AccuTouch® colonoscopy simulator (N = 4) No simulator training (N = 4)	The analysis of the staff evaluations and patient surveys were compared between the two groups of Fellows. Comparisons were made of all colonoscopies performed as well as by examining procedures in chronological groups of 15 based on the order of performance (ie colonoscopies 1 – 15, 16 – 30, 11 – 45 etc). Rates of independent procedure completion were analysed using a *t*-test. All other results were analysed using Wilcoxon rank-sum test. P was considered significant at $P < 0.05$.
Cohen *et al.* 2006b RCT Level II n = 49	Simbionix GI Mentor™ (N = 22) No simulator training (N = 23)	A 2-sample *t*-test was used to compare the difference in objective competence, subjective competence, and observed patient discomfort between the simulator group and no-simulator group at every group of 20 cases (each 20 cases called a block). All of the blocks' data were then combined, and a mixed-effects model was applied to compare the difference between groups at every block simultaneously: in the mixed-effects model, a random effect was used to take into consideration the correlations between the observations from the same Fellow over time; fixed effects included each block as a categorical variable, a group indicator (simulator and no-simulator), and the interaction between them. A log-rank test was performed to compare the two groups. A Bonferroni correction was made on multiple comparisons, and a comparison was considered to be statistically significant if the *P* value was below 0.005.
Tuggy, 1998 RCT Level II n = 10	Gastro-Sim® flexible sigmoidoscopy simulator (N = 5) No simulator training (N = 5)	The paired t-test was used to compare the differences between the mean scores of the two groups at the designated points in the study protocol. The Mann-Whitney *U* test was used to analyse the qualitative assessment of colon viewing. It was not stated for which value *P* was significant.
Chaer *et al.* 2006 RCT Level II n = 20	Procedicus VIST™ (N = 10) No simulator training (N = 10)	Wilcoxon 2-sample test and Fisher's exact test were used to evaluate for statistically significant changes in performance pre- and post-training. Mean differences were considered significant for a $P < 0.05$.
Hamilton *et al.* 2001 RCT Level II n = 21	TEP hernia repair rubber model simulator, instructional video and interactive CD ROM (N = 10) No simulation training (N = 11)	Within-group comparisons of individual and composite global assessment scores were performed using Wilcoxon signed rank tests. Between-group comparisons were performed using the Wilcox rank sum test to accomplish an analysis of covariance (ANCOVA) with pre-test scores serving as the covariate. Questionnaire data were analysed using a Fisher's exact test. Statistical significance was set at a threshold of $P < 0.05$.
Edmond, 2002 Comparative study Level III-2 n = 4	ESS simulator (N = 2) No simulator training (N = 2)	Statistical analyses not reported.
Simulation-based training *vs.* patient-based training		
Gerson and van Dam 2003 RCT Level II n = 16	VR sigmoidoscopy simulator (N = 9) Patient-based training (N = 7)	Quantitative data were analysed using the paired Student's *t*-test. Categorical data analysis was conducted using Fisher's exact test. The level of significance was set at $P < 0.05$.

4. Results

The results have been presented in two main sections. Initially, the results of participants training on the simulators are presented (ie participant simulator performance results), after which results for participant's performance on patients in the operating theatre are presented (ie the evidence pertaining to transfer of skills).

The different parameters measured during operating room performance have been grouped together as closely as possible into tables, but it should be noted that there were many differences in assessment tools and techniques, which have been described in footnotes.

Studies were categorised initially by the non-simulation-based training method (ie simulation-based training *vs.* no simulation-based training and; simulation-based training *vs.* patient-based training), then by the intervention, and the level of evidence.

The terms trainees, Fellows, residents and participants all refer to people included in the studies, who were at the beginner-level of the surgical interventions studied.

Performance on simulators

There was considerable variation in the reporting of performance data (metrics) between studies.

Simulation-based training vs. no simulation-based training

Eight studies reported simulator performance data for simulator-trained participants (Ahlberg *et al.* 2005; Chaer *et al.* 2006; Cohen *et al.* 2006b; Scott *et al.* 1999; Scott *et al.* 2000; Sedlack *et al.* 2004; Sedlack and Kolars 2004; Seymour *et al.* 2002).

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported improvements (adjusted for baseline) in simulator performance for participants using the SCMIS GEM video trainer in comparison with the group that had no training with this device (both groups received baseline and final testing on the simulator). The trained group had significantly larger median time reductions for all five video trainer tasks compared with the control group ($p < 0.05$ for all). A similar RCT reported performance time data for video trainer performance for all participants in the study at baseline and at the end of the study. The simulator-trained group performed all five video trainer tasks significantly faster than the control group ($p < 0.05$ for all, data not shown) (Scott *et al.* 1999).

Seymour *et al.* (2002) reported that residents in the MIST-VR simulator group all successfully achieved the required criterion levels of performance in 3 – 8 training sessions. No further detail was reported.

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Ahlberg *et al.* (2005) reported that all six trainees in the AccuTouch® VR endoscopy simulator group successfully reached the expert criterion level. The expert criterion level was defined in the simulator by calculating mean values from five expert colonoscopists after a period of familiarisation with the simulator.

Sedlack and Kolars (2004) reported that the four participants in the AccuTouch® colonoscopy simulator-trained group completed an average of 21 simulated cases (range 19 – 26) prior to patient-based procedures. In a similar study, Sedlack *et al.* (2004) reported that the 19 participants in the AccuTouch® sigmoidoscopy simulator-trained group completed an average of 9 simulated cases (range 7 - 19) prior to patient-based flexible sigmoidoscopies.

Cohen *et al.* (2006b) reported that participants showed longitudinal improvement in skills on the Simbionix GI Mentor™ simulator between hour 1 and hour 10 of training on the device in the second year of the study (13/22). By the 10th hour of training, participants demonstrated significantly faster times to the caecum ($p = 0.022$) and total procedure time ($p < 0.001$), and significantly improved efficiency scores ($p < 0.001$) for both case 1 and case 2. There was no significant difference for percentage of mucosal surface examined and the number of episodes of excessive pressure.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported that all residents receiving training with the Procedicus VIST™ simulator were able to complete the mentored simulator training session in less than two hours (mean of 90 ± 21 minutes).

Simulation-based training vs. patient-based training

One study reported outcomes for simulator performance for simulator trained versus patient-trained participants (Gerson and van Dam 2003).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) reported that on average, residents spent over two hours on the VR sigmoidoscopy simulator performing the simulated cases. The average distance obtained with the simulator was 42 cm, demonstrating that most residents were unable to complete the simulated cases successfully to the level of the splenic fixture. In order to determine whether residents' performance improved on the simulator over time, data regarding procedure time, insertion length and retroflexion ability was calculated for each resident during the first 3 cases performed

on the simulator compared with the last 3 cases on the simulator (data not shown). None of the parameters significantly improved over time.

Skills transfer outcomes

Overall performance of patient-based procedure

The parameter overall performance was a global summary of all the performance parameters measured (during patient-based assessment procedures) within the studies.

Simulation-based training vs. no simulation-based training

Three studies reported on the overall performance in the clinical setting between simulation-trained and non-simulation-trained participants (Cohen *et al.* 2006b; Chaer *et al.* 2006; Edmond 2002) (Table 5).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Cohen *et al.* (2006b) reported objective and subjective competence results for each session for Simbionix GI Mentor™ trained individuals compared with controls. Participants who had received training on the Simbionix GI Mentor™ scored significantly higher in objective competence (the ability to reach the transverse colon and the caecum without assistance, and the ability to correctly recognise and identify abnormalities) during sessions two ($p < 0.0001$), three ($p < 0.0001$), four ($p < 0.0001$), and five ($p = 0.03$) of the 10 session assessment. For subjective competence (measured on a 5-point scale; 1, totally unskilled to 5, competent and expedient), the trained individuals scored significantly better in sessions two ($p = 0.004$) and three ($p = 0.005$) of the 10 session assessment.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported on overall performance between Procedicus VIST™ simulator-trained and untrained participants for two catheter-based interventions for occlusive vascular disease. The trained group scored significantly higher than controls during the first ($p = 0.0015$) and second ($p = 0.0006$) endovascular intervention. A subjective evaluation of resident performance was also conducted and trained participants scored higher overall on the global rating scale of endovascular performance for the first ($p = 0.0052$) and second ($p = 0.0015$) intervention. Resident performance did not improve from the first to the second intervention for either the simulator-trained group or the untrained group.

Table 5 Patient-based assessments: overall performance results

Simulation-based training *vs.* no simulation-based training											
Colonoscopy/sigmoidoscopy											
Cohen *et al.* 2006b Level II	**Intervention**	**N = 49***	**Mean objective competence† for each session**								
			1	2	3	4	5	6	7	8	9	10
	Simbionix GI Mentor™	22	50.4	64.5	74.0	76.7	76.8	77.8	80.8	89.5	87.8	92.7
	No training	23	40.9	52.0	62.0	64.4	70.2	77.6	80.5	83.7	85.2	90.9
	P-value		0.06	<0.0001	<0.0001	<0.0001	0.03	0.91	0.89	0.01	0.02	0.04
	Intervention	**N = 49***	**Mean subjective competence‡ for each session**								
			1	2	3	4	5	6	7	8	9	10
	Simbionix GI Mentor™	22	47.6	68.6	76.3	78.0	81.3	82.0	86.1	88.8	88.9	90.8
	No training	23	36.6	57.4	68.4	75.4	79.4	82.3	84.1	86.4	86.8	90.5
	P-value		0.08	0.004	0.005	0.32	0.28	0.88	0.32	0.11	0.32	0.82
Catheter-based intervention for occlusive vascular disease											
Chaer *et al.* 2006 Level II	**Intervention**	**N = 20**	**Overall assessment, mean check list scores ± SD**								
			Intervention 1					**Intervention 2**			
	Procedicus VIST™	10	50 ± 6					53 ± 6			
	No training	10	33 ± 9					36 ± 7			
	P-value		0.0015					0.0006			
	Intervention	**N = 20**	**Overall assessment, mean global rating scale scores ± SD**								
			Intervention 1					**Intervention 2**			
	Procedicus VIST™	10	30 ± 7					33 ± 6			
	No training	10	19 ± 5					21 ± 6			
	P-value		0.0052					0.0015			
Endoscopic sinus surgery											
Edmond 2002 Level III-2	**Intervention**	**N = 4**	**Overall rating, mean ± SD**								
			Baseline					**After training**			
	ESS simulator	2	NA					6.7 ± 1.1			
	No training	2	NA					4.0 ± 6.7E-02			
	P-value		NA					NS			
Simulation-based training *vs.* patient-based training											
Colonoscopy/sigmoidoscopy											
Gerson and van Dam 2003 Level II	**Intervention**	**N = 16**	**Mean score§ ± SEM**								
								After training			
	Sigmoidoscopy simulator	9	NA					2.9 ± 0.2			
	Patient-based training	7	NA					3.8 ± 0.2			
	P-value		NA					< 0.001			

* 4 participants lost to assessment; groups not specified.
† Objective competence is the ability to reach the transverse colon and the caecum without assistance, and the ability to correctly recognise and identify abnormalities.
‡ Measured on a 5 point scale; 1 (totally unskilled) to 5 (competent and expedient).
§ Measured on a scale of 1 (unable to clear the rectum) to 5 (independent examination less than 20 minutes in duration).
NA not applicable
NS not significant
NR not reported
SEM standard error of the mean
SD standard deviation

Endoscopic sinus surgery

Non-randomised comparative studies

Edmond (2002) reported on the overall performance between endoscopic sinus simulator-trained and untrained participants for endoscopic sinus surgery. The two simulation-trained residents were rated better than the two control residents across all measures, but these differences were not significant.

Simulation-based training vs. patient-based training

One RCT reported on the overall performance on simulation-trained *vs.* patient-trained participants (Gerson and van Dam 2003) (Table 5).

Sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) reported on the overall performance between VR sigmoidoscopy-trained and patient-trained participants for sigmoidoscopy. The mean score in patient-trained participants was significantly better than the simulator-trained participants ($p < 0.001$).

Performance time

Performance time was reported as the time taken, in minutes or seconds, to conduct the patient-based assessment procedures.

Simulation-based training vs. no simulation-based training

Five studies reported performance times between simulator-trained and untrained participants for five different interventions (Grantcharov *et al.* 2004; Seymour *et al.* 2002; Ahlberg *et al.* 2005; Tuggy 1998; Sedlack and Kolars 2004) (Table 6).

Laparoscopic cholecystectomy

Randomised controlled trials

Two studies reported procedure or dissection times for laparoscopic cholecystectomy after MIST-VR trained participants and participants who did not have this training (Grantcharov *et al.* 2004; Seymour *et al.* 2002). Grantcharov *et al.* (2004) assessed the laparoscopic cholecystectomy procedure from the point at which the clips were applied to the cystic artery and cystic duct, to the dissection of the gallbladder from the liver bed. The study found that participants who received MIST-VR training performed two such laparoscopic cholecystectomy procedures significantly faster than those in the control group ($p = 0.021$). Seymour *et al.* (2002) also assessed the clip and cut part of the laparoscopic cholecystectomy procedure and found that the duration of the dissection for the MIST-VR trained group was 29% less than the standard training group, but this result was not significant.

Table 6 Patient-based assessments: performance time results

Simulation-based training *vs.* no simulation-based training						
Laparoscopic cholecystectomy						
Grantcharov *et al.* 2004	**Intervention**	**N = 20***	**Duration of procedure (minutes), median (range)**			
			Baseline	**Final**		
	MIST-VR	20	62 (46 - 80)	53 (46 - 62)		
Level II	No training	20	55 (43 - 58)	57 (43 - 74)		
	P-value		NS	0.021		
Seymour *et al.* 2002	**Intervention**	**N = 16**	**Duration of dissection (minutes)**			
			Baseline	**Final**		
Level II	MIST-VR	8	NA	14.5		
	No training	8	NA	20.5		
	P-value		NA	NS		
Colonoscopy/sigmoidoscopy						
Ahlberg *et al.* 2005	**Intervention**	**N = 12**	**Time to reach caecum (minutes), median, (IQR)**			
			Baseline	**Final**		
	AccuTouch®	6	NA	30 (17 – 38)		
Level II	No training	6	NA	40 (25 – 45)		
	P-value		NA	0.008		
Sedlack and Kolars 2004	**Intervention**	**N = 8**	**Time to reach maximum insertion (minutes), median (IQR)**			
			Baseline	**Final**		
				C† 1-15	C† 16-30	C† 31-45	C† 46-60
Level II	AccuTouch®	4	NA	23(19–30)	21(19–28)	21(18–28)	17(12–25)
	No training	4	NA	25(20-30)	22(15-30)	20(15-30)	20(15-27)
	P-value		NA	0.155	0.947	0.321	0.090
Tuggy 1998	**Intervention**	**N = 10**	**Total examination time (seconds)**			
			After 5 hr training	**After 10 hr training**		
Level II	Gastro-Sim®	5	530	323		
	No training	5	654	654		
	P-value		0.31	0.01		
	Intervention	**N = 10**	**Time to reach 30 cm (seconds)**			
			After 5 hr training	**After 10 hr training**		
	Gastro-Sim®	5	286	119		
	No training	5	357	357		
	P-value		0.52	0.03		
	Intervention	**N = 10**	**Time to reach 40 cm (seconds)**			
			After 5 hr training	**After 10 hr training**		
	Gastro-Sim®	5	341	211		
	No training	5	518	518		
	P-value		0.27	0.03		
Simulation-based training *vs.* patient-based training						
Colonoscopy/sigmoidoscopy						
Gerson and van Dam 2003	**Intervention**	**N = 16**	**Time per case, (minutes), mean ± SEM**			
			Baseline	**Final**		
Level II	Sigmoidoscopy simulator	9	NA	24 ± 1.0		
	Patient-based training	7	NA	24 ± 1.1		
	P-value		NA	> 0.05		

* 4 participants lost to assessment; Trained group n = 2, Control group n = 2.
† Colonoscopies grouped in blocks of 15
IQR interquartile range
NA not applicable
NS not significant
NR not reported
SEM standard error of the mean

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Ahlberg *et al.* (2005) reported the time required to reach the caecum during colonoscopy for participants that had been trained on the AccuTouch® endoscopy simulator compared with those that had not received this training. There was a significant decrease in procedure time in favour of the simulator-trained group ($p = 0.008$). Sedlack and Kolars (2004) reported the time to reach maximum insertion during patient-based colonoscopy after AccuTouch® colonoscopy simulator training or no simulator training, and found no significant difference between the two groups. Tuggy (1998) reported total examination time for patient-based sigmoidoscopy for participants that had been trained on the Gastro-Sim® flexible sigmoidoscopy simulator for five and 10 hours compared with those that had not received this training. There were no significant differences between the five hour-trained group and the untrained group for the following parameters: time to reach 30 cm, time to reach 40 cm, and total examination time. After 10 hours of training on the simulator, the simulator-trained group achieved significantly faster insertion times to 30 cm ($p = 0.03$), and 40 cm ($p = 0.03$), and a shorter mean length of examination ($p = 0.01$).

Simulation-based training vs. patient-based training

One study reported time outcomes for participants that had undergone simulator-training in comparison to participants that had received patient-based training (Gerson and van Dam 2003) (Table 6).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) reported the time taken per sigmoidoscopy after participants had been trained either using a VR sigmoidoscopy simulator or by patient-based methods and found no significant difference between the two groups.

Success rate of patient-based assessment

Success rate was described as the percentage of participants able to complete the patient-based assessment as specified in the study methods.

Simulation-based training vs. no simulation-based training

Two studies reported on the success rate of patient-based assessment operations (Seymour *et al.* 2002; Ahlberg *et al.* 2005) (Table 7).

Laparoscopic cholecystectomy

Randomised controlled trials

Seymour *et al.* (2002) reported that all participants in both the MIST-VR-trained group and the control group successfully completed the dissection of the gallbladder from the liver bed during laparoscopic cholecystectomy.

Table 7 Patient-based assessments: success rate results

Simulation-based training *vs.* no simulation-based training			
Laparoscopic cholecystectomy			
Seymour *et al.* 2002	**Intervention**	**N = 16**	**Success rate (%)**
	MIST-VR	8	100
	No training	8	100
Level II	*P*-value		NR
Colonoscopy/sigmoidoscopy			
Ahlberg *et al.* 2005	**Intervention**	**N = 12**	**Success rate* (%)**
	AccuTouch®	6	52
	No training	6	19
Level II	*P*-value		0.0011†

* Success rate defined as intubation of the caecum within the given time limits.

† In addition, patient gender significantly (p = 0.016) affected success rate. It was 3 times more likely that a colonoscopy would be successful in a male patient than a female patient. The number of previously performed gastroscopies influenced success rate (p = 0.006) equally in both groups. It was 3.76 times more likely that a participant with a precious experience exceeding 50 gastroscopies would succeed with colonoscopy.

NR Not reported

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Ahlberg *et al.* (2005) reported a significant increase in success rate for trainees that had trained on the AccuTouch® endoscopy simulator in comparison to controls who did not have any training (p = 0.0011).

Ability to complete procedure independently

The ability of a participant to complete the case independently referred to the ability of a participant to conduct the patient-based assessment without assistance. Studies that used a global assessment form developed by Reznick *et al.* (1997) (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001) described a parameter they called overall performance as the ability of a participant to complete the patient-based assessment procedure independently.

Simulation-based training vs. no simulation-based training

Six studies reported results in relation to the ability of participants to complete the patient-based assessment operation independently (Scott *et al.* 2000; Scott *et al.* 1999; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Chaer *et al.* 2006; Hamilton *et al.* 2001) (Table 8).

Table 8 Patient-based assessments: ability to complete the assessment operation

Simulation-based training *vs.* no simulation-based training						
Laparoscopic cholecystectomy						
Scott *et al.* 2000	**Intervention**	**N = 27***	**Overall performance†, median global assessment scores (25th – 75th percentile)**			
			Baseline		**Adjusted improvement‡**	
Level II	SCMIS GEM	13	2.5 (2.0 – 3.0)		0.7 (0.6 – 1.0)	
	No training	9	3.0 (2.3 – 3.0)		0.2 (-0.5 – 0.6)	
	P-value		0.581		0.007	
Scott *et al.* 1999	**Intervention**	**N = 22**	**Overall performance†, mean global assessment score ± SD**			
			Baseline		**Final**	
Level II	Video-trainer	9	NR		3.4 ± 0.3	
	No training	13	NR		2.8 ± 0.7	
	P-value		NS		< 0.05	
Colonoscopy/sigmoidoscopy						
Sedlack and Kolars 2004	**Intervention**	**N = 8**	**Resident completed independently (%) (IQR)**			
			C§ 1-15 (%)	**C§16-30 (%)**	**C§ 31-45 (%)**	**C§ 46-60 (%)**
Level II	AccuTouch®	4	38 (27-50)	59 (46-71)	50 (37-63)	72 (59-84)
	No training	4	20 (9-31)	34 (21-46)	50 (37-63)	58 (46-71)
	P-value		0.027	0.007	1.000	0.128
Sedlack *et al.* 2004	**Intervention**	**N = 38**	**Competent to perform endoscopy independently, median score‖ (IQR)**			
			Staff-assessed		**Self-assessed**	
Level II	AccuTouch®	19	8 (7 – 9)		8 (5 – 9)	
	No training	19	8 (7 – 9)		8 (4 – 10)	
	P-value		0.893		0.791	
Catheter-based intervention for occlusive vascular disease						
Chaer *et al.* 2006	**Intervention**	**N = 20**	**Ability to complete the case¶, mean global rating scale score**			
			Intervention 1		**Intervention 2**	
Level II	Procedicus VIST™	10	2.4		2.6	
	No training	10	1.4		1.4	
	P-value		0.03		0.01	
TEP hernia repair						
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Overall performance†, mean global assessment score ± SD**			
			Baseline		**Final**	
Level II	TEP hernia repair††	10	2.4 ± 0.8		3.6 ± 0.7	
	No training	11	1.9 ± 0.7		2.4 ± 0.9	
	P-value		NS		< 0.05	
	Intervention	**N = 21**	**Composite score** (%) ± SD**			
			Baseline		**Final**	
	TEP hernia repair††	10	44.6 ± 24.6		65.7 ± 15.5	
	No training	11	29.6 ± 15.7		41.0 ± 23.5	
	P-value		NS		NS	
Simulation-based training *vs.* patient-based training						
Colonoscopy/sigmoidoscopy						
Gerson and van Dam 2003	**Intervention**	**N = 16**	**Resident completed independently n/N (%)**			
	Sigmoidoscopy simulator	9	10/32 (29%)			
Level II	Patient-based training	7	23/32 (72%)			
	P-value		0.001			

* 5 participants lost to assessment; Trained group n = 4, Control group n = 1.
† The parameter called 'overall performance' was measured on a scale of 1 (unable to perform operation independently) to 5 (clearly superior, able to perform operation independently with confidence).
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Colonoscopies grouped in blocks of 15.
‖ Measured on a scale of 1 (strongly agree) to 10 (strongly disagree).
¶ Measured on a scale of 0 (not able to complete the case) to 4 (able to complete the case independently).
** Calculation of composite score not reported.
†† Including instructional video and interactive CD ROM.
IQR interquartile range, SD standard deviation, NR not reported, NS not significant

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported on the ability of participants to complete the case (referred to as overall performance) between SCMIS GEM trained and untrained participants for laparoscopic cholecystectomy. They found that the trained group had significantly better adjusted improvement scores than controls ($p = 0.007$). A similar RCT reported on the ability to complete the case (referred to as overall performance) between participants who were trained using a video-trainer compared with untrained participants for laparoscopic cholecystectomy (Scott *et al.* 1999). After the training period, there was a significant improvement in overall performance (the ability to complete the case) for participants who had used the video-trainer as opposed to untrained controls ($p < 0.05$).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Sedlack and Kolars (2004) reported the percentage of colonoscopies completed independently for participants who had AccuTouch® colonoscopy simulator training in comparison with participants did not have this training. Simulator-trained participants performed significantly better than non-simulator trained participants during their first 15 colonoscopies ($p = 0.027$) and the following 15 colonoscopies ($p = 0.007$). There were no significant differences for the two blocks of colonoscopies thereafter.

Sedlack *et al.* (2004) reported staff-assessed scores as well as participant self-assessed scores for being competent to perform endoscopy independently and found that there were no significant differences between AccuTouch® colonoscopy simulator-trained and untrained scores as well as no differences between staff-assessed and self-assessed scores.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported on the ability of participants to complete two catheter-based interventions for occlusive vascular disease after Procedicus VIST™ training or after no simulation-based training. A significant difference in favour of the simulator-trained group was seen for the first ($p = 0.03$) and second ($p = 0.01$) intervention.

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) reported on the ability to complete the case (referred to overall performance) between TEP hernia repair curriculum-trained and non-simulator trained participants for TEP hernia repair. After training, the trained group had significantly higher scores for the ability to complete the case than did controls ($p <$

0.05). There was no significant difference in composite scores between the two groups after the training period. When conducting an internal group comparison, both the trained and untrained groups demonstrated significant improvements in overall performance and composite scores after the training period ($p < 0.05$), but the post-training scores for the control group were significantly lower than the trained group (p value not reported).

Simulation-based training vs. patient-based training

One study reported on the ability of participants to complete live-patient assessment procedures independently after either simulator-training or patient-based training (Gerson and van Dam 2003).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) reported a significant improvement in individuals who had performed 10 live-patient sigmoidoscopic examinations with an attending gastroenterologist compared with individuals who had VR sigmoidoscopy simulator training prior to conducting assessment sigmoidoscopies ($p = 0.001$).

Supervising surgeon takeover

Supervising surgeon takeover (also called attending takeover) referred to instances where the supervising attending surgeon was required to take control of the procedure when a participant encountered difficulties.

Simulation-based training vs. no simulation-based training

Two RCTs reported on the need for assistance and supervising surgeon takeover between simulator-trained and non-simulator-trained participants when conducting live patient assessment (Seymour *et al.* 2002; Chaer *et al.* 2006) (Table 9).

Laparoscopic cholecystectomy

Randomised controlled trials

Seymour *et al.* (2002) found that the attending surgeon had to take over for six out of eight control trainees, compared with zero out of eight trainees who had trained on the MIST-VR simulator.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported a significantly better mean global rating score in supervising surgeon takeover for the two patient-based assessed interventions after participants had Procedicus VIST™ training compared with no simulator training (intervention one, $p = 0.003$ and intervention two $p = 0.006$). Likewise, the requirement for verbal prompts was significantly better in the simulator-trained group (intervention one, $p = 0.03$ and intervention two, $p = 0.01$).

Simulation-based training vs. patient-based training

One study reported on supervising surgeon takeover for participants that had either received training via a simulator, or via patient-based methods (Gerson and van Dam 2003) (Table 9).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) reported that significantly fewer participants who had patient-based training required assistance to perform a sigmoidoscopy than those in the VR sigmoidoscopy simulator-trained group (p = 0.001).

Table 9 Patient-based assessments: supervising surgeon takeover and need for assistance

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Seymour *et al.* 2002	**Intervention**	**N = 16**	**Attending takeover* (n)**	
			Baseline	**After training**
	MIST-VR	8	NA	0
Level II	No training	8	NA	6
	P-value		NA	NR
Catheter-based intervention for occlusive vascular disease				
Chaer *et al.* 2006	**Intervention**	**N = 20**	**Attending takeover†, mean global rating scale score**	
			Intervention 1	**Intervention 2**
	Procedicus VIST™	10	2.6	2.9
Level II	No training	10	1.4	1.7
	P-value		0.003	0.006
	Intervention	**N = 20**	**Need for verbal prompts‡, mean global rating scale score**	
			Intervention 1	**Intervention 2**
	Procedicus VIST™	10	2.3	2.4
	No training	10	1.0	1.4
	P-value		0.03	0.01
Simulation-based training *vs.* patient-based training				
Colonoscopy/sigmoidoscopy				
Gerson and van Dam 2003	**Intervention**	**N = 16**	**Required assistance§, n/N (%)**	
			NA	**After training**
	Sigmoidoscopy simulator	9	NA	24/34 (71%)
Level II	Patient-based training	7	NA	9/32 (28%)
	P-value		NA	0.001

* The supervising attending surgeons takes the dissecting instrument (right hand) or retracting instrument (left hand) from the resident and performs a component of the procedure.

† Measured on a scale of 0 (occurred at every stage) to 4 (ability to complete the case without attending takeover).

‡ Measured on a scale of 0 (repeatedly needed verbal prompts) to 4 (able to complete the case without prompts).

§ If the participant encountered difficulties, the attending physician was allowed to takeover the instrument in order to pass the area of difficulty and then was expected to return the colonoscope to the participant.

NA not applicable

NR not reported

Use of assistants

Simulation-based training vs. no simulation-based training

Three studies reported on the appropriate use of assistants during the assessment operation after simulation-based training or no simulation-based training (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001) (Table 10).

Table 10 Patient-based assessments: the use of assistants

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000	**Intervention**	**N = 27***	**Use of assistants†, median global assessment score (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
Level II	SCMIS GEM	13	2.0 (1.5 - 2.8)	1.0 (0.8 – 1.6)
	No training	9	2.0 (1.5 – 3.0)	0.7 (-0.4 – 1.0)
	P-value		0.785	0.035
Scott *et al.* 1999	**Intervention**	**N = 22**	**Use of assistants*, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	Video-trainer	9	NR	3.3 ± 0.6
	No training	13	NR	2.8 ± 0.7
	P-value		NS	< 0.05
TEP hernia repair				
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Use of assistants*, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	TEP hernia repair§	11	2.5 ± 1.3	3.5 ± 1.1
	No training	10	1.6 ± 0.7	2.1 ± 1.1
	P-value		NS	< 0.05

* 5 lost to assessment; Trained group n = 4, Control group n = 1.
† Measured on a scale of 1 (failed to use assistants) to 5 (strategically used assistants to the best advantage at all times).
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Including an instructional video and interactive CD ROM.
NS not significant
NR not reported
SD standard deviation

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported a significant difference in favour of the SCMIS GEM trained group compared with controls in relation to use of assistants. SCMIS GEM trained participants performed significantly better in relation to the use of assistants than did untrained controls when performing laparoscopic cholecystectomy ($p = 0.035$).

Scott *et al.* (1999) reported a significant difference in favour of the video-trainer trained group compared with controls in relation to the use of assistants. Video-trainer trained participants scored significantly higher in relation to the use of assistants than did untrained controls when performing laparoscopic cholecystectomy ($p < 0.05$).

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) reported a significant difference between baseline and mean global assessment scores for TEP hernia repair curriculum-trained and untrained groups after training compared with untrained controls for patient-based TEP hernia repair ($p < 0.05$) for use of assistants.

Performance errors

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

Simulation-based training vs. no simulation-based training

Two studies reported error data for simulator-trained and non-simulator-trained participants for laparoscopic cholecystectomy (Seymour *et al.* 2002; Tuggy 1998) (Table 11).

Table 11 Patient-based assessments: errors made during assessment operations

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Seymour *et al.* 2002	**Intervention**	**N = 16**	**Total errors*, mean**	
			Baseline	**Final**
Level II	MIST-VR	8	NA	1.19
	No training	8	NA	7.38
	P-value		NA	< 0.006
Colonoscopy/sigmoidoscopy				
Tuggy 1998	**Intervention**	**N = 10**	**Directional errors† (n)**	
			After 5 hrs simulation	**After 10 hrs simulation**
Level II	Gastro-Sim®	5	2.8	1.6
	No training	5	8.6	8.6
	P-value		0.01	< 0.01

* 8 events associated with the excisional phase of procedure defined as errors and chosen as study measurements: lack of progress, gallbladder injury, liver injury, incorrect plane dissection, burn non-target tissue, tearing tissue, instrument out of view and attending takeover.

† Defined as the inability to direct the sigmoidoscope correctly toward the lumen when it was visualised. Measurements taken from videotaped examinations.

NA not applicable

Laparoscopic cholecystectomy

Randomised controlled trials

Seymour *et al.* (2002) showed a significant improvement in total mean errors from baseline for MIST-VR trained residents compared with controls who did not have training before performing laparoscopic cholecystectomy ($p < 0.006$). This study found that the group without simulator training had significantly more errors in relation to lack of progress ($p < 0.008$), gall bladder injury ($p < 0.039$), and burning of non-target tissue ($p < 0.039$) (data not shown). There were no significant

differences for liver injury, incorrect plane of dissection, tearing tissue, instrument out of view, and non-contact cautery injury.

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Tuggy (1998) assessed hand-eye skills by the amount of directional errors that were made during a sigmoidoscopy examination. Directional errors were defined as the inability of the examiner to direct the sigmoidoscope correctly toward the lumen when it was visualised. The study showed significant reductions in directional errors in residents after five hours ($p = 0.01$) of training in the Gastro-Sim® flexible sigmoidoscopy simulator as well as after 10 hours ($p < 0.01$) of training on the simulator as compared with controls who had no simulation-based training.

Errors - combined outcomes

One study reported combined outcomes for errors (Table 12). These outcomes could not be separated so could not be reported with the other error results.

Simulation-based training vs. no simulation-based training

Laparoscopic cholecystectomy

Randomised controlled trials

Grantcharov *et al.* (2004) used a modified version of a global assessment form developed by Reznick *et al.* (1997). Errors were assessed using a combination of a parameter from the original publication, respect for tissue, and a new parameter called precision of operative technique. The study found that MIST-VR trained participants showed significantly greater improvement in their error scores than controls who had not received simulation-based training prior to performing laparoscopic cholecystectomy ($p = 0.003$).

Table 12 Patient-based assessments: error scores – combined outcomes

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Grantcharov *et al.* 2004	**Intervention**	**N = 20***	**Error† scores, median (range)**	
			Baseline	**Final**
Level II	MIST-VR	8	6.8 (4 – 9)	3 (3 – 6)
	No training	8	6 (3 – 9)	5.8 (4 – 9)
	P-value		NS	0.003

* 4 lost to assessment; Trained group n = 2, Control group n = 2.

† Error was a combination of respect for tissue and precision of operative technique from Reznick *et al.* (1997) global assessment form. It was measured on a scale of 1 (consistently handled tissue appropriately with minimal damage, and fluent, secure and correct technique in all stages of the operative procedure) to 5 (frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments, and imprecise, wrong technique in approaching operative intentions).

NS not significant

Flow of operation

Flow of operation referred to the ability of a participant to move continuously and fluently through the procedure, confident of each step.

Simulation-based training vs. no simulation-based training

Five studies reported results on the flow of the procedure for simulator-trained and non-simulator-trained participants (Scott *et al.* 2000; Scott *et al.* 1999; Schijven *et al.* 2005; Chaer *et al.* 2006; Hamilton *et al.* 2001) (Table 13).

Table 13 Patient-based assessments: flow of operation and economy of movement

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000 Level II	**Intervention**	**N = 27***	**Flow of operation†, median global assessment score (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
	SCMIS GEM	13	2.0 (2.0 – 3.2)	1.0 (0.6 – 1.2)
	No training	9	3.0 (2.0 – 3.2)	0.4 (-0.5 – 1.2)
	P-value		0.356	0.090
Scott *et al.* 1999 Level II	**Intervention**	**N = 22**	**Flow of operation† mean global assessment score ± SD**	
			Baseline	**Final**
	Video-trainer	9	NR	3.5 ± 0.4
	No training	13	NR	3.0 ± 0.8
	P-value		NS	NS
Schijven *et al.* 2005 III-2	**Intervention**	**N = 24§**	**Fluency rating‖**	
			Observer 1	**Observer 2**
	Training course¶	10	NR	NR
	No training	10	NR	NR
	P-value		NS	0.0037
Catheter-based intervention for occlusive vascular disease				
Chaer *et al.* 2006 Level II	**Intervention**	**N = 20**	**Flow of operation** mean global rating scale score**	
			Intervention 1	**Intervention 2**
	Procedicus VIST™	10	2.4	2.8
	No training	10	1.4	1.2
	P-value		NS	0.002
TEP hernia repair				
Hamilton *et al.* 2001 Level II	**Intervention**	**N = 21**	**Flow of operation†, mean ± SD**	
			Baseline	**Final**
	TEP hernia repair††	10	2.8 ± 1	3.7 ± 0.7
	No training	11	2.1 ± 0.8	2.6 ± 1.1
	P-value		NS	< 0.05

* 5 lost to assessment; Trained group n = 4, Control group n = 1.
† Measured on a scale of 1 (frequently stopped operating and seemed unsure of next move) to 5 (planned course of operation effortless flow from one move to the next).
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ 4 lost to assessment; Trained group n = 2, Control group n = 2.
‖ Measured on a scale of 0 (completely disagree) to 5 (completely agree) to statement: Pattern of movement is fluent, precise and efficient (few unnecessary/random movements).
¶ The training course consisted of a variety of teaching elements, including videos, oral presentations, table sessions, instrument displays and repetitive sessions of VR software simulations using the Xitact LS500 laparoscopy simulator platform. Both psychomotor VR simulation (MIST-VR) and procedural laparoscopic cholecystectomy simulation, including the clip-and-cut, navigation and dissection modules (Xitact) were featured.
** Measured on a scale of 0 (frequently stopped; seemed unaware of next move) to 4 (obviously planned course; effortless flow).
†† Including an instructional video and interactive CD ROM.
NS not significant
NR not reported
SD standard deviation

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported no significant difference between the median global assessment scores for the flow of laparoscopic cholecystectomy for individuals who had received SCMIS GEM training versus participants who had not had this training. Scott *et al.* (1999) also reported no significant differences in mean global assessment scores for the flow of the cholecystectomy procedure for individuals who had received video-trainer training versus participants who had not had this training.

Non-randomised comparative studies

Schijven *et al.* 2005 reported a significant improvement in fluency, as rated by one of two observers, during laparoscopic cholecystectomy for participants of a four day laparoscopic cholecystectomy training course involving a variety of teaching modalities, in comparison with the control group who did not attend this training ($p = 0.0037$). The second observer did not find a significant difference for this parameter.

Catheter-based interventions for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported a significant improvement in the flow of the operation for the second of two catheter-based interventions for occlusive vascular disease after Procedicus VIST™ training compared with controls who had no training with this device ($p = 0.002$).

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) reported a significant within-group difference for the flow of the operation between pre-training and post-training scores ($p < 0.05$) in the training group after performing a TEP hernia repair procedure, and a significant between-group difference between trained and untrained groups after training ($p < 0.05$) after performing a TEP hernia repair procedure.

Time and motion

Time and motion was reported in studies that used the global assessment form developed and validated by Reznick *et al.* (1997) (Scott *et al.* 2000; Scott *et al.* 1999; Chaer *et al.* 2006; Hamilton *et al.* 2001). This parameter measured a participant's ability to conduct the patient-based assessment procedure with efficiency and economy of movement.

Simulation-based training vs. no simulation-based training

Four studies reported time and motion for participants that received training via simulation compared with participants who did not receive simulation-based training (Scott *et al.* 2000; Scott *et al.* 1999; Chaer *et al.* 2006; Hamilton *et al.* 2001) (Table 14).

Table 14 Patient-based assessments: time and motion outcomes

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000	**Intervention**	**N = 27***	**Time and motion†, median global assessment score (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
	SCMIS GEM	13	3.0 (2.0 - 3.0)	0.3 (0.1 – 0.8)
Level II	No training	9	2.5 (2.0 - 3.0)	-0.3 (0.1 – 0.8)
	P-value		0.781	0.075
Scott *et al.* 1999	**Intervention**	**N = 22**	**Time and motion†, mean global assessment score ± SD**	
			Baseline	**Final**
	Video-trainer	9	NR	3.0 ± 0.4
Level II	No training	13	NR	2.6 ± 0.7
	P-value		NS	< 0.05
Catheter-based intervention for occlusive vascular disease				
Chaer *et al.* 2006	**Intervention**	**N = 20**	**Time and motion§, mean global rating scale scores**	
			Procedure 1	**Procedure 2**
	Procedicus VIST™	10	2.3	2.6
Level II	No training	10	1.4	1.7
	P-value		NS	0.01
TEP hernia repair				
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Time and motion†, mean global rating scale scores ± SD**	
			Baseline	**Final**
	TEP hernia repair‖	10	2.7 ± 1.1	3.5 ± 0.7
Level II	No training	11	1.8 ± 0.6	2.4 ± 0.9
	P-value		< 0.05	< 0.05

* 5 participants lost to assessment; Trained group n = 4, Control group n = 1.
† Measured on scale of 1 (many unnecessary moves) to 5 (clear economy of movement and maximum efficiency).
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Measured on a scale of 0 (many unnecessary moves) to 4 (clear economy of motion; maximum efficiency).
‖ Including an instructional video and interactive CD ROM.
NR not reported
NS not significant
SD standard deviation

Laparoscopic cholecystectomy

Randomised controlled trials

Two studies reported time and motion for participants undertaking laparoscopic cholecystectomy after receiving video-trainer training or no simulation-based training (Scott *et al.* 2000; Scott *et al.* 1999). Scott *et al.* (2000) reported no significant differences in median improvements in time and motion global assessment scores between the SCMIS GEM trained group and the control group for laparoscopic cholecystectomy. Scott *et al.* (1999) reported mean global assessment scores for time and motion for participants undertaking laparoscopic cholecystectomy after receiving video-trainer training compared with no simulation-based training. The trained residents scored significantly higher (ie better) than the controls ($p < 0.05$).

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported mean global rating scale scores for time and motion for participants that had undergone Procedicus VIST™ training and participants with no

simulator training. There was a significant difference in time and motion for the second intervention only (p = 0.01).

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) reported baseline and final mean differences for TEP hernia repair curriculum-trained participants and participants who did not receive this training, for patient-based TEP hernia repair. There was a significant difference between trained and untrained groups for time and motion prior to training ($p < 0.05$). In each group there was a significant difference between pre-training and post-training time and motion scores ($p < 0.05$). Between the groups, there was a significant difference for time and motion between trained and untrained groups after training in favour of the trained group ($p < 0.05$). For the group not trained, there was also a significant difference between pre-training and post-training scores for time and motion ($p < 0.05$) but the post-training score was significantly lower than the scores for the trained group (p value not reported).

Economy of movement - combined outcomes

One study reported combined outcomes for economy of movement (Table 15) which could not be separated and reported with the other time and motion data. (Grantcharov *et al.* 2004).

Table 15 Patient-based assessments: economy of movement – combined outcomes

Laparoscopic cholecystectomy				
Simulation-based training *vs.* no simulation-based training				
Grantcharov *et al.* 2004	**Intervention**	**N = 20***	**Economy of movement†, median rating (range)**	
			Baseline	**Final**
Level II	MIST-VR	8	5.8 (4.5 – 6)	3.3 (2 – 6)
	No training	8	6 (6 – 8)	6 (4.5 – 9)
	P-value		NS	0.003

* 4 lost to assessment. Trained group n = 2, Control group n = 2.

† Economy of movement was a combination of two parameters, unnecessary movements and confidence of movements (referred to as time and motion, and instrument handling in original global rating scale developed by Reznick *et al.* (1997)). Economy of movement was measured on a scale of 1 (clear economy of movement and maximum efficiency, and; fluent moves with instruments and no awkwardness) to 5 (many unnecessary moves, and; repeated tentative awkward or inappropriate moves with instruments).

NS not significant

Simulation-based training vs. no simulation-based training

Laparoscopic cholecystectomy

Randomised controlled trials

Economy of movement was a combination of two parameters, unnecessary movements and confidence of movements (referred to as time and motion, and instrument handling in the original published global rating scale developed by Reznick *et al.* (1997)). The study found significant improvements in economy of movement during laparoscopic cholecystectomy in participants who had received

MIST-VR training compared with controls who had not had this training ($p = 0.003$).

Procedural knowledge

The parameter procedural knowledge was reported in studies that used the global assessment form developed and validated by Reznick *et al.* (1997) (Scott *et al.* 2000; Scott *et al.* 1999; Chaer *et al.* 2006; Hamilton *et al.* 2001) and focussed on a participant's knowledge and familiarity with the patient-based assessment procedure.

Simulation-based training vs. no simulation-based training

Four studies reported on specific procedural knowledge of participants (Scott *et al.* 2000; Scott *et al.* 1999; Chaer *et al.* 2006; Hamilton *et al.* 2001) (Table 16).

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported no significant difference in adjusted improvement of knowledge of specific procedure for participants trained on SCMIS GEM compared with individuals with no simulation-based training before performing laparoscopic cholecystectomy. Similarly, Scott *et al.* (1999) reported no significant difference in mean global assessment scores for knowledge of specific procedure for participants trained on a video-trainer compared with individuals without this training prior to performing laparoscopic cholecystectomy.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006) reported a significant increase in knowledge of procedure for participants that had trained on the Procedicus VIST™ simulator compared with individuals that had no training on this device prior to conducting two catheter-based interventions for occlusive vascular disease ($p = 0.005$ for the second intervention only).

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) reported differences in participants trained with the TEP hernia repair curriculum in comparison with individuals without this training. There were significant differences in favour of the trained group between pre-training scores and post-training scores ($p < 0.05$). There was also a significant difference in favour of the trained group for knowledge of procedure between trained and untrained groups after training ($p < 0.05$).

Table 16 Patient-based assessments: knowledge of procedure

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000	**Intervention**	**N = 27***	**Knowledge of specific procedure†, median global assessment scores (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
Level II	SCMIS GEM	13	3.0 (2.5 – 3.2)	1.0 (0.4 – 1.3)
	No training	9	3.0 (2.5 – 3.6)	0.4 (0.0 – 1.1)
	P-value		0.433	0.100
Scott *et al.* 1999	**Intervention**	**N = 22**	**Knowledge of specific procedure†, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	Video-trainer	9	NR	3.8 + 0.6
	No training	13	NR	3.5 ± 0.9
	P-value		NS	NS
Catheter-based intervention for occlusive vascular disease				
Chaer *et al.* 2006	**Intervention**	**N = 20**	**Knowledge of procedure§, mean global rating score**	
			Intervention 1	**Intervention 2**
Level II	Procedicus VIST™	10	2.0	2.4
	No training	10	1.4	1.1
	P-value		NS	0.005
TEP hernia repair				
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Knowledge of procedure†, mean ± SD**	
			Baseline	**Final**
Level II	TEP hernia repair‖	10	2.5 ± 0.6	3.8 ± 0.9
	No training	11	2.0 ± 0.8	2.6 ± 0.9
	P-value		NS	< 0.05

* 5 lost to assessment; Trained group n = 4, Control group n = 1.
† Measured on a scale of 1 (required specific instruction at most steps) to 5 (familiar with all aspects of the operation).
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Measured on a scale of 0 (deficient knowledge) to 4 (familiar with all aspects of the procedure).
‖ Including an instructional video and interactive CD ROM.
NS not significant
NR not reported
SD standard deviation

Knowledge of instruments

Knowledge of instruments was described as the ability of a participant to use the correct instrument for the procedure, and was reported in three of the four studies that used the global assessment form developed and validated by Reznick *et al.* (1997) (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001).

Simulation-based training vs. no simulation-based training

Three studies reported on the knowledge of instruments between simulation-trained residents and residents that had no prior simulation-based training (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001) (Table 17).

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported a small additional improvement in knowledge of instruments between SCMIS GEM trained residents and residents that had no

simulation-based training prior to performing the final laparoscopic cholecystectomy that was of borderline significance (p = 0.058). Similarly, Scott *et al.* (1999) reported no significant differences in knowledge of instruments between video-trainer trained residents and residents that had no simulation-based training prior to performing the final live-patient laparoscopic cholecystectomy procedure.

Table 17 Patient-based assessments: knowledge of instruments*

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000	**Intervention**	**N = 27†**	**Knowledge of instruments, median global assessment score (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
Level II	SCMIS GEM	13	3.0 (2.3 – 3.3)	0.6 (0.5 – 1.5)
	No training	9	3.0 (2.5 – 3.5)	0.4 (0.0 – 1.0)
	P-value		0.433	0.058
Scott *et al.* 1999	**Intervention**	**N = 22**	**Knowledge of instruments, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	Video-trainer	9	NR	3.9 ± 0.6
	No training	13	NR	3.5 ± 0.9
	P-value		NS	NS
TEP hernia repair				
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Knowledge of instruments, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	TEP hernia repair§	10	3.0 ± 1.3	3.7 ± 0.8
	No training	11	2.9 ± 1.5	2.9 ± 0.4
	P-value		NS	0.05

* Measured on a scale of 1 (frequently asked for wrong instrument or used wrong instrument) to 5 (obviously familiar with instruments and their names).
† 5 lost to assessment; Trained group n = 4, Control group n = 1.
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Including an instructional video and interactive CD ROM.
NS not significant
NR not reported
SD standard deviation

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) found a significant difference in favour of the trained group in knowledge of instruments between trained and untrained groups after training using the TEP hernia repair curriculum for TEP hernia repair (p = 0.05).

Instrument handling

Instrument handling was described as using the appropriate instrument skilfully and correctly, and was again reported in three of the four studies that used the global assessment form developed and validated by Reznick *et al.* (1997) (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001).

Simulation-based training vs. no simulation-based training

Three studies reported on instrument handling between trained residents and residents that had no simulation-based training (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001) (Table 18).

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported a significant adjusted improvement in median global assessment scores for instrument handling when performing laparoscopic cholecystectomy in favour of the SCMIS GEM trained group compared with controls ($p = 0.005$). Scott *et al.* (1999) found a significant adjusted improvement in mean global assessment scores for instrument handling when simulator-trained participants performed laparoscopic cholecystectomy compared with untrained participants after training ($p < 0.05$).

Hamilton *et al.* (2001) found a significant difference between pre-training and post-training scores ($p \leq 0.05$), and a significant difference between trained and untrained groups after TEP hernia repair curriculum training ($p = 0.05$).

Table 18 Patient-based assessments: instrument handling*

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000	**Intervention**	**N = 27†**	**Instrument handling, median global assessment score (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
Level II	SCMIS GEM	13	3.0 (2.0 – 3.2)	0.6 (0.4 – 0.8)
	No training	9	3.0 (2.5 – 3.0)	0.3 (-0.4 – 0.3)
	P-value		0.782	0.005
Scott *et al.* 1999	**Intervention**	**N = 22**	**Instrument handling, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	Video-trainer	9	NR	3.4 ± 0.3
	No training	13	NR	2.9 ± 0.6
	P-value		NS	< 0.05
TEP hernia repair				
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Instrument handling, mean global assessment score ± SD**	
			Baseline	**Final**
Level II	TEP hernia repair§	10	2.9 ± 1.0	3.7 ± 0.7
	No training	11	2.1 ± 0.7	3.1 ± 0.9
	P-value		NS	0.05

* Measured on a scale of 1 (repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments) to 5 (fluid moves with instruments with no awkward moves).
† 5 lost to assessment; Trained group n = 4, Control group n = 1.
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Including an instructional video and interactive CD ROM.
NS not significant
NR not reported
SD standard deviation

Respect for tissue

Respect for tissue was defined as the reduction of unnecessary tissue damage during patient-based procedures.

Simulation-based training vs. no simulation-based training

Four studies reported on respect for tissue between trained residents and residents that had no simulation-based training (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001; Edmond 2002). Three of these studies used the global assessment form developed and validated by Reznick *et al.* (1997) (Scott *et al.* 2000; Scott *et al.* 1999; Hamilton *et al.* 2001) (Table 19).

Table 19 Patient-based assessments: respect for tissue

Simulation-based training *vs.* no simulation-based training				
Laparoscopic cholecystectomy				
Scott *et al.* 2000	**Intervention**	**N = 27***	**Respect for tissue†, median global assessment score (25th – 75th percentile)**	
			Baseline	**Adjusted improvement‡**
	SCMIS GEM	13	3.0 (2.8 - 3.2)	0.3 (0.3 – 0.5)
Level II	No training	9	3.0 (2.4 – 3.2)	0.1 (-0.6 – 0.5)
	P-value		0.445	0.035
Scott *et al.* 1999	**Intervention**	**N = 22**	**Respect for tissue†, mean global assessment score ± SD**	
			Baseline	**Final**
	Video-trainer	9	NR	3.3 ± 0.3
Level II	No training	13	NR	2.9 ± 0.6
	P-value		NS	< 0.05
TEP hernia repair				
Hamilton *et al.* 2001	**Intervention**	**N = 21**	**Respect for tissue†, mean global assessment score ± SD**	
			Baseline	**Final**
	TEP hernia repair§	10	3.1 ± 1.1	3.5 ± 0.9
	No training	11	2.7 ± 0.9	3.1 ± 1.0
	P-value		NS	NS
Endoscopic sinus surgery				
Edmond 2002	**Intervention**	**N = 4**	**Tissue respect‖ , mean rating ± SD**	
			Baseline	**After training**
Level III-2	ESS simulator	2	NA	5.5 ± 3.1
	No training	2	NA	2.8 ± 0.2
	P-value		NA	NS

* 5 lost to assessment, Trained group n = 4, Control group n = 1.
† Measured scale of 1 (frequently and unnecessary force on tissue or caused damage by inappropriate use of instruments) to 5 (consistently handled tissues appropriately with minimal damage).
‡ Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
§ Including an instructional video and interactive CD ROM.
‖ Measured on a scale with 1 (adequate) to 10 (perfect).
NA not applicable
NS not significant
NR not reported
SD standard deviation

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported respect for tissue for SCMIS GEM trained residents and non trained controls when performing laparoscopic cholecystectomy. There was a significant difference in median global assessment scores for respect for tissue in favour of the simulator-trained group (p = 0.035). Scott *et al.* (1999) reported respect for tissue for residents who used a video-trainer compared with non-trained controls

when performing laparoscopic cholecystectomy. There was a significant difference in mean global assessment scores in favour of the simulator-trained group ($p < 0.05$).

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) did not report a significant difference in respect for tissue for TEP hernia repair for participants that had undergone training with the TEP hernia repair curriculum compared with untrained controls when conducting TEP hernia repair procedures.

Endoscopic sinus surgery

Non-randomised comparative studies

Edmond (2002) found no significant difference in mean values for tissue respect for individuals that had trained using the ESS simulator training compared with the control group.

Ability to identify landmarks

During laparoscopic cholecystectomy, misidentification of the common duct, cystic duct, and in some cases the aberrant right hepatic duct can result in biliary injuries.

Simulation-based training vs. no simulation-based training

Two studies described the ability of simulator-trained and non-simulator trained participants to identify landmarks during patient-based assessment (Sedlack and Kolars 2004; Sedlack *et al.* 2004) (Table 20).

Table 20 Patient-based assessments: ability to identify landmarks

Simulation-based training *vs.* no simulation-based training						
Colonoscopy/sigmoidoscopy						
Sedlack and Kolars 2004 Level II	**Intervention**	**N = 8**	**Identifies landmarks, median score* (25th – 75th IQR)**			
			C† 1-15	**C† 16-30**	**C† 31-45**	**C† 46-60**
	AccuTouch®	4	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)
	No training	4	6.0 (5.0 – 7.0)	6.0 (5.0 – 7.0)	6.0 (6.0 – 7.0)	7.0 (5.5 – 7.0)
	P-value		0.041	0.044	0.166	0.439
Sedlack *et al.* 2004 Level II	**Intervention**	**N = 38**	**Identifies landmarks, median score‡ (25th – 75th IQR)**			
			Staff-evaluated		**Self-evaluated**	
	AccuTouch®	19	5 (4 – 7)		3 (2 – 4)	
	No training	19	5 (4 – 5)		4 (2 – 4)	
	P-value		0.715		0.362	

* Measured on a scale of 1 (strongly disagree) to 7 (strongly agree).
† Colonoscopies grouped in blocks of 15.
‡ Measured on a scale of 1 (strongly agree) to 10 (strongly disagree).
IQR interquartile range

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Sedlack and Kolars (2004) found that participants trained on the AccuTouch® colonoscopy simulator were able to identify landmarks significantly better during their first 30 patient-based colonoscopies than participants that had trained not using the simulator ($p = 0.041$ for colonoscopies 1 – 15; $p = 0.044$ for colonoscopies 16 – 30). There were no significant differences for the remaining colonoscopies.

Sedlack *et al.* 2004 reported no significant differences in the ability of AccuTouch® sigmoidoscopy simulator trained and non-simulator trained participants to identify landmarks during patient-based colonoscopies. They also found no significant differences between staff-evaluated scores and participant self-assessed scores.

Ability to insert scope safely

Simulation-based training vs. no simulation-based training

Two studies described outcomes for simulator-trained and non-simulator trained participants to insert the scope safely during patient-based colonoscopy assessment (Sedlack and Kolars 2004; Sedlack *et al.* 2004) (Table 21).

Table 21 Patient-based assessments: ability to insert scope safely

Simulation-based training *vs.* no simulation-based training						
Colonoscopy/sigmoidoscopy						
Sedlack and Kolars 2004			**Inserts in a safe manner, median score* (25th – 75th IQR)**			
	Intervention	**N = 8**	**C† 1-15**	**C† 16-30**	**C† 31-45**	**C† 46-60**
	AccuTouch®	4	7.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	7.0 (6.0 – 7.0)
Level II	No training	4	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)
	P-value		0.008	0.274	0.256	0.559
Sedlack *et al.* 2004	**Intervention**	**N = 38**	**Inserts scope safely, median score‡ (25th – 75th IQR)**			
			Staff-evaluated		**Self-evaluated**	
	AccuTouch®	19	3 (3 – 6)		2 (1 – 3)	
Level II	No training	19	3 (2 – 5)		2 (1 – 3)	
	P-value		0.336		0.792	

* Measured on a scale of 1 (strongly disagree) to 7 (strongly agree).
† Colonoscopies grouped in blocks of 15.
‡ Measured on a scale of 1 (strongly agree) to 10 (strongly disagree).
IQR interquartile range

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Sedlack and Kolars (2004) found that participants who had trained on the AccuTouch® colonoscopy simulator were able to insert the scope safer during their first 15 patient-based colonoscopies than participants that had not trained using the simulator ($p = 0.008$). There were no significant differences for the remaining colonoscopies. Sedlack *et al.* 2004 reported no significant differences in the ability of AccuTouch® sigmoidoscopy simulator trained and non-simulator trained

participants to insert the scope safely during patient-based colonoscopies. They also found no significant differences between staff-evaluated scores and participant self-assessed scores.

Ability to adequately visualise the mucosa on withdrawal

Adequately visualises mucosa on withdrawal was defined as the ability of a participant to manoeuvre the colonoscope or sigmoidoscope effectively inside the colon in order to view the mucosa.

Simulation-based training vs. no simulation-based training

Two studies described outcomes for simulator-trained and non-simulator trained participants on the ability to adequately visualise the mucosa on withdrawal during patient-based assessment (Sedlack and Kolars 2004; Sedlack *et al.* 2004) (Table 22).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Sedlack and Kolars (2004) found that participants who had trained on the AccuTouch® colonoscopy simulator were able to better visualise the mucosa on withdrawal during their first 15 patient-based colonoscopies than participants that had not trained using the simulator ($p = 0.009$). There were no significant differences for the remaining colonoscopies.

Sedlack *et al.* 2004 reported no significant differences in the ability of AccuTouch® sigmoidoscopy simulator-trained and non-simulator trained participants to adequately visualise mucosa on withdrawal during patient-based colonoscopies. They also found no significant differences between staff-evaluated scores and participant self-assessed scores.

Table 22 Patient-based assessments: visualisation of the mucosa on withdrawal

Simulation-based training *vs.* no simulation-based training						
Colonoscopy/sigmoidoscopy						
Sedlack and Kolars 2004 Level II	**Intervention**	**N = 8**	**Adequately visualises mucosa on withdrawal, median score* (25th – 75th IQR)**			
			C† 1-15	**C† 16-30**	**C† 31-45**	**C† 46-60**
	AccuTouch®	4	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 – 7.0)	7.0 (6.0 – 7.0)
	No training	4	6.0 (5.0 – 7.0)	6.0 (5.0 – 7.0)	6.0 (5.5 – 7.0)	6.0 (6.0 – 7.0)
	P-value		0.009	0.396	0.230	0.518
Sedlack *et al.* 2004 Level II	**Intervention**	**N = 38**	**Adequately visualises mucosa on withdrawal, median score‡ (25th – 75th IQR)**			
			Staff-evaluated		**Self-evaluated**	
	AccuTouch®	19	7 (3 – 8)		3 (2 – 5)	
	No training	19	5 (4 – 7)		3 (2 – 4)	
	P-value		0.330		0.880	

* Measured on a scale of 1 (strongly disagree) to 7 (strongly agree).
† Colonoscopies grouped in blocks of 15.
‡ Measured on a scale of 1 (strongly agree) to 10 (strongly disagree).
IQR interquartile range

Staff productivity

The number of colonoscopies/sigmoidoscopies a staff member was able to perform whilst trainees trained on the simulator was compared with the number of procedures performed whilst in the presence of trainees. This was used to determine whether training on a simulator could impact staff productivity.

Simulation-based training vs. no simulation-based training

Two studies measured the number of procedures staff were able to perform when not in the presence of trainees (ie when trainees were using the simulator) and in the presence of trainees (Sedlack and Kolars 2004; Sedlack *et al.* 2004).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Sedlack and Kolars (2004) reported that during the 2 half-days of AccuTouch® colonoscopy simulator training, faculty without an accompanying Fellow were able to complete an average of 8 colonoscopies (range 7–9) per half-day while participants worked with the simulator. Faculty of the group that started assessment straight away (ie were trained without the simulator) completed an average of 3.5 procedures (range 2–4) per half-day during the same initial training interval. This allowed an average of 9 additional colonoscopies to be performed by the staff endoscopists for each Fellow trained with the simulation curriculum. There was no significant difference between supervising staff volumes once simulator-trained participants began patient-based colonoscopy.

Sedlack *et al.* 2004 reported an increase in staff productivity as a result of simulator training. During the simulator training interval, non-teaching staff performed an average of 7 procedures per half day while operating independently compared with an average of less than 3 procedures when students began assessment procedures straight away (ie no simulation-based training). The 3 hour AccuTouch® sigmoidoscopy simulator training session allowed an average of 4 additional procedures to be performed by staff for each participant trained via the simulator.

Other patient-based assessment related outcomes

Other patient-based procedure related outcomes are described below. These outcomes were procedure-specific and could not be grouped, and therefore have been reported narratively.

Simulation-based training vs. no simulation-based training

Five studies reported patient-based procedure related outcomes between trained residents and residents that had no simulation-based training (Tuggy 1998; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Chaer *et al.* 2006; Edmond 2002).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

In Tuggy (1998), after 5 hours of Gastro-Sim® simulator training, there was a significant difference for quality of viewing 360° ($p = 0.05$) in favour of trained participants. There were significant improvements of hand-eye skill measures of the trained group in percentage of colon visualised ($p = 0.02$), and viewing quality of examination ($p = 0.03$) in the trained group after 10 hours of simulator training when compared with the control group's initial performance on live patients. There were no significant differences between the group that had 5 hours of simulator training and the untrained group for time in red-out and percentage of colon visualised. After 10 hours of training, there were no significant differences for time in red-out compared with the untrained group.

In Sedlack and Kolars (2004) an analysis of procedures broken down in chronological groups of 15 demonstrated that AccuTouch® colonoscopy simulator-trained Fellows scored significantly better in all parameters during the first 15 colonoscopies performed ($p < 0.05$) with the exception of time to reach maximum insertion. The simulation-trained Fellows inserted the endoscope further with a median depth score of 4.0 *vs.* 3.0 ($p = 0.003$) and reached the caecum independently in 38% of procedures compared with 20% ($p = 0.027$) during this initial training period. Simulator-trained Fellows continued to have a significantly greater depth of insertion throughout the first 30 colonoscopies ($p < 0.05$). Beyond 30 colonoscopies, there were no statistical differences in any of the measured parameters between the two groups.

Sedlack *et al.* 2004 compared sigmoidoscopy performance between participants that had AccuTouch® sigmoidoscopy simulator training before beginning patient-based procedures to those who did not have simulator training. Participants were evaluated at the end of the training period and participants also conducted a self-assessment. There were no significant differences between the simulator-trained or non-simulator trained groups for any of the staff-reported procedural skill scores. Resident evaluations demonstrated no significant differences between groups in any of the parameters measured.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

In Chaer *et al.* (2006) there were significant improvements in mean checklist scores for the following individual measures of performance for Procedicus VIST™ simulator-trained residents compared with non-simulator trained residents for the first intervention: advance wire atraumatically ($p = 0.05$), constantly visualise wire tip ($p = 0.005$), mount and advance catheter wire ($p = 0.01$), position imaging catheter ($p = 0.04$), advance balloon over wire ($p = 0.006$), centre balloon over stenosis ($p = 0.009$), balloon inflation ($p = 0.003$) and balloon pressure ($p = 0.003$). For the second intervention, there were significant improvements in mean checklist scores

for the following individual measures of performance for simulation-trained residents compared with non-simulator strained residents: Advance wire atraumatically ($p = 0.03$), constantly visualise wire tip ($p = 0.001$), position imaging catheter ($p = 0.03$), knowledge of anatomy ($p = 0.04$), walk catheter back over wire ($p = 0.05$), advance balloon over wire ($p = 0.02$), centre balloon over stenosis ($p = 0.003$), balloon inflation ($p = 0.003$), balloon pressure ($p =0.002$), walk balloon back over wire ($p = 0.006$), advance stent over wire ($p =0.01$), centre stent over stenosis ($p = 0.01$), accurately deploy stent ($p = 0.01$), walk stent shaft out over wire ($p = 0.006$), and completion angiogram ($p = 0.04$). For the first intervention, there were significant improvements in mean endovascular global rating scale scores for wire and catheter handling ($p = 0.002$), awareness of wire position ($p = 0.005$), and precision of wire/catheter technique ($p = 0.03$). For the second intervention, there were significant improvements in mean endovascular global rating scale scores for wire and catheter handling ($p = 0.009$), awareness of wire position ($p = 0.005$), wire stability ($p = 0.04$), fluoroscopy usage ($p = 0.003$) and precision of wire/catheter technique ($p = 0.005$). Resident performance did not improve from the first to the second intervention.

Endoscopic sinus surgery

Non-randomised comparative studies

In Edmond (2002), there were no significant differences between mean rating scores for any procedure-related outcomes for residents trained with the ESS simulator compared with those residents not trained with the simulator.

Simulator training vs. patient-based training

One study reported procedure-related outcomes between simulator-trained residents and patient-trained residents for sigmoidoscopy (Gerson and van Dam 2003).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

In Gerson and van Dam (2003) there was no significant improvement in flexure recognition, but there was a significant improvement in retroflexion completed ($p = 0.02$). With regards to patients' assessments of satisfaction between VR sigmoidoscopy simulator-trained and patient-trained residents, there were no significant differences in general care and technical competence.

Patient morbidity/mortality

Patient morbidity and mortality were regarded as complications or deaths as a result of procedures being conducted on patients by trained or untrained participants during each study period. There were no reported deaths in any of the studies. Studies instead focussed on complications that occurred at the time of the procedures.

Simulation-based training vs. no simulation-based training

Four studies reported mortality/morbidity outcomes for the patients used in assessment procedures (Seymour *et al.* 2002; Ahlberg *et al.* 2005; Sedlack *et al.* 2004; Chaer *et al.* 2006).

Laparoscopic cholecystectomy

Randomised controlled trials

Seymour *et al.* (2002) used an assessment technique that emphasised operative errors. They defined errors as specific events that represented significant deviations from optimal performance, but stated that they did not link these events with adverse outcomes or proximate causes.

Colonoscopy/sigmoidoscopy

Randomised controlled trail

No major complications were reported by Ahlberg *et al.* (2005). Sedlack *et al.* (2004) also reported no adverse events with any of the procedures during their study.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trail

Chaer *et al.* (2006) stated that no peri-operative complications developed because of the study.

Simulation-based training vs. patient-based training

One study reported mortality/morbidity outcomes for the patients used in assessment procedures (Gerson and van Dam 2003).

Colonoscopy/sigmoidoscopy

Randomised controlled trail

Gerson and van Dam (2003) stated that no adverse events occurred during the test examinations.

Patient discomfort

Patient discomfort was described as either the pain felt by the patient undergoing the procedure, or the pain felt by the patient as determined by the assessor.

Simulation-based training vs. no simulation-based training

Five studies reported patient discomfort outcomes for patients undergoing assessment procedures after participants had received simulation-based training or no simulation-based training (Ahlberg *et al.* 2005; Sedlack and Kolars 2004; Sedlack *et al.* 2004; Cohen *et al.* 2006b; Tuggy 1998) (Table 23).

Table 23 Patient-based assessments: patient discomfort

Simulation-based training *vs.* no simulation-based training											
Colonoscopy/sigmoidoscopy											
Ahlberg *et al.* 2005 Level II	**Intervention**	**N = 12**	**Patient assessed discomfort on visual analog scale, median (IQR)**								
	AccuTouch®	6	4 (2.5 – 6)								
	No training	6	5 (4 – 7)								
	P-value		0.02*								
Sedlack and Kolars 2004 Level II	**Intervention**	**N = 8**	**Patient reported pain score†, median score (25th-75th IQR)**								
			C‡ 1 – 15	**C‡ 16 – 30**	**C‡ 31 – 45**	**C‡ 46 – 60**					
	AccuTouch®	4	2.0(1.0-4.0)	2.0(1.0-4.0)	2.0(1.0-4.0)	1.5(1.0-4.0)					
	No training	4	4.0(1.5-5.0)	2.0(1.0-4.0)	2.5(1.0-4.3)	2.0(1.0-3.0)					
	P-value		0.019	0.343	0.531	0.731					
	Intervention	**N = 8**	**Responds appropriately to patient discomfort§ median score (25th-75th IQR)**								
			C‡ 1 – 15	**C‡ 16 – 30**	**C‡ 31 – 45**	**C‡ 46 – 60**					
	AccuTouch®	4	6.5 (6.0 – 7.0)	6.0 (5.8 – 7.0)	6.0 (5.0 – 7.0)	7.0 (6.0 – 7.0)					
	No training	4	6.0 (5.3 – 7.0)	6.0 (6.0 – 7.0)	6.0 (6.0 - 7.0)	7.0 (5.0 – 7.0)					
	P-value		0.019	0.560	0.137	0.771					
Sedlack *et al.* 2004 Level II	**Intervention**	**N = 38**	**Patient reported pain score†, median score (25th-75th IQR)**								
	AccuTouch®	19	3 (2 - 5)								
	No training	19	4 (2 - 6)								
	P-value		< 0.01								
	Intervention	**N = 38**	**Responds to patient discomfort, median score (25th-75th IQR)**								
			Supervisor-assessed ‖					**Self-assessed¶**			
	AccuTouch®	19	3 (3 – 5)					2 (1 – 3)			
	No training	19	3 (1 – 6)					2 (2 – 3)			
	P-value		0.278					0.394			
Cohen *et al.* 2006b Level II	**Intervention**	**N = 49****	**Proctor assessed patient discomfort score†† for each session**								
			1	2	3	4	5	6	7	8	9	10
	Simbionix GI Mentor™	22	25.7	23.2	16.7	16.0	16.7	13.4	11.9	10.5	10.7	8.9
	No training	23	31.4	19.1	19.5	18.2	16.5	13.9	11.3	10.4	11.8	9.2
	P-value		0.42	0.14	0.22	0.39	0.94	0.85	0.74	0.99	0.55	0.81
Tuggy 1998 Level II	**Intervention**	**N = 10**	**Patient reported pain scale (units NR)**								
			Pain								
	Gastro-Sim®	5	NR								
	No training	5	NR								
	P-value		NS								

(table continued over page)

Table 23. Patient-based assessments: patient discomfort continued

Simulation-based training *vs.* patient-based training				
Colonoscopy/sigmoidoscopy				
Gerson and van Dam 2003 Level II	**Intervention**	**N = 16**	**Had a lot of pain**	
			Agree‡‡ (%)	**Disagree§§ (%)**
	Sigmoidoscopy simulator	9	53	37
	Patient-based training	7	42	45
	P-value			NS
	Intervention	**N = 16**	**More comfortable than expected**	
			Agree‡‡ (%)	**Disagree§§ (%)**
	Sigmoidoscopy simulator	9	33	47
	Patient-based training	7	42	29
	P-value			NS
	Intervention	**N = 16**	**Caused great discomfort**	
			Agree‡‡ (%)	**Disagree§§ (%)**
	Sigmoidoscopy simulator	9	43	53
	Patient-based training	7	31	61
	P-value			NS

* In addition, male patients reported less pain (p = 0.001) compared with female patients.
† Measured on scale of 1(no pain) to 10 (worst pain of life).
‡ Colonoscopies grouped in blocks of 15.
§ Score rendered by supervising faculty. Measured on scale of 1 (strongly disagree) to 7 (strongly agree).
‖ Score rendered by supervising faculty. Measured on scale of 1 (strongly agree) to 10 (strongly disagree).
¶ Score rendered by participant. Measured on scale of 1 (strongly agree) to 10 (strongly disagree).
** 4 lost to assessment; groups not specified.
†† Measured on scale of 1 (very comfortable) to 5 (severe pain to patient).
‡‡ Measured on scale of 1(strongly agree) to 7(strongly disagree). Includes patients who agreed or strongly agreed with the question.
§§ Measured on scale of 1(strongly agree) to 7(strongly disagree). Results include patients who disagreed or strongly disagreed with the questions.
IQR interquartile range
NR not reported
NS not significant

Colonoscopy/sigmoidoscopy

Randomised controlled trials

One study reported median patient-assessed discomfort outcomes for patients undergoing colonoscopy by AccuTouch® endoscopy simulator-trained and untrained participants (Ahlberg *et al.* 2005). Significantly less patient discomfort was reported in the simulator-trained group than controls (p = 0.02).

Sedlack and Kolars (2004) reported patient-reported pain scores for patients undergoing colonoscopy by AccuTouch® colonoscopy trained Fellows or Fellows that had no exposure to the simulator. Patient surveys demonstrated a lower median discomfort score during the first 15 colonoscopies performed by simulator trained Fellows (p = 0.019). This study also reported median supervisor scores for whether the participant responded adequately to patient discomfort during the procedure. The simulator-trained participants responded better to patient discomfort than the non-simulator trained participants (p = 0.019).

Sedlack *et al.* (2004) reported pain scores for patients undergoing sigmoidoscopy by AccuTouch® sigmoidoscopy simulator-trained participants or participants that had no exposure to the simulator. Median patient-reported discomfort scores were significantly lower for simulator-trained residents ($p < 0.01$). There were no statistical differences between the trained and untrained groups and there were no significant differences between supervisor-assessed and self-assessed scores.

Cohen *et al.* (2006b) reported no significant differences in proctor-assessed patient discomfort between the Simbionix GI Mentor™-trained group or the untrained controls during patient colonoscopies at any time during the study.

Tuggy (1998) stated that there were no significant differences between the Gastro-Sim® flexible sigmoidoscopy simulator-trained residents compared with non-trained controls in patient pain scores after five hours of simulator training (data not reported).

Simulation-based training vs. patient-based training

One study reported patient discomfort outcomes for patients undergoing sigmoidoscopy (Gerson and van Dam 2003) (Table 23).

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam (2003) reported pain outcomes for patients undergoing sigmoidoscopy by VR sigmoidoscopy simulator-trained trainees or patient-trained trainees. Almost half of the patients in both groups experienced discomfort, but there were no significant differences between the simulator-trained and patient-trained groups.

Participant satisfaction of training

To determine participants' opinions of the simulation-based training they had received, some authors administered questionnaires to participants at the end of the study periods, or observed participants during the study.

Simulation-based training vs. no simulation-based training

Five studies reported outcomes for participant satisfaction after training (Ahlberg *et al.* 2005; Cohen *et al.* 2006b; Tuggy 1998; Schijven *et al.* 2005; Hamilton *et al.* 2001).

Colonoscopy

Randomised controlled trials

Ahlberg *et al.* (2005) stated that none of the trainees complained about the amount of AccuTouch® endoscopy simulator practice needed. They reported that participants were eager to continue with training until the criterion level was reached (no data reported).

Cohen *et al.* (2006b) administered a survey and found that respondents rated the overall satisfaction with the Simbionix GI Mentor™ simulator training as moderately useful to useful, with a mean score of 3.5 (range, 1 [no use] to 5 [very useful]).

Tuggy (1998) stated that the residents' survey responses indicated they strongly agreed that the Gastro-Sim® flexible sigmoidoscopy simulator training was valuable and would enhance the likelihood of their mastering the skill later in practice. No statistical analysis was done on the survey data. Tuggy (1998) further reported that participants showed a willingness to commit to the necessary hours of using the simulator during the study and that all residents voluntarily trained on the simulator on their own time despite normal work schedules. They also reported that many trainees commented that some colons in the simulator were more challenging than the live examinations.

Laparoscopic cholecystectomy

Non-randomised comparative studies

Schijven *et al.* (2005) reported that the experimental group valued their training course highly in terms of their laparoscopic surgical skills improvement, and the use of VR simulators in the surgical curriculum, but no statistical analysis was shown.

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) reported that questionnaire outcomes after training revealed that 10 of 10 residents in the TEP hernia repair curriculum trained group felt their ability to perform a laparoscopic TEP hernia repair improved over the study period compared with 5 of 11 in the control group (100% *vs.* 45.5%, P= 0.01). Additionally, all the residents in the trained group (100%) reported that their understanding of the operation improved over the month of the training compared with 5 of 11 residents (45.5%) in the control group (p = 0.01). Compared with controls, residents in the trained group also expressed an increased willingness to offer laparoscopic TEP hernia repairs to patients with concurrent, unilateral hernias (p = 0.02).

Surgical confidence

To determine the level of surgical confidence of participants included in the studies, investigators either administered a survey to participants after the training period or evaluated them during the patient-based assessment operation.

Simulation-based training vs. no simulation-based training

Three studies reported outcomes for the surgical confidence of trainees (Scott *et al.* 2000; Tuggy 1998; Edmond 2002) (Table 24).

Table 24 Patient-based assessments: surgical confidence

Simulation-based training *vs.* no simulation-based training							
Laparoscopic cholecystectomy							
Scott *et al.* 2000	**Intervention**	**N = 27***	**Number of residents comfortable with laparoscopic skills (n/N)**				
			Baseline			**Final**	
	SCMIS GEM	13	5/9			8/9	
Level II	No training	9	3/13			6/13	
	P-value		NR			NS	
			After training, 9/9 felt that training had improved their video-eye-hand coordination and 8/9 felt that training had improved their operating skills.				
Colonoscopy/sigmoidoscopy							
Tuggy 1998	**Intervention**	**N = 10**	**Gained confidence for live patient examination**				
			Strongly agree (%)	**Agree (%)**	**Neutral**	**Disagree (%)**	**Strongly disagree (%)**
Level II	Gastro-Sim®	5	55	45	0	0	0
	No training	5	NA	NA	NA	NA	NA
Endoscopic sinus surgery							
Edmond 2002	**Intervention**	**N = 4**	**Surgical confidence†, mean rating ± SD**				
			Baseline			**After training**	
Level III-2	ESS simulator	2	NA			6.5 ± 0.7	
	No training	2	NA			2.8 ± 0.2	
	P-value		NA			0.09	

* 5 lost to assessment; Trained group n = 4, Control group n = 1.
† Measured by assessing surgeons on a scale of 1 (adequate) to 10 (perfect).
NA not applicable
NR not reported
NS not significant
SD standard deviation

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) reported that when initially asked if they felt comfortable with their laparoscopic skills, 3 of 13 control residents and 5 of 9 SCMIS GEM trained residents replied 'yes'. On the completion questionnaire, 6 of 13 control residents and 8 of 9 trained residents felt comfortable with their laparoscopic skills at the end of the rotation. Of those who were not comfortable with their laparoscopic skills at baseline, 3 of 10 in the control group were comfortable at the end of the rotation, in contrast to 3 of 4 in the trained group, however this difference was not statistically significant. After training, 9 of 9 residents improved their hand-eye-hand coordination and 8 of 9 felt that the training had improved their skills in the operating room.

Colonoscopy

Randomised controlled trials

Tuggy (1998) administered a survey to participants that underwent Gastro-Sim® flexible sigmoidoscopy simulator training. Responses showed that 55% strongly agreed and 45% agreed with the statement 'gained confidence for live patient examination'. No statistical analysis was done on this survey data.

Endoscopic sinus surgery

Non-randomised comparative studies

Edmond (2002) reported no significant differences between the simulator-trained group and controls for surgical confidence as rated by the assessing surgeons.

Training costs

The reporting of costs in relation to the purchase of surgical simulators, as well as cost savings, if any, associated with their use, varied between studies.

Simulation-based training vs. no simulation-based training

Laparoscopic cholecystectomy

Randomised controlled trials

Scott *et al.* (2000) indicated that the price range for the SCMIS GEM ranged from $US215,000 to $US285,000 (depending on the quality of video-imaging equipment installed). The authors stated that at the University of Texas Southwestern Medical centre, 186 residents trained in general surgery, urology, and gynaecology and the cost of training residents using the video-trainer was $US270 per graduating resident. They then compared this figure to a published study by Bridges and Diamond 1999 that estimated the cost of operating room time to train residents to be approximately $US48,000 per graduating resident, and stated that training outside of the operating room, using a video-trainer such as the SCMIS GEM seems cost effective.

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Ahlberg *et al.* (2005) believe that the eventual training effect from using a simulator similar to the one used in the study (the AccuTouch® flexible endoscopy simulator at a unit cost of approximately $US100,000) is obvious and substantial.

Cohen *et al.* (2006b) found that the benefits of simulator training levelled off after approximately 80 cases, and subsequently asked whether the purchase cost of a simulator of $US50,000 or more can be justified. The authors suggest that simulators may have a greater value in training practitioners who have access to fewer cases during their formal training than colonoscopy residents.

Sedlack and Kolars (2004) observed an increase in staff productivity during the time that Fellows trained on the AccuTouch® colonoscopy simulator. They equated that the nine additional staff-performed colonoscopies translated into approximately $US4,200 of recaptured education dollars per Fellow (calculations based on 2002 Medicare reimbursement data of $US466 per colonoscopy [$US216 physician fee, plus $US250 facility fee]). The cost of colonoscopy simulators, was estimated to be between $US30,000 and $US45,000 (these costs include the simulator platform and lower endoscopy software only), which needs to be balanced with the recaptured procedure revenue.

Catheter-based intervention for occlusive vascular disease

Randomised controlled trials

Chaer *et al.* (2006b) discussed the cost versus benefit of Procedicus VIST™ simulation. The average cost of available simulators is approximately $US200,000 to $US400,000 which includes a variety of software material for carotid, renal and iliofemoral interventions. They used the published estimated training cost of approximately $US50,000 per surgical resident over a training period of four years (due to increased operative time and decreased efficiency that occur when operating with a trainee) (Bridges and Diamond 1999) to argue that a durable benefit of simulation (even beyond the two cases described here) would likely need to be established before any cost advantage would be recognised. They suggested shared use of regional simulation centres would be a practical alternative to all institutions having a simulator.

TEP hernia repair

Randomised controlled trials

Hamilton *et al.* (2001) stated that after the initial cost of the training system is met, this type of *ex vivo* system is relatively inexpensive, due to the reusable nature of all components of the curriculum.

Simulation-based training vs. patient-based training

Colonoscopy/sigmoidoscopy

Randomised controlled trials

Gerson and van Dam, estimated that the cost of VR sigmoidoscopy simulator was $US50,000 (cited as personal communication, Immersion Medical Inc in article).

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. It has included studies irrespective of the type of patient-based assessment procedure, or type of simulation. Many factors determine the transfer of skills, including those that relate to the simulator design and functionality (see Appendix E for a summary) and the way that it 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 transference reported in this review cannot be attributed to the simulators alone.

Simulation-based training was, in most cases, an add-on to normal surgical training programs. In one instance only (Gerson and van Dam 2003) was simulation-based training compared with structured patient-based training. This study design was quite different to the other studies, and resulted in the study being treated separately in this review.

Determining the training methods used in some studies was made 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. In addition to this, it was difficult to delineate between different phases of some studies (Cohen *et al.* 2006b; Sedlack and Kolars 2004; Sedlack *et al.* 2004), because patient-based procedures that were assessed were referred to as training and not assessment.

Reporting of methodological detail in the included studies was generally inadequate (Appendix F provides a summary of the critical appraisal). More than half of the RCTs did not report the method of randomisation, allocation concealment, intention to treat, power calculations, losses to assessment, study period, and exclusion criteria. In addition to this, three of the RCTs did not blind assessors to the training status of the participants (Sedlack and Kolars 2004; Tuggy 1998; Gerson and van Dam 2003).

Only three studies reported power calculations for the minimum number of patients or procedures required to detect significant differences between the trained and untrained groups (Scott *et al.* 2000; Cohen *et al.* 20006; Gerson and van Dam 2003). The sample sizes within the included studies were generally quite small, with the study groups in all but one RCT (Cohen *et al.* 2006b) having less than 20 participants per group. Six of the included studies had less than 10 participants per group. 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 (Table 3). This increases the likelihood of type I error. The small sample sizes were acknowledged within many of the

studies and were attributed to the relatively small numbers of surgical trainees available to participate. This may be an unavoidable limitation in studies of psychomotor skill training in surgery.

Two studies (Grantcharov *et al.* 2004; Seymour *et al.* 2002) modified a previously validated global rating scale for the assessment of surgical performance, which may have compromised the validity of the results. One of these studies (Gratcharov *et al.* 2004) modified the global rating scale substantially by combining performance characteristics.

There were large variations in the length of time participants were trained (Table 4) and varied from less than two hours (Chaer *et al.* 2006) to unlimited access to the simulator (Gerson and van Dam 2003). It can be argued that the short duration of simulation-based training may have resulted in a positive transfer effect not being evident, although simulator-trained groups did not always show superiority over groups who did not have the training. The end-points of training were often ill-defined and were not consistent between studies, making it difficult to comment on the skill level at the end of the training.

Statistical comparisons between studies was made difficult because of other factors that were not consistent between the studies. Variables in the operating room such as differences in the severity of patient disease, the degree of independence granted by clinicians and various staff assistants, the mentoring given to residents 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. In addition to this, there may have been pre-existing faculty assessor bias about individual resident's abilities that may have influenced their evaluations and the results of the studies (Hamilton *et al.* 2001). Inherent differences between participants used for these studies will always exist, and include variations in hand-eye skills between trainees (Ahlberg *et al.* 2005) and the ability of some trainees to learn and master techniques faster than others (Chaer *et al.* 2006).

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 device used for training, or the performance of the assessment procedure, or both. Using the simulator, or a patient for baseline testing in itself allows a participant to gain familiarity with the procedure or device, and hence can lead to improvements in clinical performance. This was evident in Hamilton *et al.* (2001) where untrained controls also improved their performance over the study period. The authors attributed the improvement in the control group over time to a progression in skill and knowledge just by performing resident duties, and because they may have studied the subject more than they normally would have to compare favourably with their trained counterparts. Not withstanding these considerations, the improvements seen in the simulation-trained group significantly exceeded those observed in the

untrained group for almost all the measured parameters. Similarly, Scott *et al.* (2000) stated that improvement in the control group was expected because residents were exposed to each task three times during the initial testing session and because they were undergoing 'on-the-job' training while performing different operative cases on their surgical rotation. Likewise, the improvements seen in the simulation-trained group significantly outweighed those observed in the control group for many of the parameters measured.

Although training methods and training duration varied, participants who trained using simulation improved their simulator performance over time, as indicated by direct improvements in performance parameters, reaching criterion levels, or the completion of a certain number of simulated cases.

Transfer outcomes

Simulation-based training vs. no simulation-based training

Laparoscopic cholecystectomy

Laparoscopic surgery is particularly suited to technical skills training as it requires a skill set based on instrumentation, depth perception, and fine motor control (Valentine and Rege 2004; Villegas *et al.* 2003). Five studies were included for laparoscopic cholecystectomy, covering three different modes of simulation (MIST-VR, video-trainers and training course). The simulators in the studies taught simple tasks and were of low fidelity, which are sufficient for novice learners, like those included in the studies. There were large variations in assessment methods within this intervention, with Scott *et al.* (2000) and Scott *et al.* (1999) assessing the performance of the entire laparoscopic cholecystectomy procedure, Schijven *et al.* 2005 and Grantcharov *et al.* (2004) assessing the clip and cut part of the procedure, and Seymour *et al.* (2002) assessing only the gallbladder excision from the liver. 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 however, but the untrained group never outperformed the trained group.

Gallagher *et al.* (2005) suggest that the most valuable metrics that simulation-based training can provide involve the measurement of errors. Two studies (with small sample sizes) reported outcomes for errors during laparoscopic cholecystectomy (Seymour *et al.* 2002; Grantcharov *et al.* 2004), and it was found that the trained groups generally made fewer errors than untrained groups.

Supervising surgeon takeover was measured in one study (Seymour *et al.* 2002) and referred to instances where a surgeon had to take control of the procedure. Events such as these have large clinical significance, as they represent catastrophic failures in technique, and the point where patient safety is compromised. Simulation-trained

participants had fewer instances of supervising surgeon takeover than untrained participants (level of significance, if any, was not reported). The flow-on effects of skills failure in these settings would likely indicate other (both technical and cognitive) breakdowns, but were not measured for ethical reasons. It is important that surgical trainees are provided with sufficient technical and non-technical skills training to reduce the incidence of error, which translates to improved patient safety.

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.

Colonoscopy/sigmoidoscopy

Five studies were included for this intervention, covering three different modes of simulation (AccuTouch®, GI Mentor™ and Gastro-Sim®) and a variety of training durations. Simulation-based training prior to patient-based colonoscopy/ sigmoidoscopy appeared to provide participants some advantage over their untrained colleagues. Sedlack and Kolars (2004), Cohen *et al.* (2006b) and Ahlberg *et al.* (2005) showed that simulation-trained residents performed patient-based procedures significantly better than untrained controls during the initial stages of learning, after which there were no differences. This is an important finding, indicating that the learning curve can be reduced by simulation-based training, and hence can reduce risks to the patient of training.

The participants in the included studies had between six and 20 hours of training on their respective simulators. In accordance with this, Tuggy (1998) showed that there was a greater improvement in patient-based sigmoidoscopy skill after participants received 10 hours of simulator training compared to when they had only received five hours of training. One study (Sedlack *et al.* 2004) found no significant differences between the simulator-trained and untrained groups for all but one measured parameter (patient discomfort) during patient-based assessment. Participants in this study received only 3 hours of training and this may not have been sufficient to translate into a positive transfer effect. The results of these studies indicate that the duration of training is very likely to be an important factor to the transfer of simulator-based acquired skills to patients during colonoscopy/ sigmoidoscopy. The minimum case number recommended for competency in patient-based flexible endoscopy by international endoscopic societies ranges from as low as 50 to as high as 300 cases (Cass 1999).

Results for patient-assessed discomfort during colonoscopy/sigmoidoscopy appeared favourable for the simulator-trained group, with two out of three studies reporting significant differences in patient discomfort, and one study showing a significant difference in patient discomfort for the first block of 15 colonoscopies. When assessed by the proctor, two studies reported no significant differences for patient discomfort, whereas one study showed significant improvement in discomfort scores for the initial stages of training. Drawing conclusions from this data is difficult, as the pain relief given to patients before and during the procedures was not standardised or documented in the studies. Sedlack and Kolars (2004) have

questioned patient clarity when completing questionnaires after a procedure performed under sedation, and it has been suggested (Cohen *et al.* 2000) that the recollection of pain may be influenced by the amount of sedation used, reflecting operator skill and/or the practice of the supervising gastroenterologist.

Two studies measured differences in staff productivity when experienced staff were in the presence and absence of trainees (Sedlack and Kolars 2004; Sedlack *et al.* 2004). These studies found that staff were able to conduct more procedures whilst not in the presence of trainees (when trainees were training on the simulator), but no statistical significance was reported for either study. Sedlack *et al.* (2004) reported an increase in staff productivity during the time that participants trained on the simulator, but at the same time found no significant difference in procedural skills between the simulator-trained participants and controls. This could indicate that even if simulation-based training provides no benefit greater than traditional training in terms of procedural skill acquisition, endoscopy units and patients could still benefit from increased productivity.

Catheter-based intervention for occlusive vascular disease

For catheter-based intervention for occlusive vascular disease, training using simulation (Procedicus VIST™ simulator) prior to patient-based procedures appeared to show benefits indicating that transfer of skills is possible from a simulator to a patient for this procedure. Improvement was seen in the completion of specific steps during two catheter-based interventions, including improvements in catheter and guidewire manipulation, catheter exchange and over-the-wire balloon technique. There were also significant improvements in terms of supervising surgeon takeover and the need for verbal prompts, further indicating the benefits of simulation-based training.

TEP hernia repair

For TEP hernia repair, residents trained using the simulation-based curriculum (consisting of a moulded rubber simulator, an instructional video and an interactive CD ROM) showed a significant improvement in operative performance compared with their untrained controls, suggesting that training using this curriculum can allow the transfer of skills to the patient-based setting.

Endoscopic sinus surgery

There were no differences in performance of ESS simulator trained residents compared with controls. This study had an unclear study methodology, a small number of participants and did not clearly define the duration of training for the included participants. The authors of this study stated that skills transfer was occurring, but the lack of statistical significance for any of the measured parameters does not allow for the same conclusion to be drawn for this review.

Simulation-based training vs. patient-based training

Colonoscopy/sigmoidoscopy

Only one study compared structured patient-based training with simulation-based training (Gerson and van Dam 2003). This simulation-based training was not supervised and participants were not mentored. This was the only study to compare the methods currently being used to train surgeons with a completely non-patient-based method. Participants who received training in the assessment procedure exhibited better performance than those who had trained exclusively on a simulator without any mentoring or supervision. In addition to reporting no clinical improvement after simulation-based training, the authors also reported no improvement in simulator performance during training (results not shown). The fact that learning did not actually take place during the training period means that transference could not be demonstrated.

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, to address some risk management concerns, and to improve operating theatre efficiency.

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 any indication of the quality of the task performed, and caution should be taken when interpreting it without any additional objective quality data.

Gallagher *et al.* (2005) suggest that simulation-based training allows for the development of the 'pre-trained novice'; an individual who has been trained to the point where many psychomotor skills and spatial judgments have been automated, allowing them to focus more on learning operative strategy and how to handle intraoperative complications, rather than wasting valuable operating room time on the initial refinement of psychomotor skills.

With adequate pre-training, the trainee can gain maximum advantage from the supervised opportunities for training in the operating room. Many skills (including the cognitive skills of anatomical recognition, decision making, leadership, and communication) must be incorporated into the training of a surgeon. The place of simulation-based training in a rich curriculum incorporating these other skills, has yet to be defined. Such a curriculum will use a range of skill training modalities, didactic teaching methods and evaluation techniques (Alhberg *et al.* 2005). The aim of a training program should be to integrate training for technical and non-technical skills

with team skills training in order to make the most effective use of available resources and to maximise patient safety.

At the present time simulation-based training programs are often 'add-ons' to traditional surgical training and they are often voluntary, which is reflected in the included studies. Although data (survey data or subjective evidence) regarding the participants' views on training were positive (Schijven *et al.* 2005; Ahlberg *et al.* 2005; Cohen *et al.* 2000; Tuggy 1998) (data not shown), they were volunteers, indicating a pre-study willingness to participate. A study by Chang *et al.* (2007) found that when given the opportunity to train in a simulation laboratory, many residents chose not to use it for reasons relating to time, location and lack of interest. To increase participation rates, and for simulation to be an effective part of a training curriculum, the authors suggest that simulation-based training will need to be made mandatory, and fully integrated within the curriculum (Gaba 2004). Other approaches to improve participation in simulation-based training include training in a variety of settings, the use of multi-disciplinary approaches (Dunkin *et al.* 2007), and to increase motivation by using defined performance criteria to define specific training goals for which to aim (Ahlberg *et al.* 2005).

The cost of providing simulation-based training is often used as an argument against its use. It is difficult to comment on the cost of simulator-based training from the included studies because of the variations in the devices and the cost of current training methods (Haluck *et al.* 2005). In addition to this, no cost estimations were reported for training curricula that used more than just a simulator as the training tool; however, most authors agreed that it was important to justify the cost of simulation-based training. Data is not yet available to demonstrate the money saved through improved operating room efficiency and/or the reduced risk to patients. The cost associated with the integration of simulation-based training into surgical training programs will depend on the approach taken by organisations responsible for surgical training (Gaba 2004). It is expected that the costs of simulation-based training will decrease as it becomes viewed as an integral part of surgical training (Villegas *et al.* 2003; Haluck *et al.* 2005).

Future research

The challenges to surgical training are substantial and simulation has the potential to make a significant contribution to the evolution of surgical curriculum. It is very important that further studies be undertaken to provide the best evidence to determine how simulation-based training can be used in the most beneficial way.

It is recommended that further research be done into the transfer of skills acquired via surgical simulation to the patient-based setting to strengthen the current evidence base. While critical of the quality of study design in the included studies, it is acknowledged that there are challenges in conducting RCTs in this area. One of these is to bring the appropriate statistical and design discipline, as there is a tendency to use criteria common within educational research, but which many in the surgical

community would view as being less stringent than those normally used in clinical trials. Such challenges however, should not be used as an excuse to undertake inferior studies, or provide a justification for inadequate reporting of methodological detail. Similarly, they should not be deemed to justify avoiding a decision on the utility of simulation-based training strategies.

Future studies will have the opportunity to explore other important dimensions to the issue of skills transfer. 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, and
- changes in staff productivity as a result of surgical simulation-based training.

6. Conclusions and Recommendations

The aim of this systematic review was to determine whether skills acquired through simulation-based training are transferable to the operative setting. The studies included in this review were of variable quality and design, which limited the strength of the conclusions. 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.

Classification and Recommendations

The evidence-base in this review is rated as average. The included randomised controlled trials and non-randomised comparative studies were limited by small sample sizes and poor reporting of methodological detail. Outcomes were often not comparable between studies.

Clinical and Research Recommendations

It is recommended that further research be done into the transfer of skills acquired via surgical simulation-based training to the patient-based setting. Well designed studies should be conducted to strengthen the current evidence base. Consistency in training and assessment methods across studies would help provide further insight into the benefits of surgical simulation-based training.

Acknowledgments

The authors wish to acknowledge Dr Karen Facey and Ms Prema Thavaneswaran for their assistance during the preparation of this review. The ASERNIP-S project is funded by the Australian Government Department of Health and Ageing and the South Australian Department of Health.

References

Aggarwal R, Grantcharov T, Moorthy K, Hance J, Darzi A. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *American Journal of Surgery* 2006; **191**: 128-133.

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

Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, Hutchison C. Assessment of technical skills transfer from the bench training model to the human model. *American Journal of Surgery* 1999; **177**(2): 167-170.

Bar-Meir S. A new endoscopic simulator. *Endoscopy* 2000; **32**(11): 898-900.

Bridges M and Diamond DL. The financial impact of teaching surgical residents in the operating room. *American Journal of Surgery* 1999; **177**: 28-32.

Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ. Consensus guidelines for validation of virtual reality surgical simulators. *Surgical Endoscopy* 2005; **19**(12): 1523-1532.

Cass OW. Training to competence in gastrointestinal endoscopy: a plea for continuous measuring of objective end points. *Endoscopy* 1999; **9**: 751-754.

Cass OW, Freeman ML, Cohen J, Zuckerman G, Watkins J, Nord J, Locke GR, Jensen D, Diehl D, Cerulli M, Lyche K, Fennerty M, Edmundowicz S, Etzkorn K, Al-Kawas D, Cave D, Lehman G. Acquisition of competency in endoscopic skills (ACES) during training: a multicentre study [abstract]. *Gastrointestinal Endoscopy* 1996; **43**(3): 308.

Chaer RA, Derubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D, Morrissey NJ, Faries PL, McKinsey JF, Kent KC. Simulation improves resident performance in catheter-based intervention: results of a randomised, controlled study. *Annals of Surgery* 2006; **244**(3): 343-352.

Chang L, Petros J, Hess DT, Rotondi C, Babineau TJ. Integrating simulation into a surgical residency program. *Surgical Endoscopy* 2007; **21**: 418-421.

Chapman DM, Rhee KJ, Marx JA, Honigman B, Panacek EA, Martinez D, Brofeldt BT, Cavanaugh SH. Open thoracotomy procedural competency: Validity study of teaching and assessment modalities. *Annals of Emergency Medicine.* 1996; **28**(6): 641-647.

Cisler JJ and Martin JA. Logistical Considerations for Endoscopy Simulators. *Gastrointestinal Endoscopy Clinics of North America* 2006; **16**(3): 565-575.

Cohen J, Nuckolls L, Mourant RR. Endoscopy Simulators: Lessons from the Aviation and Automobile Industries. *Gastrointestinal Endoscopy Clinics of North America* 2006a; **16**(3): 407-423.

Cohen J, Cohen SA, Vora KC, Xue X, Burdick JS, Bank S, Bini EJ, Bodenheimer H, Cerulli M, Gerdes H, Greenwald D, Gress F, Grosman I, Hawes R, Mullen G, Schnoll-Sussman F, Starpoli A, Stevens P, Tenner S, Villanueva G. Multicentre, randomised, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointestinal Endoscopy* 2006b; **64**(3): 361-368.

Cooper JB and Taqueti VR. A brief history of the development of mannequin simulators for clinical education and training. *Quality & Safety in Health Care* 2004; **13**: I11-I18.

Datta V, Mandalia M, Mackay S, Darzi A. The PreOp flexible sigmoidoscopy trainer. Validation and early evaluation of a virtual reality based system. *Surgical Endoscopy* 2002; **16**(10): 1459-1463.

Dayal R, Faries PL, Lin SC, Bernheim J, Hollenbeck S, DeRubertis B, Trocciola S, Rhee J, McKinsey J, Morrissey NJ, Kent KC. Computer simulation as a component of catheter-based training. *Journal of Vascular Surgery.* 2004; **40**(6): 1112-1117.

Dent J. Current trends and future implications in the developing role of clinical skills centres. *Medical Teacher* 2001; **35**: 909-915.

Deziel DJ, Milikan KW, Economou SG, Doolas A, Ko ST, Airan MC. Complications of laparoscopic cholecystectomy: a national survey of 4,292 hospitals and an analysis of 77,604 cases. *American Journal of Surgery* 1993; **165**: 9-14.

Dunkin B, Andrales GL, Apelgren K, Mellinger JD. Surgical simulation: a current review. *Surgical Endoscopy* 2007; **21**(3): 357-366.

Dutta S, Gaba D, Krummel TM. To simulate of not to simulate. What is the question? *Annals of Surgery* 2006; **243**(3): 301-303.

Edmond CV, Jr. Impact of the endoscopic sinus surgical simulator on operating room performance. *Laryngoscope* 2002; **112**(7 Pt 1): 1148-1158.

Feldman LS, Sherman V, Fried GM. Using simulators to assess laparoscopic competence: Ready for widespread use? *Surgery* 2004; **135**(1): 28-42.

Fried GM, Derossis AM, Bothwell J, Sigman HH. Comparison of laparoscopic performance *in vivo* with performance measured in a laparoscopic simulator. *Surgical Endoscopy* 1999; **13**(11): 1077-1081.

Fried GM. Lessons from the Surgical Experience with Simulators: Incorporation into Training and Utilisation in Determining Competency. *Gastrointestinal Endoscopy Clinics of North America* 2006; **16**(3): 425-434.

Gaba DM. The future vision of simulation in health care. *Quality and Safety in Health Care* 2004; **12**: 2-10.

Gallagher AG, McLure N, Ritchie K, Sheehy NP. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscope* 1998; **30**(7): 617-620.

Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery* 2005; **241**(2): 364-372.

Gerson LB. Evidence-Based Assessment of Endoscopic Simulators for Training. *Gastrointestinal Endoscopy Clinics of North America* 2006; **16**(3): 489-509.

Gerson LB and van Dam J. A prospective randomised trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. *Endoscopy* 2003; **35**(7): 569-575.

Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomised clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery* 2004; **91**(2): 146-150.

Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Annals of Surgery* 2004; **240**(2): 374-381.

Haluck RS. Computer-based surgical simulation is too expensive. Or is it? *Surgical Endoscopy* 2005; **19**(2): 159-160.

Hamdorf JM and Hall JC. Acquiring surgical skills. *British Journal of Surgery* 2000; **87**(1): 28-37.

Hamilton EC, Scott DJ, Fleming JB, Rege RV, Laycock R, Bergen PC, Tesfay ST, Jones DB. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical Endoscopy* 2002; **16**(3): 406-411.

Hamilton EC, Scott DJ, Kapoor A, Nwariaku F, Bergen PC, Rege RV, Tesfay ST, Jones DB. Improving operative performance using a laparoscopic hernia simulator. *American Journal of Surgery* 2001; **182**(6): 725-728.

Haque S and Srinivasan S. A meta-analysis of the training effectiveness of virtual reality surgical simulators. *IEEE Transactions on Information Technology in Biomedicine* 2006; **10**(1): 51-58.

Herbella FAM and Del Grande JC. Human cadavers as an experimental model for esophageal surgery. *Diseases of the Esophagus* 2001; **14**(3-4): 218-222.

Higgins G, Green S. In: Handbook for systematic reviews of interventions 4.2.5. 2005. The Cochrane Library, Issue 3, 2005. Chichester, UK: John Wiley & Sons, Ltd.

Hochberger J and Maiss J. Currently Available Simulators: *Ex vivo* Models. *Gastrointestinal Endoscopy Clinics of North America* 2006; **16**(3): 435-449.

Hyltander A. Simulation as a teaching alternative: Utopia or reality? *CAL-laborate* 2003; **June**: 9-12.

Hyltander A, Liljegren E, Rhodin PH, Lonroth H. The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surgical Endoscopy* 2002; **16**(9): 1324-1328.

Kaufmann CR. Computers in surgical education and the operating room. *Annales Chirurgiae et Gynaecologiae* 2001; **90**(2): 141-146.

Kirk RM. Teaching the craft of operative surgery. *Annals of the Royal College of Surgeons of England* 1996; **78 (Suppl 1)**: 25-28.

Korndorffer JR, Jr., Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *Journal of the American College of Surgeons* 2005; **201**(1): 23-29.

Krummel TM. Surgical simulation and virtual reality: the coming revolution. *Annals of Surgery* 1998; **228**(5): 635-637.

Levine RL, Kives S, Cathey G, Blinchevsky A, Acland R, Thompson C, Pasic R. The use of lightly embalmed (fresh tissue) cadavers for resident laparoscopic training. *Journal of Minimally Invasive Gynecology* 2006; **13**(5): 451-456.

Maddern GJ. The challenge of reduced trainee hours in Australia on surgical teachers. *Surgery* 2003; **134**(1): 23-26.

Maiss J, Dumser C, Zopf Y, Naegel A, Krauss N, Hochberger J, Matthes K, Hahn EG, Schwab D. "Hemodynamic efficacy" of two endoscopic clip devices used in the treatment of bleeding vessels, tested in an experimental setting using the compact Erlangen Active Simulator for Interventional Endoscopy (compactEASIE) training model. *Endoscopy* 2006; **38**(6): 575-580.

McDougall EM, Corica FA, Boker JR, Sala LG, Stoliar G, Borin JF, Chu FT, Clayman RV. Construct validity testing of a laparoscopic surgical simulator. *Journal of the American College of Surgeons* 2006; **202**(5): 779-787.

Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *BMJ* 2003; **327**(7422): 1032-1037.

Parker LM. Anatomical dissection: why are we cutting it out? Dissection in undergraduate teaching. *ANZ Journal of Surgery* 2002; **72**: 910-912.

Reznek M, Harter T, Krummel T. Virtual reality and simulation: training the future emergency physician. *Academic Emergency Medicine* 2002; **9**: 78-87.

Reznick RK and MacRae H. Teaching surgical skills - changes in the wind. *The New England Journal of Medicine* 2006; **355**(25): 2664-2669.

Reznick RK, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative 'bench station' examination. *American Journal of Surgery*. 1997; **173**(3): 226-230.

Rivron RP and Maran AGD. The Edinburgh FESS trainer: A cadaver-based bench-top practise system for endoscopic ethmoidal surgery. *Clinical Otolaryngology & Allied Sciences.* 1991; **16**(4): 426-429.

Roberts KE, Bell RL, Duffy AJ. Evolution of surgical skills training. *World Journal of Gastroenterology* 2006; **12**(20): 3219-3224.

Royal Australasian College of Surgeons (RACS) (Eds). The Surgical Workforce 2005. Based on the 2005 census of the surgical workforce. March 2005.

Satava RM. Accomplishments and challenges of surgical simulation. *Surgical Endoscopy* 2001; **15**(3): 232-241.

Satava RM, Gallagher AG, Pellegrini CA. Surgical competence and surgical proficiency: Definitions, taxonomy, and metrics. *Journal of the American College of Surgeons.* 2003; **196**(6): 933-937.

Schijven MP and Jakimowicz J. The learning curve on the Xitact LS 500 laparoscopic simulator: profiles of performance. *Surgical Endoscopy* 2003; **18**: 121-127.

Schijven MP and Jakimowicz JJ. Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. *Minimally Invasive Therapy & Allied Technologies: Mitat.* 2005; **14**(4-5): 244-246.

Schijven MP, Jakimowicz JJ, Broeders IA, Tseng LN. The Eindhoven laparoscopic cholecystectomy training course - improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. *Surgical Endoscopy* 2005; **19**(9): 1220-1226.

Scott DJ, Bergen PC, Euhus DM, Guo WA, Jeyarajah DR, Laycock R, Rege RV, Tesfay ST, Thompson WM, Valentine RJ, Jones DB. Intense laparoscopic skills training improves operative performance of surgery residents. *American College of Surgeons Surgical Forum* 1999; **50**(L): 670-671.

Scott DJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Valentine RJ, Euhus DM, Jeyarajah DR, Thompson WM, Jones DB. Laparoscopic training on bench models: Better and more cost effective than operating room experience? *Journal of the American College of Surgeons* 2000; **191**(3): 272-283.

Sedlack RE and Kolars JC. Colonoscopy curriculum development and performance-based assessment criteria on a computer-based endoscopy simulator. *Academic Medicine* 2002; **77**: 750-751.

Sedlack RE and Kolars JC. Validation of a computer-based colonoscopy simulator. *Gastrointestinal Endoscopy* 2003; **57**(2): 214-218.

Sedlack RE and Kolars JC. Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. *American Journal of Gastroenterology* 2004; **99**(1): 33-37.

Sedlack RE, Kolars JC, Alexander JA. Computer simulation training enhances patient comfort during endoscopy. *Clinical Gastroenterology and Hepatology* 2004; **2**: 348-352.

Sewell C, Morris D, Blevins N, Barbagli F, Salisbury K. An Event-Driven Framework for the Simulation of Complex Surgical Procedures. Proceedings of the International Conference on Medical Imaging Computing and Computer Assisted Intervention (MICCAI). Springer Lecture Notes in Computer Science, Vol II. St. Malo, France: 2004.

Seymour NE. Virtual reality in general surgical training. *European Surgery - Acta Chirurgica Austriaca Supplement.* 2005; **37**(5): 298-303.

Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance: results of a randomised, double-blinded study. *Annals of Surgery* 2002; **236**(4): 458-463.

Seymour NE and Rotnes JS. Challenges to the development of complex virtual reality surgical simulations. *Surgical Endoscopy* 2006; **20**(11): 1774-1777.

Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL. Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. *Journal of Shoulder and Elbow Surgery* 2004; **13**(2): 196-205.

Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, Maddern GJ. Surgical simulation: a systematic review. *Annals of Surgery* 2006; **243**(3): 291-300.

Tuggy ML. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *The Journal of the American Board of Family Practice* 1998; **11**(6): 426-433.

Valentine RJ and Rege RV. Integrating technical competency into the surgical curriculum: doing more with less. *Surgical Clinics of North America* 2004; **84**(6): 1647-1667.

Van Sickle KR, Ritter EM, Smith CD. The pretrained novice: Using simulation-based training to improve learning in the operating room. *Surgical Innovation.* 2006; **13**(3): 198-204.

Villegas L, Schneider BE, Callery MP, Jones DB. Laparoscopic skills training. *Surgical Endoscopy* 2003; **17**(12): 1879-1888.

Wagh MS and Waxman I. Animal Models for Endoscopic Simulation. *Gastrointestinal Endoscopy Clinics of North America* 2006; **16**(3): 451-456.

Wong K and Stewart F. Competency-based training of basic surgical trainees using human cadavers. *ANZ Journal of Surgery* 2004; **74**: 639-642.

Woodrow SI, Segouin C, Armbruster J, Hamstra SJ, Hodges B. Duty hours reforms in the United States, France, and Canada: Is it time to refocus our attention on education? *Academic Medicine* 2006; **81**(12): 1045-1051.

APPENDIX A – EXCLUDED STUDIES

Appendix A - Excluded studies

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

Excluded Studies

Study	Reason for exclusion
Adrales GL, Chu UB, Witzke DB, Donnelly MB, Hoskins D, Mastrangelo MJ, Jr., Gandsas A, Park AE. Evaluating minimally invasive surgery training using low-cost mechanical simulations. *Surgical Endoscopy* 2003; **17**(4): 580-585.	Assessed in a simulator/via simulation
Aggarwal R, Grantcharov T, Moorthy K, Hance J, Darzi A. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *American Journal of Surgery* 2006; **191**: 128-133.	Assessed in a simulator/via simulation
Ahlberg G, Heikkinen T, Iselius L, Leijonmarck CE, Rutqvist J, Arvidsson D. Does training in a virtual reality simulator improve surgical performance? *Surgical Endoscopy* 2002; **16** (1): 126-129.	Assessment in animals
Ahlberg G, Kruuna O, Leijonmarck C-E, Ovaska J, Rosseland A, Sandbu R, Stromberg C, Arvidsson D. Is the learning curve for laparoscopic fundoplication determined by the teacher or the pupil? *American Journal of Surgery.* 2005; **189**(2): 184-189.	Focus not on skills training
Alinier G. Determining the value of simulation in nurse education: study design and initial results. *Nurse Education in Practice* 2004; **4**(3): 200-207.	Not surgical
Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, Hutchison C. Assessment of technical skills transfer from the bench training model to the human model. *American Journal of Surgery* 1999; **177**(2): 167-170.	Assessment in cadaver
Andreatta PB, Woodrum DT, Birkmeyer JD, Yellamanchilla RK, Doherty GM, Gauger PG, Minter RM. Laparoscopic skills are improving with LapMentor training: results of a randomised, double-blinded study. *Annals of Surgery* 2006; **243**(6): 854-863.	Assessment in animals
Avinash S, Abhay D, Ramkrishna P, Chetan K, Pritha B. Cadaver as a model for laparoscopic training. *Indian Journal of Gastroenterology* 2005; **24**(2): 111-113.	Assessment in animals
Beard JD, Jolly BC, Newble DI, Thomas WE, Donnelly J, Southgate LJ. Assessing the technical skills of surgical trainees. *British Journal of Surgery* 2005; **92**(6): 778-782.	Focus not on skills training
Biasutto SN, Caussa LI, Criado del Rio LE. Teaching anatomy: cadavers *vs.* computers? *Annals of Anatomy* 2006; **188**(2): 187-190.	Assessment in cadaver
Blum MG, Powers TW, Sundaresan S. Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Annals of Thoracic Surgery* 2004; **78**(1): 287-291.	Not surgical
Broe D, Ridgway PF, Johnson S, Tierney S, Conlon KC. Construct validation of a novel hybrid surgical simulator. *Surgical Endoscopy* 2006; **20**(6): 900-904.	Assessed in a simulator/via simulation
Burgoyne L and Cyna A. Laryngeal mask vs intubating laryngeal mask: insertion and ventilation by inexperienced resuscitators. *Anaesthesia & Intensive Care* 2001; **29**(6): 604-608.	Not surgical
Carter MB, Wesley G, Larson GM. Didactic lecture versus instructional standardised patient interaction in the surgical clerkship. *American Journal of Surgery* 2005; **189**(2): 243-248.	Not surgical
Caversaccio M, Eichenberger A, Hausler R. Virtual simulator as a training tool for endonasal surgery. *American Journal Rhinology* 2003; **17**(5): 283-290.	Focus not on skills training

Chapman DM, Rhee KJ, Marx JA, Honigman B, Panacek EA, Martinez D, Brofeldt BT, Cavanaugh SH. Open thoracotomy procedural competency: Validity study of teaching and assessment modalities. *Annals of Emergency Medicine.* 1996; **28**(6): 641-647.	Assessment in animals
Clark JA, Volchok JA, Hazey JW, Ssadighi PJ, Fanelli RD. Initial experience using an endoscopic simulator to train surgical residents in flexible endoscopy in a community medical centre residency program. *Current Surgery* 2005; **62**(1): 59-63.	Assessed in a simulator/via simulation
Datta V, Bann S, Beard J, Mandalia M, Darzi A. Comparison of bench test evaluations of surgical skill with live operating performance assessments. *Journal of the American College of Surgeons* 2004; **199**(4): 603-606.	No comparison between groups
Datta V, Bann S, Mandalia M, Darzi A. The surgical efficiency score: a feasible, reliable, and valid method of skills assessment. *American Journal of Surgery* 2006; **192**(3): 372-378.	Assessed in a simulator/via simulation
Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *Journal of the American College of Surgeons* 2001; **193**(5): 479-485.	Assessed in a simulator/via simulation
Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos P, Fowler DL. Construct validity for the LAPSIM laparoscopic surgical simulator. *Surgical Endoscopy.* 2005; **19**(3): 401-405.	Assessed in a simulator/via simulation
Fearn SJ, Burke K, Hartley DE, Semmens JB, Lawrence-Brown MMD. A laparoscopic access technique for endovascular procedures: Surgeon training in an animal model. *Journal of Endovascular Therapy* 2006; **13**(3): 350-356.	Assessment in animals
Ford GS, Mazzone MA, Taylor K. Effect of computer-assisted instruction versus traditional modes of instruction on student learning of musculoskeletal special tests. *Journal of Physical Therapy Education* 2005; **19**(2): 22-30.	Not surgical
Fried GM, Derossis AM, Bothwell J, Sigman HH. Comparison of laparoscopic performance *in vivo* with performance measured in a laparoscopic simulator. *Surgical Endoscopy* 1999; **13**(11): 1077-1081.	No comparison between groups
Friedlich M, MacRae H, Oandasan I, Tannenbaum D, Batty H, Reznick R, Regehr G. Structured assessment of minor surgical skills (SAMSS) for family medicine residents. *Academic Medicine.* 2001; **76**(12): 1241-1246.	Assessed in a simulator/via simulation
Gallagher AG, Smith CD, Bowers SP, Seymour NE, Pearson AMS, Hananel D, Satava RM. Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. *Journal of the American College of Surgeons* 2003; **197**: 479-488.	Assessed in a simulator/via simulation
Goldmann K and Steinfeldt T. Acquisition of basic fiberoptic intubation skills with a virtual reality airway simulator. *Journal of Clinical Anaesthesia* 2006; **18**(3): 173-178.	Case series
Grantcharov TP, Rosenberg J, Pahle E, Funch-Jensen P. Virtual reality computer simulation: An objective method for the evaluation of laparoscopic surgical skills. *Surgical Endoscopy* 2001; **15**(3): 242-244.	Assessment in animals
Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA. Laboratory based training in urological microsurgery with bench model simulators: a randomised controlled trial evaluating the durability of technical skill. *Journal of Urology* 2004; **172**(1): 378-381.	Assessment in animals
Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Annals of Surgery* 2004; **240**(2): 374-381.	Assessment in animals
Halvorsen FH, Elle OJ, Dalinin VV, Mork BE, Sorhus V, Rotnes JS, Fosse E. Virtual reality simulator training equals mechanical robotic training in improving robot-assisted basic suturing skills. *Surgical Endoscopy.* 2006; **20**(10): 1565-1569.	Assessment in animals

Hamad MA, Mentges B, Buess G. Laparoscopic sutured anastomosis of the bowel. *Surgical Endoscopy* 2003; **17**(11): 1840-1844.	Assessment in animals
Hamilton EC, Scott DJ, Fleming JB, Rege RV, Laycock R, Bergen PC, Tesfay ST, Jones DB. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical Endoscopy* 2002; **16**(3): 406-411.	No untrained group
Hance J, Aggarwal R, Moorthy K, Munz Y, Undre S, Darzi A. Assessment of psychomotor skills acquisition during laparoscopic cholecystectomy courses. *American Journal of Surgery* 2005; **190**(3): 507-511.	Assessment in animals
Hariri S, Rawn C, Srivastava S, Youngblood P, Ladd A. Evaluation of a surgical simulator for learning clinical anatomy. *Medical Education* 2004; **38**(8): 896-902.	Not skills training
Hart R, Doherty DA, Karthigasu K, Garry R. The value of virtual reality-simulator training in the development of laparoscopic surgical skills. *Journal of Minimally Invasive Gynaecology* 2006; **13**(2): 126-133.	Assessment in animals
Heinrich M, Tillo N, Kirlum HJ, Till H. Comparison of different training models for laparoscopic surgery in neonates and small infants. *Surgical Endoscopy* 2006; **20**: 641-644.	Assessment in animals
Henkel TO, Potempa DM, Rassweiler J, Manegold BC, Alken P. Lap simulator, animal studies, and the Laptent. Bridging the gap between open and laparoscopic surgery. *Surgical Endoscopy* 1993; **7**(6): 539-543.	Focus not on skills training
Hochberger J, Matthes K, Maiss J, Koebnick C, Hahn EG, Cohen J. Training with the compactEASIE biologic endoscopy simulator significantly improves hemostatic skill of gastroenterology Fellows: a randomised controlled comparison with clinical endoscopy training alone. *Gastrointestinal Endoscopy* 2005; **61**(2): 204-215.	Assessed in a simulator/via simulation
Hyltander A, Liljegren E, Rhodin PH, Lonroth H. The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surgical Endoscopy* 2002; **16**(9): 1324-1328.	Assessment in animals
Jacomides L, Ogan K, Cadeddu JA, Pearle MS. Use of a virtual reality simulator for ureteroscopy training. *Journal of Urology* 2004; **171**(1): 320-323.	Assessed in a simulator/via simulation
Kimara T, Kawabe A, Suzuki K, Wada H. Usefulness of a virtual reality simulator or training box for endoscopic surgery training. *Surgical Endoscopy* 2006; **20**: 656-659.	Assessment in animals
Knudson MM and Sisley AC. Training residents using simulation technology: experience with ultrasound for trauma. *Journal of Trauma* 2000; **48**(4): 659-665.	Not surgical
Korndorffer J, Jr., Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *Journal of the American College of Surgeons.* 2005; **201**(1): 23-29.	Assessment in animals
Korndorffer JR, Jr., Hayes DJ, Dunne JB, Sierra R, Touchard CL, Markert RJ, Scott DJ. Development and transferability of a cost-effective laparoscopic camera navigation simulator. *Surgical Endoscopy* 2005; **19**(2): 161-167.	Assessment in animals
Korndorffer JR, Stefanidis D, Scott DJ. Laparoscopic skills laboratories: current assessment and a call for resident training standards. *American Journal of Surgery* 2006; **191**(1): 17-22.	Focus not on skills training
Kothari SN, Kaplan BJ, DeMaria EJ, Broderick TJ, Merrell RC. Training in laparoscopic suturing skills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. *Journal of Laparoendoscopic & Advanced Surgical Techniques. Part A.* 2002; **12**(3): 167-173.	Assessed in a simulator/via simulation
Larsson A. Intracorporeal suturing and knot tying in surgical simulation. *Studies in Health Technology Information* 2001; **81**: 266-271.	Focus not on skills training
Lee SK. Trauma assessment training with a patient simulator: a prospective, randomised study. *Journal of Trauma* 2003; **55**(4): 651-657.	Not surgical

Lehmann KS, Ritz JP, Maass H, Cakmak HK, Kuehnapfel UG, Germer CT, Bretthauer G, Buhr HJ. A prospective randomised study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Annals of Surgery* 2005; **241**(3): 442-449.	Assessed in a simulator/via simulation
Madam AK, Frantzides CT, Shervin N, Tebbit CL. Assessment of individual hand performance in box trainers compared to virtual reality trainers. *The American Surgeon* 2003; **69**: 1112-1114.	Assessed in a simulator/via simulation
Madan AK, Frantzides CT, Tebbit C, Quiros RM. Participants' opinions of laparoscopic training devices after a basic laparoscopic training course. *American Journal of Surgery* 2005; **189**(6): 758-761.	Assessment in animals
Maiss J, Dumser C, Zopf Y, Naegel A, Krauss N, Hochberger J, Matthes K, Hahn EG, Schwab D. "Hemodynamic efficacy" of two endoscopic clip devices used in the treatment of bleeding vessels, tested in an experimental setting using the compact Erlangen Active Simulator for Interventional Endoscopy (compactEASIE) training model. *Endoscopy* 2006; **38**(6): 575-580.	Assessed in a simulator/via simulation
Maiss J, Wiesnet J, Proeschel A, Matthes K, Prat F, Cohen J, Chaussade S, Sautereau D, Naegel A, Krauss N, Peters A, Hahn EG, Hochberger J. Objective benefit of a 1-day training course in endoscopic hemostasis using the "compactEASIE" endoscopy simulator. *Endoscopy* 2005; **37**(6): 552-558.	Assessed in a simulator/via simulation
Martin JA, Regehr G, Reznick RK, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skills (OSATS) for surgical residents. *British Journal of Surgery* 1997; **84**(2): 273-278.	Focus not on skills training
Matsumoto ED, Kondraske GV, Ogan K, Jacomides L, Wilhelm DM, Pearle MS, Cadeddu JA. Assessment of basic human performance resources predicts performance of ureteroscopy. *American Journal of Surgery* 2006; **191**(6): 817-820.	Assessment in cadaver
McClusky DA, III, Ritter EM, Lederman AB, Gallagher AG, Smith CD. Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. *The American Surgeon* 2005; **71**(1): 13-20.	Assessed in a simulator/via simulation
Molinas CR, Binda MM, Mailova K, Koninckx PR. The rabbit nephrectomy model for training in laparoscopic surgery. *Human Reproduction* 2004; **19**(1): 185-190.	Case series
Moorthy K, Mansoori M, Bello F, Hance J, Undre S, Munz Y, Darzi A. Evaluation of the benefit of VR simulation in a multi-media web-based educational tool. *Studies in Health Technology Information* 2004; **98**: 247-252.	Assessed in a simulator/via simulation
Moorthy K, Munz Y, Adams S, Pandey V, Darzi A, Imperial C. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theatre. *American Journal of Surgery* 2006; **192**(1): 114-118.	Assessed in a simulator/via simulation
Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers: is one superior to the other? *Surgical Endoscopy* 2004; **18**(3): 485-494.	Assessed in a simulator/via simulation
Nadu A, Olsson LE, Abbou CC. Simple model for training in the laparoscopic vesicourethral running anastomosis. *Journal of Endourology* 2003; **17**(7): 481-484.	Results for two groups not reported separately
Ogan K, Jacomides L, Shulman MJ, Roehrborn CG, Cadeddu JA, Pearle MS. Virtual ureteroscopy predicts ureteroscopic proficiency of medical students on a cadaver. *Journal of Urology* 2004; **172**(2): 667-671.	Assessment in cadaver
Ost D, DeRosiers A, Britt EJ, Fein M, Lesser L, Mehta AC. Assessment of a bronchoscopy simulator. *American Journal of Respiratory and Critical Care* 2001; **164**: 2248-2255.	Not surgical

O'Toole RV, Playter RR, Krummel TM, Blank WC, Cornelius NH, Roberts WR, Bell WJ, Raibert M. Measuring and developing suturing technique with a virtual reality surgical simulator. *Journal of the American College of Surgeons* 1999; **189**(1): 114-127.	Assessed in a simulator/via simulation
Paisley AM, Baldwin PJ, Paterson-Brown S. Validity of surgical simulation for the assessment of operative skill. *British Journal of Surgery* 2001; **88**(11): 1525-1532.	Results for two groups not reported separately
Peugnet F, Dubois P, Rouland JF. Virtual reality versus conventional training in retinal photocoagulation: a first clinical assessment. *Computer Aided Surgery* 1998; **3**(1): 20-26.	Not surgical
Richards C, Rosen J, Hannaford B, Pellegrini C, Sinanan M. Skills evaluation in minimally invasive surgery using force/torque signatures. *Surgical Endoscopy* 2000; **14**(9): 791-798.	No comparison between groups
Ritter ME, McClusky DA, Gallagher AG, Enochsson L, Smith D. Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. *American Journal of Surgery* 2006; **192**: 379-384.	Assessed in a simulator/via simulation
Rosenthal R, Gantert WA, Scheidegger D, Oertli D. Can skills assessment on a virtual reality trainer predict a surgical trainee's talent in laparoscopic surgery? *Surgical Endoscopy* 2006; **20**(8): 1286-1290.	Assessed in a simulator/via simulation
Rossi JV, Verma D, Fujii GY, Lakhanpal RR, Wu SL, Humayun MS, De Juan E Jr. Virtual vitreoretinal surgical simulator as a training tool. *Retina* 2004; **24**(2): 231-236.	Assessed in a simulator/via simulation
Rulli F, Cina G, Galata G, Cina A, Vincenzoni C, Fiorentino A, Farinon AM. Teaching subfascial perforator veins surgery: Survey on a 2-day hands-on course. *ANZ Journal of Surgery.* 2004; **74**(12): 1116-1119.	Not skills training
Scerbo MW, Schmidt EA, Bliss JP. Comparison of a virtual reality simulator and simulated limbs for phlebotomy training. *Journal of Infusion Nursing* 2006; **29**(4): 214-224.	No untrained group
Schijven MP and Jakimowicz J. The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. *Surgical Endoscopy* 2004; **18**(1): 121-127.	Assessed in a simulator/via simulation
Scott DJ, Rege RV, Bergen PC, Guo WDA, Laycock R, Tesfay ST, Valentine RJ, Jones DB. Measuring operative performance after laparoscopic skills training: Edited videotape versus direct observation. *Journal of Laparoendoscopic & Advanced Surgical Techniques -Part A* 2000; **10**(4): 183-190.	Focus not on skills training
Seymour NE. Integrating simulation into a busy residency program. *Minimally Invasive Therapy and Allied Technology* 2005; **14**(4): 280-286.	Assessment in animals
Smeak DD, Beck ML, Shaffer CA, Gregg CG. Evaluation of video tape and a simulator for instruction of basic surgical skills. *Veterinary Surgery* 1991; **20**(1): 30-36.	Assessment in animals
Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL. Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. *Journal of Shoulder and Elbow Surgery* 2004; **13**(2): 196-205.	Assessed in a simulator/via simulation
Stefanidis D, Sierra R, Korndorffer JR, Jr., Dunne JB, Markley S, Touchard CL, Scott DJ. Intensive continuing medical education course training on simulators results in proficiency for laparoscopic suturing. *American Journal of Surgery* 2006; **191**(1): 23-27.	Case series
Strom P, Hedman L, Sarna L, Kjellin A, Wredmark T, Fellander-Tsai L. Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomised crossover study in surgical residents. *Surgical Endoscopy* 2006; **20**(9): 1383-1388.	Assessed in a simulator/via simulation

Suzuki N, Thomas-Gibson S, Vance M, Fraser C, Swain D, Schofield G, Saunders BP. Efficacy of an accelerated colonoscopy training week: Audit from one national colonoscopy training centre in the UK. *Digestive Endoscopy.* 2006; **18**(4): 288-293.	Case series
Torkington J, Smith SG, Rees BI, Darzi A. Skill transfer from virtual reality to a real laparoscopic task. *Surgical Endoscopy* 2001; **15**(10): 1076-1079.	Assessed in a simulator/via simulation
Van Sickle KR, Ritter EM, Smith CD. The pre-trained novice: Using simulation-based training to improve learning in the operating room. *Surgical Innovation* 2006; **13**(3): 198-204.	Assessment in animals
Verdaasdonk EG, Stassen LP, van Wijk RP, Dankelman J. The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. *Surgical Endoscopy* 2006;	Assessed in a simulator/via simulation
Watterson JD, Beiko DT, Kuan JK, Denstedt JD. Randomised prospective blinded study validating acquisition of ureteroscopy skills using computer based virtual reality endourological simulator. *Journal of Urology* 2002; **168**(5): 1928-1932.	Assessed in a simulator/via simulation
Wong K and Stewart F. Competency-based training of basic surgical trainees using human cadavers. *ANZ Journal of Surgery* 2004; **74**(8): 639-642.	Case series
Woodrum DT, Andreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM. Construct validity of the LapSim laparoscopic surgical simulator. *American Journal of Surgery* 2006; **191**(1): 28-32.	Case series
Youngblood PL, Srivastava S, Curet M, Heinrichs WL, Dev P, Wren SM. Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. *Journal of the American College of Surgeons* 2005; **200**(4): 546-551.	Assessment in animals

APPENDIX B – HIERARCHY OF EVIDENCE

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.

APPENDIX C – METHODOLOGICAL ASSESSMENT AND STUDY DESIGN TABLES

Appendix C – Methodological assessment and study design tables

Appendix C.1 Study design tables – MIST-VR training *vs.* no simulator training

<table>
<tr><th>Author</th><th>Intervention</th><th>Study Design</th><th>Study population</th></tr>
<tr>
<td>Grantcharov et al. 2004

Location
Department of Surgical Gastroenterology, Aarhus University, and Department of Surgical Gastroenterology, University of Copenhagen at Hvidovre, Glostrup, and Gentofte Hospitals.

Denmark</td>
<td>Objective: To investigate whether laparoscopic skills acquired on MIST-VR transfer to performance of laparoscopic cholecystectomy.

Pre-Test:
16 surgical residents performed laparoscopic cholecystectomy on patients under supervision of experienced surgeon.

Intervention:
Participants then randomised:
• Training: MIST-VR: 10 repetitions of all 6 tasks: 1) grasp virtual sphere and place in virtual box, 2) grasp virtual sphere, transfer instrument and place in virtual box, 3) grasp alternative segments of virtual pipe, 4) grasp virtual sphere, touch it with tip of other instrument, withdraw and reinsert, and once more touch sphere, 5) grasp sphere, touch virtual plates by other instrument and virtual diathermy them away with pedal 6) combines actions of 4) and 5).
• Control: No simulator training.

Time to assessment: within 14 days of training.

Assessment:
All participants again performed a supervised laparoscopic cholecystectomy on patients.
Both operative procedures recorded on video tape and assessed independently by 2 senior surgeons using predefined rating scales. Only part of the procedure was assessed starting from the point at which clips were applied to the cystic artery and cystic duct, finishing with dissection of the gallbladder from the liver bed.

Device:
MIST-VR: Mentice Medical Simulation, Gothenburg, Sweden</td>
<td>Randomised controlled trial

Method of randomisation: not stated

Allocation concealment: sealed envelopes

Level of evidence: II

Blinding: supervising surgeon blinded during assessment operation. Assessors blinded to training status and performed evaluation independently.

Intention to treat: not stated

Power calculation: not stated

Lost to assessment: n = 4 (video recorder malfunction)
• Training: n = 2
• Control: n = 2

Study period: August 2000 – August 2002

Outcome measures: rating scale (validated) only assessed psychomotor skills:
• Time to complete procedure
• Error score
• Economy of movement</td>
<td>Sample size: n = 20
• Training: n = 10
• Control: n = 10

Baseline characteristics of participants:
<table>
<tr><th></th><th>Training</th><th>Control</th></tr>
<tr><td>Sex ratio (M:F)</td><td>5 : 3</td><td>5 : 3</td></tr>
<tr><td>Age (yrs)</td><td>36.5 (31 – 40)</td><td>36.5 (32 – 44)</td></tr>
<tr><td>Time since graduation (yrs)*</td><td>7 (4-10)</td><td>7 (5 – 14)</td></tr>
<tr><td>No of previous laparoscopic cholecystectomies performed*</td><td>6 (2 -7)</td><td>4 (0 – 8)</td></tr>
</table>
*Values are median (range)

Inclusion: general surgical residents

Exclusion: not stated

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

Appendix C.1 Study design tables – Procedicus VIST™ training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Chaer *et al.* 2006 **Location** Columbia University, Weill Cornell Division of Vascular Surgery, New York Presbyterian Hospital, New York. USA	Objective: To measure the effectiveness of simulator training on the performance of catheter-based interventions by surgical residents. All participants: All participants were provided with didactic teaching in the form of reading material and a lecture on catheter-based intervention. An entrance survey was performed to determine demographics as well as previous experiences. All participants were administered a visuospatial evaluation consisting of a card rotation and a cube comparison (Educating Testing Service, Princeton, NJ) to evaluate pre-study differences between the capability of residents to perceive 3 dimensional structures. Intervention: Participants then randomised: • **Training**: Procedicus VIST™, computer-based, haptic simulator for standardised iliofemerol angioplasty/stenting. Supervision present. Endpoint of simulation training was independent completion of the procedure by the participant with proficiency in all basic endovascular techniques. Training was not allowed to exceed 2 hours. • **Control**: No simulator training. Time to assessment: within 14 days of randomisation. Assessment: All participants performed 2 consecutive catheter based interventions for lower extremity occlusive disease supervised by a blinded attending vascular surgeon. Assessment was done by the attending vascular surgeon using an 18 step checklist of the required steps for a standard catheter based intervention as well as a global rating scale of the participant's endovascular technique and ability (validated). All participants were verbally guided through the steps of the procedure by the attending surgeon while their technical skills were assessed. Device: Procedicus VIST™ (Mentice, Inc., Evanston, IL). Desktop PC (Intel Xeon 2.66 GHz, 1 GB RAM, nVIDIA GeForce4 Ti 4200 with AGP 8X) with 3-D software of the human arterial system. This is coupled to a haptic module utilising a force feedback system. The instructional system is displayed on a touch screen monitor and a simulated fluoroscopic image is displayed on a second monitor).	**Randomised controlled trial** Method of randomisation: not stated Allocation concealment: sealed envelopes Level of evidence: II Intention to treat: not stated Power calculation: not stated Lost to assessment: not stated Study period: not stated Blinding: assessing vascular surgeon blinded to training status Outcome measures: criteria assessed included: • Time and motion • Wire and catheter handling • Awareness of wire position • Maintenance of wire stability • Awareness of fluoroscopy usage • Precision or wire/catheter technique • Flow of operation • Knowledge of procedure • Quality of final product • Ability to complete the case • Need for verbal prompts • Attending takeover	Sample size: n = 20 • **Training**: n = 10 • **Control**: n = 10 Baseline characteristics of participants: information collected but data not shown Inclusion: general surgical residents with no prior endovascular experience Exclusion: not stated Details of patients for live assessment: not stated

Appendix C.1 Study design tables – Total extraperitoneal (TEP) hernia repair curriculum training *vs.* no TEP hernia repair curriculum training

Author	Intervention	Study Design	Study population
Hamilton *et al.* 2001 **Location** Southwestern Centre for Minimally Invasive Surgery, Department of Surgery, University of Texas Southwestern Medical Centre, Dallas, Texas. USA	Objective: To evaluate the impact of a laparoscopic hernia repair curriculum on resident surgeons' operative performance and confidence with a laparoscopic TEP hernia repair. Participants: All participants completed a questionnaire regarding baseline operative experience and familiarity and comfort with total extraperitoneal (TEP) hernia repair. Pre-test: All performed a laparoscopic TEP hernia repair during week 1 of a general surgery rotation; baseline performance evaluated by 1 of 4 faculty members using a validated global assessment tool. Each operation performed under guidance of a faculty surgeon who served as both first assistant and assessor. Participants were encouraged to work as independently as possible. Intervention: Participants randomised to two groups: • **Training**: 10 separate 30 minute sessions over a 2 week period. Training included a detailed instructional video (watched once at beginning and once at end of study); an interactive CD-ROM; and a moulded rubber hernia simulator (alternate daily instruction via CD-ROM and simulator). • **Control**: TEP hernia repair curriculum training. Time to assessment: after completion of 2 weeks of training. Assessment: Participants were tested after the training period in an operating room with real TEP hernia repair under conditions similar to the pre-test evaluation by the same group of assessors. Device: TEP simulator designed by Southwestern Centre for Minimally Invasive Surgery in conjunction with GSI. Consists of a rubber model of a human pelvis including all relevant pelvic anatomy. The model accepts a laparoscope and trocars and a replaceable rubber insert that allows participants to practise repeated mesh fixation over indirect, direct, or femoral defects.	**Randomised controlled trial** Randomisation method: not stated Allocation concealment: not stated Level of evidence: II Intention to treat: not stated Power calculation: not stated Follow-up: not stated Lost to assessment: not stated Study period: January 2000 – March 2001 Blinding: evaluators blinded to participants training status Outcome measures: performance measures for laparoscopic TEP hernia repair: • Respect for tissue • Time and motion • Instrument handling • Knowledge of instruments • Flow of operation • Use of assistants • Knowledge of specific procedure • Overall performance.	Sample Size: 21 • **Training:** n = 10 • **Control:** n = 11 Baseline characteristics of participants: not stated Gender Mix: not stated Inclusion Criteria: 3rd and 4th year surgery residents on surgical rotation. Exclusion Criteria: not stated Details of patients for live assessment: not stated

Appendix C.1 Study design tables – MIST-VR training *vs.* no simulator training

Authors	Intervention	Study Design	Study Population
Seymour *et al.* 2002 **Location** Department of Surgery, Yale University School of Medicine, New Haven, Connecticut, USA, and Department of Psychology, Queen's University, Belfast. Northern Ireland	Objective: To demonstrate that virtual reality training transfers technical skills to the operating room environment. Pre-test: All residents completed a series of previously validated tests to assess fundamental abilities – visuospatial (Card Rotation Cube Comparison and Map Plan tests), perceptual (Pictorial Surface Orientation test), and psychomotor (MIST-VR at medium level of difficulties). Intervention: Participants randomised: • **MIST-VR training + standard training (ST):** criterion levels first established by 4 experienced surgeons on Manipulate and Diathermy task at difficult level. Participants were then required to perform the same task equally as well with both hands on 2 consecutive trials at this criterion level. Training lessons lasted approximately 1 hour, and sessions were continued until expert criterion levels were achieved. Training was supervised. • **Control:** Standard programmatic training – no simulator training. All participants watched a standard video. This video defined specific deviations from optimal performance that would be considered errors. After the viewing, all residents were given an 8-question multiple-choice exam that tested the recognition of these errors. Time to assessment: NR Assessment: All residents performed laparoscopic cholecystectomy (gall bladder excision after division of the identified cystic structures) using a standardised 2-handed method with 1 of the surgeon-investigators. This part of procedure was video recorded with voice audio. Procedures with attending takeover were flagged for examination of audio. Each procedural video was viewed without audio by 2 surgeon-investigators. Device: MIST-VR, Mentice AB, Gothenburg, Sweden. Frameset v.1.2. MIST-VR system (frameset v 1.2) was run on a desktop PC (400 MHz Pentium II, 64 MB RAM). The video subsystem employed (Matrox Mystique, * MB SDRAM) delivered a frame rate of approximately 15 frames per second, permitting near-real-time translations of instrument movements to the video screen. The laparoscopic interface input device (Immersion	**Randomised controlled trial** Randomisation method: not stated Participants stratified by postgraduate year Allocation concealment: not stated Level of Evidence: II Intention to treat: not stated Power calculation: not stated Blinding: assessors blinded to participant's training status Follow-up: not stated Lost to assessment: not stated Study Period: not stated Outcome measures: from archive footage, 8 events associated with the excisional phase of procedure were identified as errors and chosen as study measurements: • Lack of progress • Gallbladder injury • Liver injury • Incorrect plane of dissection • Burn non-target tissue • Tearing tissue • Instrument out of view • Attending takeover Gall bladder excision phase scored on a minute by minute basis using a scoring matrix. Errors recorded using fixed-interval time span sampling.	Sample Size: n = 16 • **Training:** n = 8 • **Control:** n = 8 Baseline characteristics of participants: Female n = 5; Male n = 11 Inclusion Criteria: surgical residents (years 1-4) Exclusion Criteria: not stated Details of patients for live assessment: not stated

	Corporation, San Jose, CA) consisted of 2 laparoscopic instruments at a comfortable surgical height relative to the operator, mounted in a frame by position-sensing gimbals that provided 6 degrees of freedom, as well as a foot pedal to activate simulated electrosurgery instruments.		

Appendix C.1 Study design tables – Video trainer (SCMIS GEM) training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Scott *et al.* 2000 **Location** Department of Surgery, University of Texas Southwestern Medical Centre, Dallas, Texas. USA	Objective: To develop a model to provide intense laparoscopic skills training to surgical residents and to determine if improvement of skills on a video trainer translates into an improvement in operative performance. Participants: All participants randomised. All participants completed a baseline questionnaire regarding earlier laparoscopic experience and competency in laparoscopic skills. Pre-test: During week 1, all residents were tested on all tasks on the video-trainer (each task was performed 3 times). No practice was allowed before testing except for the suture foam drill so participants could get familiar with device. All participants performed a laparoscopic cholecystectomy in the operating theatre. Operations performed in a one-handed or two-handed fashion according to the supervising surgeon's preference. Participants allowed to perform as much of the operation as possible. Operations were supervised by 1 of 3 designated faculty surgeons (blinded) who also acted as first assistant during the operation. Participants were asked about key anatomic landmarks and their operative plan. Operations were videotaped for investigations outside the scope of this study (used in separate study Scott *et al.* 2000 – reference below). Intervention: • **Training**: SCMIS GEM video trainer: During weeks 2 and 3, residents in training group met for at least 30 minutes a day for 10 days to perform the 5 established laparoscopic drills on a video trainer. Tasks were: bean drop, running string, checkerboard, block move, and suture foam. • **Control**: no simulator training. Time to assessment: During week 4. Assessment: All residents again tested on the 5 tasks on the video trainer (each task performed 3 times), and average completion time recorded. All residents performed another laparoscopic cholecystectomy in the operating theatre with the same first assistant surgeon. The same 3 faculty evaluators performed global assessments based on direct observation. Global assessments performed by 3 additional faculty surgeons who did not participate in operations and were blinded to training status.	**Randomised controlled trial** Method of randomisation: random digits table Allocation concealment: not stated Level of evidence: II Intention to treat: not stated Power calculation: nonparametric power analysis performed to determine sample size needed to detect an effect of training. 27 or more participants provide a power of at least 0.8 with a type I error of 0.05. Blinding: assessing surgeons and faculty evaluators blinded Lost to assessment: scheduling difficulties • **Training**: n = 4 • **Control**: n = 1 Study period: August 1998 – January 1999 Outcome measures: improvement defined as baseline minus post training score calculated for each resident, adjusted by linear analysis of covariance for differences in baseline scores. • Respect for tissue • Time and motion • Instrument handling • Knowledge of instruments • Flow of operation • Knowledge of specific procedure • Overall performance Scott *et al.* 2000* (not the included study, see reference below) compared the global assessment data from direct observations in this study with global assessments of the videotaped procedures. Global assessment of the videotapes was conducted by the same assessors as global assessments from direct observation. Correlation coefficients for videotape versus direct observation for five global assessment criteria were <0.33 for both raters (NS for all values). The correlation coefficient for interrater reliability for the overall score was 0.57 (P =	Sample size: n = 27 • **Training**: n = 13 • **Control**: n = 14 Baseline characteristics of participants: **Training**: Mean no. previous laparoscopic cholecystectomies performed*: 15 **Control**: 18 * as either the surgeon or first assistant (P = 0.501). Inclusion: 2nd and 3rd year general surgical residents rotating for one month periods on the general surgery services at Parkland Memorial Hospital. Exclusion: not stated Details of patients for live assessment: patients with the diagnosis of symptomatic cholelithiasis for whom elective cholecystectomy was indicated were scheduled for the observed cases.

	At the completion of the rotation, all residents were asked to complete a questionnaire regarding their laparoscopic experience. After the rotation, the control group was offered video trainer experience (results not included as part of this study). <u>Device</u>: Southwestern Centre for Minimally Invasive Surgery Guided Endoscopic Module (SCMIS GEM). Six-station video-trainer (Karl Storz Endoscopy, Culver City, CA).	0.01) for direct observation *vs.* 0.28 (NS) for videotape. The trained group had significantly better overall performance than the control group according to the assessment by direct observation (P = 0.02) but not by videotape assessment (NS).	

* Scott DJ, Rege RV, Bergen PC, Guo WDA, Laycock R, Tesfay ST, Valentine RJ, Jones DB. Measuring operative performance after laparoscopic skills training: Edited videotape versus direct observation. *Journal of Laparoendoscopic & Advanced Surgical Techniques - Part A* 2000; 10(4): 183-190.
NS not significant

Appendix C.1 Study design tables – Video trainer training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Scott *et al.* 1999 **Location** Southwestern Centre for Minimally Invasive Surgery, University of Texas Southwestern Medical Centre, Dallas, Texas. USA	Objective: To develop a model for intense laparoscopic skills training and to determine if improvement of skill level on a video-trainer translates into an improvement in operative performance. Participants: All participants randomised. Pre-test: All residents underwent a validated global assessment of their ability to perform laparoscopic cholecystectomy in the operating theatre. All residents tested on 5 standardised video-trainer tasks based on time needed to complete each task. Intervention: • **Training**: Video trainer: practised 5 established laparoscopic tasks for 30 minutes a day for 10 days. Tasks were: bean drop, running string, checkerboard, block move, and suture foam. • **Control**: no video-trainer training. Time to assessment: after 10 days of training. Assessment: Laparoscopic cholecystectomies conducted and scored on same global assessment scale. Rated on a scale of 1 – 5 in 8 areas by the same 3 raters. All residents repeated the video-trainer test. Device: Video-trainer (Karl Storz Endoscopy).	**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 Blinding: raters of operative procedures blinded Lost to assessment: not stated Study period: not stated Outcome measures: • Respect for tissue • Time and motion • Instrument handling • Knowledge of instruments • Flow of operation • Knowledge of specific procedure • Overall performance	Sample size: n = 22 • **Training**: n = 9 • **Control**: n = 13 Baseline characteristics of participants: not stated Inclusion: 2nd and 3rd year residents Exclusion: not stated Details of patients for live assessment: not stated

Appendix C.1 Study design tables – AccuTouch® flexible endoscopy simulator *vs.* no simulator training

Author	Intervention	Study Design	Study population
Ahlberg *et al.* 2005 **Location** Department of Surgery and Centre for Advanced Medical Simulation, Karolinska Hospital, Stockholm. Sweden	Objective: To investigate whether the use of the AccuTouch® flexible endoscopy simulator improves the early part of the learning curve in colonoscopy training. Participants: All participants were given the same theoretical study material, containing a booklet on colonoscopy, and a free sample instructive CD on colonoscopy. Intervention: Participants randomised: • **Training**: AccuTouch® flexible endoscopy simulator. An expert level of performance was established by assessing 5 expert colonoscopists. Participants instructed to train under supervision until reaching criterion level. To reach criterion level in simulator, the participants had to be able to intubate the caecum within 7 minutes without the use of sedation, a 'virtual attending', simulation tips and external view. More than 97% of procedure time had to be without patient discomfort and there had to be no period of severe or extreme discomfort. Navigation to the caecum had to be performed with less than 1500ml of air insufflated and with less than 15% of procedure time being in red-out. Training was initially under supervision. Feedback given at any time. • **Control**: no simulator training. Time to assessment: began within 1 week after training. Assessment: All participants performed their first 10 colonoscopic procedures in patients under supervision. Assessing supervisors instructed not to guide the participant. The colon was divided into 9 consecutive segments and the subjects were given a maximum of 15 minutes to pass each segment, with a maximum overall procedure time of 60 minutes. The supervisors were instructed to take over the procedure if the time limits were reached, if the patient experienced excessive discomfort or if the colon was poorly prepared. Patient gender and diagnosis were also noted. In addition, a patient form was completed in which the maximum discomfort during the procedure was graded on a visual analog scale. Device: AccuTouch® System virtual reality endoscopy simulator (version 1.3). Simulator comprises a computer, a display monitor and the	**Randomised controlled trial** Method of randomisation: not stated Allocation concealment: sealed envelopes Level of evidence: II Intention to treat: not stated Power calculation: not stated Blinding: patients and assessing supervisors blinded to training status of participants Lost to assessment: not stated Study period: not stated Outcome measures: • Total procedure time • Segment of colon where investigation stopped • Reason for stopping (if applicable) • Analgesic drugs given • Complications	Sample size: n = 12 (surgical n = 10; gastroenterological n = 2) • **Training**: n = 6 • **Control**: n = 6 Baseline characteristics of participants: 10 men; 2 women No previous experience in colonoscopy. All had experience in gastroscopy, with a minimum of 20 individually performed procedures. Inclusion: not stated Exclusion: not stated Details of patients for live assessment: patients designated to undergo diagnostic colonoscopy, on an all-comer basis, without a history of previous abdominal surgery were asked to participate.

	AccuTouch® endoscopic interface device. The simulator also includes a multi media function where instructive video clips and an anatomy/pathology atlas are demonstrated.		

Appendix C.1 Study design tables – AccuTouch® colonoscopy simulator training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Sedlack and Kolars 2004 **Location** Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota. USA	Objective: To determine if simulation training improves patient-based colonoscopy performance of novice colonoscopists. All participants: Participants randomised. Intervention: • **Training**: 6 hours of AccuTouch® colonoscopy simulator training over a 2 day period, comprising a multimedia tutorial followed by the performance of 20 – 25 simulated colonoscopies prior to beginning patient based training. Simulator and curriculum previously validated. • **Control:** No simulation training. Time to assessment: after 2 half days of simulator training Assessment: Participants underwent 4 – 8 weeks of patient based colonoscopy training during which they were supervised by 1 of 38 faculty gastroenterologists during one half day (ie 4 hour) assignment intervals. Each patient-based procedure was graded by the supervising staff utilising 7 performance parameters (both subjective and objective). Each patient was also asked to complete a questionnaire grading the discomfort experienced during the procedure. Device: AccuTouch® colonoscopy simulator (Immersion Medical, Gaithersburg, MD, Simulator engine 1.1.1). This computer based colonoscopy simulator consists of a specialised colonoscope that is inserted into a computer based module and provides 6 colonoscopy scenarios of varying complexity.	**Randomised controlled trial** Method of randomisation: not stated Allocation concealment: not stated Level of evidence: II Blinding: evaluating staff not blinded to participant training status Intention to treat: not stated Power calculation: not stated Lost to assessment: not stated Study period: not stated Outcome measures: • Time to reach maximum insertion • Depth of unassisted insertion • Independent procedure completion • Identify endoscopic landmarks • Insert in a safe manner • Adequately visualise mucosa on withdrawal • Respond appropriately to patient discomfort (using Likert scale)	Sample size: n = 8 • **Simulator training**: n = 4 • **Control**: n = 4 Baseline characteristics of participants: Gender: 5 male; 3 female None of the participants had prior colonoscopy training or simulator experience. Each participant completed 2 months of esophagoggastroduodenoscopy training prior to their colonoscopy training. Inclusion: First year GI Fellows Exclusion: not stated Details of patients for live assessment: not stated

Appendix C.1 Study design tables – AccuTouch® flexible sigmoidoscopy simulator training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Sedlack *et al.* 2004 **Location** Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota. USA	Objective: To determine whether computer based endoscopy training results in measurable benefit to trainees' endoscopic performance in flexible sigmoidoscopy on the basis of staff evaluations of procedural skill and patient evaluations of discomfort. All participants: Participants randomised. Intervention: • **Simulator training**: independent 3 hour AccuTouch® flexible sigmoidoscopy simulator based training curriculum under the supervision of a gastroenterology Fellow. Curriculum entailed a brief multimedia tutorial followed by 8 – 10 hands on simulated scenarios. • **No training**: no simulator training. Assessment: **The following aspects of the study were referred to as training, but as these procedures were used for the assessment component, they have been called assessment.** • After simulation training, the trained group conducted staff supervised patient based endoscopies for 3 additional afternoons (3 hours each day). • The no simulator training group had staff conducted supervised patient based colonoscopies for 4 afternoons. During the final day of training, supervising staff completed a standardised evaluation for all participants that graded performance. This examination evaluated 8 performance parameters on a Likert Scale. Participants completed a self evaluation on the final day of training grading their own performance using the same parameters and scoring system. Every patient who underwent flexible sigmoidoscopy (by resident or staff) was asked to complete a questionnaire grading the discomfort experienced during endsocopy by using a 1 – 10 Likert Scale. Device: AccuTouch® flexible sigmoidoscopy simulator (Immersion Medical, Gaithersburg, MD, Simulator engine 1.1.1). This computer based colonoscopy simulator consists of a specialised endoscope that is inserted into a computer based module and provides 6 sigmoidoscopy scenarios of varying complexity. Simulator utilises internal mechanics that provide tactile feedback as well as a computer-generated voice to simulate patient discomfort.	**Randomised controlled trial** Method of randomisation: not stated Allocation concealment: not stated Level of evidence: II Blinding: evaluating staff not blinded to participant training status Intention to treat: not stated Power calculation: not stated Follow-up (years): not stated Lost to assessment: not stated Study period: July 2001 and June 2002 Outcome measures: Supervisors and participants assessed: • Ability to perform procedure independently • Identify pathology • Identify landmarks • Respond to patient discomfort • Insert scope safely • Adequately visualise mucosa on withdrawal • Routinely reach 40 cm • Perform biopsies	Sample size: n = 38 • **Simulator training**: n = 19 • **No simulator training**: n = 19 Baseline characteristics of participants: none had prior endoscopy training Inclusion: 2nd year internal medicine residents Exclusion: not stated Details of patients for live assessment: not stated

Appendix C.1 Study design tables – 4-day VR laparoscopic cholecystectomy training course *vs.* no training course

Author	Intervention	Study Design	Study population
Schijven *et al.* 2005 **Location** IJsselland Hospital, 2900 AR Capelle, IJssel. The Netherlands	Objective: To investigate operating room performance of surgical residents after participating in the Eindhoven virtual reality laparoscopic cholecystectomy training course. Participants: All participants attended a Basic Surgical Skills Course before participating in this study. Intervention: • **Training**: trainees participated in a 4-day virtual reality laparoscopic cholecystectomy training course. Videos, oral presentations, interactive sessions incorporated. Repetitive training using open Xitact LS500 laparoscopic simulator platform, VR simulation (MIST-VR) and procedural laparoscopic cholecystectomy simulation including the clip-and-cut, navigation, and dissection modules (Xitact) were featured. On days 2 and 3, participants in course attended the operating theatre in conjunction with their VR sessions to act as assistant surgeon or camera assistant on laparoscopic cholecystectomy by expert surgeon. • **Control**: no training course. Time to assessment: on day 4. Assessment: On day 4, all participants performed a full laparoscopic cholecystectomy under supervision. Procedure was videotaped. Only the clip and cut part of the procedure, the clipping and cutting of the cystic artery and cystic duct, was assessed. The procedure was assessed starting from the moment the laparoscopic clip applier was introduced and ending at the moment the laparoscopic scissors were removed from the operative field. Video fragments from both the trained and control groups were independently evaluated by 2 laparoscopic engaged surgeons from different academic training hospitals. Participant's video fragments were mixed in random order before being copied to the reviewer's videotape. Devices: Xitact LS500 laparoscopy simulator platform (Xitact SA, Morges, Switzerland). MIST-VR	**Non-randomised comparative study** Level of evidence: III-2 Lost to assessment: technical recording failures • **Training**: n = 2 • **Control**: n = 2 Blinding: video tape assessors blinded to training status and video tape segments in random order Study period: April 2003 - March 2004 Outcome measures: structured questionnaire using a 5-point Likert rating scale used for assessment. Parameters measured: • Fluency • Carefulness • Sumscore: focused on clip and cut part of procedure. Performance judged by integration of psychomotor skills, procedural knowledge of anatomy and decision making. Constructed primarily by metrics of Xitact's clip and cut simulation. • Judgment • Time to complete clip and cut procedure.	Sample size: n= 24 • **Training**: n = 12 • **Control**: n = 12 Baseline characteristics of participants: • Mean age in both groups: 31 years • Mean years of training for both groups: 1.8 years • All participants right handed **Training** / **Control** Sex ratio (M:F): 8 :4 / 10 : 2 No. previous laparoscopic cholecystectomies, mean (range): 0.3 (0 – 1) / 1.8 (0 – 3)* No. participants training to be general surgeons: 6 / 8.4 No. participants who have conduced similar training course†: 0 / 3 *p = 0.008 in favour of control group † either animal or non animal training courses Inclusion: trainee surgeons and novices who had performed less than 4 laparoscopic cholecystectomies. Exclusion: not stated Details of patients for live assessment: patients selected for assessment were American Society of Anesthesiology class 1 with a history of uncomplicated cholelithiasis and no previous abdominal complications or surgery.

Appendix C.1 Study design tables – Endoscopic sinus surgical simulator training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Edmond 2002 **Location** Madigan Army Medical Centre, the Departments of Surgery and Otolaryngology – Head and Neck Surgery, University of Washington, Seattle and the Department of Surgery, Tripler Army Medical Centre Honolulu, Hawaii. USA	Objective: To evaluate an endoscopic sinus surgical simulator (ESS) as a training device and to introduce a methodology to assess its impact on actual operating room performance. Intervention: • **Training**: novice, intermediate and advanced models of ESS simulator. Performance of each trial recorded by the system. Hazards created in all 3 models. Subtask scores were calculated as accuracy (navigation, injections and dissection accuracy), optimal time/completed time. To normalise across models, optimal times were derived from performance of 2 expert surgeons on each subtask for each model. • **Control**: no simulator training. Time to assessment: not stated Assessment: 4 subjects performed a routine endoscopic sinus surgery procedure on a patient. Assessors evaluated participants on a 10 point scale. Procedures were videotaped. A blinded panel of 4 experienced sinus surgeons rated 5 videotapes (4 first time surgeries and 1 operation performed by an experienced staff surgeon). The panel rated each videotape on the same 10 point scale. Device: ESS simulator incorporates 3 computer systems linked by an Ethernet interface. A Silicon Graphics Inc (SGI) Onyx 2 computer serves as the simulation host platform. In contains the virtual patient model and is responsible for the simulation of the endoscopic image, the surgical interface, and the user interface. The Onyx is configured with two R10000 CPUs, IR graphics hardware, and a soundboard. The second computer, a 333 MHz Pentium PC is dedicated to control the electromechanical hardware. A 3rd computer allows voice recognition and provides virtual instruction while training.	**Non-randomised comparative study** Level of evidence: III-2 Lost to assessment: not stated Study period: not stated Blinding: videotape assessors blinded to training status Outcome measures: performance criteria measured on 10 point scale: • Navigation • Injection • Uncinectomy • Anterior ethmoidectomy • Maxillary antrostomy • Orientation of video image • Image-to-task alignment • Proper depth of image for task • Tool manipulation • Tool selection • Tool-tool dexterity • Tissue respect • Surgical confidence • Case difficulty	Sample size: 10 • **Training**: n = 2 • **Control**: n = 2 Baseline characteristics of participants: • 2 subjects had trained on the simulator. • 2 had no prior simulator experience • None had prior operating room sinus surgical experience Inclusion: first-year (junior) ear-nose-throat residents Exclusion: not stated Details of patients for live assessment: not stated

Appendix C.1 Study design tables – Gastro-Sim® flexible sigmoidoscopy simulator training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Tuggy 1998 **Location** Swedish Family Medicine Residency, Seattle. USA	Objective: To determine whether a virtual reality flexible sigmoidoscopy simulator improves the hand-eye coordination and various performance parameters in a live patient. All participants: Participants randomised. Intervention: • **Simulator training**: 5 hours training on the Gastro-Sim® flexible sigmoidoscopy simulator. Participants not given any training or guidance on the skills required for sigmoidoscopy other than what was encountered during the simulation. • **Control**: no simulator training or preparation before performing first live patient examinations. Time to assessment: not stated Assessment: Examinations performed on 2 live patient volunteers. Before each set of sigmoidoscopies, each patient received a brief examination by a supervising physician to ensure the colon was adequately prepared. The air was then removed from the colon before the study participants performed their examinations. Each matched pair of residents then performed examinations sequentially on the same patient to reduce the risk of encountering different colon structures. A sigmoidoscopist monitored the examinations. The sigmoidoscopist inserted or retracted the sigmoidoscope. The examinations were videotaped and used for assessment. The patients completed a pain scale, rated the perceived confidence of the examiner, and evaluated the duration of the examination. All participants completed a survey on the effect of the simulator on their perception of their hand-eye skills. After the first set of live patient examinations, each of the 5 residents in the control group was then allowed access to the simulator and completed 5 hours of training. The experimental group continued to train on the simulator for up to 5 additional hours. Once this training was completed, the matched resident pairs again performed the procedure on the volunteer patients. During this second phase of the trial, the paired residents examined the alternate patient. Device: Gastro-Sim® flexible sigmoidoscopy simulator (Interact Medical).	**Randomised controlled trial** Method of randomisation: not stated Allocation concealment: not stated Level of evidence: II Blinding: patient blinded to experience and training status of participant Intention to treat: not stated Power calculation: not stated Lost to assessment: not stated Study period: not stated Outcome measures: • Time to reach 30 cm, 40 cm, and maximal insertion • Total time of examination • Total time in red-out • Quality of visualisation of the colon walls, • Estimated percentage of the colon visualised • Hand eye skills assessed by the amount of directional errors made during procedure	Sample size: n = 10 • **Simulator training**: n = 5 • **Control**: n = 5 Baseline characteristics of participants: not stated Inclusion: residents with no experience in flexible sigmoidoscopy Exclusion: not stated Details of patients for live assessment: 2 healthy men aged 25 – 35 who were compensated for their participation in the study.

Appendix C.1 Study design tables – Simbionix GI Mentor™ sigmoidoscopy simulator training *vs.* no simulator training

Author	Intervention	Study Design	Study population
Cohen *et al.* 2006b **Location** New York NY, Charleston SC, Dallas Texas. USA	Objective: To determine whether a 10 hour structured training program using the GI Mentor™ simulator provides an objective benefit to novice gastroenterology Fellows learning to perform colonoscopy. All participants: All participants completed a questionnaire outlining demographics including age, gender, and year of graduation; gastroenterology training program; and the number of flexible sigmoidoscopies performed before the GI Fellowship. All participants attended general lectures on colonoscopy as part of a didactic endoscopy course. Intevention: Participants randomised: • **Simulator training**: participants were given a supervised orientation to the Simbionix GI Mentor™ simulator. Each hour of training followed a standard protocol of activities. In all, 10 different cases were used during the simulator training. Each participant completed five 2-hour private sessions on the simulator over 8 weeks. Participants kept a log of attempted simulated procedures performed during each 2-hour session and a log all actual sigmoidoscopies and gastroscopies performed during the study. During the second year of the study, the simulator was able to record performance variables. Participants did not perform any real colonoscopies until they completed all 10 hours on the simulator. Participants filled out a questionnaire at the end of the training regarding usefulness of training. • **No simulator training (referred to as traditional training in study, but does not involve a pre-assessment training element)**: supervised colonoscopies starting approximately 8 weeks into the Fellowship (so that both groups commenced the actual colonoscopies at the same time). Time to assessment: 8 weeks after the beginning of simulator training. Assessment: For each procedure, participants were responsible for getting their proctor to fill out an evaluation form. For those cases where the participant could not reach the caecum without assistance, the proctor was asked to indicate whether or not the procedure was difficult for the proctor to complete. Proctors also asked to rate patient discomfort. Forms were collected each	**Randomised controlled trail** Method of randomisation: a random number table Allocation concealment: not stated Level of evidence: II Blinding: proctors filling out individual evaluation forms were blinded to resident's training status. Participant's names did not appear on study evaluation forms (only a code number). Only principal investigator had access to the key to code numbers. Intention to treat: not stated Power calculation: Kaplan-Meier curves were generated to determine the number of blocks of 20 cases needed for each group to reach a median of 90% objective and subjective competency. Lost to assessment: n = 4 (protocol violations during training phase) Study period: not stated Outcome measures: • Ability to reach the transverse colon and caecum without assistance • Ability to correctly recognise and identify abnormalities • Overall subjective rating of competency on a 5 point scale	Sample size: n = 49 • **Simulator training**: n = 22 • **No training**: n = 23 Baseline characteristics of participants: **Precolonoscopy experience***: Mean gastroscopies: Simulator training 67, No training 80, Total 147; Mean flexible sigmoidoscopies: 4, 5, 9; Total: 71, 85, 156 *No significant differences between the 2 groups Inclusion: the participant's training director had to agree to adhere to the protocol and to delay any first year performance of colonoscopy for the first 8 weeks of the Fellowship Exclusion: previous formal training in colonoscopy (> 10 cases) and an inability to comply with the training schedule Details of patients for live assessment: not stated

	month until the participant reached 200 procedures or until the study period was over. Device: Simbionix GI Mentor™ (Simbionix USA Corporation, Gathersburg, Md)		

Appendix C.2 Study design tables – Virtual reality sigmoidoscopy simulator training *vs.* patient-based training

<table>
<tr><th>Author</th><th>Intervention</th><th>Study Design</th><th>Study population</th></tr>
<tr>
<td>Gerson and van Dam 2003

Location
Division of Gastroenterology and Hepatology, Stanford University School of Medicine, Stanford, California.

USA</td>
<td>Objective: To compare patient-based teaching of sigmoidoscopy with that provided by a virtual reality endoscopy simulator.

Participants:
All participants completed a questionnaire which inquired about age, gender, amount of training in gastroenterology, and prior experience with computers.

Intervention:
• Simulator training: participant's permitted unlimited time and access to the VR sigmoidoscopy simulator. Residents were instructed to review all didactic modules and complete all 6 cases prior to the test cases. Residents were allowed unlimited time on the simulator during the 2 week period prior to the scheduled test cases. Participants were not allowed to view live cases as part of their training prior to using simulator. Performance on the simulator was not observed and coaching not provided.
• Patient-based training: Participants were required to perform 10 sigmoidoscopic examinations with an attending gastroenterologist. Participants were trained with a video endoscope. The attending physician was instructed to teach the residents using his or her own teaching preferences and techniques. Residents were expected to learn how to advance the colonoscope independently by the end of the 10 sessions. The teaching sessions were scheduled to occur during a consecutive 2 week period.

Time to assessment: two weeks after commencement of training.

Assessment:
All residents performed 5 test examinations under supervision and evaluation of an attending gastroenterologist.
Residents were instructed to complete the examination to the splenic flexure and to notify the attending physician when the flexure was identified. Residents were also required to notify if any pathology was encountered during the examination. Residents were expected to perform retroflexion at the completion of the sigmoidoscopy. If the participant encountered difficulties, the attending physician was allowed to take over until the resident could continue.
Attending physician used a standardised form for data collection.
Patients were interviewed by a medical assistant at the completion of the sigmoidoscopy using a standard questionnaire.</td>
<td>Randomised controlled trial

Method of randomisation: sequentially allocated by investigator

Allocation concealment: none

Level of evidence: II

Intention to treat: not stated

Power calculation: Using a comparison of the means of 2 independent samples, the sample size required to detect a magnitude difference between arms of 30%, a power of 90% and an alpha of 0.05, was calculated to be 30 examinations in each arm of the study (assuming 5% of examinations would not be completed, and another 5% due to patient intolerance, 33 patients would need to be recruited per arm).

Blinding: neither the investigators nor the participants were blinded to the group assignment. Participating patients were blinded to residents' training status. Assessing physicians not blinded.

Lost to assessment: none

Study period: not stated

Outcome measures: standardised form for data collection:
• Examination duration and extent
• Splenic flexure recognition
• Ability to recognise pathology
• Completion of retroflexion</td>
<td>Sample size: n = 16
• Simulator training: n = 9
• Patient-based training: n = 7

Baseline characteristics of participants:
<table>
<tr><th></th><th>Simulation training</th><th>Patient-based training</th></tr>
<tr><td>Internal medicine residents (n)</td><td>9</td><td>7</td></tr>
<tr><td>Mean resident age ± SEM (yrs)</td><td>28 ± 0.8</td><td>29.4 ± 1.1</td></tr>
</table>
10 residents had extensive experience with the use of computers, and the remaining 6 described themselves as 'somewhat experienced'.

Inclusion: internal medicine residents

Exclusion: participants with prior experience with flexible sigmoidoscopy, observation of sigmoidoscopy as part of a rotation, or prior use of an endoscopic simulator.

Details of patients for live assessment:
Asymptomatic patients referred for routine colorectal cancer screening via flexible sigmoidoscopy were asked to participate.
<table>
<tr><th></th><th>Simulation training</th><th>Patient-based training</th></tr>
<tr><td>Patient mean age ± SEM (yrs)</td><td>54 ± 1.4</td><td>56 ± 1.5</td></tr>
<tr><td>Patient gender (M/F, % male)</td><td>10/34 (29%)</td><td>16/32 (50%)</td></tr>
</table>
</td>
</tr>
</table>

	Device: Virtual reality sigmoidoscopy simulator (Immersion Medical Inc., Gathersburg, Maryland, USA).		

SEM Standard error of the mean

APPENDIX D – RESULTS TABLES

Appendix D – Results tables

Simulation training vs. no simulation training

MIST-VR simulator training *vs.* no simulator training for laparoscopic cholecystectomy (Grantcharov *et al.* 2004)

Table D.1. Performance of laparoscopic cholecystectomy between VR trained and non-VR trained participants

Grantcharov *et al.* 2004	**Simulator training (n = 8)**		**No training (n = 8)**		***P*-value**
Performance parameter	**Operation 1 median (range)**	**Operation 2 median (range)**	**Operation 1 median (range)**	**Operation 2 median (range)**	
Time taken to perform operation (min)	62 (46 – 80)	53 (46 – 62)	55 (43 – 58)	57 (43 – 74)	0.021
Error scores	6.8 (4 – 9)	3 (3 – 6)	6 (3 – 9)	5.8 (4 – 9)	0.003
Economy of movement score	5.8 (4.5 – 6)	3.3 (2 – 6)	6 (6 – 8)	6 (4.5 – 9)	0.003

Participants who received VR training performed laparoscopic cholecystectomy significantly faster than those in the control group (p = 0.021, t test).
The VR trained group showed significantly greater improvement in their error (p = 0.003) and economy of movement (p = 0.003) scores.

SCMIS GEM video trainer *vs.* no simulator training for laparoscopic cholecystectomy (Scott *et al.* 2000)

Table D.2. Baseline video-trainer scores: time (seconds) for task completion

Scott *et al.* 2000	**No training (n = 13)**	**Simulator training (n = 9)**	***P*-value***
Checker-board	144 (122 - 152)	146 (126 - 187)	0.616
Bean drop	58 (48 - 69)	53 (45 - 66)	0.471
Running string	62 (46 - 74)	74 (67 - 84)	0.096
Block move	50 (38 - 55)	48 (39 - 58)	0.815
Suture foam	58 (49 - 90)	56 (42 - 94)	0.695

Values are medians with 25th – 75th percentiles in parentheses.
*Trained versus control groups, one-tailed Wilcoxon rank-sum test.

Table D.3. Baseline laparoscopic cholecystectomy global assessment scores

Scott *et al.* 2000	**No training (n = 13)**	**Simulator training (n = 9)**	***P*-value***
Respect for tissue	3.0 (2.4 - 3.2)	3.0 (2.8 - 3.2)	0.445
Time and motion	2.5 (2.0 - 3.0)	3.0 (2.0 - 3.0)	0.781
Instrument handling	3.0 (2.5 - 3.0)	3.0 (2.0 - 3.2)	0.782
Knowledge of instruments	3.0 (2.5 - 3.5)	3.0 (2.3 - 3.3)	0.433
Flow of operation	3.0 (2.0 - 3.2)	2.0 (2.0 - 3.2)	0.356
Use of assistants	2.0 (1.5 - 3.0)	2.0 (1.5 - 2.8)	0.785
Knowledge of specific procedure	3.0 (2.5 - 3.6)	3.0 (2.5 - 3.2)	0.632
Overall performance	3.0 (2.3 - 3.0)	2.5 (2.0 - 3.0)	0.581

Values are medians with 25th – 75th percentiles in parentheses. 1 = worst possible score, 5 = best possible score.
* Trained versus control groups, two-tailed Wilcoxon rank-sum test.

Table D.4. Adjusted improvement* in video-trainer scores: time (seconds) for completion of task

Scott *et al.* 2000	**No training (n = 13)**	**Simulator training (n = 9)**	***P*-value†**
Checker-board	31 (-1 – 37)	63 (23 - 75)	0.014
Bean drop	14 (10 - 18)	24 (18 - 26)	0.002
Running string	3 (-13 - 16)	26 (21 - 38)	0.001
Block move	9 (-2 - 14)	22 (11 - 25)	0.015
Suture foam	26 (18 - 38)	48 (44 - 50)	0.001

Values are medians with 25th – 75th percentiles in parentheses.
*Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
†Trained versus control groups, one-tailed Wilcoxon rank-sum test.

Table D.5. Adjusted improvement* in laparoscopic cholecystectomy global assessment scores

Scott *et al.* 2000	**No training (n = 13)**	**Simulator training (n = 9)**	***P*-value†**
Respect for tissue	0.1 (-0.6 - 0.5)	0.3 (0.3 - 0.5)	0.035
Time and motion	-0.3 (-0.5 - 0.6)	0.3 (0.1 - 0.8)	0.075
Instrument handling	0.3 (-0.4 - 0.3)	0.6 (0.4 - 0.8)	0.005
Knowledge of instruments	0.4 (0.0 - 1.0)	0.6 (0.5 - 1.5)	0.058
Flow of operation	0.4 (-0.5 - 1.2)	1.0 (0.6 - 1.2)	0.090
Use of assistants	0.7 (-0.4 - 1.0)	1.0 (0.8 - 1.6)	0.035
Knowledge of specific procedure	0.4 (0.0 - 1.1)	1.0 (0.4 - 1.3)	0.100
Overall performance	0.2 (-0.5 - 0.6)	0.7 (0.6 - 1.0)	0.007

Values are medians with 25th – 75th percentiles in parentheses.
*Improvement defined as post-training minus baseline scores, calculated individually for each participant, adjusted by linear analysis of covariance for differences in baseline scores.
†Trained versus control groups, one-tailed Wilcoxon rank-sum test.

All 9 residents in the training group completed 10 practice sessions lasting 30 minutes.
On average, residents practised 138 video trainer tasks (range 94 – 171). Each of the 5 tasks was practised 28 times (range 19 – 34).
Laparoscopic experience in the operating room during the study interval was comparable for both groups (p = 0.612).
The trained group had significantly larger median time reductions for all five video trainer tasks compared with the control group.
On global assessment, the trained group had significantly larger median increases in four of eight performance criteria, compared with the control group.

When initially asked if they felt comfortable with their laparoscopic skills, 3 of 13 control residents and 5 of 9 trained residents replied ‘yes’.
On the completion questionnaire, 6 of 13 control residents and 8 of 9 trained residents felt comfortable with their laparoscopic skills at the end of the rotation.
Of those who were not comfortable with their laparoscopic skills at baseline, 3 of 10 in the control group were comfortable at the end of the rotation, in contrast to 3 of 4 in the trained group (p = 0.175). After training, 9 of 9 residents improved their hand-eye-hand coordination and 8 of 9 felt that the training had improved their skills in the operating room.

Table D.6. Participant self-assessment improvement after simulator training

Scott *et al.* 2000	**No simulator training (n = 13)**		**Simulator training (n = 9)**	
Self assessment	**baseline**	**End of rotation**	**baseline**	**End of rotation**
No. participants comfortable with skills	3/13	6/13	5/9	8/9
Skills improvement			After training, 9/9 felt that VT had improved their video-eye-hand coordination and 8/9 felt that VT had improved their operating skills.	

SCMIS GEM video trainer *vs.* no video-trainer training for laparoscopic cholecystectomy (Scott *et al.* 1999)

Table D.7. Video trainer skills: time (seconds) for task completion after training

Scott *et al.* 1999	No simulator training (mean ± SD) (n = 13)	Simulator training (mean ± SD) (n = 9)	*P*-value
Checker-board	125 ± 31	94 ± 27	< 0.05
Bean drop	45 ± 11	34 ± 7	< 0.05
Running string	63 ± 23	46 ± 14	< 0.05
Block move	42 ± 15	28 ± 7	< 0.05
Suture foam	42 ± 15	20 ± 6	< 0.05

Values obtained at repeat examination after training.

Table D.8. Laparoscopic cholecystectomy global assessment score after training

Scott *et al.* 1999	**No simulator training (mean ± SD) (n = 13)**	**Simulator training (mean ± SD) (n = 9)**	***P*-value**
Respect for tissue	2.9 ± 0.6	3.3 ± 0.3	< 0.05
Time and motion	2.6 ± 0.7	3.0 ± 0.4	< 0.05
Instrument handling	2.9 ± 0.6	3.4 ± 0.3	< 0.05
Knowledge of instruments	3.5 ± 0.9	3.9 ± 0.6	NS
Flow of operation	3.0 ± 0.8	3.5 ± 0.4	NS
Use of assistants	2.8 ± 0.7	3.3 ± 0.6	< 0.05
Knowledge of specific procedure	3.5 ± 0.9	3.8 ± 0.6	NS
Overall performance	2.8 ± 0.7	3.4 ± 0.3	< 0.05

Values obtained at repeat examination after training. 1 indicates worst possible score; 5, best possible score.

MIST-VR computer simulation *vs.* no simulator training for laparoscopic cholecystectomy (Seymour *et al.* 2002)

Table D.9. Outcomes of the study

Seymour *et al.* 2002	**Standard training (n=8)**	**MIST-VR (n=8)**	***P*-value**
Operative performance	final	final	
Errors			
Lack of progress	10	1	p < 0.008
Gall bladder injury	5	1	p < 0.039
Liver injury	1	1	
Intraperitoneal, incorrect plane of dissection	3	0	
Burn non-target tissue	10	2	p < 0.039
Tearing tissue	0	0	
Instrument out of view	9	4	
Attending takeover	6	0	
Non-contact cautery injury	0	0	
Mean total errors	7.38	1.19	p < 0.006
Time			
Duration of dissection (min)	20.5	14.5	NS

NS not significant

There were no significant differences in any of the initial battery of assessment tests noted between the groups.
All residents in both groups successfully completed the dissection of the gallbladder from the liver bed.
All residents in the VR group successfully achieved the required criterion levels of performance in 3 – 8 training sessions.
The duration of the dissection for the VR trained group was 29% less than the standard training group (p = NS).

4-day VR laparoscopic cholecystectomy training course *vs.* no training course for laparoscopic cholecystectomy (Schijven *et al.* 2005)

Participants of both experimental and control groups did not differ in demographic parameters, except for the number of laparoscopic cholecystectomies in favour of the control group ($p = 0.008$).
Both observers judged that there was a significant improvement for 'judgment' (Observer 1, $p = 0.004$ and Observer 2, $p = 0.013$).
Observer 1 found a significant improvement for 'fluency' ($p = 0.0037$).

AccuTouch® colonoscopy simulator training *vs.* no simulator training for colonoscopy (Ahlberg *et al.* 2005)

Table D.10. Results of the study

Ahlberg *et al.* 2005	**Simulator training (n = 6)**	**No simulator training (n = 6)**	***P*-value**
Training time, median hours (range)	20 (15 – 25)	0	
No colonoscopies assessed (n = 120) by gender of patient	32 M; 28 F	29 M; 29 F (gender not indicated for 2 patients)	
Success* rate of colonoscopy (%)	52	19	0.0011
Time to reach caecum in successful cases median minutes (IQR)†	30 (17 - 38)	40 (25 – 45)	0.008
Patient discomfort median (IQR)	4 (2.5 – 6)	5 (4 – 7)	0.02‡

* A successful colonoscopy was defined as intubation of the caecum within the given time limits.
† The time difference was also significant (p = 0.037) in unsuccessful cases and the simulator trained group managed to reach further in a shorter time compared with controls.
‡ In addition, male patients reported less pain (p = 0.001) compared with female patients.

Median total training time per resident was 20 hours (range 15 – 25). Training sessions lasted 1 – 2 hours on each occasion, and could be repeated several times per day and continued over at least 4 days. After completion of the training, participants had to do their colonoscopies within 1 week.
Trainees in the control group started after studying the theoretical material.
A successful colonoscopy was defined as intubation of the caecum within given time limits.

The success rate was 52% in the simulator trained group compared with 19% in controls, which was a significant difference (p = 0.0011). See Table.
There was a significant difference (p = 0.008) in procedure time in favour of the simulator trained group. The time to reach the caecum in successful cases was a median of 30 minutes (interquartile range (IQR) 17 – 38 minutes) in the trained group compared with 40 minutes (IQR 25 – 45 minutes) in controls.

There was a significant learning curve (p = 0.039) for all participants during the study, and the likelihood of a participant reaching the caecum was 2.57 times higher for procedures 6 – 10 compared with procedures 1 – 5 in both test groups.
Patient gender significantly (p = 0.016) affected success rate. It was 3 times more likely that a colonoscopy would be successful in a male patient than a female patient.
The number of previously performed gastroscopies influenced success rate (p = 0.006) equally in both groups. It was 3.76 times more likely that a participant with a previous experience exceeding 50 gastroscopies would succeed with colonoscopy.

The post graduate year of the participant did not significantly influence the level of success rate.
Significantly less patient discomfort was reported in the simulator trained group (median 4, IQR 2.5 – 6) compared with controls (median 5, IQR 4 – 7) (p = 0.02). In addition it was noted that male patients reported less pain (p = 0.001) compared with female patients.

No major complications were reported in patients used in the assessment process.

AccuTouch® colonoscopy simulator training *vs.* no simulator training for colonoscopy (Sedlack and Kolars 2004)

Table D.11. Median Performance Scores (25–75% IQR) Rendered by Supervising Faculty are Shown for Each Parameter

Sedlack and Kolars 2004	Colon 1–15	Colon 16–30	Colon 31–45	Colon 46–60
N = Number of faculty evaluations recorded in each group				
Simulator	60	58	58	60
Regular	60	59	60	60
Time to reach maximum insertion (minutes)				
Simulator	23 (19–30)	21 (19–28)	21 (18–28)	17 (12–25)
Regular	25 (20–30)	22 (15–30)	20 (15–30)	20 (15–27)
P-value	0.155	0.947	0.321	0.090
Depth of unassisted insertion (1 = rectum, 6 = terminal ileum)				
Simulator	4.0 (3.0–5.0)	5.0 (4.0–5.0)	4.5 (4.0–5.0)	5.0 (4.0–5.0)
Regular	3.0 (2.0–4.0)	4.0 (3.0–5.0)	4.5 (3.0–5.0)	5.0 (4.0–5.0)
P-value	0.003	0.006	0.905	0.085
Percentage completed independently by the fellow (%)				
Simulator	38% (27–50)	59% (46–71)	50% (37–63)	72% (59–84)
Regular	20% (9–31)	34% (21–46)	50% (37–63)	58% (46–71)
P-value	0.027	0.007	1.000	0.128
Identifies landmarks (1 = strongly disagree, 4 = neutral)				
7 = strongly agree)				
Simulator	6.0 (6.0–7.0)	6.0 (6.0–7.0)	6.0 (6.0–7.0)	7.0 (6.0–7.0)
Regular	6.0 (5.0–7.0)	6.0 (5.0–7.0)	6.0 (6.0–7.0)	7.0 (5.5–7.0)
P-value	0.041	0.044	0.166	0.439
Inserts in a safe manner (1–7)				
Simulator	7.0 (6.0–7.0)	6.0 (6.0–7.0)	6.0 (6.0–7.0)	7.0 (6.0–7.0)
Regular	6.0 (6.0–7.0)	6.0 (6.0–7.0)	6.0 (6.0–7.0)	7.0 (6.0–7.0)
P-value	0.008	0.274	0.256	0.559
Adequately visualises mucosa withdrawal (1–7)				
Simulator	6.0 (6.0–7.0)	6.0 (6.0–7.0)	6.0 (6.0–7.0)	7.0 (6.0–7.0)
Regular	6.0 (5.0–7.0)	6.0 (5.0–7.0)	6.0 (5.5–7.0)	6.0 (6.0–7.0)
P-value	0.009	0.396	0.230	0.518
Responds appropriately to patient discomfort (1–7)				
Simulator	6.5 (6.0–7.0)	6.0 (5.8–7.0)	6.0 (5.0–7.0)	7.0 (6.0–7.0)
Regular	6.0 (5.3–7.0)	6.0 (6.0–7.0)	6.0 (6.0–7.0)	7.0 (5.0–7.0)
P-value	0.019	0.560	0.137	0.771

"Percentage of endoscopies completed independently" shown are averages (95% CI).
Columns represent the chronological order of procedures being compared in blocks of 15.

Table D.12. Median scores (25–75% IQR) of patient discomfort

Sedlack and Kolars 2004	Colon 1–15	Colon 16–30	Colon 31–45	Colon 46–60
N = Number of patient evaluations recorded in each group				
Simulator	60	58	58	60
Regular	60	59	60	60
Patient reported pain score: (1 = no pain, 10 = worst pain of life)				
Simulator	2.0 (1.0–4.0)	2.0 (1.0–4.0)	2.0 (1.0–4.0)	1.5 (1.0–4.0)
Regular	4.0 (1.5–5.0)	2.0 (1.0–4.0)	2.5 (1.0–4.3)	2.0 (1.0–3.0)
P-value	0.019	0.343	0.531	0.731

There appeared to be a significant reduction in pain scores during the first 15 colonoscopies performed by each simulator trained fellow. The absence of sedation data, however, makes these results difficult to interpret.

The 4 simulator trained participants completed an average of 21 simulated cases (range 19 – 26) prior to their patient based experience.
The 4 simulator trained participants performed an average of 117 (range 62 – 195) patient based colonoscopies over an average of 7 weeks (range 4 – 8) of training.
The 4 traditionally trained participants completed an average of 108 (range 76 – 157) patient based colonoscopies over an average of 7 weeks (range 4 – 8) of training.

An analysis of procedures broken down in chronological groups of 15 demonstrates that simulator trained fellows scored significantly better in all parameters during the first 15 colonoscopies ($p < 0.05$) performed with the exception of 'time to reach maximum insertion'. The simulation trained fellows inserted the endoscope further with a median depth score of 4.0 vs 3.0 ($p = 0.003$) and reached the caecum independently in 38% of procedures vs 20% ($p = 0.027$) during this initial training period. Simulator trained fellows continued to have a significantly greater 'depth of insertion', higher percentage of 'independently completed procedures' and greater ability to 'identify landmarks' throughout the first 30 colonoscopies ($p < 0.05$). Beyond 30 colonoscopies, none of the parameters showed a statistical difference between groups.
Patient surveys also demonstrated a lower median discomfort score during the first 15 colonoscopies performed by the simulator trained fellows; with colonoscopies by trained fellows achieving a median discomfort score of 2.0 (range 1 – 8) vs 4.0 (range 1 – 7) for traditionally trained fellows ($p = 0.019$). The degree of sedation during each procedure was not assessed.

Patient surveys also demonstrated a lower median discomfort score during the first 15 colonoscopies performed by the simulator-trained fellows; with colonoscopies by CBCS-trained fellows achieving a median discomfort score of 2.0 (range 1–8) *vs.* 4.0 (range 1–7) for traditionally trained fellows ($p= 0.019$). The degree of sedation during each procedure was not assessed, therefore, no comparisons of the adequacy of sedation could be made between the 2 groups nor could a correlation be made of sedation levels and individual patient discomfort scores.

During the 2 half-days of the CBCS training, faculty without an accompanying fellow were able to complete an average of 8 colonoscopies (range 7–9) per half-day while simulator fellows worked with the CBCS. Faculty of the traditional fellow group completed an average of 3.5 procedures (range 2–4) per half-day during the same initial training interval. This allowed an average of 9 additional colonoscopies to be performed by the staff endoscopists for each fellow trained with this CBCS curriculum. There was no significant difference between supervising staff volumes once CBCS fellows began patient-based colonoscopy.

AccuTouch® flexible sigmoidoscopy simulator training *vs.* no simulator training for sigmoidoscopy (Sedlack *et al.* 2004)

Table D.13. Staff and resident evaluations

Sedlack *et al.* 2004	**Staff evaluations**			**Resident self evaluations**		
Parameter	**Simulator training (n = 19)**	**Traditional training (n = 19)**	***P*-value**	**Simulator training (n = 19)**	**Traditional training (n = 19)**	***P*-value**
Competent to perform endoscopy independently	8 (7 – 9)	8 (7 – 9)	0.893	8 (5 – 9)	8 (4 – 10)	0.791
Identifies pathology	6 (5 – 8)	5 (5 – 7)	0.624	3 (2 – 4)	3 (2 – 5)	0.522
Identifies landmarks	5 (4 – 7)	5 (4 – 5)	0.715	3 (2 – 4)	4 (2 – 4)	0.362
Responds to patient discomfort	3 (3 – 5)	3 (1 - 6)	0.278	2 (1 – 3)	2 (2 – 3)	0.394
Inserts scope safely	3 (3 – 6)	3 (2 – 5)	0.336	2 (1 – 3)	2 (1 – 3)	0.792
Adequately visualises the mucosa on withdrawal	7 (3 – 8)	5 (4 – 7)	0.330	3 (2 – 5)	3 (2 – 4)	0.880
Routinely reaches 40 cm	3 (2 – 8)	3 (1 – 6)	0.459	7 (6 – 9)	8 (6 – 9)	0.285
Competent to perform biopsies	7 (5 - 8)	8 (5 - 9)	0.659	4.5 (2 – 7.8)	6 (3 - 10)	0.218

Median resident performance scores (25^{th} – 75^{th} % IQR) at the end of training are shown based on a 10 point Likert Scale (1, strongly agree; 5, neutral; 10 strongly disagree). None of the staff evaluated parameters of residents' skill reached statistical significance.

The 19 simulator trained residents performed an average of 9 simulated procedures (range 6 – 11) followed by an average of 11 patient based procedures each (range 7 – 19) for a total of 212 patient based procedures. Of the simulator trained residents, 150 (71%) patients agreed to complete surveys

The residents that had no simulator training performed an average of 12 procedures each (range (7 – 17) for a total of 230 patient based procedures.
Of the non-simulator trained residents, 175 (76%) patients completed surveys.

During the same period, staff endoscopists completed 780 procedures with 585 (75%) patient surveys returned.

No adverse events were reported with any of the procedures during this study.

Analysis of patient surveys demonstrated that median patient reported discomfort scores were significantly lower for simulator trained residents than for traditionally trained resident, 3 (IQR, 2 – 5) vs 4 (IQR, 2 – 6) ($p < 0.01$).

An increase in staff productivity was shown as a result of simulator training. During the simulator training interval, non-teaching staff performed an average of 7 procedures per half day while operating independently compared with an average of less than 3 procedures during direct patient-based teaching. The 3 hour simulator training session allowed an average of 4 additional procedures to be performed by staff for each participant trained via the simulator.

Simbionix GI Mentor™ simulator training *vs.* no simulator training for colonoscopy (Cohen *et al.* 2006b)

Table D.14. Comparison between simulator and no-simulator group in objective and subjective, competence and patient discomfort

Cohen *et al*. 2006b	Objective* competence			Subjective† competence			Patient discomfort‡		
	Simulator (n = 23) Mean	No training (n = 22) Mean	*P*-value (t-test)	Simulator (n = 23) Mean	No training (n = 22) Mean	*P*-value (t-test)	Simulator (n = 23) Mean	No training (n = 22) Mean	*P*-value (t-test)
Session 1	50.4	40.9	0.06	47.6	36.6	0.08	25.7	31.4	0.42
Session 2	64.5	52.0	< 0.0001	68.6	57.4	0.004	23.2	19.1	0.14
Session 3	74.0	62.0	< 0.0001	76.3	68.4	0.005	16.7	19.5	0.22
Session 4	76.7	64.4	< 0.0001	78.0	75.4	0.32	16.0	18.2	0.39
Session 5	76.8	70.2	0.03	81.3	79.4	0.28	16.7	16.5	0.94
Session 6	77.8	77.6	0.91	82.0	82.3	0.88	13.4	13.9	0.85
Session 7	80.8	80.5	0.89	86.1	84.1	0.32	11.9	11.3	0.74
Session 8	89.5	83.7	0.01	88.8	86.4	0.11	10.5	10.4	0.99
Session 9	87.8	85.2	0.02	88.9	86.8	0.32	10.7	11.8	0.55
Session 10	92.7	90.9	0.04	90.8	90.5	0.82	8.9	9.2	0.81

* Objective competency is the ability to reach the transverse colon and the caecum without assistance, and the ability to correctly recognise and identify abnormalities.
† Subjective competency is on a 5-point scale; 1 (totally unskilled) to 5 (competent and expedient).
‡ Patient discomfort rated on a scale of 1 (very comfortable) to 5 (severe pain to the patient).

There were no significant differences in colonoscopy experience between the two groups.

The respondents rated the overall satisfaction with the simulator training as moderately useful to useful, with a mean score of 3.5 (range, 1 [no use] to 5 [very useful]).

Gastro-Sim® flexible sigmoidoscopy simulator training *vs.* no simulator training for sigmoidoscopy (Tuggy 1998)

Table D.15. Performance comparisons and quality 360-degree visualisation technique between control and experimental groups

Tuggy 1998	No training (n = 5)	5 hrs simulator training (n = 5)	*P*-value	No training (n = 5)	10 hrs simulator training (n = 5)	*P*-value
Time to 30 cm (sec)	357	286	0.52	357	119	0.03
Time to 40 cm (sec)	518	341	0.27	518	211	0.03
Total examination time (sec)	654	530	0.31	654	323	0.01
Directional errors (n)	8.6	2.8	0.01	8.6	1.6	< 0.01
Time in red-out (sec)	70	27	0.16	70	14	0.07
Percentage of colon visualised	45	55	0.60	45	79	0.02
Quality of viewing 360° *	2.4	1.3	0.05	2.4	1.4	0.03

* Based on a rating scale of 1 – organised, 2 – adequate, 3 – haphazard.

Table D.16. Resident survey responses to using virtual reality sigmoidoscopy simulator

Tuggy 1998	Strongly agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Strongly disagree (%)
Simulator resembles live sigmoidoscopy	0	89	11	0	0
Graphics resembled actual colon	22	67	11	0	0
Tactile feedback was similar to colon	0	44	44	11	0
Learned hand-eye skills on simulator	45	55	0	0	0
Learned more with more practice	22	55	11	11	0
Gained confidence for live patient examination	55	45	0	0	0
Likely to perform in practice in simulator available in training	55	33	11	0	0
Tutorial component was helpful	0	33	33	11	22
Enhanced features on simulator would make me skilled in flexible sigmoidoscopy	44	44	11	0	0

Training on the virtual reality simulator produced substantial improvements in examination times and hand-eye skill measures.
After 6 to 10 hours of training on the simulator, the experimental group achieved significantly faster insertion times to 30 cm ($p = 0.03$), 40 cm ($p = 0.03$), and a shorter mean length of examination ($p = 0.01$). There was also significant improvement of hand-eye skill measures of the experimental group in directional errors ($p < 0.01$), percentage of colon visualised ($p = 0.02$), and viewing quality of examination when compared with the control group's initial performance on live patients.

Resident survey findings after the study confirmed the trainee's perception of the benefit of the simulator training.
There were no observed significant differences in procedure-assessed patient discomfort between the 2 groups at any time during the training.

Procedicus VIST™ simulator training *vs.* no simulator training for catheter based intervention for occlusive vascular disease (Chaer *et al.* 2006)

Table D.17. Mean checklist scores on individual measures of performance for simulator and non simulator trained residents

Chaer *et al.* 2006	**Simulator training (n = 10)**	**No training (n = 10)**	***P*-value**	**Simulator training (n = 10)**	**No training (n = 10)**	***P*-value**
	Intervention 1			Intervention 2		
Advance femoral wire	2.4	1.4	NS	2.6	2.0	NS
Advance wire atraumatically	2.6	1.8	0.05	2.8	2.0	0.03
Constantly visualise wire tip	2.9	1.4	0.005	3.1	1.9	0.001
Mount and advance catheter wire	2.9	2.0	0.01	3.1	2.9	NS
Position imaging catheter	2.1	1.2	0.04	2.4	1.8	0.03
Knowledge of anatomy	2.4	1.3	NS	2.5	1.6	0.04
Walk catheter back over wire	2.9	2.0	NS	3.4	2.7	0.05
Advance balloon over wire	3.1	2.2	0.006	3.4	2.6	0.02
Centre balloon over stenosis	3.0	2.0	0.009	2.9	1.9	0.003
Balloon inflation	3.0	2.0	0.003	3.0	1.7	0.003
Balloon pressure	2.6	1.3	0.003	2.3	1.1	0.002
Walk balloon back over wire	3.0	2.2	NS	3.3	2.0	0.006
Image after PTA	2.5	1.8	NS	2.6	1.9	NS
Advance stent over wire	3.0	2.3	NS	3.4	2.2	0.01
Centre stent over stenosis	2.6	2.1	NS	2.9	1.8	0.01
Accurately deploy stent	2.6	1.4	NS	3.0	1.7	0.01
Walk stent shaft out over wire	3.0	2.4	NS	3.3	2.0	0.006
Completion angiogram	2.2	1.9	NS	2.7	1.7	0.04
Overall assessment	50 ± 6	33 ± 9	0.0015	53 ± 6	36 ± 7	0.0006

NS not significant
PTA percutaneous transluminal angioplasty

Table D.18. Mean endovascular global rating scale scores on individual measures of performance for simulator and non simulator trained residents

Chaer *et al.* 2006	**Simulator training (n = 10)**	**No training (n = 10)**	***P*-value**	**Simulator training (n = 10)**	**No training (n = 10)**	***P*-value**
	Intervention 1			Intervention 2		
Time and motion	2.3	1.4	NS	2.6	1.7	0.01
Wire and catheter handling	2.8	1.6	0.002	3.0	1.9	0.009
Awareness of wire position	2.6	1.7	0.005	3.0	1.8	0.01
Wire stability	2.6	1.9	NS	3.0	2.1	0.04
Fluoroscopy usage	1.5	1.1	NS	2.0	1.1	0.003
Precision of wire/catheter technique	2.8	1.7	0.03	2.8	1.7	0.005
Flow of operation	2.4	1.4	NS	2.8	1.2	0.002
Knowledge of procedure	2.0	1.4	NS	2.4	1.1	0.005
Quality of final product	3.0	3.0	0.03	3.3	3.2	NS
Ability to complete the case	2.4	1.4	0.03	2.6	1.4	0.01
Need for verbal prompts	2.3	1.0	0.03	2.4	1.4	0.01
Attending takeover	2.6	1.4	0.003	2.9	1.7	0.006
Overall assessment	30 ± 7	19 ± 5	0.0052	33 ± 6	21 ± 6	0.0015

NS not significant

There were no pre-study differences between the training and control groups. Both groups were comparable in terms of their age and gender as well as previous experiences that might be relevant to a resident's ability to assimilate catheter techniques. Performance on the visuospatial test was not different between the 2 groups (p = not significant).
All residents in the training group were able to complete the simulator training session in less than 2 hours (mean of 90 ± 21 minutes).
There were no intraoperative complications that developed as a result of resident participation in the study. There were no differences in peri-operative morbidity or mortality between patients treated by the 2 groups.

The trained group scored higher than controls during the first (50 ± 6 vs 33 ± 9, p = 0.0015) and second (53 ± 6 vs 36 ± 7, p = 0.0006) endovascular intervention. For the first intervention, training led to a numerical enhancement of each of the individual measures of performance (not all statistically significant). For the second intervention, the trained group was significantly better in all but 3 variables. There were 2 procedural steps where a significant difference was not found with either the first or the second intervention (advance femoral wire and image after percutaneous angioplasty). The first is not well taught by a simulator. The second measures the ability of a resident to remember to perform a completion angiogram after angioplasty. Resident performance did not improve from the first to the second intervention.

All residents in the simulator-trained group were able to complete the mentored simulator training session in less than 2 hours (mean of 90 ± 21 minutes).

A more subjective evaluation of resident performance was also conducted by attending surgeons. Trained participants scored higher overall on the global rating scale of endovascular performance for the first (30 ± 7 vs 19 ± 5, p = 0.052) and second (33 ± 6 vs 21 ± 6, p = 0.0015) intervention. For both procedures, training led to numerical enhancement in all the individual measures of performance. For the first intervention, 4 of these were not statistically significant and 1 was not statistically significant for the second intervention. Resident performance did not improve from the first to the second intervention

No peri-operative complications developed because of the study.

TEP hernia repair simulator curriculum training *vs.* no TEP hernia repair curriculum training (Hamilton *et al.* 2001)

Table D.19. Composite score and individual categories comprising composite score of global assessment tool

Hamilton *et al.* 2001	**Simulator training (n = 10)**		**No training (n =11)**	
Individual assessment parameter	Baseline mean ±SD	Final mean ±SD	Baseline mean ±SD	Final mean ±SD
Respect for tissue	3.1 ± 1.1	3.5 ± 0.9	2.7 ± 0.9	3.1 ± 1.0
Time and motion	2.7 ± 1.1‡	3.5 ± 0.7*§	1.8 ± 0.6	2.4 ± 0.9*
Instrument handling	2.9 ± 1.0	3.7 ± 0.7†‖	2.1 ± 0.7	3.1 ± 0.9*
Knowledge of instrument	3.0 ± 1.3	3.7 ± 0.8‖	2.9 ± 1.5	2.9 ± 0.4
Flow of operation	2.8 ± 1.0	3.7 ± 0.7*§	2.1 ± 0.8	2.6 ± 1.1
Use of assistants	2.5 ± 1.3	3.5 ± 1.1§	1.6 ± 0.7	2.1 ± 1.1
Knowledge of procedures	2.5 ± 0.9	3.8 ± 0.9*§	2.0 ± 0.8	2.6 ± 0.9*
Overall performance	2.4 ± 0.8	3.6 ± 0.7*§	1.9 ± 0.7	2.4 ± 0.9*
composite score (%)	44.6 ± 24.6	65.7 ± 17.5*	29.6 ± 15.7	41.0 ± 23.5*

Between group comparisons made using Wilcox rank sum test. Within group comparisons made using Wilcoxon signed rank test.

***** Significant difference between pre-training and post training scores at $p < 0.05$.

† Significant difference between pre-training and post training scores at $p = 0.05$.

‡ Significant difference between trained and untrained groups prior to training at $p < 0.05$.

§ Significant difference between trained and untrained groups after training at $p < 0.05$.

‖ Significant difference between trained and untrained groups after to training at $p = 0.05$.

Questionnaire data responses were similar for the control and trained group: at baseline (p = NS). After training, 10 of 10 residents in the trained group felt their ability to perform a laparoscopic TEP hernia repair improved over the study period compared with 5 of 11 in the control group (100% *vs.* 45.5%, P= 0.01). Additionally, all the residents in the trained group (100%) reported that their understanding of the operation improved over the month compared with 5 of 11 residents (45.5%) in the control group (p = 0.01). Compared with controls, residents in the trained group also expressed an increased willingness to offer laparoscopic TEP hernia repairs to patients with concurrent, unilateral hernias (p = 0.02).

Endoscopic sinus surgery simulator training *vs.* no simulator training for catheter for endoscopic sinus surgery (Edmond 2002)

Table D.20. Mean Rating for Residents with and without prior simulation training across videotape rating criteria for their first operating room procedure.

Edmond 2002	No training (n = 2) Mean ± SD	Simulator training (n = 2) Mean ± SD	*P*-value
Navigation	7. 7 ± 0	6.5 ± 1.2	0.40
Injection	2.8 ± 1.6	6.8 ± 1.6	0.14
Uncinectomy	2.83 ± 1.2	6.7 ± 0.5	0.15
Anterior ethmoidectomy	3.3 ± 0.9	7 ± 0.9	0.0.6
Maxillary antrostomy	3. 7 ± 0.5	6.3 ± 0.9	0.17
Orientation of image	4.8 ± 0.2	6.8 ± 1.6	0.34
Image-task alignment	4.8 ± 0.2	7 ± 1.9	0.35
Proper depth of image	5. 3 ± 0.5	6.8 ± 1.2	0.34
Tool manipulation	3 ± 0	7 ± 0.9	0.11
Tool selection	3.8 ± 0.2	6.8 ± 1.6	0.24
Tool-tool dexterity	3.5 ± 1.2	6.5 ± 2.1	0.33
Tissue respect	2.8 ± 0.2	5.5 ± 3.1	0.44
Surgical confidence	2.8 ± 0.2	6.5 ± 0.7	0.09
Case difficulty	5 ± 0	7.2 ± 0.7	0.14
Overall mean rating	4.0 ± 6.7E-02	6.7 ± 1.1	0.19

NS not significant

Four participants: 2 had extensive simulator experience. 2 had none. None had prior operating room sinus surgery experience.
While the 2 simulation-trained residents were rated consistently better than the other 2 residents across all measures, these differences approached statistical significance for only two items (most likely as a result of the small number of subjects): anterior ethmoidectomy ($p = 0.06$; $p < 0.05$) and surgical confidence ($p = 0.09$; $p < 0.05$).

Simulator experience could be a strong predictor of first-time operating room performance as determined by rating videos. This result approaches but does not achieve significance ($r = 0.911$, $p < 0.1$).

Appendix D – Results tables

Simulation training vs. patient-based training

Sigmoidoscopy simulator training *vs.* patient-based training for sigmoidoscopy (Gerson and van Dam 2003)

Table D.21. Results of the study

Gerson and van Dam 2003	Simulator training (n = 9)	Traditional training (n = 7)	*P*-value
Test cases (n ± SEM)	38 ± 0.5	4.6 ± 0.5	> 0.05
Time per case (min ± SEM)	24 ± 1.0	24 ± 1.1	> 0.05
Resident completed independently	10/34 (29%)	23/32 (72%)	0.001
Required assistance	24/34 (71%)	9/32 (28%)	0.001
Flexure recognition	10/12 (83%)	22/30 (73%)	> 0.05
Retroflexion completed	19/34 (56%)	27/32 (84%)	0.02
Mean score (± SEM)	2.9 ± 0.2	3.8 ± 0.2	< 0.001
Mean score, only in female patients	2.9 ± 0.2	3.9 ± 0.2	< 0.001

Table D.22. Results of patients' assessments of satisfaction

Gerson and van Dam 2003	Simulator training		Traditional training		*P*-value
	Agree (%)*	Disagree (%)†	Agree (%)*	Disagree (%)†	
General satisfaction					
Satisfied with care	93	3	90	10	> 0.05
Recommend to friends	70	17	65	19	> 0.05
Willing to have another	77	13	81	6	> 0.05
Pain/discomfort					
Had a lot of pain	53	37	42	45	> 0.05
More comfortable than expected	33	47	42	29	> 0.05
Caused great discomfort	43	53	31	61	> 0.05
Technical competence					
Physician too rough	10	83	13	77	> 0.05
Confident that examination was performed properly	73	13	84	16	> 0.05

*Includes patients who agreed or strongly agreed with the question

†Includes patients who disagreed or strongly disagreed with the questions. The remainder of the patients were uncertain or neutral.

An analysis of the results according to the patient's gender was carried out. There was no significant difference between the mean score when the examination was performed in female patiens (3.3 ± 0.8) compared to male patients (3.4 ± 1.8; p = 0.7). The mean scores for each group did not significantly differ from the overall scores when examinations performed in female patiens were compared.

No significant differences in patient discomfort was found between the two arms of the study.

Residents in the simulator trained group were asked to rate the performance of the simulator and to provide feedback about its performance after they performed the test cases. On average, residents spent over 2 hours on the simulator performing the practise cases (excluding the time spent watching the educational videos, which is not assessed). The average distance obtained with the simulator was 42 cm, demonstrating that most residents were unable to complete the simulated cases successfully to the level of the splenic fixture.

In order to determine whether residents' performance improved on the simulator over time, data regarding procedure time, insertion length and retroflexion ability was calculated for each resident during the first 3 cases performed on the simulator compared with the last 3 cases on the simulator. None of the parameters significantly improved over time. No association was found between time spent on the simulator and performance during the test cases.

At the end of the test cases, the simulator trained residents were asked to rate their training experience. In general, the residents trained on the simulator felt that the educational videos and teaching cases were useful. However, all the residents stated that the test cases were more difficult than the simulated examinations. Specific critiques of the simulator included the fact that the rectum appeared insufflated upon insertion of the endoscope, and that the negotiation of the rectosigmoid area was not challenging enough during the simulated cases. In addition, the sumilator was unable to simulate a realistic sensation of resistance to passage of the sigmoidoscope. Both attending physicians noted considerable anxiety when the residents in the simulator trained group attempted the test cases, although this impression was not quantified.

No adverse events occurred during the test examinations.

APPENDIX E – DESCRIPTIVE SUMMARY OF SIMULATORS AND PATIENT-BASED TRAINING

Appendix E – Summary of simulators and patient-based training

Study	Simulator overview
Grantcharov *et al.* 2004 Schijven *et al.* 2005 Seymour *et al.* 2002	The **Minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR)** system is a low fidelity virtual reality simulator with combined metrics systems to provide feedback to novice learners during practice. The system trains 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).
Scott *et al.* 2000	The **Southwestern Centre for Minimally Invasive Surgery Guided Endoscopic Module (SCMIS GEM)** video-trainer is composed of frames supporting traditional laparoscopic video monitors, light sources, and camera systems (Hamilton *et al.* 2002). The frame forms a box inside which pre-manufactured tasks are performed. Speed is measured by the trainee and is the measure of performance (Hamilton *et al.* 2002). Video trainers 'shape' performance by training skills of progressive difficulty (Valentine and Rege 2004; Gallagher *et al.* 2005).
Ahlberg *et al.* 2005 Sedlack and Kolars 2004 Sedlack *et al.* 2004	The **AccuTouch® simulator system** is a full-procedure simulator and includes training for flexible sigmoidoscopy, colonoscopy and bronchoscopy (bronchoscopy was not part of this review). 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).
Chaer *et al.* 2006	The **Procedicus VIST™** simulator is a multimedia device designed to simulate endovascular techniques in a variety of clinical settings. It consists of a three-dimensional representation of the human arterial system, coupled to a haptic module that uses a force feedback system that provides tactile sensory information when the user inserts and manipulates standard angiographic catheters and guide wires (Dayal *et al.* 2004). Separate devices are attached that simulate the injection of contrast dye, performance of angioplasty, deployment of stents, and performance of fluoroscopy with digital subtraction angiography (Dayal *et al.* 2004). It is one of the most sophisticated VR simulators in the world (Gallagher *et al.* 2005).
Cohen *et al.* 2006b	The **Simbionix GI Mentor™** is a real-time interactive computer simulator that replicates both diagnostic and therapeutic procedures (Barr-Meir 2000). It includes a life-sized plastic head and torso with apertures for upper and lower endoscopy (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. 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 give a the beginning of training, which are gradually faced out until the trainee can perform the task without support (Gallagher *et al.* 2005).
Tuggy 1998	The **Gastro-Sim® flexible sigmoidoscopy** simulator is no longer commercially available (Gerson 2006) and widespread information does not exist on this simulator (United States Patent 4907973).
Hamilton *et al.* 2001	The **TEP hernia repair simulator** 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 (Valentine and Rege 2004).
Edmond 2002	The **endoscopic sinus surgery simulator** 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 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 give a the beginning of training, which are gradually faced out until the trainee can perform the task without support (Gallagher *et al.* 2005).
Schijven *et al.* 2005	The **Xitact LS500** is a hybrid simulator that combines a physical object with computer software simulation providing visual images and tactile feedback. It is a modular VR training platform that trains a variety of laparoscopic skills (Schijven and Jakimowicz 2003).

Appendix E – Summary of simulators and patient-based training continued

Study	Patient-based training overview
Gerson and van Dam 2003	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.

Simulator device name used for simulator training in Scott *et al.* (1999) and Gerson and van Dam (2003) not specified.

APPENDIX F – SUMMARY OF CRITICAL APPRAISAL

Appendix F – 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 (Level II)											
Grantcharov *et al.* 2004	×	✓	✓	×	×	✓	✓	✓	✓	×	✓
Scott *et al.* 2000	✓	×	✓	×	✓	✓	✓	✓	✓	×	✓
Scott *et al.* 1999	×	×	✓	×	×	×	×	✓	✓	×	×
Seymour *et al.* 2002	×	×	✓	×	×	×	×	×	✓	×	✓
Ahlberg *et al.* 2005	×	✓	✓	×	×	×	×	×	✓	×	✓
Sedlack and Kolars 2004	×	×	✓×	×	×	×	×	×	✓	×	✓
Sedlack *et al.* 2004	×	×	✓×	×	×	×	✓	×	✓	×	✓
Cohen *et al.* 2006b	✓	×	✓	×	✓	✓	×	✓	✓	✓	✓
Tuggy 1998	×	×	×	×	×	×	×	×	✓	×	×
Chaer *et al.* 2006	×	✓	✓	×	×	×	×	✓	✓	×	×
Hamilton *et al.* 2001	×	×	✓	×	×	×	✓	✓	✓	×	×
Gerson and van Dam 2003	✓	×	✓×	×	✓	✓ (none)	×	×	✓	✓	✓
Non-randomised comparative studies (Level III-2)											
Edmond 2002	•	•	✓	•	•	×	×	×	✓	×	✓
Schijven *et al.* 2005	•	•	✓	•	•	✓	✓	×	✓	×	✓

× not reported

✓ reported

✓× reported in the study, but not done eg it was reported that assessors were not blinded

• not applicable