Cochrane Library

Cochrane Database of Systematic Reviews

Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy (Review)

Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM

Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM.
Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy.
Cochrane Database of Systematic Reviews 2018, Issue 8. Art. No.: CD008237.
DOI: 10.1002/14651858.CD008237.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy

Rishad Khan[1], Joanne Plahouras[2], Bradley C Johnston[3], Michael A Scaffidi[4], Samir C Grover[4], Catharine M Walsh[5]

[1]Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Canada. [2]University of Toronto, Toronto, Canada. [3]Department of Community Health and Epidemiology, Dalhousie University, Halifax, Canada. [4]Department of Medicine, Division of Gastroenterology, St. Michael's Hospital, University of Toronto, Toronto, Canada. [5]Division of Gastroenterology, Hepatology, and Nutrition, The Hospital for Sick Children, Toronto, Canada

Contact: Catharine M Walsh, Division of Gastroenterology, Hepatology, and Nutrition, The Hospital for Sick Children, 555 University Ave, Toronto, ON, M5G 1X8, Canada. catharine.walsh@mail.utoronto.ca.

Editorial group: Cochrane Colorectal Cancer Group.
Publication status and date: New search for studies and content updated (conclusions changed), published in Issue 8, 2018.

Citation: Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. *Cochrane Database of Systematic Reviews* 2018, Issue 8. Art. No.: CD008237. DOI: 10.1002/14651858.CD008237.pub3.

ABSTRACT

Background

Endoscopy has traditionally been taught with novices practicing on real patients under the supervision of experienced endoscopists. Recently, the growing awareness of the need for patient safety has brought simulation training to the forefront. Simulation training can provide trainees with the chance to practice their skills in a learner-centred, risk-free environment. It is important to ensure that skills gained through simulation positively transfer to the clinical environment. This updated review was performed to evaluate the effectiveness of virtual reality (VR) simulation training in gastrointestinal endoscopy.

Objectives

To determine whether virtual reality simulation training can supplement and/or replace early conventional endoscopy training (apprenticeship model) in diagnostic oesophagogastroduodenoscopy, colonoscopy, and/or sigmoidoscopy for health professions trainees with limited or no prior endoscopic experience.

Search methods

We searched the following health professions, educational, and computer databases until 12 July 2017: the Cochrane Central Register of Controlled Trials, Ovid MEDLINE, Ovid Embase, Scopus, Web of Science, BIOSIS Previews, CINAHL, AMED, ERIC, Education Full Text, CBCA Education, ACM Digital Library, IEEE Xplore, Abstracts in New Technology and Engineering, Computer and Information Systems Abstracts, and ProQuest Dissertations and Theses Global. We also searched the grey literature until November 2017.

Selection criteria

We included randomised and quasi-randomised clinical trials comparing VR endoscopy simulation training versus any other method of endoscopy training with outcomes measured on humans in the clinical setting, including conventional patient-based training, training using another form of endoscopy simulation, or no training. We also included trials comparing two different methods of VR training.

Data collection and analysis

Two review authors independently assessed the eligibility and methodological quality of trials, and extracted data on the trial characteristics and outcomes. We pooled data for meta-analysis where participant groups were similar, studies assessed the same intervention and comparator, and had similar definitions of outcome measures. We calculated risk ratio for dichotomous outcomes with

95% confidence intervals (CI). We calculated mean difference (MD) and standardised mean difference (SMD) with 95% CI for continuous outcomes when studies reported the same or different outcome measures, respectively. We used GRADE to rate the quality of the evidence.

Main results

We included 18 trials (421 participants; 3817 endoscopic procedures). We judged three trials as at low risk of bias. Ten trials compared VR training with no training, five trials with conventional endoscopy training, one trial with another form of endoscopy simulation training, and two trials compared two different methods of VR training. Due to substantial clinical and methodological heterogeneity across our four comparisons, we did not perform a meta-analysis for several outcomes. We rated the quality of evidence as moderate, low, or very low due to risk of bias, imprecision, and heterogeneity.

Virtual reality endoscopy simulation training versus no training: There was insufficient evidence to determine the effect on composite score of competency (MD 3.10, 95% CI -0.16 to 6.36; 1 trial, 24 procedures; low-quality evidence). Composite score of competency was based on 5-point Likert scales assessing seven domains: atraumatic technique, colonoscope advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure, and overall performance. Scoring range was from 7 to 35, a higher score representing a higher level of competence. Virtual reality training compared to no training likely provides participants with some benefit, as measured by independent procedure completion (RR 1.62, 95% CI 1.15 to 2.26; 6 trials, 815 procedures; moderate-quality evidence). We evaluated overall rating of performance (MD 0.45, 95% CI 0.15 to 0.75; 1 trial, 18 procedures), visualisation of mucosa (MD 0.60, 95% CI 0.20 to 1.00; 1 trial, 55 procedures), performance time (MD -0.20 minutes, 95% CI -0.71 to 0.30; 2 trials, 29 procedures), and patient discomfort (SMD -0.16, 95% CI -0.68 to 0.35; 2 trials, 145 procedures), all with very low-quality evidence. No trials reported procedure-related complications or critical flaws (e.g. bleeding, luminal perforation) (3 trials, 550 procedures; moderate-quality evidence).

Virtual reality endoscopy simulation training versus conventional patient-based training: One trial reported composite score of competency but did not provide sufficient data for quantitative analysis. Virtual reality training compared to conventional patient-based training resulted in fewer independent procedure completions (RR 0.45, 95% CI 0.27 to 0.74; 2 trials, 174 procedures; low-quality evidence). We evaluated performance time (SMD 0.12, 95% CI -0.55 to 0.80; 2 trials, 34 procedures), overall rating of performance (MD -0.90, 95% CI -4.40 to 2.60; 1 trial, 16 procedures), and visualisation of mucosa (MD 0.0, 95% CI -6.02 to 6.02; 1 trial, 18 procedures), all with very low-quality evidence. Virtual reality training in combination with conventional training appears to be advantageous over VR training alone. No trials reported any procedure-related complications or critical flaws (3 trials, 72 procedures; very low-quality evidence).

Virtual reality endoscopy simulation training versus another form of endoscopy simulation: Based on one study, there were no differences between groups with respect to composite score of competency, performance time, and visualisation of mucosa. Virtual reality training in combination with another form of endoscopy simulation training did not appear to confer any benefit compared to VR training alone.

Two methods of virtual reality training: Based on one study, a structured VR simulation-based training curriculum compared to self regulated learning on a VR simulator appears to provide benefit with respect to a composite score evaluating competency. Based on another study, a progressive-learning curriculum that sequentially increases task difficulty provides benefit with respect to a composite score of competency over the structured VR training curriculum.

Authors' conclusions

VR simulation-based training can be used to supplement early conventional endoscopy training for health professions trainees with limited or no prior endoscopic experience. However, we found insufficient evidence to advise for or against the use of VR simulation-based training as a replacement for early conventional endoscopy training. The quality of the current evidence was low due to inadequate randomisation, allocation concealment, and/or blinding of outcome assessment in several trials. Further trials are needed that are at low risk of bias, utilise outcome measures with strong evidence of validity and reliability, and examine the optimal nature and duration of training.

PLAIN LANGUAGE SUMMARY

Virtual reality simulators for training in gastrointestinal endoscopy

Review question

Can virtual reality simulation training supplement and/or replace early patient-based training in gastrointestinal endoscopy?

Background

Traditionally, trainees have learned to perform gastrointestinal endoscopy (a tubular camera used to visualise structures within the bowel or stomach) in the clinical setting under the supervision of a trained endoscopist. Virtual reality computer simulators use computer technology to create a three-dimensional image or environment that can be interacted with in a seemingly real or physical way. This technique is becoming popular as a way of providing trainees with an opportunity to practice skills in a risk-free environment. However, simulation-based training can be expensive. It is therefore important to ensure that skills gained through simulation-based training translate to the clinical environment.

Search date

The evidence is current to 12 July 2017.

Study characteristics

We included 18 trials with 421 participants and 3817 endoscopy procedures. Ten trials compared virtual reality training with no training; five compared virtual reality training with patient-based endoscopy training; one compared virtual reality training with another form of endoscopy simulation training; and two compared two different methods of virtual reality training. Ten trials studied colonoscopy, three studied sigmoidoscopy, and five studied oesophagogastroduodenoscopy. Participants included medical trainees with limited or no endoscopy training from gastroenterology, medicine, family medicine, or general surgery, along with nurses.

Key results

Compared to no training, virtual reality training appears to provide trainees with an advantage as measured by the ability to complete procedures independently, overall rating of performance, and visualisation of the colon or oesophagus. We found no conclusive evidence that virtual reality training, as compared with traditional patient-based training or another method of endoscopy simulation training, provided benefit, although data were limited. Existing virtual reality simulation curricula can be improved by applying educational theory such as a progressive learning strategy, whereby trainees complete increasingly difficult cases. The results of this review have shown that virtual reality endoscopy training can be used to supplement early traditional endoscopy training for trainees with limited or no endoscopic experience.

Quality of the evidence

Overall, the quality of the evidence was poor based on potential bias due to poor methodological reporting in trials and imprecision due to few participants and endoscopic procedures. Future studies must adhere to quality standards, such as proper randomisation, along with using valid metrics to measure endoscopic performance. Researchers should also compare the effectiveness of different simulation curricula that are based on educational theories.

Summary of findings for the main comparison. Virtual reality endoscopy simulation training versus no training

Virtual reality endoscopy simulation training versus no training for health professions trainees in gastrointestinal endoscopy

Patient or population: health professions trainees in gastrointestinal endoscopy
Setting: 4 single-centre studies from Canada, USA, and South Korea, and 2 multicentre European studies
Intervention: virtual reality endoscopy simulation training
Comparison: no training

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of procedures** (studies)	Quality of the evidence (GRADE)	Comments
	Risk with no training	Risk with virtual reality endoscopy simulation training				
Composite score of competency	-	The mean composite score of competency was 3.10 MD higher (0.16 lower to 6.36 higher).	-	24 (1 trial)	⊕⊕⊖⊖ LOW [1, 2]	The composite score of competency was based on 5-point Likert scales assessing 7 domains: atraumatic technique, colonoscope advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure, and overall performance. The range of scores was from 7 to 35, with a higher score representing a higher level of competence.
Independent procedure completion	Study population		RR 1.62 (1.15 to 2.26)	815 (6 trials)[5]	⊕⊕⊕⊖ MODERATE [1]	Independent procedure completion refers to the number of endoscopic procedures that trainees completed without assistance from a supervisor. A higher number of independent procedure completions represents a more positive outcome.
	465 per 1000	754 per 1000 (535 to 1000)				
Performance time (minutes)	-	The mean performance time was 0.20 MD lower (0.71 lower to 0.30 higher).	-	29 (2 trials)[6]	⊕⊖⊖⊖ VERY LOW [2, 3]	7 trials reported performance time, but only 2 provided sufficient data for quantitative analysis. Performance time refers to the time required to complete a given endoscopic procedure. A shorter performance time indicates a positive outcome.
Complication or critical flaw occurrence	See comment	See comment	-	550 (3 trials)	⊕⊕⊕⊖ MODERATE [1]	All trials reporting this outcome reported no incidence of procedure-related complications or critical flaws in either group. Complications or critical flaws are procedure-related adverse events such as bleeding, luminal perforation, and infection.

Patient discomfort	-	The mean patient discomfort was 0.16 SMD lower (0.68 lower to 0.35 higher).	-	145 (2 trials)[6]	⊕⊖⊖⊖ VERY LOW [2,3,4]	7 trials reported patient discomfort, but only 2 provided sufficient data for quantitative analysis.
Overall global rating of performance or competency	-	The mean overall global rating was 0.45 MD higher (0.15 higher to 0.75 higher).	-	18 (1 trial)[7]	⊕⊖⊖⊖ VERY LOW [2, 3]	4 trials reported overall global ratings, but only 1 with 2 data sets (from 2 types of assessor) provided sufficient data for quantitative analysis. Overall global ratings represent a single rating of endoscopic performance as rated by an external assessor. The range of scores was from 1 to 5, with a higher score representing a better endoscopic performance.
Visualisation of mucosa	-	The mean visualisation of mucosa was 0.60 MD higher (0.20 higher to 1.00 higher).	-	55 (1 trial)[7]	⊕⊖⊖⊖ VERY LOW [2, 3]	3 trials reported visualisation of mucosa, but only 1 provided sufficient data for quantitative analysis. Higher mucosal visualisation represents a more successful endoscopic procedure.

*The basis for the **assumed risk** is provided in footnotes. **The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

**The unit of analysis is an individual endoscopic procedure, as opposed to a study participant. For example, the outcome 'independent procedure completion' should be interpreted as virtual reality training leading to a 1.62x increased likelihood of completion of an endoscopic procedure.

CI: confidence interval; **MD:** mean difference; **RR:** risk ratio; **SMD:** standardised mean difference

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate quality: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low quality: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low quality: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

[1]Downgraded one level for serious risk of bias (due to unclear or inadequate methods of randomisation, allocation sequence generation, and/or blinding of outcome assessment).
[2]Downgraded one level for serious imprecision (due to few participants and endoscopic procedures under study).
[3]Downgraded two levels for very serious risk of bias (due to inadequate methods of randomisation, allocation sequence generation, and/or blinding of outcome assessment).
[4]Downgraded due to unexplained heterogeneity.
[5]Analysis based on randomised trials and two quasi-randomised trials.
[6]Analysis based on two quasi-randomised trials.
[7]Analysis based on one quasi-randomised trial.

Summary of findings 2. Virtual reality endoscopy simulation training versus conventional patient-based training

Virtual reality endoscopy simulation training versus conventional patient-based training for health professions trainees in gastrointestinal endoscopy

Patient or population: health professions trainees in gastrointestinal endoscopy
Setting: 1 single-centre study from the USA and 2 multicentre European studies
Intervention: virtual reality endoscopy simulation training
Comparison: conventional patient-based training

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of procedures** (studies)	Quality of the evidence (GRADE)	Comments
	Risk with conventional patient-based training	Risk with virtual reality endoscopy simulation training				
Composite score of competency	-	See comment	-	(0 studies)	-	1 trial reported composite score of competency but did not provide sufficient data for quantitative analysis.
Independent procedure completion	Study population 337 per 1000	152 per 1000 (91 to 250)	RR 0.45 (0.27 to 0.74)	174 (2 trials)	⊕⊕⊖⊖ LOW [1]	
Performance time (minutes)	-	The mean performance time was 0.12 SMD higher (0.55 lower to 0.80 higher).	-	34 (2 trials)	⊕⊖⊖⊖ VERY LOW [1,2]	4 trials reported performance time, but only 2 provided sufficient data for quantitative analysis. Performance time refers to the time required to complete a given endoscopic procedure. A shorter performance time indicates a positive outcome.
Complication or critical flaw occurrence	See comment	See comment	-	72 (3 trials)	⊕⊖⊖⊖ VERY LOW [1,2]	All trials reporting this outcome reported no incidence of procedure-related complications or critical flaws in either group. Complications or critical flaws are procedure-related adverse events such as bleeding, luminal perforation, and infection.
Patient discomfort	-	See comment	-	(0 studies)	-	2 trials reported patient discomfort, but neither provided sufficient data for quantitative analysis.
Overall global rating of performance or competency	-	The mean overall global rating was 0.90 MD lower	-	16 (1 trial)	⊕⊖⊖⊖ VERY LOW [1,2]	3 trials reported overall global ratings, but only 1 provided sufficient data for quantitative analysis. Overall global ratings represent a single rating of endoscopic performance as rated by an external asses-

		(4.40 lower to 2.60 higher).				sor. The range of scores was from 1 to 5, with a higher score representing a better endoscopic performance.
Visualisation of mucosa	-	The mean visualisation of mucosa was 0 MD (6.02 lower to 6.02 higher).	-	18 (1 trial)	⊕⊖⊖⊖ VERY LOW [1, 2]	2 trials reported visualisation of mucosa, but only 1 provided sufficient data for quantitative analysis. Higher mucosal visualisation represents a more successful endoscopic procedure.

*The basis for the **assumed risk** is provided in footnotes. **The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

**The unit of analysis is an individual endoscopic procedure, as opposed to a study participant. For example, the outcome 'independent procedure completion' should be interpreted as virtual reality training leading to a 1.62x increased likelihood of completion of an endoscopic procedure.

CI: confidence interval; **MD:** mean difference; **RR:** risk ratio; **SMD:** standardised mean difference

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate quality: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low quality: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low quality: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

[1]Downgraded two levels for very serious risk of bias (due to inadequate methods of randomisation, allocation sequence generation, and/or blinding of outcome assessment).
[2]Downgraded one level for serious imprecision.

BACKGROUND

Over the last two decades, there has been a push to integrate simulation-based training into health professions education to facilitate novice skill acquisition in a low-risk environment (Issenberg 1999; Issenberg 2005), and potentially increase the capacity to train individuals at a time where there is a critical shortage of health professionals worldwide (WHO 2013).

Description of the condition

Gastrointestinal endoscopy is an important diagnostic and therapeutic tool used in the evaluation and treatment of gastrointestinal disorders (Faigel 2005). The procedure is technically challenging and requires substantial training for competent performance. Traditionally, novice endoscopists have acquired procedural proficiency through the apprenticeship model, whereby they learn skills under the supervision of experienced preceptors in the clinical setting. This poses several challenges. First, patients are often partially sedated or fully awake during procedures. Second, there is an 'all-or-none' phenomenon requiring the instructor to give up complete control of the endoscope to allow the trainee to master the technique (Dunkin 2003). Third, the finding of pathology during a case is intermittent and unpredictable. A trainee must therefore complete a large volume of procedures to acquire the knowledge and skill necessary to correctly identify, interpret, and manage findings (Dunkin 2003). Fourth, clinical training adds time to each procedure, which has implications with regard to capacity and economics (McCashland 2000). Additionally, clinical demands and time restrictions often limit a preceptor's capacity to provide detailed instruction and feedback.

Description of the intervention

Virtual reality (VR) computer simulators are widely used to enhance traditional endoscopy teaching. We define VR simulation as an educational tool that uses computer technology to create a three-dimensional image or environment that can be interacted with in a seemingly real or physical way (Kim 2001). The use of simulation to teach gastrointestinal endoscopy dates back to 1969, with VR simulators becoming commercially available in 1998 (Bar-Meir 2000; Dunkin 2003; Dunkin 2007). A combination of visual and haptic (tactile) interfaces allows VR simulators to present learners with situations that resemble reality (Krummel 1998; Sturm 2007). In this environment, trainees can practice the technical, cognitive, and non-technical skills of a procedure under varying conditions with no risk of patient harm or discomfort (Sturm 2007). In addition, VR simulators can provide users with objective measures of performance, such as procedural completion time, per cent of mucosa visualised, and degree of patient pain. Such measures can be used to help analyse trainees' actions and identify errors and may allow for the assessment of competence (Walsh 2016).

How the intervention might work

Simulated environments purportedly allow learners to acquire knowledge and build a framework of basic skills through sustained deliberate practice of relevant tasks, with the aim of better preparing novices for patient-based training (Grantcharov 2003; Issenberg 2005). In addition, simulation-based instruction has the potential to improve patient safety as performance of skills on patients by novices may lead to inappropriate applications of procedures, incorrect diagnosis, lower rates of success, and higher rates of complications, all of which put patient safety in jeopardy (Issenberg 2005; Matharoo 2017; Ziv 2003). Furthermore, the simulated setting may provide a more learner-centred educational experience, as supervisors have more time to focus on the needs of the trainee (rather than having to focus on the patient). In addition, errors can be allowed to progress in order to allow the trainee to learn from their mistakes. This can potentially serve to organise future behaviours, as trainees can use the information gained as a basis for change (Blumenthal 1994; Rasmussen 2003; Ziv 2003). Simulation also permits individualised learning, as cases can be adapted to a trainee's unique needs, and the nature and difficulty of the simulation tasks can be systematically varied over time to adapt to the skill level of the learner.

Why it is important to do this review

The growing awareness of the need for patient safety has brought the issue of simulation-based training to the forefront. Because of ethical and medicolegal considerations, gaining experience on patients is becoming increasingly unacceptable during the early stages of training (Kneebone 2001). Virtual reality simulators are becoming popular as a means of providing trainees with the opportunity to rehearse psychomotor, cognitive, and non-technical skills in a risk-free environment, so that they may attain some degree of proficiency prior to performance in the clinical setting. Furthermore, there has been a paradigm shift towards outcomes-based education throughout the healthcare professions, with increasing emphasis on the use of simulation modalities for competency-based evaluation (Brydges 2014; Cook 2013; Frank 2015; Hatala 2005; Holmboe 2010; Langsley 1991; Scalese 2008; Swing 2002).

Simulation technology has the potential to reduce training costs, as staff endoscopists are more productive when performing procedures independently (as compared with supervising trainees) (McCashland 2000). However, it is possible that simulation training carried out on VR simulators does not save money due to the high costs associated with acquiring and maintaining such equipment. It is therefore important to ensure that skills gained through simulation-based training positively transfer to the clinical environment.

This systematic review is an update of our previous review published in 2012 (Walsh 2012). While other systematic reviews have been published more recently, they have not performed comprehensive searches of the educational and computer literature databases and conference proceedings (Dawe 2014; Ekkelenkamp 2016; Qiao 2014; Singh 2014). Additionally, several trials have been published since the most recent systematic review, and these studies have now been assessed for inclusion and presented in this update (Ende 2012; Gomez 2015; Grover 2015; Grover 2017; McIntosh 2014).

OBJECTIVES

To determine whether virtual reality simulation training can supplement and/or replace early conventional endoscopy training (apprenticeship model) in diagnostic oesophagogastroduodenoscopy, colonoscopy, and/or sigmoidoscopy for health professions trainees with limited or no prior endoscopic experience.

METHODS

Criteria for considering studies for this review

Types of studies

We considered randomised controlled trials and quasi-randomised trials (method of allocating participants to treatment not strictly random), irrespective of language, blinding, or publication status. In addition, we considered conference abstracts reporting randomised controlled trials and quasi-randomised trials presented since January 2009. We only considered studies published in abstract format if original outcome data could be retrieved from the abstract or following contact with the authors.

Types of participants

We included health professions trainees, such as physicians (medical students, residents, fellows, and practitioners), nurses, and physician assistants with limited or no prior endoscopy experience. Health professionals are defined as those who study, advise on, or provide preventive, curative, rehabilitative, and promotional health services based on an extensive body of theoretical and factual knowledge in diagnosis and treatment of disease and other health problems (WHO 2008). For the purposes of this review, we defined limited endoscopic experience as:

1. previous performance of no greater than 20 cases of the procedure under study in the clinical or simulated setting; and/or
2. any level of experience in performing other gastrointestinal endoscopic procedures (oesophagogastroduodenoscopy, colonoscopy, and sigmoidoscopy).

Types of interventions

We included trials comparing VR endoscopy (oesophagogastroduodenoscopy, colonoscopy, and sigmoidoscopy) simulation training versus any other method of endoscopy training, including conventional patient-based training, training using another form of endoscopy simulation (e.g. low-fidelity simulator), or no training (however defined by authors). We also included trials comparing one method of VR training versus another method of VR training (e.g. comparison of two different VR simulators, comparison of two different VR curricula). We did not include virtual patient computer-based simulations (interactive computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment) (Ellaway 2006; Kononowicz 2016).

Types of outcome measures

We included only trials measuring outcomes on humans (as opposed to animals or simulators) in the clinical setting.

Primary outcomes

1. Composite score of competency in performing endoscopy (as defined by authors).

The outcome 'composite score of competency' reflects an overall aggregate score derived from various workplace-based assessment tools that can be used to assess competence in performing an endoscopic procedure within the real clinical setting. Workplace-based assessment tools are reliant on an external rater to directly observe and assess a learner using predefined criteria that are built around an assessment framework (Walsh 2016). The individual components that make up different assessment tools vary but are similar in that the item scores are aggregated to produce an overall score. Published validity evidence for each individual workplace-based assessment tool is variable (Walsh 2016). These tools allow for structured assessment at the 'does' level of Miller's pyramid of assessment of clinical competence, reflective of what an individual does during a real clinical encounter, thus providing a high degree of authenticity (Miller 1990).

Secondary outcomes

1. Independent procedure completion (objective measure).
2. Performance time (objective measure of the time taken to perform the evaluation task(s) post-training (minutes)).
3. Complication or critical flaw occurrence related to the endoscopic procedure (e.g. bleeding, luminal perforation, and infection) (ASGE 2011).
4. Patient discomfort (as defined by authors).
5. A single measure providing an overall global rating of performance or competency in performing endoscopy (as defined by the authors).
6. Visualisation of mucosa (as defined by authors).

Search methods for identification of studies

Electronic searches

We searched the following electronic health professions, educational, and computer literature databases for publications addressing the above clinical problem. We have presented all search strategies in Appendix 1 including information on the time span for the searches.

1. The Cochrane Central Register of Controlled Trials (CENTRAL; 2017, Issue 6) in the Cochrane Library (searched 12 July 2017)
2. MEDLINE (1946 to 12 July 2017)
3. Embase (1947 to 12 July 2017)
4. Scopus (1960 to 12 July 2017)
5. Web of Science
   a. Science Citation Index Expanded (1900 to 12 July 2017)
   b. Social Sciences Citation Index (1956 to 12 July 2017)
   c. Arts and Humanities Citation Index (1975 to 12 July 2017)
   d. Conference Proceedings Citation Index - Science (1990 to 12 July 2017)
   e. Conference Proceedings Citation Index - Social Science & Humanities (1990 to 12 July 2017)
6. Biosis Previews (1980 to 12 July 2017)
7. CINAHL (Cumulative Index to Nursing and Allied Health Literature) (1981 to 12 July 2017)
8. AMED (Allied and Complementary Medicine Database) (1985 to 12 July 2017)
9. ERIC (1966 to 12 July 2017)
10. Education Full Text (1969 to 12 July 2017)
11. CBCA Education (1933 to 12 July 2017)
12. ACM Digital Library (1948 to 12 July 2017)
13. IEEE Xplore (1950 to 12 July 2017)
14. Abstracts in New Technologies and Engineering (1981 to 12 July 2017)

15. Computer and Information Systems Abstracts (1981 to 12 July 2017)
16. ProQuest Dissertations and Theses Global (1997 to 12 July 2017)

Searching other resources

We handsearched the reference lists of the studies and review articles identified using the computer-assisted search to identify further relevant studies.

We searched abstracts and proceedings of major gastrointestinal, educational, and surgical meetings

1. Gastrointestinal
   a. Digestive Diseases Week (2009-17)
   b. Canadian Digestive Diseases Week (2009-17)
   c. British Society of Gastroenterology (2009-17)
   d. United European Gastroenterology Week (2009-17)
2. Educational
   a. The Association for Medical Education in Europe Conference (2009-17)
   b. Canadian Conference on Medical Education (2009-17)
   c. Research in Medical Education Conference (2009-17)
3. Surgical
   a. American College of Surgery Clinical Congress (2009-17)
   b. The Society of American Gastrointestinal and Endoscopic Surgeons Conference (2009-17)
   c. European Association for Endoscopic Surgery Congress (2009-17)).

We searched the grey literature including: metaRegister of controlled trials (active and archived registers) (12 November 2017).

Data collection and analysis

We collected data on customised data extraction forms and performed analyses as described below.

Selection of studies

After completing the literature searches, we merged the search results using the software package EndNote X8 (reference management software) and removed duplicate records (Endnote 2016). In this updated review, two review authors (RK and JP) independently reviewed all titles and abstracts identified by the literature search for inclusion. We retrieved the full text for further assessment if the inclusion criteria were unclear from the abstract. We documented excluded trials, with the reasons for exclusion. A third review author (CMW) resolved any discrepancies between the first two review authors.

Data extraction and management

We used a standard data collection form that was updated from the previous version of this review as per updated Methodological Expectations of Cochrane Intervention Reviews (MECIR) standards (Higgins 2016). Two review authors (RK and JP) independently extracted the data listed below.

1. General article information: title, authors, publication year, language of publication, country where study was performed
2. Year of conduct of trial
3. Funding source of trial
4. Declarations of interest for primary investigators
5. Study design: randomisation process, allocation concealment, blinding
6. Sample size and sample size calculation
7. Study participants: inclusion/exclusion criteria, years participants were enrolled, health profession (physicians (medical students, residents, fellows, and practitioners), nurses, or physician assistants), training programme (e.g. gastroenterology, general surgery) level of training, endoscopy experience, numbers randomised, baseline characteristics (age, gender)
8. Endoscopic procedure under study (oesophagogastroduodenoscopy, colonoscopy, and/or sigmoidoscopy)
9. Intervention: learning theory used to design intervention (if any), name of VR endoscopy simulator, name of non-VR simulators, training task, duration of training, description of intervention, nature of observation, instruction, and feedback (if applicable)
10. Comparison: nature of comparison group (conventional patient-based training, training using another form of endoscopy simulation (e.g. low-fidelity simulator), no training, training using another method of VR training), name of VR endoscopy simulator(s) (if applicable), name of non-VR simulator(s) (if applicable), training task (if applicable), duration of training (if applicable), description of intervention (if applicable), nature of observation, instruction, and feedback (if applicable)
11. Outcomes assessed, assessment method, and time to assessment
12. Assessment scoring (if applicable), and validation of instrument used for assessment scoring (if applicable)
13. Data on the primary outcome measures (as described above)
14. Data on the secondary outcome measures (as described above)
15. Methodological quality (as described below) Intention-to-treat analysis

Assessment of risk of bias in included studies

Two review authors (RK and JP) independently assessed the methodological quality of included studies, without masking of the study names, using the Cochrane domain-based tool for assessing risk of bias (Higgins 2011). We assessed the following factors: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other sources of bias (Appendix 2).

We judged each domain as low risk, high risk, or unclear risk of bias according to the criteria used in the Cochrane 'Risk of bias' tool (Higgins 2011). We considered a trial to be at low risk of bias if we assessed the trial as at low risk of bias across all domains. Otherwise, we considered trials at unclear risk of bias or at high risk of bias if we assessed one or more domains as at unclear or high risk of bias, respectively. If the published data provided inadequate information we sought clarification from the trial authors. Two review authors (RK and JP) independently assessed the risk of bias. Any unresolved discrepancies between review authors were resolved through discussion with a third review author (CMW).

Measures of treatment effect

When abstracting data from studies reporting learning curves (multiple points across time) (Cohen 2006; Ferlitsch 2010; Sedlack 2004; Sedlack 2007), we used the first assessment interval for analysis and plots in order to minimise the potential effect of variable clinical training on the outcomes over time. We performed a meta-analysis according to the recommendations in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). We used the statistical package Review Manager 5 provided by The Cochrane Collaboration to analyse and synthesise data (RevMan 2014). For dichotomous data, such as independent procedure completion (yes/no), we expressed the impact of the intervention as a risk ratio with 95% confidence intervals. We used risk ratio due to its ease of interpretation. For continuous data such as performance time, composite score, independent insertion depth, and patient discomfort, we estimated the effect size by computing the mean difference with corresponding 95% confidence intervals when studies reported the same outcome measures, or standardised mean difference with corresponding 95% confidence intervals when studies reported different outcome measures.

Unit of analysis issues

The unit of analysis was each patient-based gastrointestinal endoscopic procedure performed (e.g. oesophagogastroduodenoscopy, colonoscopy, and sigmoidoscopy) on which an outcome measure was assessed.

Dealing with missing data

If outcome data were missing, we contacted the trial authors for further details and asked them to provide original data if the published paper or abstract contained insufficient or unclear information. If it was unclear whether trials shared the same participants, completely or partially (by identifying common authors or centres), we contacted the authors of the trials to clarify whether the trial had been duplicated. A third review author (CMW) resolved any differences in opinion through discussion.

Assessment of heterogeneity

Two review authors (RK and JP) independently evaluated eligible studies for clinical and methodological heterogeneity. We assessed heterogeneity using the Cochran Chi^2 test (Q-test) with the alpha level of significance set at 0.10. We also estimated the degree of heterogeneity using the I^2 statistic, which describes the percentage of total variation across studies that results from heterogeneity rather than chance. We quantified heterogeneity using the I^2 statistic with the following interpretations: 0% to 40% low heterogeneity, 30% to 60% moderate heterogeneity, 50% to 90% substantial heterogeneity, and 75% to 100% considerable heterogeneity. We applied this for all outcomes as suggested in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

Assessment of reporting biases

We had planned to examine potential publication bias by means of a funnel plot (Egger 1997; Macaskill 2001). However, this was not done due to the low number of trials reporting similar outcomes (Sterne 2011).

Data synthesis

We performed the meta-analysis using Review Manager 5 (RevMan 2014). We planned to pool data a priori for meta-analysis if participant groups were similar and the studies assessed the same or similar interventions with the same comparator, and had similar definitions of outcome measures (determined by consensus). We used a random-effects or fixed-effect model depending on the presence or absence of heterogeneity. For the fixed-effect model, we performed weighting using the Mantel-Haenszel method (Higgins 2016). For the random-effects model, we weighted studies using the DerSimonian and Laird method (Higgins 2016). For studies with three or more arms (Ende 2012; Gomez 2015), we excluded groups that included combination training (e.g. VR simulation training followed by patient-based training) from the meta-analysis to allow for a direct comparison of VR simulation to a control, such as no intervention or conventional patient-based training.

Subgroup analysis and investigation of heterogeneity

We performed the following subgroup analyses.

1. Type of endoscopic procedure under study (oesophagogastroduodenoscopy, colonoscopy, and sigmoidoscopy)
2. Level of participant endoscopy experience (no prior versus limited endoscopy experience)

Sensitivity analysis

We planned to perform the following sensitivity analyses, but due to too few included trials for each outcome analysis we did not carry them out.

1. Excluding studies at high or unclear risk of bias (trials with adequate methodology compared to trials with unclear or inadequate methodologies)
2. Excluding studies that were published only in abstract form and that required contact with authors to retrieve full methodology and original outcome data

'Summary of findings' table

We evaluated the quality of evidence using the GRADE approach for each outcome including any subgroup analysis for each of the following comparisons (Schünemann 2013).

1. Virtual reality endoscopy simulation training versus no training
2. Virtual reality endoscopy simulation training versus conventional patient-based training

We used GRADEpro GDT to present the quality of evidence in 'Summary of findings' tables (see Summary of findings for the main comparison and Summary of findings 2) (GRADEpro 2017). We downgraded the quality of evidence by one level (serious concern) or two levels (very serious concern) for the following reasons: risk of bias, inconsistency (unexplained heterogeneity, inconsistency of results), indirectness (indirect population, intervention, control, outcomes), imprecision (wide confidence intervals, single trial, few events or patients randomised across trials), and publication bias. We also upgraded the quality by one level due to a large summary effect or a large training response (as training increased, the effect increased).

RESULTS

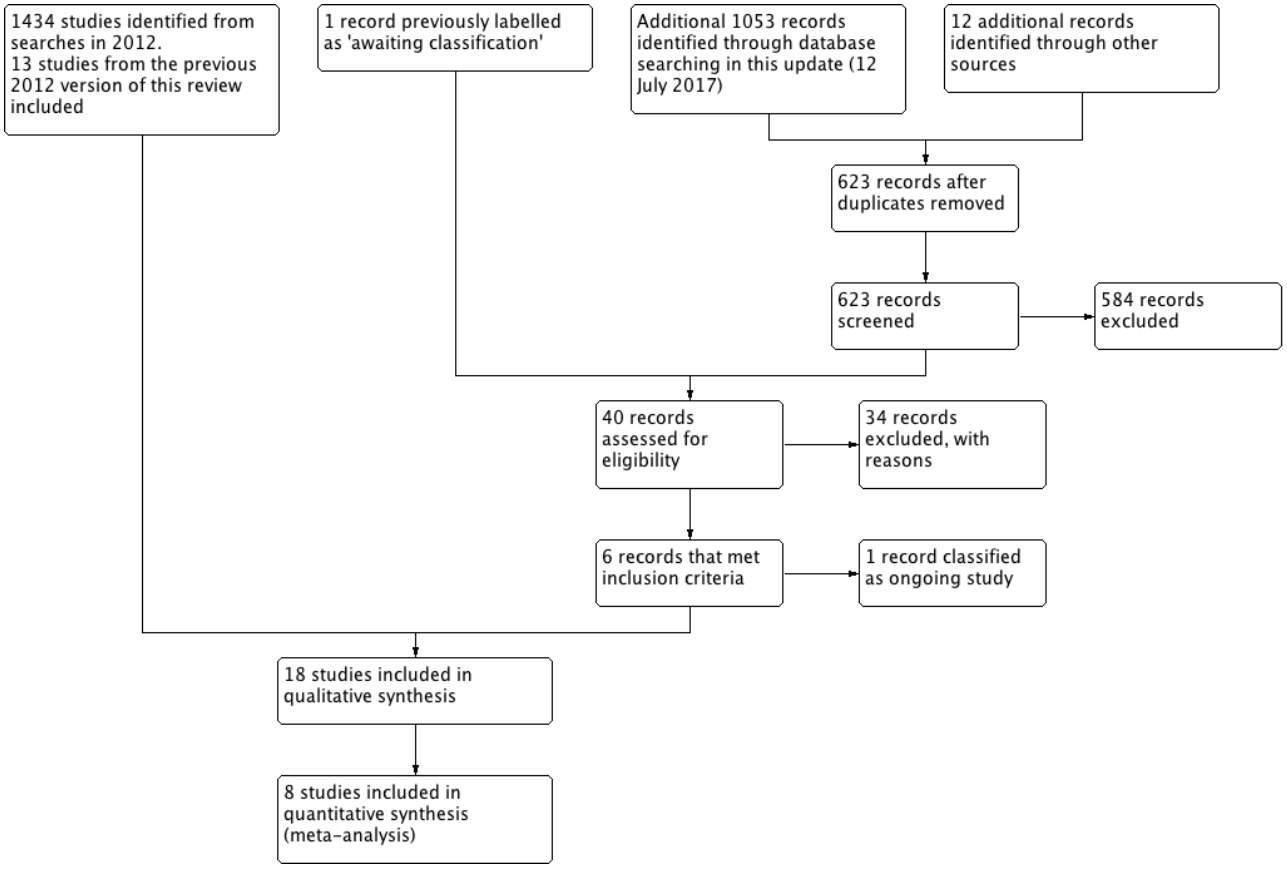
Description of studies

Details of the included and excluded studies are listed in the Characteristics of included studies, Characteristics of excluded studies, and Characteristics of studies awaiting classification tables.

Results of the search

The searches from the first published version of this review in 2012 yielded a total of 1434 references. In the updated search performed 12 July 2017, we identified an additional 1065 references, after exclusion of the 1434 references identified in the 2012 search (Figure 1). We identified 1053 abstracts through electronic searches of the Cochrane Central Register of Controlled Trials (n = 143), MEDLINE (n = 154), Embase (n = 341), Scopus (n = 160), Web of Science (n = 120), and other databases (n = 135). We identified an additional 12 abstracts through searching the conference proceedings of major gastrointestinal, educational, and surgical conferences. We removed 442 duplicate references and excluded 584 references through review at the title and abstract level. We identified one abstract that was labelled as 'awaiting classification' in the previous version of this review (NCT01405443).

Figure 1. Study flow diagram.



We retrieved 40 full-text articles and/or conference abstracts for further assessment. We did not identify any additional references though a manual search of the reference lists of the identified trials. We excluded 34 references for the reasons listed in the Characteristics of excluded studies table. We classified one study as an ongoing trial. This updated review included five new trials and 13 trials from the previous version of this review, for a total of 18 trials. The study flow diagram is provided in Figure 1.

Included studies

We included 18 trials with 421 participants, and a total of 3817 endoscopic procedures. Ten trials compared VR training versus no intervention (Ahlberg 2005; Cohen 2006; Di Giulio 2004; Ferlitsch 2010; McIntosh 2014; Park 2007; Sedlack 2004; Sedlack 2007; Tuggy 1998; Yi 2008), while five trials compared VR training versus conventional patient-based endoscopy training (apprenticeship model) (Ende 2012; Gerson 2003; Haycock 2010; Sedlack 2004a; Shirai 2008). One trial compared VR training to another form of endoscopy simulation (Gomez 2015), and two trials compared different methods of VR training (Grover 2015; Grover 2017). Two of the above trials included three arms (Ende 2012; Gomez 2015). In one trial (Ende 2012), the intervention arm received VR training in addition to conventional patient-based training, while the two comparator arms received VR training only and conventional patient-based training only, respectively. In the second trial (Gomez 2015), the intervention arm received VR training in addition to another form of endoscopy simulation training, while the two comparator arms received VR training only and another form of endoscopy simulation training only, respectively.

Ten trials studied training in colonoscopy (Ahlberg 2005; Cohen 2006; Gomez 2015; Grover 2015; Grover 2017; Haycock 2010;

McIntosh 2014; Park 2007; Sedlack 2004; Yi 2008); three studied sigmoidoscopy (Gerson 2003; Sedlack 2004a; Tuggy 1998); and five studied oesophagogastroduodenoscopy (Di Giulio 2004; Ende 2012; Ferlitsch 2010; Sedlack 2007; Shirai 2008). Details of the trials such as methodological quality, inclusion and exclusion criteria, and the outcomes measured are shown in the Characteristics of included studies table.

Four trials included gastroenterology trainees (medical residents or fellows or both) only (Cohen 2006; Di Giulio 2004; Sedlack 2004; Sedlack 2007). Three trials included trainees in gastroenterology, medicine, and general surgery (Grover 2015; Grover 2017; McIntosh 2014); one trial included gastroenterology and general surgery trainees (Ahlberg 2005); and one trial included general surgery residents only (Gomez 2015). Two trials stated that the participants were residents or fellows or both but did not state their discipline (Shirai 2008; Yi 2008). One trial included participants from any healthcare background (e.g. physicians, nurses) or position recognised by the training institution as appropriate for training in colonoscopy (Haycock 2010). The remaining six trials included internal medicine, family medicine, and/or surgical residents without any prior experience in endoscopy (Ende 2012; Ferlitsch 2010; Gerson 2003; Park 2007; Sedlack 2004a; Tuggy 1998).

Two trials that studied training in colonoscopy included participants with prior experience in oesophagogastroduodenoscopy (Ahlberg 2005; Sedlack 2004), and one study included trainees who had previously performed fewer than 25 colonoscopies or flexible sigmoidoscopies; however, none of the participants had performed more than 1 of the procedures under study (colonoscopy or flexible sigmoidoscopy or both) (Haycock 2010). Four studies included trainees who had been the primary endoscopist for fewer than 3, Park 2007, 10, McIntosh 2014, or 20, Grover 2015; Grover 2017, procedures of any type, respectively. One study included trainees who had prior experience in oesophagogastroduodenoscopy and flexible sigmoidoscopy, but had performed fewer than 10 previous colonoscopies (the procedure under study) (Cohen 2006). Another study did not state participants' previous endoscopy experience (Yi 2008). The remaining nine trials included participants with no prior endoscopy experience (Di Giulio 2004; Ende 2012; Ferlitsch 2010; Gerson 2003; Gomez 2015; Sedlack 2004a; Sedlack 2007; Shirai 2008; Tuggy 1998).

Further details regarding the simulators used, training tasks, and outcomes evaluated are shown in Table 1.

Excluded studies

We excluded 34 studies for the reasons provided in the Characteristics of excluded studies table. We excluded one study that is an ongoing trial; see Characteristics of ongoing studies.

Risk of bias in included studies

See the 'Risk of bias' tables in Characteristics of included studies.

We considered three trials to be at low risk of bias (Ahlberg 2005; Grover 2015; Grover 2017). We considered nine trials to be at high risk of bias as sequence generation was not random; there was no description of allocation concealment methods; and/or there was no blinding of outcome assessment (Di Giulio 2004; Ende 2012; Ferlitsch 2010; Gerson 2003; Gomez 2015; McIntosh 2014; Sedlack 2004; Sedlack 2004a; Yi 2008). We considered the remaining six trials to be at unclear risk of bias as the method of randomisation and/or blinding of outcome assessment was unclear; and/or an assessment instrument with no evidence of validity was used (Cohen 2006; Haycock 2010; Park 2007; Sedlack 2007; Shirai 2008; Tuggy 1998). We assessed 'Risk of bias' domains as unclear when despite attempts to contact study authors, information was insufficient to make a clear judgement about risk of bias. Risk of bias is summarised in Figure 2 and Figure 3.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

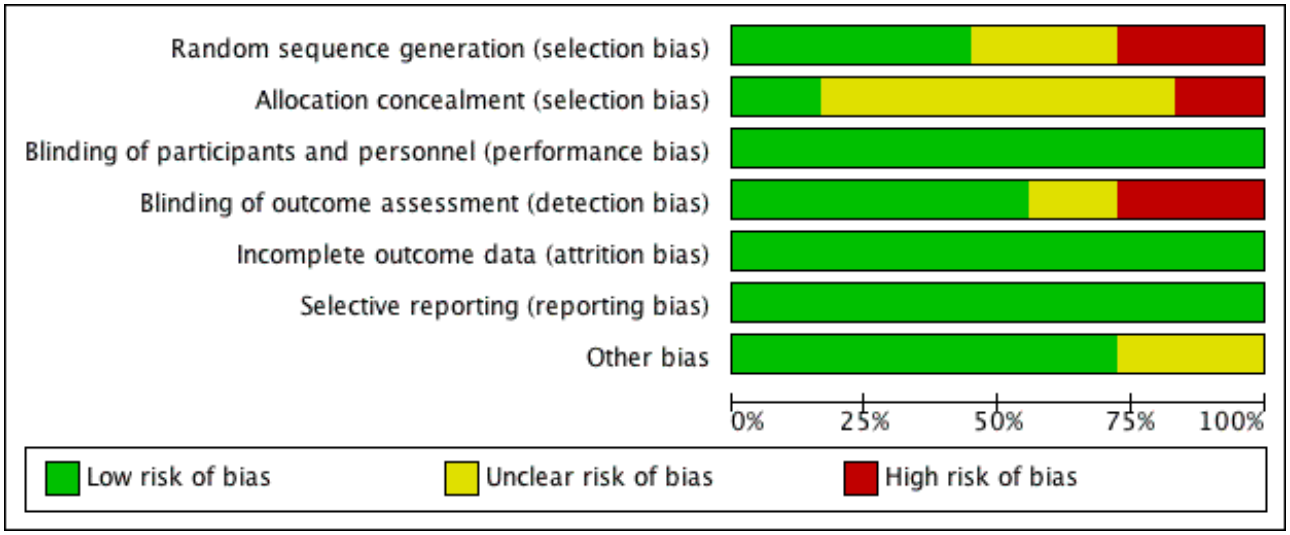
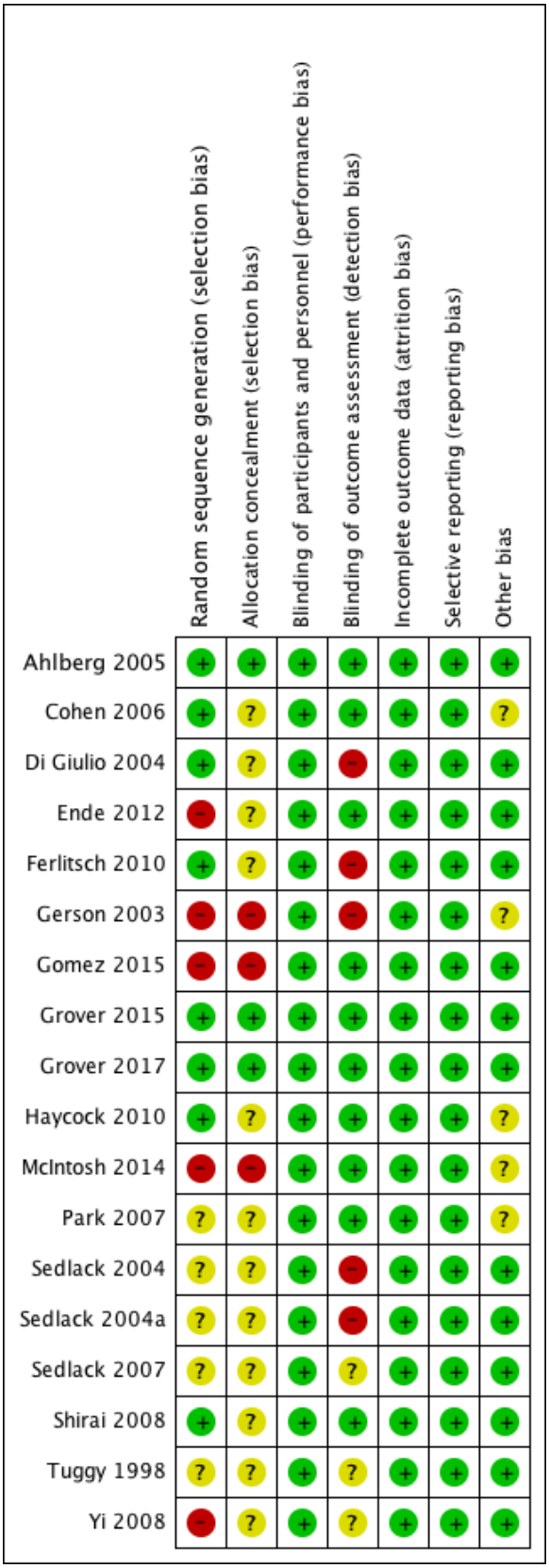


Figure 3. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.



Allocation

Six trials reported adequate generation of the allocation sequence (Cohen 2006; Di Giulio 2004; Ferlitsch 2010; Grover 2015; Grover 2017; Haycock 2010). Five trials reported inadequate methods of sequence generation (Ende 2012; Gerson 2003; Gomez 2015; McIntosh 2014; Yi 2008). The other seven trials did not describe the sequence generation process utilised (Ahlberg 2005; Park 2007; Sedlack 2004; Sedlack 2004a; Sedlack 2007; Shirai 2008; Tuggy 1998). Three trials reported using appropriate procedures to minimise or eliminate bias in allocation concealment (Ahlberg 2005; Grover 2015; Grover 2017). Three trials reported inadequate allocation concealment (Gerson 2003; Gomez 2015; McIntosh 2014). The remaining 12 trials did not report on allocation concealment (Cohen 2006; Di Giulio 2004; Ende 2012; Ferlitsch 2010; Haycock 2010; Park 2007; Sedlack 2004; Sedlack 2004a; Sedlack 2007; Shirai 2008; Tuggy 1998; Yi 2008).

Blinding

Due to the nature of the intervention, it was not possible to blind the participants and personnel administering the intervention; however, the outcome was not likely to have been influenced by the lack of blinding. Ten trials reported adequate blinding of the outcome assessment (Ahlberg 2005; Cohen 2006; Ende 2012; Gomez 2015; Grover 2015; Grover 2017; Haycock 2010; McIntosh 2014; Park 2007; Shirai 2008). Five trials reported inadequate assessor blinding (Di Giulio 2004; Ferlitsch 2010; Gerson 2003; Sedlack 2004; Sedlack 2004a). The remaining three trials did not report on assessor blinding or provided insufficient information to permit judgement for this domain (Sedlack 2007; Tuggy 1998; Yi 2008).

Incomplete outcome data

All 18 trials addressed incomplete outcome data (Ahlberg 2005; Cohen 2006; Di Giulio 2004; Ende 2012; Ferlitsch 2010; Gerson 2003; Gomez 2015; Grover 2015; Grover 2017; Haycock 2010; McIntosh 2014; Park 2007; Sedlack 2004; Sedlack 2004a; Sedlack 2007; Shirai 2008; Tuggy 1998; Yi 2008).

Selective reporting

All 18 trials were free of selective outcome reporting (Ahlberg 2005; Cohen 2006; Di Giulio 2004; Ferlitsch 2010; Gerson 2003; Haycock 2010; Park 2007; Sedlack 2004; Sedlack 2004a; Sedlack 2007; Shirai 2008; Tuggy 1998; Yi 2008).

Other potential sources of bias

None of the trials reported intention-to-treat analysis. Only eight trials reported a sample size calculation (Ende 2012; Ferlitsch 2010; Gerson 2003; Grover 2015; Grover 2017; Haycock 2010; McIntosh 2014; Park 2007). While the authors of one study reported the use of a "validated" Global Performance Score (Park 2007), no reference or details of validity evidence were provided. Another study utilised the same Global Performance Score (Haycock 2010). In addition, one study utilised subsections of the UK Joint Advisory Group's Colonoscopy Direct Observation of Procedural Skills (Haycock 2010), which has good validity evidence (Barton 2008; Barton 2012); however, validity evidence of the abbreviated version has not been examined. Another trial, Cohen 2006, utilised a previously developed outcome instrument (Cass 1996); however, there is no literature to suggest that the validity of this instrument has been assessed. Finally, one trial utilised a 5-point Likert scale to evaluate trainees' technique in colonoscopy, for which no validity evidence could be found (McIntosh 2014).

Effects of interventions

See: **Summary of findings for the main comparison** Virtual reality endoscopy simulation training versus no training; **Summary of findings 2** Virtual reality endoscopy simulation training versus conventional patient-based training

We included 18 trials with 421 participants in this review. We reported only outcomes assessed on humans in the clinical setting. There was substantial clinical and methodological heterogeneity, therefore it was not possible to perform a meta-analysis for several outcomes among our four comparisons. In addition, several trials did not provide sufficient data for inclusion in a meta-analysis. Specifically, trials did not provide data with respect to central tendency (mean) and variability (standard deviation) to allow for quantitative analysis. Where a meta-analysis was not performed, we have presented the results of the studies, categorised by outcome measure, in tabular form. We have reported the level of statistical significance across groups where available.

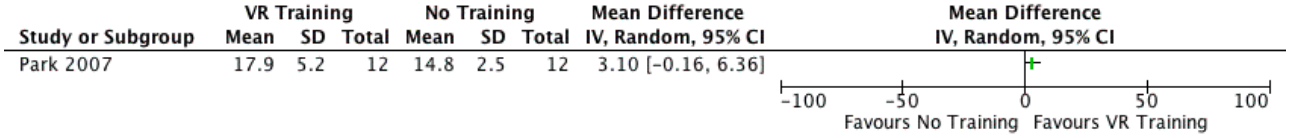
1. Virtual reality endoscopy simulation training versus no training

Primary outcome

1.1 Composite score of competency in performing endoscopy (as defined by authors)

One trial comparing VR endoscopy simulation training versus no training reported a composite score of competency (as defined by authors), and showed no statistically significant difference in composite score of competency in the VR training group as compared with the no-training group (mean difference (MD) 3.10, 95% confidence interval (CI) -0.16 to 6.36; 1 trial (n = 24 procedures); Analysis 1.1) (Park 2007). We downgraded this finding to low quality due to serious risk of bias and serious imprecision. The results are summarised in Summary of findings for the main comparison, Table 2, Analysis 1.1, and Figure 4.

Figure 4. Analysis 1.1. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.1 Composite score of competency.



Secondary outcomes

1.2 Independent procedure completion (objective measure)

Six trials comparing VR endoscopy simulation training versus no training reported independent procedure completion (Ahlberg 2005; Di Giulio 2004; McIntosh 2014; Park 2007; Sedlack 2004; Yi 2008). The meta-analysis showed that the VR training group had a significantly higher number of independent procedure completions than the no-training group (risk ratio (RR) 1.62, 95% CI 1.15 to 2.26; 6 trials (n = 815); Analysis 1.2, Analysis 1.3). Heterogeneity was statistically significant (P = 0.030) and moderate (I^2 = 61%). We performed subgroup analyses for type of endoscopic procedure under study (colonoscopy and oesophagogastroduodenoscopy) and for level of participant endoscopy experience (no prior versus limited endoscopy experience). The VR training groups had significantly more independent procedure completions compared to no-training groups for studies in colonoscopy (RR 1.84, 95% CI 1.35 to 2.50; I^2 = 11%; 5 trials (n = 408)) and oesophagogastroduodenoscopy (RR 1.25, 95% CI 1.13 to 1.39; 1 trial (n = 407); Analysis 1.2). Tests for interaction showed statistically significant procedure-related heterogeneity (P = 0.020). In addition, the VR training groups had significantly more independent procedure completions compared to no-training groups when participants had limited prior endoscopy experience (RR 1.82, 95% CI 1.07 to 3.12; I^2 = 54%; 3 trials (n = 329); Analysis 1.3) and no prior experience (1.32, 95% CI 1.09 to 1.61; I^2 = 13%; 3 trials (n = 486); Analysis 1.3). However, tests for interaction showed no statistically significant prior experience-related heterogeneity (P = 0.27). We downgraded this finding to moderate quality due to serious risk of bias. The results are summarised in Summary of findings for the main comparison, Table 3, Analysis 1.2, Analysis 1.3, Figure 5, and Figure 6.

Figure 5. Analysis 1.2. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.2 Independent procedure completion: type of endoscopic procedure under study.

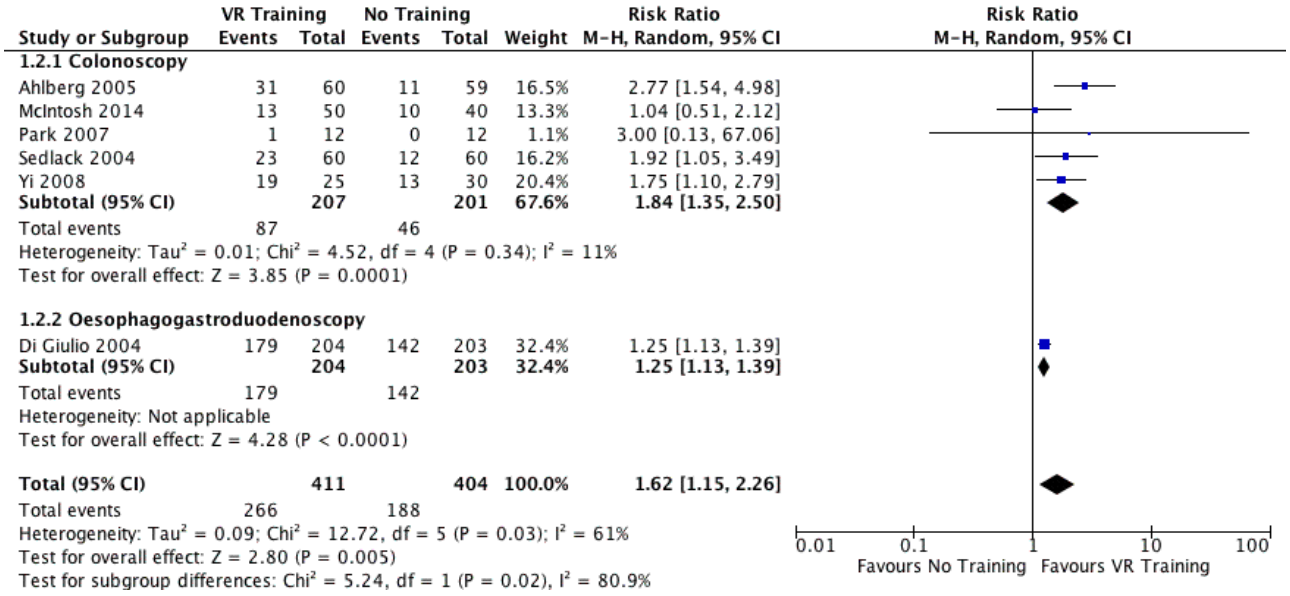
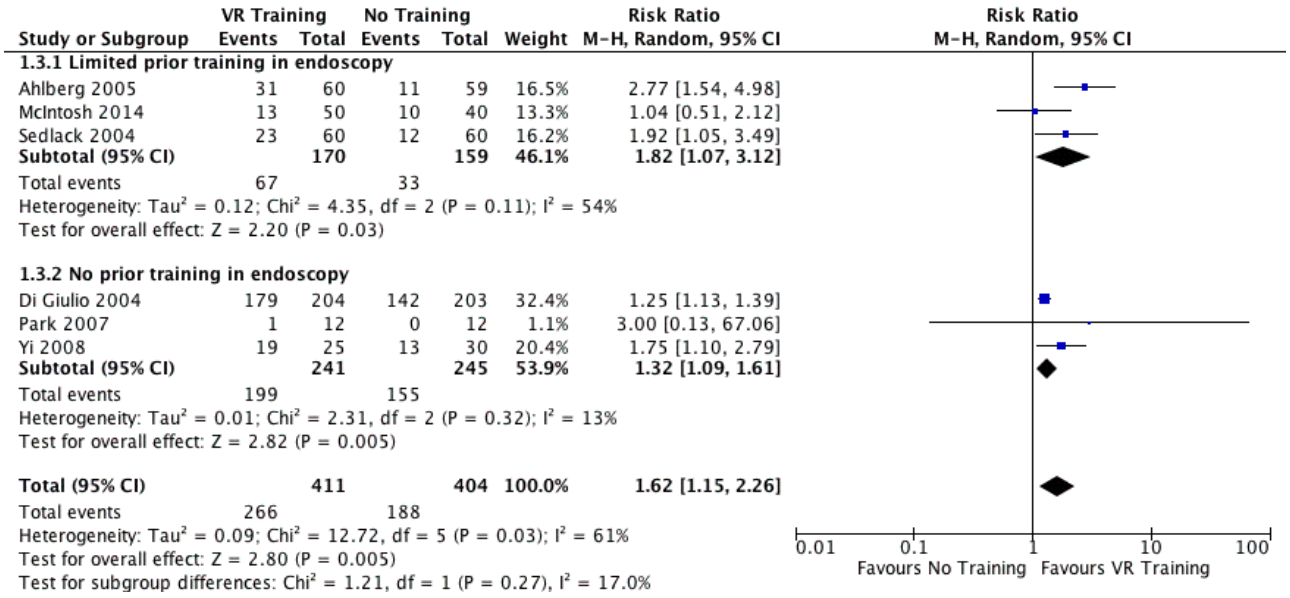


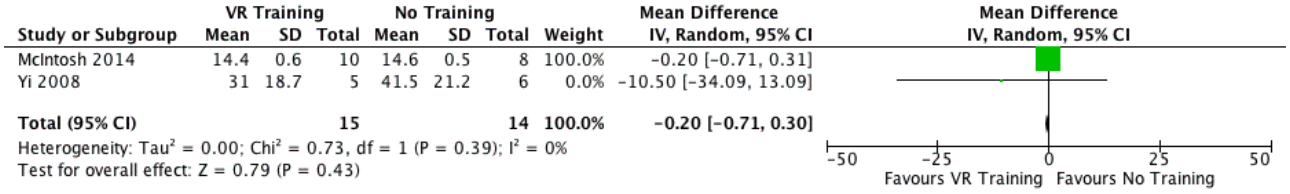
Figure 6. Analysis 1.3. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.3 Independent procedure completion: level of participant endoscopy experience.



1.3 Performance time (objective measure of the time taken to perform the evaluation task(s) post-training)

Seven trials comparing VR endoscopy simulation training versus no training reported performance time (time taken to perform the evaluation task(s)) (Ahlberg 2005; Di Giulio 2004; Ferlitsch 2010; McIntosh 2014; Sedlack 2004; Tuggy 1998; Yi 2008). We included only two trials in the meta-analysis due to insufficient central tendency and variability data (McIntosh 2014; Yi 2008), which showed no significant difference between the VR training group and no-training group with respect to performance time (MD -0.20, 95% CI -0.71 to 0.30; 2 trials (n = 29); Analysis 1.4). Heterogeneity was not statistically significant (P = 0.39) and was low (I^2 = 0%). Among the remaining five trials that reported this outcome, three showed a statistically significantly faster time for the VR training group as compared to the no-training group (Ahlberg 2005; Ferlitsch 2010; Tuggy 1998), and two showed no significant difference (Di Giulio 2004; Sedlack 2004). We downgraded this finding to very low quality due to serious risk of bias and serious imprecision. The results are summarised in Summary of findings for the main comparison, Table 4, Analysis 1.4, and Figure 7.

Figure 7. Analysis 1.4. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.4 Performance time.



1.4 Complication or critical flaw occurrence

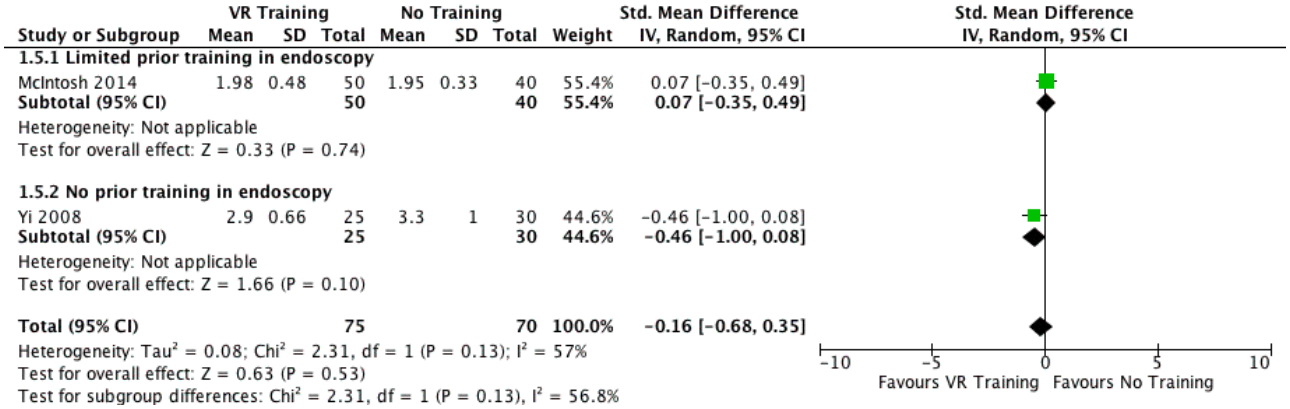
Three trials (550 procedures) comparing VR endoscopy simulation training versus no training reported the occurrence of procedure-related complications or critical flaws (Ahlberg 2005; Di Giulio 2004; Park 2007). All three trials reported no complications or critical flaws in any of the study groups. We downgraded this finding to moderate quality due to serious risk of bias. The results are summarised in Summary of findings for the main comparison and Table 5.

1.5 Patient discomfort (as defined by authors)

Seven trials comparing VR endoscopy simulation training versus no training reported patient discomfort (as defined by authors) (Ahlberg 2005; Cohen 2006; Ferlitsch 2010; McIntosh 2014; Sedlack 2004; Tuggy 1998; Yi 2008). We included only two trials in the meta-analysis due to insufficient central tendency and variability data (McIntosh 2014; Yi 2008), which showed no significant difference between the VR training group and the no-training group with respect to performance time (standardised mean difference (SMD) -0.16, 95% CI -0.68 to 0.35; 2 trials (n = 145); Analysis 1.5). Heterogeneity was not statistically significant (P = 0.13) and was moderate (I^2 = 57%). We performed subgroup analysis for level of participant endoscopy experience (no prior versus limited endoscopy experience). There were no significant differences with respect to patient discomfort when participants had limited prior endoscopy experience (SMD 0.07, 95% CI -0.35 to 0.49; 1 trial (n = 90); Analysis 1.5) and no prior experience (SMD -0.46, 95% CI -1.00 to 0.08; 1 trial (n = 55); Analysis 1.5). Tests for interaction showed no statistically significant prior experience-related heterogeneity

(P = 0.13). Among the remaining five trials that reported this outcome (Ahlberg 2005; Cohen 2006; Ferlitsch 2010; Sedlack 2004; Tuggy 1998), there was no significant difference between groups with respect to patient discomfort. We downgraded this finding to very low quality due to serious risk of bias, serious imprecision, and inconsistency (unexplained heterogeneity). The results are summarised in Summary of findings for the main comparison, Table 6, Analysis 1.5, and Figure 8.

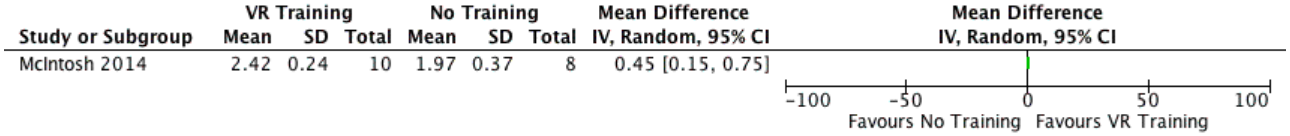
Figure 8. Analysis 1.5. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.5 Patient discomfort.



1.6 A single measure providing an overall global rating of performance or competency in performing endoscopy (as defined by the authors)

Four trials comparing VR endoscopy simulation training versus no training reported an overall rating of performance or competency (Cohen 2006; Di Giulio 2004; McIntosh 2014; Sedlack 2007). We did not perform a meta-analysis as only one trial had sufficient central tendency and variability data (McIntosh 2014). This trial showed statistically significantly more positive ratings in the VR training group compared to the no-training group (MD 0.45, 95% CI 0.15 to 0.75; 1 trial (n = 18); Analysis 1.6) (McIntosh 2014). Two other trials showed statistically significantly more positive ratings in the VR training group (Cohen 2006; Di Giulio 2004), and one trial showed no significant difference between groups (Sedlack 2007). We downgraded this finding to very low quality due to very serious risk of bias and serious imprecision. The results are summarised in Summary of findings for the main comparison, Table 7, Analysis 1.6, and Figure 9.

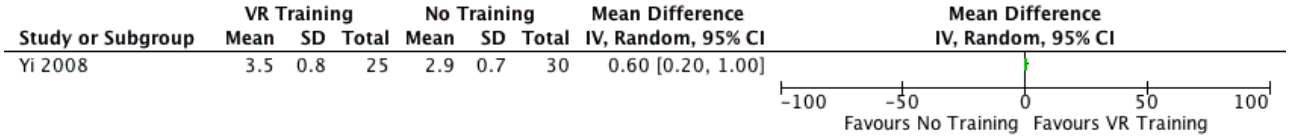
Figure 9. Analysis 1.6. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.6 Overall global rating of performance or competency.



1.7 Visualisation of mucosa (as defined by authors)

Three trials comparing VR endoscopy simulation training versus no training reported visualisation of the mucosa (as defined by the authors) (Sedlack 2004; Tuggy 1998; Yi 2008). We did not perform a meta-analysis as only one trial had sufficient central tendency and variability data (Yi 2008). Visualisation was significantly greater in this trial in the VR training group (MD 0.60, 95% CI 0.20 to 1.00; 1 trial (n = 55); Analysis 1.7) (Yi 2008). Visualisation was also significantly greater in the VR training group in the other two trials (Sedlack 2004; Tuggy 1998). We downgraded this finding to very low quality due to very serious risk of bias and serious imprecision. The results are summarised in Summary of findings for the main comparison, Table 8, Analysis 1.7, and Figure 10.

Figure 10. Analysis 1.7. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1.7 Visualisation of mucosa.



2. Virtual reality endoscopy simulation training versus conventional patient-based training

Primary outcomes

2.1 Composite score of competency in performing endoscopy (as defined by authors)

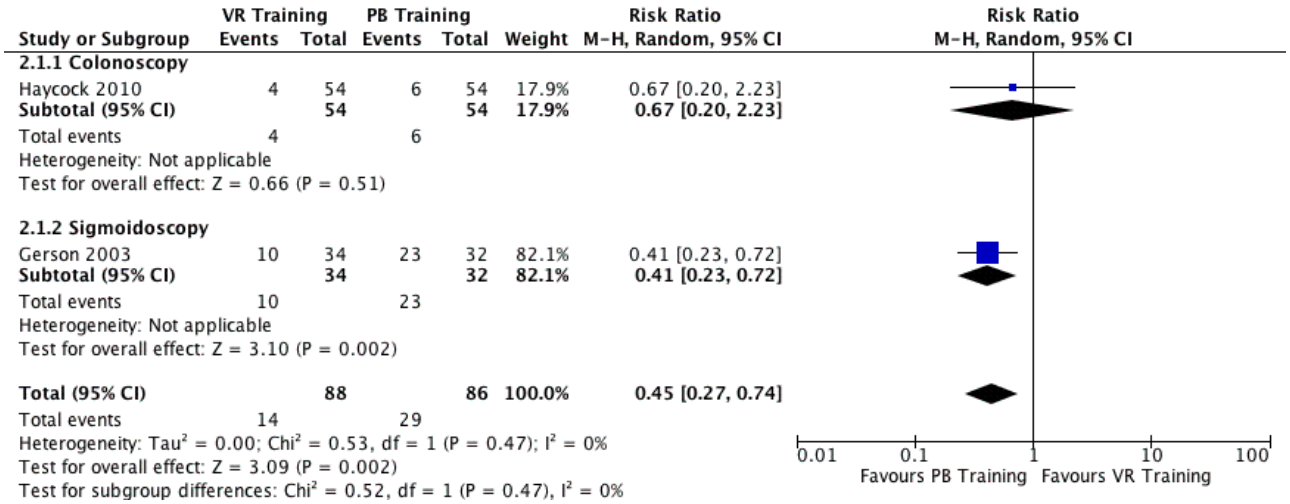
One trial comparing VR endoscopy simulation training versus conventional patient-based training reported a composite score of competency (as defined by authors) (Haycock 2010). There was no significant difference between groups. The results are summarised in Summary of findings 2 and Table 2.

Secondary outcomes

2.2 Independent procedure completion (objective measure)

Two trials comparing VR endoscopy simulation training versus conventional patient-based training reported independent procedure completion (Gerson 2003; Haycock 2010). The meta-analysis showed that the VR training group had a significantly lower number of independent procedure completions than the conventional training group (RR 0.45, 95% CI 0.27 to 0.74; 2 trials (n = 174); Analysis 2.1). Heterogeneity was not statistically significant (P = 0.47) and was low (I^2 = 0%). We performed subgroup analyses for the type of endoscopic procedure under study (colonoscopy, sigmoidoscopy). There were no statistically significant differences between groups in the colonoscopy study (RR 0.67, 95% CI 0.20 to 2.23; 1 trial (n = 108); Analysis 2.1). The VR training group had significantly fewer independent procedure completions compared to the conventional training group for the sigmoidoscopy study (RR 0.41, 95% CI 0.23 to 0.72; 1 trial (n = 66); Analysis 2.1). Tests for interaction showed no statistically significant procedure-related heterogeneity (P = 0.47). We downgraded this finding to low quality due to very serious risk of bias. The results are summarised in Summary of findings 2, Table 3, Analysis 2.1, and Figure 11.

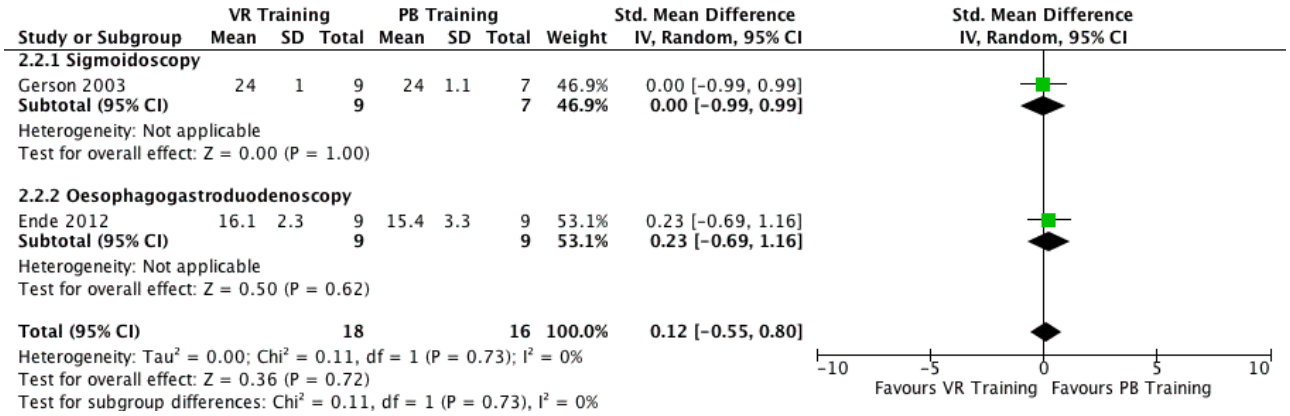
Figure 11. Analysis 2.1. Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 2.1 Independent procedure completion.



2.3 Performance time (objective measure of the time taken to perform the evaluation task(s) post-training)

Four trials comparing VR endoscopy simulation training versus conventional patient-based training reported performance time (time taken to perform the evaluation task(s)) (Ende 2012; Gerson 2003; Haycock 2010; Shirai 2008). We included only two trials in the meta-analysis due to insufficient central tendency and variability data (Ende 2012; Gerson 2003), which showed no significant difference between the VR training group and conventional training group with respect to performance time (SMD 0.12, 95% CI -0.55 to 0.80; 2 trials (n = 34); Analysis 2.2). Heterogeneity was not statistically significant (P = 0.73) and was low (I^2 = 0%). We performed a subgroup analysis for type of endoscopic procedure under study (sigmoidoscopy, oesophagogastroduodenoscopy). There were no statistically significant differences between groups in the sigmoidoscopy study (SMD 0.0 minutes, 95% CI -0.99 to 0.99; 1 trial (n = 16); Analysis 2.2) or the oesophagogastroduodenoscopy study (SMD 0.23 minutes, 95% CI -0.69 to 1.16; 1 trial (n = 18); Analysis 2.2). Tests for interaction showed no statistically significant procedure-related heterogeneity (P = 0.73). Among the remaining two trials reporting this outcome (Haycock 2010; Shirai 2008), there were no significant differences between groups with respect to performance time. We downgraded this finding to very low quality due to very serious risk of bias and serious imprecision. The results are summarised in Summary of findings 2, Table 4, Analysis 2.2, and Figure 12.

Figure 12. Analysis 2.2 Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 2.2 Performance time.



2.4 Complication or critical flaw occurrence

Three trials (72 procedures) comparing VR endoscopy simulation training versus conventional patient-based training reported the occurrence of procedure-related complications or critical flaws (Ende 2012; Gerson 2003; Sedlack 2004a). All three trials reported no complications or critical flaws in any of the study groups. We downgraded this finding to very low quality due to very serious risk of bias and serious imprecision. The results are summarised in Summary of findings 2 and Table 5.

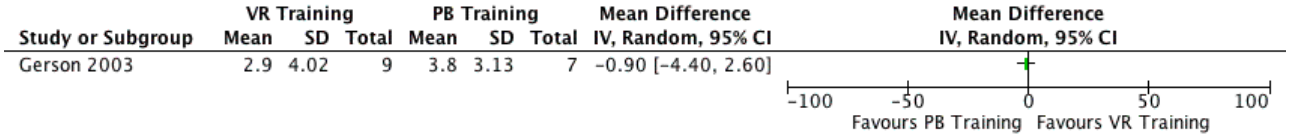
2.5 Patient discomfort (as defined by authors)

Two trials comparing VR endoscopy simulation training versus conventional patient-based training reported patient discomfort (as defined by authors) (Gerson 2003; Sedlack 2004a). We did not perform a meta-analysis as neither trial had sufficient central tendency and variability data. Patient discomfort was statistically significantly lower in the VR training group in one trial (Sedlack 2004a). No significant difference was found between the two groups in the other trial (Gerson 2003). The results are summarised in Summary of findings 2 and Table 6.

2.6 A single measure providing an overall global rating of performance or competency in performing endoscopy (as defined by the authors)

Three trials comparing VR endoscopy simulation training versus conventional patient-based training reported an overall rating of performance or competency as an outcome (Ende 2012; Gerson 2003; Sedlack 2004a). We did not perform a meta-analysis as only one trial had sufficient central tendency and variability data (Gerson 2003). This trial showed statistically significantly fewer positive ratings in the VR training group compared to the conventional training group (MD -0.90, 95% CI -4.40 to 2.60; 1 trial (n = 16); Analysis 2.3) (Gerson 2003). Another trial showed no significant difference between groups (Sedlack 2004a). The third trial showed statistically significantly more positive ratings in the VR plus conventional training group compared to the VR training-only group (Ende 2012), but no significant difference compared to the conventional training-only group. We downgraded this finding to very low quality due to very serious risk of bias and serious imprecision. The results are summarised in Summary of findings 2, Table 7, Analysis 2.3, and Figure 13.

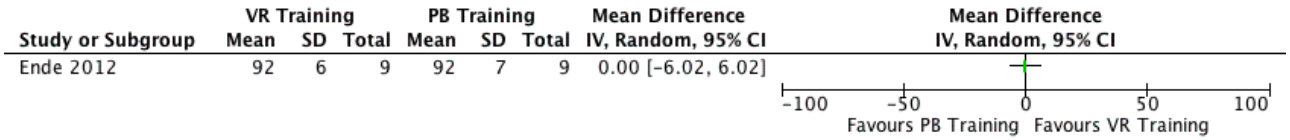
Figure 13. Analysis 2.3 Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 2.3 Overall global rating of performance or competency.



2.7 Visualisation of mucosa (as defined by authors)

Two trials comparing VR endoscopy simulation training versus conventional patient-based training reported visualisation of the mucosa (as defined by the authors) (Ende 2012; Sedlack 2004a). We did not perform a meta-analysis as only one trial had sufficient central tendency and variability data (Ende 2012). This trial showed no significant difference in visualisation between groups (MD 0.0, 95% CI -6.02 to 6.02; 1 trial (n = 18); Analysis 2.4). The other trials also showed no significant difference in visualisation between groups. We downgraded this finding to very low quality due to very serious risk of bias and serious imprecision. The results are summarised in Summary of findings 2, Table 8, Analysis 2.4, and Figure 14.

Figure 14. Analysis 2.4 Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 2.4 Visualisation of mucosa.



3. Virtual reality endoscopy simulation training versus another form of endoscopy simulation

Primary outcome

3.1 Composite score of competency in performing endoscopy (as defined by authors)

One trial comparing VR endoscopy simulation training versus another form of endoscopy simulation reported a composite score of competency (as defined by authors) (Gomez 2015), which showed no significant difference between groups. The results are summarised in Table 2.

Secondary outcomes

3.2 Independent procedure completion (objective measure)

No trials comparing VR endoscopy simulation training versus another form of endoscopy simulation reported this outcome.

3.3. Performance time (objective measure of the time taken to perform the evaluation task(s) post-training)

One trial comparing VR endoscopy simulation training versus another form of endoscopy simulation reported performance time (time taken to perform the evaluation task(s)) (Gomez 2015), with no significant difference in performance time between groups. The results are summarised in Table 4.

3.4 Complication or critical flaw occurrence

No trials comparing VR endoscopy simulation training versus another form of endoscopy simulation reported this outcome.

3.5 Patient discomfort (as defined by authors)

No trials comparing VR endoscopy simulation training versus another form of endoscopy simulation reported this outcome.

3.6 A single measure providing an overall global rating of performance or competency in performing endoscopy (as defined by the authors)

No trials comparing VR endoscopy simulation training versus another form of endoscopy simulation reported this outcome.

3.7 Visualisation of mucosa (as defined by authors)

One trial comparing VR endoscopy simulation training versus another form of endoscopy simulation reported visualisation of the mucosa (as defined by the authors) (Gomez 2015), which showed no significant difference in mucosal visualisation between groups. The results are summarised in Table 8.

4. Two methods of virtual reality simulation training

Primary outcomes

4.1 Composite score of competency in performing endoscopy (as defined by authors)

Two trials comparing two methods of VR simulation training reported a composite score of competency (as defined by authors) (Grover 2015; Grover 2017). Both trials showed a statistically significant increased composite score of competency in the interventional VR training group as compared with the control VR training group. We did not perform a meta-analysis as the studies did not have similar interventions and comparators. Participants in the interventional VR training group in one trial, Grover 2015, received a similar curriculum as the control VR training group in the other trial (Grover 2017). The results are summarised in Table 2.

Secondary outcomes

4.2 Independent procedure completion (objective measure)

No trials comparing two methods of VR simulation training reported this outcome.

4.3 Performance time (objective measure of the time taken to perform the evaluation task(s) post-training)

No trials comparing two methods of VR simulation training reported this outcome.

4.4. Complication or critical flaw occurrence

No trials comparing two methods of VR simulation training reported this outcome.

4.5 Patient discomfort (as defined by authors)

No trials comparing two methods of VR simulation training reported this outcome.

4.6 A single measure providing an overall global rating of performance or competency in performing endoscopy (as defined by the authors)

No trials comparing two methods of VR simulation training reported this outcome.

4.7 Visualisation of mucosa (as defined by authors)

No trials comparing two methods of VR simulation training reported this outcome.

Other reported outcomes

The 18 studies reported a number of other outcomes (e.g. whether analgesic drugs were given (yes/no), number of times manual assistance was required (n), completion of retroflexion (yes/no), ability to recognise pathology (yes/no), ability to insert in a safe manner (1-to-5 Likert scale), and outcomes in the simulated setting). We did not include the data for these outcomes as we

considered them to be of minimal clinical significance and thus did not include them a priori. Additionally, these outcome measures do not have adequate validity evidence.

Sensitivity analysis

We planned a sensitivity analysis a priori including and excluding studies at high or unclear risk of bias. However, we did not perform sensitivity analysis due to the few trials available in each category. We also planned sensitivity analysis including and excluding studies that were only published in abstract form and which required contact with authors to retrieve full methodological details and original outcome data. We did not perform this analysis either as there were no trials published in abstract form for which there was successful retrieval of necessary data from authors.

Funnel plot

Given the heterogeneity of the outcomes reported and the low number of trials reporting similar outcomes across each of our four comparisons, we did not construct a funnel plot to assess for publication bias.

DISCUSSION

Training of new endoscopists has primarily followed the time-honoured concept of 'see one, do one, teach one,' with novices learning basic skills under the supervision of experienced preceptors in the clinical setting. However, over the last two decades there has been an increasing push to incorporate simulation-based instruction into medical training as a means for novices to master basic skills in a low-risk controlled environment prior to performance on real patients. As Vozenilek and colleagues point out, "the concept of 'learning by doing' has become less acceptable, particularly when invasive procedures and high-risk care are required" (Vozenilek 2004).

This systematic review was undertaken to determine whether VR simulation training can supplement and/or replace early conventional endoscopy training (apprenticeship model) in diagnostic oesophagogastroduodenoscopy, colonoscopy, and/or sigmoidoscopy for health professions trainees with limited or no prior endoscopic experience. The results of this review indicate that the use of VR endoscopy training can effectively supplement early conventional endoscopy training (apprenticeship model). However, there is insufficient evidence to advise for or against the use of VR simulation-based training as a replacement for early conventional endoscopy training for health professions trainees with limited or no prior endoscopic experience.

Summary of main results

Eighteen trials with 421 participants met the inclusion criteria.

Virtual reality training versus no training

Ten studies, evaluating oesophagogastroduodenoscopy, colonoscopy, and sigmoidoscopy, compared simulation-based training with no intervention. Virtual reality training compared to no training appears to provide some benefit as measured by our a priori outcomes. Data from one trial showed no statistically significant difference for composite score of competency between the two groups (Park 2007). Pooled data from six studies showed a statistically significant increased number of procedures completed independently among trainees from the VR training group compared to the no-training group, regardless of the procedure under study or prior endoscopy experience (Ahlberg 2005; Di Giulio 2004; McIntosh 2014; Park 2007; Sedlack 2004; Yi 2008). Data from one trial showed a statistically significantly higher overall rating of performance among trainees from the VR training group compared to the no-training group (McIntosh 2014). Data from another trial showed a statistically significantly better visualisation of mucosa among trainees from the VR training group compared to the no-training group (Yi 2008). Pooled data from two trials showed no statistically significant difference between groups with respect to performance time or patient discomfort (McIntosh 2014; Yi 2008). Three trials reported no procedure-related complications of critical flaws in either study group (Ahlberg 2005; Di Giulio 2004; Park 2007). We assessed the quality of the evidence as moderate, low, or very low owing to risk of bias, imprecision, and/or unexplained heterogeneity.

Several trials reporting performance time, patient discomfort, overall global rating of competency, and visualisation of mucosa did not provide sufficient data for quantitative analysis, therefore these outcomes are further discussed qualitatively. Four of the seven trials that reported the outcome of performance time showed that trainees who received VR training were able to complete procedures significantly faster than the no-training group (Ahlberg 2005; Ferlitsch 2010; Tuggy 1998; Yi 2008). Three of the four trials that reported an overall rating of performance or competency showed statistically significantly more positive ratings for VR-trained participants (Cohen 2006; Di Giulio 2004; McIntosh 2014). Finally, all three of the trials that reported mucosal visualisation as an outcome showed that trainees who received simulation-based training had greater visualisation (Sedlack 2004; Tuggy 1998; Yi 2008).

Virtual reality training versus conventional patient-based endoscopy training

Five studies, evaluating oesophagogastroduodenoscopy, colonoscopy, and sigmoidoscopy, compared VR training with conventional patient-based endoscopy training (apprenticeship model). We found no conclusive evidence that VR training provides benefit compared to conventional patient-based endoscopy training. The one trial that reported composite score of competency showed no statistically significant difference in scores in the VR training group compared to the conventional training group (Haycock 2010). Pooled data from two studies showed a statistically significantly lower number of procedures completed independently among trainees from the VR training group compared to the conventional training group (Gerson 2003; Haycock 2010), though this difference was only significant where sigmoidoscopy was the procedure under study (Gerson 2003), and not colonoscopy (Haycock 2010). Pooled data from two trials showed no statistically significant difference between groups with respect to performance time (Ende 2012; Gerson 2003). Three trials reported no procedure-related complications or critical flaws in either study group (Ende 2012; Gerson 2003; Sedlack 2004a). Data from one trial showed no statistically significant difference with respect to overall rating of performance between groups (Gerson 2003). We assessed the quality of the evidence as low or very low owing to risk of bias or imprecision or both.

Several trials reporting performance time, patient discomfort, overall global rating of competency, and visualisation of mucosa did not provide sufficient data for quantitative analysis, therefore

these outcomes are further discussed qualitatively. There was no significant difference between groups as measured by performance time (Ende 2012; Gerson 2003; Haycock 2010; Shirai 2008), procedure-related complication or critical flaw occurrence (Ende 2012; Gerson 2003; Sedlack 2004a), and visualisation of mucosa (Ende 2012; Sedlack 2004a). One of the two studies that reported patient discomfort as an outcome measure found a significant training advantage for the VR group (Sedlack 2004a). One of the three studies that reported an overall global rating of competency found that trainees who received VR training received statistically significantly more negative overall ratings of performance as compared to those receiving conventional patient-based endoscopy training (Gerson 2003). Results from one trial suggest that VR training in combination with conventional training may confer benefit compared to VR training alone with respect to overall global rating of competency (Ende 2012).

Virtual reality training versus another form of endoscopy simulation

One study comparing VR training with another form of endoscopy simulation training found no statistically significant differences between groups with respect to composite score of competency, performance time, or visualisation of mucosa (Gomez 2015). Virtual reality training in combination with another form of endoscopy simulation training did not appear to confer any benefit compared to VR training alone. No other a priori outcomes were reported in this trial.

Two methods of virtual reality training

Two studies evaluating colonoscopy compared two methods of VR training. One trial compared a structured VR endoscopy simulation curriculum to unstructured, self regulated learning on a VR simulator (Grover 2015). Trainees in the structured VR curriculum group had statistically significantly higher composite scores of competency compared to the self regulated group. Another trial compared the same structured VR curriculum to a VR curriculum that applied a progressive learning strategy, whereby trainees completed increasingly difficult cases (Grover 2017). Trainees in the progressive-learning group had statistically significantly higher composite scores of competency compared to the structured curriculum group. Neither trial reported other a priori outcomes. These trials suggest that educational-theory-based strategies, such as structured curricula and progressive learning, can confer benefit and lead to improved outcomes in the clinical setting.

Overall completeness and applicability of evidence

While we included 18 trials assessing the effect of VR simulation-based training, our findings were limited by small sample sizes and considerable variability in outcome measures across studies. In addition, few trials utilised outcomes with adequate validity evidence. We also found considerable insufficiencies with respect to data for meta-analyses. Where quantitative analysis was possible, we downgraded recommendations to moderate, low, or very low due to risk of bias, imprecision, and/or unexplained heterogeneity. Furthermore, the VR training interventions varied considerably between studies, making comparisons difficult. The simulation-based training sessions may not have been intensive or long enough to provide benefit. Tuggy and colleagues examined outcomes after 5 hours and 6 to 10 hours of simulation-based training (Tuggy 1998). However, this trial only demonstrated a training benefit after 6 to 10 hours of simulation-based training, indicating that there may be a minimum length of training required to achieve benefit. In addition, only five studies provided trainees with instruction during the entirety of simulation-based training (Ahlberg 2005; Grover 2015; Grover 2017; Sedlack 2004a). Simply providing trainees with access to simulators does not guarantee that they will be used optimally, as shown by Grover and colleagues (Grover 2015), who compared a structured VR curriculum to self regulated learning on a VR simulator. It is clear from the literature that augmented (extrinsic) feedback and instruction are needed for the acquisition of gastrointestinal endoscopy skills (Grover 2015; Issenberg 2005; Walsh 2009). Mahmood and colleagues (Mahmood 2004), who examined whether novices were able to learn the skill of colonoscopy through the use of a simulator in the absence of structured external feedback, found no improvement in performance on the simulator over successive trials in the absence of augmented feedback. This indicates that extrinsic feedback is essential to facilitate clinical skill acquisition. In addition, in three recent reviews of simulation-based medical education, feedback was identified as a critical feature for effective learning in a simulated setting (Cook 2013; Hatala 2014; Issenberg 2005).

Quality of the evidence

The results of this review should be interpreted with caution. Overall, the methodological quality of included studies was moderate to very low for outcomes for which we could assess the quality of evidence (Summary of findings for the main comparison; Summary of findings 2). We downgraded the quality of evidence mainly for risk of bias. The major sources of bias were inadequate randomisation, lack of allocation concealment or lack of reporting with respect to allocation concealment, lack of assessor blinding, and the use of outcome measures with inadequate validity evidence. Only six trials used adequate methods for randomisation (Cohen 2006; Di Giulio 2004; Grover 2015; Grover 2017; Haycock 2010). Only three trials reported allocation concealment (Ahlberg 2005; Grover 2015; Grover 2017). Assessors were blinded in only 10 trials (Ahlberg 2005; Cohen 2006; Ende 2012; Gomez 2015; Grover 2015; Grover 2017; Haycock 2010; McIntosh 2014; Park 2007; Shirai 2008). Only three studies utilised outcome measures with good validity evidence (Gomez 2015; Grover 2015; Grover 2017). We also downgraded the quality of evidence due to unexplained heterogeneity and imprecision. There were too few trials to permit sensitivity analysis. Based on qualitative findings, however, the relationship between study quality and findings is unclear. While the three studies assessed as at low risk of bias reported largely positive outcomes for the intervention group as compared with the control group, these studies were heterogenous with respect to methodology (Ahlberg 2005; Grover 2015; Grover 2017). We did not assess publication bias as there were too few trials.

Potential biases in the review process

Limitations in study quality, inadequate reporting of methodological detail, sparse data for most outcomes, important inconsistencies across trials, and a high or unclear risk of bias in all but three studies decrease the overall quality of the evidence. Consequently, the conclusions of this review should be interpreted with caution. Variability in the training regimens as well as the timing and definitions of outcome measurements, and the absence of objective measures of performance with strong validity evidence for use in evaluating the competence of clinicians performing endoscopy, would all contribute to inaccuracies in the assessment of the intervention effects.

Agreements and disagreements with other studies or reviews

Four recent reviews have explored VR endoscopy simulation-based training (Dawe 2014; Ekkelenkamp 2016; Qiao 2014; Singh 2014). Our findings are in agreement with the most recent review (Ekkelenkamp 2016), which concluded that the use of VR simulators in early training accelerates the learning of practical skills; however, the results were not overwhelmingly conclusive. Two other reviews concluded that simulation-based training prior to patient-based training is associated with improved performance in clinical practice during the initial stages of learning and patient outcomes as compared to no intervention (Dawe 2014; Singh 2014). A further review reported that VR training is effective for oesophagogastroduodenoscopy, but the data remain limited for colonoscopy (Qiao 2014).

Our review builds on these previous reviews in several respects. First, we have conducted a broad search that includes computer and educational literature databases and conference proceedings. Second, we have included several newer trials that were published since the most recent previous review. Third, we have included only randomised and quasi-randomised trials, rather than observational studies, which are at very serious risk of bias. Finally, we have used the GRADE approach to inform the quality and applicability of our findings and subsequent recommendations.

AUTHORS' CONCLUSIONS

Implications for practice

Despite moderate- to very low-quality evidence, we can conclude that VR training, as compared with no training, generally appears to provide participants with some advantage over their untrained peers as measured by independent procedure completion, overall rating of performance or competency, and mucosal visualisation. Results from this systematic review indicate that VR endoscopy training can be used to effectively supplement early conventional endoscopy training (apprenticeship model) in diagnostic oesophagogastroduodenoscopy, colonoscopy, and/or sigmoidoscopy for health professions trainees with limited or no prior endoscopic experience. Alternatively, we found no conclusive evidence that simulation-based training compared with conventional patient-based endoscopy training (apprenticeship model) provides benefit, although data were limited. Consequently, there is insufficient evidence to advise for or against the use of VR simulation-based training as a replacement for early conventional endoscopy training (apprenticeship model) for health professions trainees with limited or no prior endoscopic experience. There is also insufficient evidence to recommend VR training over another form of endoscopy simulation training. Results from trials comparing two VR curricula suggest that using educational-theory-based approaches such as structured curricula or progressive learning can improve endoscopic performance. As mentioned previously, outcome data are limited, training was of short duration in all trials, and only three studies were at low risk of bias, therefore these results should be interpreted with caution.

Implications for research

Further research is needed to help establish the potential use of VR simulation-based training to supplement and/or replace conventional endoscopy training.

1. Future trials must adhere to strict quality standards such as adequate randomisation and allocation concealment along with the use of measures of performance in endoscopy with strong validity evidence.
2. Randomised trials assessing broader non-technical competencies relevant to the skill of endoscopy, such as communication skills and clinical reasoning, are needed.
3. Future trials should compare the impact of different educational-theory-based endoscopy simulation curricula on the acquisition of endoscopic competence in the clinical setting.
4. Studies comparing the cost of simulation-based training with other forms of training are needed.
5. What is the impact of non-technical skills-specific training in endoscopy on performance in the clinical setting?
6. What are the characteristics of instruction and feedback required to optimise skill transfer to the clinical setting?
7. What is the nature and duration of endoscopy simulation-based training required to optimise skill transfer to the clinical setting?

ACKNOWLEDGEMENTS

Ms Thomasin Adams-Webber (Manager, Library and Archives Services, Learning Institute, the Hospital for Sick Children) for assisting with the electronic search strategy.

Cochrane Colorectal Cancer Editorial office and editors for their support throughout the review process.

REFERENCES

References to studies included in this review

Ahlberg 2005 *{published and unpublished data}*

* 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):1198-204. [DOI: 10.1055/s-2005-921049; PUBMED: 16329017]

Cohen 2006 *{published data only}*

* Cohen J, Cohen SA, Vora KC, Xue X, Burdick JS, Bank S, et al. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointestinal Endoscopy* 2006;**64**(3):361-8. [DOI: 10.1016/j.gie.2005.11.062; PUBMED: 16923483]

Di Giulio 2004 *{published data only}*

* Di Giulio E, Fregonese D, Casetti T, Cestari R, Chilovi F, D'Ambra G, et al. Training with a computer-based simulator achieves basic manual skills required for upper endoscopy: a randomized controlled trial. *Gastrointestinal Endoscopy* 2004;**60**(2):196-200. [PUBMED: 15278044]

Ende 2012 *{published data only}*

Ende A, Zopf Y, Konturek P, Naegel A, Hahn EG, Matthes K, et al. Strategies for training in diagnostic upper endoscopy: a prospective, randomized trial. *Gastrointestestinal Endoscopy* 2012;**75**(2):254-60. [DOI: 10.1016/j.gie.2011.07.063; PUBMED: 22153875]

Ferlitsch 2010 *{published data only}*

* Ferlitsch A, Schoefl R, Puespoek A, Miehsler W, Schoeniger-Hekele M, Hofer H, et al. Effect of virtual endoscopy simulator training on performance of upper gastrointestinal endoscopy in patients: a randomized controlled trial. *Endoscopy* 2010;**42**(12):1049-56. [DOI: 10.1055/s-0030-1255818; PUBMED: 20972956]

Gerson 2003 *{published data only}*

* Gerson LB, Van Dam J. A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. *Endoscopy* 2003;**35**(7):569-75. [DOI: 10.1055/s-2003-40243; PUBMED: 12822091]

Gomez 2015 *{published data only}*

Gomez PP, Willis RE, Van Sickle K. Evaluation of two flexible colonoscopy simulators and transfer of skills into clinical practice. *Journal of Surgical Education* 2015;**72**(2):220-7. [DOI: 10.1016/j.jsurg.2014.08.010; PUBMED: 25239553]

Grover 2015 *{published data only}*

Grover SC, Garg A, Scaffidi MA, Yu JJ, Plener IS, Yong E, et al. Impact of a simulation training curriculum on technical and nontechnical skills in colonoscopy: a randomized trial. *Gastrointestinal Endoscopy* 2015;**82**(6):1072-9. [DOI: 10.1016/j.gie.2015.04.008; PUBMED: 26007221]

Grover 2017 *{published data only}*

Grover SC, Scaffidi MA, Khan R, Garg A, Al-Mazroui A, Alomani T, et al. Progressive learning in endoscopy simulation training improves clinical performance: A blinded randomized trial. *Gastrointestinal Endoscopy* 2017;**85**(5):881-9. [DOI: 10.1016/j.gie.2017.03.1529; PUBMED: 28366440]

Haycock 2010 *{published and unpublished data}*

* Haycock A, Koch AD, Familiari P, van Delft F, Dekker E, Petruzziello L, et al. Training and transfer of colonoscopy skills: a multinational, randomized, blinded, controlled trial of simulator versus bedside training. *Gastrointestinal Endoscopy* 2010;**71**(2):298-307. [DOI: 10.1016/j.gie.2009.07.017; PUBMED: 19889408]

McIntosh 2014 *{published data only}*

McIntosh KS, Gregor JC, Khanna NV. Computer-based virtual reality colonoscopy simulation improves patient-based colonoscopy performance. *Canadian Journal of Gastroenterology and Hepatology* 2014;**28**(4):203-6. [PUBMED: 24729994]

Park 2007 *{published data only}*

Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, et al. Randomized controlled trial of virtual reality simulator training: transfer to live patients. *American Journal of Surgery* 2007;**194**(2):205-11. [DOI: 10.1016/j.amjsurg.2006.11.032; PUBMED: 17618805]

Sedlack 2004 *{published data only}*

* Sedlack RE, 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):33-7. [PUBMED: 14687137]

Sedlack 2004a *{published data only}*

* Sedlack RE, Kolars JC, Alexander JA. Computer simulation training enhances patient comfort during endoscopy. *Clinical Gastroenterology and Hepatology* 2004;**2**(4):348-52. [PUBMED: 15067632]

Sedlack 2007 *{published data only}*

* Sedlack RE. Validation of computer simulation training for esophagogastroduodenoscopy: pilot study. *Journal of Gastroenterology and Hepatology* 2007;**22**(8):1214-9. [DOI: 10.1111/j.1440-1746.2007.04841.x; PUBMED: 17559386]

Shirai 2008 *{published data only}*

* Shirai Y, Yoshida T, Shiraishi R, Okamoto T, Nakamura H, Harada T, et al. Prospective randomized study on the use of a computer-based endoscopic simulator for training in esophagogastroduodenoscopy. *Journal of Gastroenterology and Hepatology* 2008;**23**(7 Pt 1):1046-50. [DOI: 10.1111/j.1440-1746.2008.05457.x; PUBMED: 18554236]

Tuggy 1998 *{published data only}*

* Tuggy ML. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *Journal of the American Board of Family Practice* 1998;**11**(6):426-33. [PUBMED: 9875997]

Yi 2008 *{published data only}*

* Yi SY, Ryu KH, Na YJ, Woo HS, Ahn W, Kim WS, et al. Improvement of colonoscopy skills through simulation-based training. *Studies in Health Technology and Informatics* 2008;**132**:565-7. [PUBMED: 18391369]

References to studies excluded from this review

Ahad 2011 *{published data only}*

Ahad S, Advani V, Boehler ML, Schwind C, Hassan I. The impact of simulator fidelity on colonoscopic skill acquisition. A randomized trial between high and low fidelity colonoscopic simulators. *Journal of the American College of Surgeons* 2011;**1**:S127-8.

Ahad 2013 *{published data only}*

Ahad S, Boehler M, Schwind CJ, Hassan I. The effect of model fidelity on colonoscopic skills acquisition. A randomized controlled study. *Journal of Surgical Education* 2013;**70**(4):522-7.

Ahn 2016 *{published data only}*

Ahn JY, Lee JS, Lee GH, Lee JW, Na HK, Jung KW, et al. The efficacy of a newly designed, easy-to-manufacture training simulator for endoscopic biopsy of the stomach. *Gut and Liver* 2016;**10**(5):764-72.

Ansell 2013 *{published data only}*

Ansell J, Arnaoutakis K, Goddard S, Hawkes N, Leicester R, Dolwani S, et al. The WIMAT colonoscopy suitcase model: a novel porcine polypectomy trainer. *Colorectal Disease* 2013;**15**(2):217-23.

Bai 2011 *{published data only}*

Bai Y, Zhi F, Du Q, Liu S, Zhang Q, Pan D, et al. Optimization study of virtual reality simulator training methods for colonoscopy. [Chinese]. *Chinese Journal of Gastroenterology* 2011;**16**(6):345-7.

Bai 2012 *{published data only}*

Bai Y, Zhi FC, Liu SD, Chen CL, Pan DS, Du XF, et al. Control study on colonoscopy skills acquiring from endoscopic simulation system transferring to patients. *National Medical Journal of China* 2012;**92**(18):1285-7.

Carot 2015 *{published data only}*

Carot L, Hernandez C, Balaguer F, Alvarez C, Lanas A, Cubiella J, et al. Rate of detection of serrated lesions in proximal colon by simulated sigmoidoscopy. *United European Gastroenterology Journal* 2015;**3**(5S):A626-7 (Abstract P1648).

Carot 2016 *{published data only}*

Carot L, Castells A, Hernandez C, Alvarez-Urturi C, Balaguer F, Lanas A, et al. Rate of detection of serrated lesions in proximal colon by simulated sigmoidoscopy: comparison with colonoscopy and faecal immunochemical testing in a multicentre pragmatic, randomised controlled trial. *Gastroenterology* 2016;**1**:S750-1.

Castells 2014 *{published data only}*

* Castells A, Quintero A, Alvarez E, Bujanda C, Cubiella L, Salas J, et al. Rate of detection of advanced neoplasms in proximal colon by simulated sigmoidoscopy vs fecal immunochemical tests. Clinical Gastroenterology and Hepatology 2014; Vol. 12, issue 10:1708-16.e4.

Ekkelenkamp 2016 *{published data only}*

Ekkelenkamp VE, Koch AD, de Man RA, Kuipers EJ. Training and competence assessment in GI endoscopy: a systematic review. *Gut* 2016;**65**(4):607-15.

Elvevi 2012 *{published data only}*

Elvevi A, Cantu P, Maconi G, Conte D, Penagini R. Evaluation of hands-on training in colonoscopy: is a computer-based simulator useful?. *Digestive & Liver Disease* 2012;**44**(7):580-4.

Grover 2016 *{published data only}*

Grover S, Scaffidi M, Chana B, Gupta K, Zasowski M, Zarghom O, et al. A virtual reality curriculum in non-technical skills improves performance in colonoscopy: a randomized trial. *Canadian Journal of Gastroenterology and Hepatology* 2016;**4792898**:8-9 (Abstract A10).

Hritz 2013 *{published data only}*

Hritz I, Dubravcsik Z, Szepes A, Szepes Z, Kruglikova I, Funch-Jensen P, et al. Assessment of the effectiveness of ERCP mechanical simulator (EMS) exercise on trainees' ERCP performance in the initial learning period: multicenter randomized controlled trial. *United European Gastroenterology Journal* 2013;**1**(1S):A333 (Abstract P751).

Jirapinyo 2014 *{published data only}*

Jirapinyo P, Bing V, Kumar N, Ryan MB, Aihara H, Imaeda AB, et al. A randomized trial of endoscopic simulator training in first year gastroenterology fellows. *Gastrointestinal Endoscopy* 2014;**79**(5S):AB218 (Abstract Su1571).

Jirapinyo 2015 *{published data only}*

Jirapinyo P, Kumar N, Tintara S, Bing V, Aihara H, Perencevich M, et al. Endoscopic part-task simulator training improves endoscopic performance in gastroenterology fellows. *Gastroenterology* 2015;**148**(4 Suppl 1):S-202 (Abstract Sa1032).

Jun 2013 *{published data only}*

Jun W. Upper gastrointestinal endoscopy training with a computer-based simulator. *Journal of Gastroenterology and Hepatology* 2013;**28**(Suppl 3):725 (Abstract PR0087).

Kaltenbach 2011 *{published data only}*

Kaltenbach T, Leung C, Wu K, Yan K, Friedland S, Soetikno R. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. *Digestive Diseases and Sciences* 2011;**56**(5):1496-502.

Koch 2015 *{published data only}*

Koch AD, Ekkelenkamp VE, Haringsma J, Schoon EJ, de Man RA, Kuipers EJ. Simulated colonoscopy training leads to improved performance during patient-based assessment. *Gastrointestinal Endoscopy* 2015;**81**(3):630-6.

Li 2012 *{published data only}*

Li Z, Xu AG, Ma QY, Li BS, Du QF, Liu SD, et al. Effect of mental imagery rehearsal on gastroscopy training with virtual reality endoscopic simulator. *World Chinese Journal of Digestology* 2012;**20**(24):2276-80.

Liao 2013 *{published data only}*

Liao WC, Leung JW, Wang HP, Chang WH, Chu CH, Lin JT, et al. Coached practice using ERCP mechanical simulator improves trainees' ERCP performance: a randomized controlled trial. *Endoscopy* 2013;**45**(10):799-805.

Lim 2011 *{published data only}*

Lim BS, Leung JW, Lee J, Yen D, Beckett L, Tancredi D, et al. Effect of ERCP mechanical simulator (EMS) practice on trainees' ERCP performance in the early learning period: US multicenter randomized controlled trial. *American Journal of Gastroenterology* 2011;**106**(2):300-6.

Meng 2016 *{published data only}*

Meng W, Leung JW, Yue P, Wang Z, Wang X, Wang H, et al. Simulation practice with ERCP mechanical simulator (EMS) improves basic skills of novice surgical trainees - a progress report. *Journal of Gastroenterology and Hepatology* 2016;**31**:321-2.

NCT01405443 *{unpublished data only}*

NCT01405443. Simulator training for gastrointestinal endoscopy [Simulator training for gastrointestinal endoscopy - how much simulator training is required to acquire proficiency in gastrointestinal endoscopy]. clinicaltrials.gov/ct2/show/NCT01405443 (first received 29 July 2011).

Nehme 2013 *{published data only}*

Nehme J, Sodergren MH, Sugden C, Aggarwal R, Gillen S, Feussner H, et al. A randomized controlled trial evaluating endoscopic and laparoscopic training in skills transfer for novices performing a simulated NOTES task. *Surgical Innovation* 2013;**20**(6):631-8.

Plooy 2016 *{published data only}*

Plooy AM, Hill A, Horswill MS, Cresp ASG, Karamatic R, Riek S, et al. The efficacy of training insertion skill on a physical model colonoscopy simulator. *Endoscopy International Open* 2016;**4**(12):E1252-60.

Qiao 2014 *{published data only}*

Qiao W, Bai Y, Lv R, Zhang W, Chen Y, Lei S, et al. The effect of virtual endoscopy simulator training on novices: a systematic review. *PLoS ONE* 2014;**9**(2):e89224.

Santos 2017 *{published data only}*

Santos N, Carter J, He F, Linsk A, Lungarini A, Nemani A, et al. A learning curve study using the Virtual Translumenal Endoscopic Surgery Trainer (VTEST). *Surgical Endoscopy and Other Interventional Techniques* 2017;**31**(1S):S217 (Abstract P306).

Scaffidi 2018 *{published data only}*

Scaffidi MA, Al Mazroui A, Lin P, Kalaichandran R, Lyn R, Walsh CM, et al. A45 Impact of an ergonomic intervention on simulated colonoscopy performance. *Canadian Journal of Gastroenterology and Hepatology* 2018;**1**(Suppl 1):78.

Seshadri 2014 *{published data only}*

Seshadri D, Barkel D, Riggs T, Wasvary H. Endoscopic simulation training for colonoscopy. *Diseases of the Colon and Rectum* 2014;**57**(5):e340 (Abstract P369).

Singh 2014 *{published data only}*

Singh S, Sedlack RE, Cook DA. Effects of simulation-based training in gastrointestinal endoscopy: a systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology* 2014;**12**(10):1611-23.e4.

Snyder 2011 *{published data only}*

Snyder CW, Vandromme MJ, Tyra SL, Porterfield JR, Clements RH, Hawn MT. Effects of virtual reality simulator training method and observational learning on surgical performance. *World Journal of Surgery* 2011;**35**(2):245-52.

Strosberg 2017 *{published data only}*

Strosberg DS, Osayi SN, Drosdeck J, Dettorre R, Suzo A, Hazey J. Virtual reality simulation in flexible endoscopy: implications for resident training. *Surgical Endoscopy and Other Interventional Techniques* 2017;**31**(S1):S207 (Abstract P273).

Van Sickle 2011 *{published data only}*

Van Sickle KR, Buck L, Willis R, Mangram A, Truitt MS, Shabahang M, et al. A multicenter, simulation-based skills training collaborative using shared GI Mentor II systems: results from the Texas Association of Surgical Skills Laboratories (TASSL) flexible endoscopy curriculum. *Surgical Endoscopy* 2011;**25**(9):2980-6.

Williams 2015 *{published data only}*

Williams MR, Crossett JR, Cleveland EM, Smoot CP, Aluka KJ, Coviello LC, et al. Equivalence in colonoscopy results between gastroenterologists and general surgery residents following an endoscopy simulation curriculum. *Journal of Surgical Education* 2015;**72**(4):654-7.

References to ongoing studies

Grover 2017a *{published data only}*

Grover SC, Scaffidi MA, Khan R, Chana B, Iqbal S, Lin PC, et al. A virtual reality curriculum in non-technical skills improves colonoscopic performance: a randomized trial. *Gastrointestinal Endoscopy* 2017;**85**(5S):AB181.

Additional references

ASGE 2011

ASGE Standards of Practice Committee. Complications of colonoscopy. *Gastrointestinal Endoscopy* 2011;**74**(4):745-52.

Bar-Meir 2000

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

Barton 2008

Barton R. Validity and reliability of an accreditation assessment for colonoscopy [abstract]. *Gut* 2008;**57**(Suppl 1):A2.

Barton 2012

Barton JR, Corbett S, van der Vleuten CP, for the English Bowel Cancer Screening Programme and UK Joint Advisory Group for Gastrointestinal Endoscopy. The validity and reliability of a direct observation of procedural skills assessment tool: assessing colonoscopic skills of senior endoscopists. *Gastrointestinal Endoscopy* 2012;**75**:591-7.

Blumenthal 1994

Blumenthal D. Making medical errors into "medical treasures''. *JAMA* 1994;**272**(23):1867-8.

Brydges 2010

Brydges R, Carnahan H, Rose D, Rose L, Dubrowski A. Coordinating progressive levels of simulation fidelity to maximize educational benefit. *Academic Medicine* 2010;**85**:806-12.

Brydges 2014

Brydges R, Hatala R, Zendejas B, Erwin PJ, Cook DA. Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Academic Medicine* 2015;**90**(2):246-56.

Brydges 2015

Brydges R, Manzone J, Shanks D, Hatala R, Hamstra SJ, Zendejas B, et al. Self-regulated learning in simulation-based training: a systematic review and meta-analysis. *Medical Education* 2015;**49**(4):368-78.

Cass 1996

Cass OW, Freeman ML, Cohen J, Zuckerman G, Watkins J, Nord J, et al. Acquisition of competency in endoscopic skills (ACES) during training: a multicentre study [abstract]. *Gastrointestinal Endoscopy* 1996;**43**:308.

Chen 2001

Chen G, Gully SM, Eden D. Validation of a new general self-efficacy scale. *Organizational Research Methods* 2001;**4**(1):62-83.

Classen 1974

Classen M, Rupin H. Practical endoscopy training using a new gastrointestinal phantom. *Endoscopy* 1974;**6**(2):127-31.

Cook 2013

Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Academic Medicine* 2013;**88**(6):872-83.

Dawe 2014

Dawe SR, Windsor JA, Broeders JA, Cregan PC, Hewett PJ, Maddern GJ. A systematic review of surgical skills transfer after simulation-based training: laparoscopic cholecystectomy and endoscopy. *Annals of Surgery* 2014;**259**(2):236-48.

Dedy 2015

Dedy NJ, Szasz P, Louridas M, Bonrath EM, Husslein H, Grantcharov TP. Objective structured assessment of nontechnical skills: reliability of a global rating scale for the in-training assessment in the operating room. *Surgery* 2015;**157**(6):1002-13.

Dunkin 2003

Dunkin BJ. Flexible endoscopy simulators. *Seminars in Laparoscopic Surgery* 2003;**10**(1):29-35.

Dunkin 2007

Dunkin B, Adrales GL, Apelgren K, Mellinger JD. Surgical simulation: a current review. *Surgical Endoscopy* 2007;**21**(3):357-66.

Egger 1997

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**(7109):629-34.

Ellaway 2006

Ellaway R, Candler C, Greene P, Smothers V. An architectural model for MedBiquitous Virtual Patients: MedBiquitous white paper. 2016. groups.medbiq.org/medbiq/display/VPWG/MedBiquitous+Virtual+Patient+Architecture (accessed 10 March 2018).

Endnote 2016 [Computer program]

Clarivate Analytics. Endnote X8. Version 8.1. Philadelphia: Clarivate Analytics, 2016.

Faigel 2005

Faigel DO, Baron TH, Lewis B, Petersen B, Petrini J, Popp JW, et al. for the ASGE Taskforce on Ensuring Competence in Endoscopy and the American College of Gastroenterology Executive and Practice Management Committees. Ensuring competence in endoscopy. 2005. s3.gi.org/physicians/EnsuringCompetence.pdf (accessed 10 December 2017).

Frank 2015

Frank JR, Snell L, Sherbino J, editors. The CanMEDS 2015 Physician Competency Framework. Ottawa: The Royal College of Physicians and Surgeons of Canada, 2015.

GRADEpro 2017 [Computer program]

McMaster University (developed by Evidence Prime). GRADEpro GDT. Version accessed 2 December 2017. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015.

Grantcharov 2003

Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *American Journal of Surgery* 2003;**185**(2):146-9.

Guadagnoli 2012

Guadagnoli M, Morin MP, Dubrowski A. The application of the challenge point framework in medical education. *Medical Education* 2012;**46**(5):447-53.

Hatala 2005

Hatala R, Kassen BO, Nishikawa J, Cole G, Issenberg SB. Incorporating simulation technology in a Canadian internal medicine specialty examination: a descriptive report. *Academic Medicine* 2005;**80**(6):554-6.

Hatala 2014

Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. *Advances in Health Sciences Education* 2014;**19**(2):251-72.

Higgins 2011

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2016

Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological Expectations of Cochrane Intervention Reviews.. Methodological Expectations of Cochrane Intervention Reviews. Cochrane: London, 2016.

Hochberger 2004

Hochberger J, Euler K, Naegel A, Hahn E G, Maiss J. The compact Erlangen Active Simulator for interventional endoscopy: a prospective comparison in structured team-training courses on "endoscopic hemostasis" for doctors and nurses to the "Endo-Trainer" model. *Scandinavian Journal of Gastroenterology* 2004;**39**(9):895-902.

Hodges 2003

Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Medical Education* 2003;**37**(11):1012-6.

Holmboe 2010

Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Medical Teacher* 2010;**32**(8):676-82.

Issenberg 1999

Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, et al. Simulation technology for health care professional skills training and assessment. *JAMA* 1999;**282**(9):861-6.

Issenberg 2005

Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 2005;**27**(1):10-28.

JAG Central Office 2010

Joint Advisory Group Central Office. Summative DOPS assessment form – colonoscopy and flexible sigmoidoscopy. www.thejag.org.uk/Downloads/DOPS%20Forms%20For%20International%20and%20reference%20use%20only/Summative%20DOPS_Colonoscopy%20and%20Flexible%20sigmoidoscopy.pdf (accessed 10 December 2017).

Kim 2001

Kim JH, Park S, Lee H, Yuk KC, Lee H. Virtual reality simulations in physics education. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning 2001; Vol. 3, issue 2.

Kneebone 2001

Kneebone R, ApSimon D. Surgical skills training: simulation and multimedia combined. *Medical Education* 2001;**35**(9):909-15.

Kononowicz 2016

Kononowicz AA, Woodham L, Georg C, Edelbring S, Stathakarou N, Davies D, et al. Virtual patient simulations for health professional education. *Cochrane Database of Systematic Reviews* 2016, Issue 5. [DOI: 10.1002/14651858.CD012194]

Krummel 1998

Krummel TM. Surgical simulation and virtual reality: the coming revolution. *Annals of Surgery* 1998;**228**(5):635-7.

Langsley 1991

Langsley DG. Medical competence and performance assessment. A new era. *JAMA* 1991;**266**(7):977-80.

LeBlanc 2009

LeBlanc VR, Tabak D, Kneebone R, Nestel D, MacRae H, Moulton CA. Psychometric properties of an integrated assessment of technical and communication skills. *American Journal of Surgery* 2009;**197**(1):96-101.

Macaskill 2001

Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 2001;**20**(4):641-54.

Mahmood 2004

Mahmood T, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surgical Endoscopy* 2004;**18**(8):1224-30.

Matharoo 2017

Matharoo M, Haycock A, Sevdalis N, Thomas-Gibson S. A prospective study of patient safety incidents in gastrointestinal endoscopy. *Endoscopy International Open* 2017;**5**(1):E83-9.

McCashland 2000

McCashland T, Brand R, Lyden E, de Garmo P. The time and financial impact of training fellows in endoscopy. CORI Research Project. Clinical Outcomes Research Initiative. *American Journal of Gastroenterology* 2000;**95**(11):3129-32.

Miller 1990

Miller GE. The assessment of clinical skills/competence/performance. *Academic Medicine* 1990;**65**(9 Suppl):S63-7.

Murad 2010

Murad MH, Coto-Yglesias F, Varkey P, Prokop LJ, Murad AL. The effectiveness of self-directed learning in health professions education: a systematic review. *Medical Education* 2010;**44**:1057-68.

Palter 2013

Palter VN, Orzech N, Reznick RK, Grantcharov TP. Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery: a randomized controlled trial. *Annals of Surgery* 201;**257**:224-30.

Rasmussen 2003

Rasmussen J. The role of error in organizing behaviour. 1990. *Quality and Safety in Health Care* 2003;**12**(5):377-83.

RevMan 2014 [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rostom 2013

Rostom A, Ross ED, Dubé C, Rutter MD, Lee T, Valori R, et al. Development and validation of a nurse-assessed patient comfort score for colonoscopy.. *Gastrointestinal Endoscopy* 2013;**77**(2):255-61.

Scalese 2008

Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. *Journal of General Internal Medicine* 2008;**23**(Suppl 1):46-9.

Schünemann 2013

Schünemann H, Brozek J, Oxman A, editor(s). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach (updated October 2013). GRADE Working Group, 2013. Available from gdt.guidelinedevelopment.org/app/handbook/handbook.html. The GRADE Working Group.

Sedlack 2002

Sedlack RE. Development of a colonoscopy curriculum and performance based assessment criteria on a computer-based endoscopy simulator. *Academic Medicine* 2002;**77**(7):750-1.

Sterne 2011

Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting bias. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Sturm 2007

Sturm L, Windsor J, Cregan P, Hewett P, Cosman P, Maddern G. Surgical simulation training: skills transfer to the operating room. ASERNIP-S Report No. 61. Adelaide, South Australia: ASERNIP-S, 2007. www.surgeons.org/media/300327/Surgicalsimulation_systematicreview.pdf (accessed 10 December 2017).

Swing 2002

Swing SR. Assessing the ACGME general competencies: general considerations and assessment methods. *Academic Emergency Medicine* 2002;**9**(11):1278-88.

Vassiliou 2010

Vassiliou MC, Kaneva PA, Poulose BK, Dunkin BJ, Marks JM, Sadik R, et al. Global assessment of gastrointestinal endoscopic skills (GAGES): a valid measurement tool for technical skills in flexible endoscopy. *Surgical Endoscopy* 2010;**24**(8):1834-41.

Vozenilek 2004

Vozenilek J, Huff JS, Reznek M, Gordon JA. See one, do one, teach one: advanced technology in medical education. *Academic Emergency Medicine* 2004;**11**(11):1149-54.

Walsh 2008

Walsh CM, Coopper MA, Rabeneck L, Carnahan H. Bench-top versus virtual reality simulation training in endoscopy: expertise discrimination. *Canadian Journal Gastroenterology and Hepatology* 2008;**22**(Suppl A):164.

Walsh 2009

Walsh CM, Ling SC, Wang CS, Carnahan H. Concurrent versus terminal feedback: it may be better to wait. *Academic Medicine* 2009;**84**(10 Suppl):S54-7.

Walsh 2016

Walsh CM. In-training gastrointestinal endoscopy competency assessment tools: types of tools, validation and impact. *Best Practice and Research: Clinical Gastroenterology* 2016;**30**(3):357-74.

WHO 2008

World Health Organization. Classifying health workers: Mapping occupations to the international standard classification. www.who.int/hrh/statistics/Health_workers_classification.pdf?ua (accessed prior to 30 July 2018).

WHO 2013

World Health Organization. Transforming and scaling up health professionals' education and training: World Health Organization Education Guidelines 2013. www.who.int/hrh/resources/transf_scaling_hpet/en/ (accessed 20 March 2018).

Ziv 2003

Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Academic Medicine* 2003;**78**(8):783-8.

References to other published versions of this review

Walsh 2012

Walsh CM, Sherlock ME, Ling SC, Carnahan H. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. *Cochrane Database of Systematic Reviews* 2012, Issue 6. [DOI: 10.1002/14651858.CD008237.pub2]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Ahlberg 2005

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Multicentre (8). **Year(s) of conduct of trial:** Not stated. **Generation of the allocation sequence:** Blinded random draw of numbers contained within sealed envelopes. **Allocation concealment:** Adequate (sealed envelope). **Blinding of assessors:** Adequate (physician assessors and participants blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** Sweden. **Year(s) participants randomised:** Not stated. **Number:** 12 randomised and analysed. **Inclusion criteria:** Surgical and gastroenterology residents (postgraduate years 2 to 5) with experience in EGD (minimum of 20 individually performed procedures) who were designated to start colonoscopy training. **Exclusion criteria:** Prior experience in colonoscopy (performing or assisting). **Health profession:** Medical trainees (surgery residents (n = 10) and gastroenterology fellows (n = 2)). **Level of training:** Postgraduate years 2 to 5. **Endoscopy experience:** Minimum of 20 individually performed EGDs. **Sex:** 10 male, 2 female. **Age:** Not stated.
Interventions	**Learning theory:** None stated. Prior to undergoing the training task, all participants were given the same theoretical study material, containing a booklet on colonoscopy together with a free sample instructive CD on colonoscopy (New Technology and Technique by Williams, Way, and Sakai). Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 6)** 1. **VR simulator:** Simulator: AccuTouch virtual reality endoscopy simulator version 1.3 (Immersion Medical, Inc., Gaithersburg, Maryland, USA). 2. **Duration of training and/or training endpoint:** Participants practiced until predefined expert level of performance reached (see below).

Ahlberg 2005 *(Continued)*

3. **Description of intervention:** Participants practiced "under strict supervision" on the simulator for a median time of 20 hours (range 15 to 25) during 1- to 2-hour sessions, over at least 4 days. All patient cases in the introduction, biopsy, and polypectomy modules were used. Participants practiced until a predefined expert level of performance was reached on an examination case (case 6 in the introductory series). Expert level of performance was defined as:
   a. ability to intubate the caecum within 7 minutes without the use of sedation, a "virtual attending", simulation tips, and external view. The use of assistance tools (e.g. abdominal pressure, shifting patient position) were allowed;
   b. More than 97% of the procedure time without patient discomfort and no period of severe or extreme discomfort;
   c. navigation to the caecum with less than 1500 mL of air insufflated; and
   d. navigation to the caecum with less than 15% of procedure time being in "red-out."
   - Expert level of performance was defined by assessing 5 experienced endoscopists (> 1000 procedures each) and calculating the mean performance quality parameters on case 6 in the introductory section from all experts after a period of familiarisation with the simulator. Participants could attempt the examination case (case 6 in the introductory section) at any time, but they had to fulfil all parameters in the expert criterion in order to pass.
4. **Observation, instruction, and feedback:** Participants practiced on the simulator "under strict supervision." Feedback was given to the trainee after each completed trial and at any given time comparison with expert level of performance could be made. A safe technique for manoeuvring the scope was taught. Use of the instructional aides from the simulator (e.g. sedation, "virtual attending", simulation tips, external view, "find scope tip", shifting position of patient, and assistance with local pressure) were allowed during practice. It was not stated whether participants had access to the performance quality parameters generated by the simulator during practice.

GROUP 2: No intervention (n = 6)

1. **Description of intervention:** No intervention.
2. **Observation, instruction, and feedback:** None.

Outcomes

Time to assessment: After completion of training, participants in the simulator-trained group began their individual colonoscopies within 1 week. Participants in the control group started after studying the theoretical material.

Assessment model: 10 colonoscopies were completed (maximum 60 minutes overall procedure time and/or maximum 15 minutes per segment: rectosigmoid angle, sigmoid colon sigmoid-descending colon junction, descending colon, left flexure, transverse colon, right flexure, ascending colon, caecum) under the supervision and evaluation of a blinded supervisor who was instructed not to guide the participant.

Details of patients used for live assessment: All patients, without a history of previous abdominal surgery, designated to undergo diagnostic colonoscopy.

Outcome measures:

1. Time to reach caecum (min) or total procedure time in unsuccessful cases (min)
2. Completed procedure rate (intubation of caecum within given time limits) (n)
3. Segment of colon where procedure was stopped (9 consecutive segments: rectosigmoid angle, sigmoid colon, sigmoid-descending colon junction, descending colon, left flexure, transverse colon, right flexure, ascending colon, caecum)
4. Reason for stopping (if applicable)
5. Analgesic drugs given (yes/no)
6. Complications (n)
7. Maximum discomfort (rated by patient, visual analogue scale)

Notes

Funding: Not stated.

Declarations of conflicts of interest for primary investigators: None stated.

Ahlberg 2005 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Blinded random draw of numbers contained within sealed envelopes. Quote: "...a series of envelopes in a numbered sequence and with every second designated to training. Envelopes were drawn in a blinded fashion when each trainee was randomised." (personal correspondence)
Allocation concealment (selection bias)	Low risk	Adequate: Sealed envelopes. Quote: "...using the sealed envelope method."
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind resident participants due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians and patients were blinded to residents training method. Quote: "The patients were blinded concerning the pupils training status." Quote: "The supervisors were blinded concerning the pupils training status."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: Accounted for missing outcome data from the 1 procedure in the control group that was not analysed. Quote: “One procedure in the control group series was excluded because of poor bowel preparation” and “in one patient examined in the trained group series, an obstructive tumour was found in the transverse colon; this procedure was registered as successful.”
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Cohen 2006

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Multicentre (16). **Year(s) of conduct of trial:** Not stated (2 years). **Generation of the allocation sequence:** Random-number table. **Allocation concealment:** Not stated. **Blinding of assessors:** Adequate (physician assessors blinded).

Cohen 2006 *(Continued)*

Inclusion of all randomised participants: (45/49) 91.84%.

Sample size calculation: None.

Intention-to-treat analysis: Not stated.

Participants	**Country:** USA. **Year(s) participants randomised:** Not stated. **Number:** 45 analysed (49 randomised, but 4 participants withdrew after randomisation because of protocol violations during the training phase). **Inclusion criteria:** First-year gastroenterology fellows starting fellowship at teaching institutions in the New York metropolitan area over 2 years whose training director agreed to adhere to the protocol and to delay any performance of colonoscopy for the first 8 weeks of the fellowship. **Exclusion criteria:** Previous formal training in colonoscopy (> 10 cases) and an inability to comply with the training schedule. **Health profession:** Medical trainees (gastroenterology fellows). **Level of training:** First-year fellows. **Endoscopy experience (average number of procedures):** 1. VR simulator training group: 67 EGDs and 4 sigmoidoscopies. 2. No intervention group: 80 EGDs and 5 sigmoidoscopies. **Sex:** Not stated. **Age:** Not stated.
Interventions	**Learning theory:** None stated. Prior to undergoing the training task, all participants attended general lectures on colonoscopy as part of a didactic endoscopy course given to all incoming fellows, which emphasised key principles, such as application of torque, reduction of loops, and careful examination of pathology during scope withdrawal. Participants were randomly assigned to two groups: **GROUP 1: VR simulator training (n = 22)** 1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA Corp., Cleveland, OH, UA 2. **Duration of training and/or training endpoint:** 10 hours over 8 weeks (5, 2-hour private simulator sessions). 3. **Description of intervention:** Received supervised orientation to the simulator during the first week of fellowship. Over the next 8 weeks, fellows had 5, 2-hour private simulator training sessions. Each hour of training followed a standard protocol of activities (warm-up hand-eye co-ordination exercises and performance of 2 specific simulated procedures each hour). In total, 10 different cases were used during the simulator training programme. Fellows kept a log of attempted procedures and performed no colonoscopies in the clinical setting prior to completion of their simulation training. 4. **Observation, instruction, and feedback:** Supervised orientation to GI Mentor simulator during the first week of fellowship, along with instructions about the simulator training sessions to be completed. Simulation training was unsupervised. It was not stated whether participants had access to the performance quality parameters generated by the simulator during practice. **GROUP 2: No intervention (n = 23)** 1. **Description of intervention:** No intervention. 2. **Observation, instruction, and feedback:** None.

Cohen 2006 *(Continued)*

Outcomes	**Time to assessment:** Approximately 8 weeks after starting fellowship. Participants in the no-intervention group who were from an individual training programme did not begin performing supervised colonoscopy training until the same time that the fellows in the VR simulator training group at their institution completed their simulation training. **Assessment model:** 200 colonoscopies were performed on live patients (or number performed prior to study completion, whichever happened first), under the supervision and evaluation of an attending endoscopist. Fellows were responsible for having their attending fill out the evaluation form. Participants kept a log of colonoscopies completed. Outcomes were compared between groups for every group of 20 cases (i.e. procedures 0 to 20, 21 to 40, 41 to 60, etc.). **Details of patients used for live assessment:** Not specified. **Outcome measures:** 1. Objective competency defined as a. Ability to reach the transverse colon and caecum without assistance b. Ability to correctly recognise and identify abnormalities. 2. Overall rating of competency (rated by attending, 1-to-5 Likert scale: 1 = totally unskilled, 5 = competent and expedient). 3. Patient discomfort level (rated by attending, 1-to-5 Likert scale: 1 = very comfortable to 5 = severe pain). 4. Median number cases required to reach 90% competency (n).Usefulness of simulation training (self rated, questionnaire).
Notes	**Funding:** None stated (simulator donated). **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Random-number table. Quote: "Those who met entry criteria and consented to participate were randomised into 2 groups, with a 50% chance of being placed in either group. The method of sequence generation was a random-number table."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians were blinded to the training status of participants. Quote: "Proctors filling out the individual evaluation forms remained blinded as to whether the particular fellows did or did not receive prior simulator training."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: Accounted for missing outcome data. Quote: "51 first-year gastroenterology fellows, from 16 hospitals, were approved to participate. Two were excluded because of prior colonoscopy experience, and 4 others dropped out after randomisation because of protocol violations during the training phase, leaving 45 who completed the study."

Cohen 2006 *(Continued)*

Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Unclear risk	Unclear: Use of an assessment instrument with no evidence of validity (there is insufficient evidence to suggest that this will introduce bias). No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Di Giulio 2004

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** EGD. **Language of publication:** English. **Number of centres:** Multicentre (7). **Year(s) of conduct of trial:** 2000 (March to May). **Generation of the allocation sequence:** Randomisation list for each site. **Allocation concealment:** Not stated. **Blinding of assessors:** Inadequate (physician assessors not blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** Italy. **Year(s) participants randomised:** Not stated. **Number:** 22 randomised and analysed. **Inclusion criteria:** Gastroenterology trainees. **Exclusion criteria:** Prior direct experience with performance of endoscopy. **Health profession:** Medical trainees (gastroenterology trainees). **Level of training:** Participants were in the "early phase of training" of a 5-year program. **Endoscopy experience:** None. **Sex:** Not stated. **Age:** Not stated.
Interventions	**Learning theory:** None stated. Prior to undergoing the training task, all participants took part in a 2-hour session in which the workings of the endoscope were explained to them by an expert endoscopist and correct methods for performance of upper endoscopy were described. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 11)**

Di Giulio 2004 *(Continued)*

1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix Ltd., Lod, Israel).
2. **Duration of training and/or training endpoint:** 10 hours over 3 to 5 sessions.
3. **Description of intervention:** Participants received basic directions by an instructor with regard to use of the simulator and then completed 10 hours of training in 3 to 5 sessions without supervision. Participants were permitted to try each of the 10 available simulated cases within the times and in the sequence they preferred.
4. **Observation, instruction, and feedback:** Simulation-based training was not supervised. It was not stated whether participants had access to performance quality parameters generated by the simulator during practice.

GROUP 2: No intervention (n = 11)

1. **Description of intervention:** No intervention.
2. **Observation, instruction, and feedback:** None.

Outcomes

Time to assessment: Not stated.

Assessment model: 20 consecutive EGDs on patients scheduled for diagnostic endoscopy, under the supervision and evaluation of an attending physician. Participants were required to keep a procedural logbook detailing procedure duration, number of attempts at intubation, and in event of failure, the reasons for interruption of the procedure and/or the need for assistance in completing the procedure.

Details of patients used for live assessment: Patients were premedicated with midazolam (2.5 mg intravenously) or diazepam (5 mg intravenously), and topical anaesthesia was induced by spraying lidocaine. Patients were excluded if they were:

1. Less than 18 years of age
2. Pregnant
3. Had prior digestive surgery
4. Major risk factors for the procedure, defined as:
   a. Severe respiratory failure
   b. Severe cardiac failure
   c. Patients in an intensive care unit
   d. Gastrointestinal bleeding
5. Coagulation abnormalities
6. Dysphagia.

Outcome measures:

1. Completeness of procedure, rated by attending physician as "complete" or "incomplete", “complete” defined as:
   a. Oesophageal intubation achieved
   b. Participant identified, within 20 minutes, all anatomical landmarks (oesophagogastric mucosal junction, gastric angulus, pylorus)
   c. Participant performed certain basic manoeuvres (aspiration of gastric juice, pylorus intubation in no more than 3 attempts, duodenal bulb exploration, intubation of the second part of the duodenum and retroflexion) with or without verbal direction
2. Overall judgement of performance based on “completeness” of the examination, the need for assistance, and the presumed difficulty of the procedure (rated by attending, 0-to-10 Likert scale with a procedure receiving a score of 5 or less being classified as “negative” and a procedure receiving a score of 6 or more as “positive”: 0 = bad; 10 = good).
3. Number of times manual assistance was required and reason (n).
4. Number of times verbal assistance was required and reason (n).
5. Number of identified or missed lesions (n).
6. Number of complications (n).
7. Failure to effect oesophageal intubation (yes/no).

Di Giulio 2004 *(Continued)*

	8. Number of attempts at oesophageal intubation (n).
Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Randomisation list. Quote: "...trainees were randomised into two groups by using randomisation lists created independently in each hospital."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	High risk	Inadequate: Assessing physicians were not blinded to the training status of participants. Quote: "The instructors were not blinded as to whether trainees had or had not used the simulator."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: Missing outcome data accounted for. Quote: "6 trainees in the SIM group and 7 in the non-SIM group performed one or two procedures less than planned because of the temporary assignment to other clinical activities." and "No attempted procedure was excluded from statistical analysis."
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Ende 2012

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** EGD. **Language of publication:** English. **Number of centres:** Multicentre (15). **Year(s) of conduct of trial:** 2005 to 2006. **Generation of the allocation sequence:** Stratified allocation based on participant performance on a baseline endoscopic skills assessment. **Allocation concealment:** Not stated.

Ende 2012 *(Continued)*

Blinding of assessors: Adequate (1 physician assessor blinded, and 1 unblinded. No significant differences in scoring between the 2 assessors).

Inclusion of all randomised participants: 100%.

Sample size calculation: Yes.

Intention-to-treat analysis: Not stated.

Participants	**Country:** Germany. **Year(s) participants randomised:** Not stated. **Number:** 28 randomised and analysed. **Inclusion criteria:** Medicine or surgery residents interested in training in diagnostic EGD from regional hospitals associated with the institution. **Exclusion criteria:** Any prior endoscopic experience. **Health profession:** Medical trainees (medicine and surgery residents). **Level of training:** Not stated. **Endoscopy experience:** None. **Sex:** 19 males, 9 females. **Age:** Not stated.
Interventions	**Learning theory:** None stated. All participants had 4, 90-minute sessions on endoscope handling, theory of endoscopy, pictures and videos of pathology, use of endoscopic accessories, and patient care. They also underwent a 4-hour course on diagnostic upper gastrointestinal endoscopy led by 2 expert endoscopists using 3 different simulators (Plastic Phantom (Classen 1974), GI Mentor (Simbionix USA, Cleveland, OH, USA), or compactErlangen Active Simulator for Interventional Endoscopy (compactEASIE) (Hochberger 2004)) and received a CD-ROM with video clips of the most important diagnostic findings and lecture notes. At the end of the 4-hour session, participants completed a manual skills test on the compactEASIE for assessment of baseline endoscopic skills. Participants underwent stratified randomisation into 1 of 3 groups based on similar baseline skills level, as assessed by the manual skills test. **GROUP 1: VR simulator training followed by conventional patient-based endoscopy training (n = 10)** 1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA, Cleveland, OH, USA). 2. **Non-VR simulators:** Plastic Phantom and compactEASIE (Classen 1974; Hochberger 2004). 3. **Duration of training and/or training endpoint:** 18 to 20 hours over 9 to 10 sessions and conventional patient-based training over 4 months (29 ± 21 EGDs). 4. **Description of intervention:** Participants received training on 3 simulators once weekly for 2 hours. Only the GI Mentor was a VR simulator. Trainees were supervised by 2 experienced tutors during simulator sessions. Participants were required to attend at least 9 of the 10 sessions offered. Participants also received standard clinical education at their home institution. 5. **Observation, instruction, and feedback:** Participants received training and supervision during simulated and clinical procedures from 2 supervised, experienced tutors. It was not stated whether participants had access to performance quality parameters generated by the simulator during practice. **GROUP 2: Conventional patient-based endoscopy training (n = 8)** 1. **Duration of training and/or training endpoint:** 4 months (19 ± 18 EGDs).

Ende 2012 *(Continued)*

2. **Description of intervention:** Participants received standard clinical education at their home institution over 4 months.
3. **Observation, instruction, and feedback:** Participants received training and supervision during clinical procedures from 2 supervised, experienced tutors.

GROUP 3: VR simulator training only (n = 9)

1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA, Cleveland, OH, USA).
2. **Non-VR simulators:** Plastic Phantom and compactEASIE (Classen 1974; Hochberger 2004).
3. **Duration of training and/or training endpoint:** 18 to 20 hours over 9 to 10 sessions.
4. **Description of intervention:** Participants received training on 3 simulators once weekly for 2 hours. Only the GI Mentor was a VR simulator. Trainees were supervised by 2 experienced tutors during simulator sessions. Participants were required to attend at least 9 of the 10 sessions offered.
5. **Observation, instruction, and feedback:** Participants received training and supervision during simulated procedures from 2 supervised, experienced tutors. It was not stated whether participants had access to performance quality parameters generated by the simulator during practice.

Outcomes	**Time to assessment:** Assessment began the day following the conclusion of the 4-month training period and continued for up to 2 months. **Assessment model:** A final evaluation of a manual skills test on the compactEASIE simulator upon completion of the 4-month training period, and evaluation of 3 clinical EGDs during a 2-month period after the final evaluation under the supervision of an unblinded expert endoscopist and a blinded expert endoscopist. **Details of patients used for live assessment:** There was no restriction regarding patients. The chief of the endoscopy department selected 3 appropriate clinical cases. **Outcome measures:** 1. Time to reach the descending portion of the duodenum (seconds). 2. Endoscopic skills rated using a 10-point visual analogue scale (rated by expert endoscopists, 1 = worst performance; 10 = optimal performance). 3. Mean procedure times (time for oesophageal intubation, time to pass the pylorus, time to reach the descending duodenum, overall procedure time) (seconds). 4. Mean percentage of estimated visualised mucosal surface. 5. Incidence of complications (n).
Notes	**Funding:** Yes (peer-reviewed research grant)**.** **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Inadequate. Not completely random allocation. Stratified randomisation based on baseline endoscopic skills level. Quote: “stratified randomisation was performed placing participants into 3 groups with a similar skills level”
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).

Ende 2012 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: 1 physician assessor was blinded, 1 physician assessor was unblinded, with no significant difference between any mean ratings assigned by the 2 raters. Quote: "The overall clinical evaluation, performed by a blinded expert and an unblinded expert, was not statistically significantly different."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate. Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No intention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

Ferlitsch 2010

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** EGD. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2003 to 2007 **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Inadequate (physician assessors not blinded, patients blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** Yes. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** Austria. **Year(s) participants randomised:** Not stated. **Number:** 28 enrolled and analysed. **Inclusion criteria:** At least third-year residents in internal medicine. **Exclusion criteria:** Previous endoscopy training. **Health profession:** Medical trainees (internal medicine residents). **Level of training:** At least third-year residents. **Endoscopy experience:** None. **Sex:** 1. VR simulator training group: 7 males, 7 females. 2. No-intervention group: 12 males, 2 females.

Ferlitsch 2010 *(Continued)*

	Age (median (IQR)): 31 (28 to 37).
Interventions	**Learning theory:** None stated. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 14)** 1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA Corp., Cleveland, OH, USA). 2. **Duration of training and/or training endpoint:** 2 hours per day of structured training for 5 to 20 hours total (their choice). Median training time was 10 hours (range 5 to 20 hours). 3. **Description of intervention:** 2 hours per day of structured training (5 to 20 hours total) on the VR simulator. Participants were permitted to practice using 20 virtual EGD cases, haptic (targeted steering) training games “Endobasket” and “Endobubble". 4. **Observation, instruction, and feedback:** Trainers were present for the first 2 hours of simulator training. It was not stated whether participants had access to performance quality parameters generated by the simulator during practice. **GROUP 2: No intervention (n = 14)** 1. **Description of intervention:** No intervention. 2. **Observation, instruction, and feedback:** None. After the training task, all participants received equal instruction and training in EGD including instruction in handling the endoscope, observing 5 to 10 EGD examinations by experts, and withdrawing the endoscope 3 to 5 times from the descending duodenum in patients. Participants were introduced to pathological findings of the upper gastrointestinal tract, using an endoscopic atlas and CD. Participants were trained in 1-hand steering technique; were allowed to try to intubate the oesophagus twice before the attending physician took over the scope; were allowed to try to perform pyloric passage twice before they were assisted by the attending; and performed routine biopsies.
Outcomes	**Time to assessment:** Not stated. **Assessment model:** Observed and evaluated by expert endoscopists (performed > 5000 EGD) performing their first 10 EGD on consecutive patients who met inclusion criteria (listed below). 14 of 28 participants were assessed while performing their 51st to 60th EGD on consecutive patients who met inclusion criteria. **Details of patients used for live assessment:** Patients scheduled for diagnostic EGD and unwilling to undergo sedation. Patients wanting to have concomitant sedation or requiring therapeutic interventions were excluded. **Outcome measures:** 1. Time from the first attempt at oesophageal intubation until the descending part of the duodenum reached. 2. Time between the first attempt at oesophageal intubation and the end of the investigation. 3. Technical accuracy (evaluated by recording whether the novice endoscopist was able to intubate the oesophagus (“unaided”), whether manual help by the expert was needed (“expert help”), or if the expert had to take over (“expert takeover”)). 4. Pyloric passage (evaluated as “unaided", requiring “expert help", or requiring “expert takeover”). 5. Retroflexion (J-manoeuvre) in the gastric fundus (evaluated as “unaided", requiring “expert help", or requiring “expert takeover”). 6. Diagnostic accuracy (evaluated as the number of pathological entities found or missed). 7. Discomfort and pain (evaluated immediately after EGD using patient questionnaire that used 2, 100-millimetre visual analogue scales for discomfort and pain).
Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Ferlitsch 2010 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Random number draw. Quote: "Randomization was performed by a member of the department not involved into the study. A group of 4–6 residents started every 6 months. Their names, each written on a piece of paper, were drawn out of a box after calling of "group C" or group S"."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	High risk	Inadequate: Assessing physicians were not blinded to the training status of participants. Assessing patients were blinded. Quote: "The experts were informed about the training status of the endoscopic novices (i.e., which were simulator-trained), but the patients were not." and "Patients were blind to the training status of the trainee (i.e., whether they had simulator training or not, and the number of patient endoscopies they had performed)."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No intention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

Gerson 2003

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Sigmoidoscopy. **Language of publication:** English. **Number of centres:** Single centre (2 sites). **Year(s) of conduct of trial:** 2001. **Generation of the allocation sequence:** Sequential allocation. **Allocation concealment:** No. **Blinding of assessors:** Inadequate (physician assessors not blinded, patients blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** Yes.

Gerson 2003 *(Continued)*

	Intention-to-treat analysis: Not stated.
Participants	**Country:** USA. **Year(s) participants randomised:** Not stated. **Number:** 16 enrolled and analysed. **Inclusion criteria:** Internal medicine residents. **Exclusion criteria:** Any prior experience with flexible sigmoidoscopy, observation of sigmoidoscopy as part of a clinical rotation, or prior use of an endoscopy simulator. **Health profession:** Medical trainees (internal medicine residents). **Level of training:** 8/16 first-year residents (VR group: 2/9, control group: 6/7). **Endoscopy experience:** None. **Sex:** 12 males, 4 females (no significant difference between groups). **Age (mean ± SD):** 1. VR simulator training group: 29.4 ± 1.1. 2. Conventional endoscopy training group: 28 ± 0.8 (no significant difference between groups).
Interventions	**Learning theory:** None stated. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 9)** 1. **VR simulator:** AccuTouch VR endoscopy simulator (Immersion Medical, Inc., Gaithersburg, MD, USA). 2. **Duration of training and/or training endpoint:** 2 weeks (unlimited simulator access). 3. **Description of intervention:** Unlimited simulator use during a 2-week period (average time (mean ± SEM): 138 ± 28 minutes; average number cases (mean ± SEM): 12.8 ± 2.9). Participants were instructed to review all didactic modules and complete all 6 practice cases on the simulator. 4. **Observation, instruction, and feedback:** Not observed and no external instruction provided. Participants permitted to use simulator teaching features ("virtual attending physician" and external view of colon) during each examination. Performance quality parameters were provided to participants by the simulator after each procedure, including: procedure time, insertion length, degree of air insufflation, percentage of mucosa visualised, time in red-out, patient discomfort, recognition of pathology, occurrence of perforation, performance of retroflection. **GROUP 2: Conventional patient-based endoscopy training (n = 7)** 1. **Duration of training and/or training endpoint:** 2 weeks (10 sigmoidoscopic examinations). 2. **Description of intervention:** 10 sigmoidoscopic examinations during a 2-week period (average time: 300 minutes) performed with a video colonoscope. 3. **Observation, instruction, and feedback:** An attending gastroenterologist observed each participant's procedures and was instructed to teach the resident using his or her own teaching preferences and techniques. Participants were expected to learn how to advance the colonoscope independently by the end of the 10 sessions.
Outcomes	**Time to assessment:** Not stated. **Assessment model:** 5 sigmoidoscopic examinations (insertion and withdrawal) were completed, under the supervision and evaluation of an attending gastroenterologist who provided no coaching during the test examinations. Participants were expected to perform retroflexion at the completion of the sigmoidoscopy and were required to notify the attending when the splenic flexure was identified and if any pathology was encountered. If the participant encountered difficulty, the attending was allowed to take over until the resident could continue.

Gerson 2003 *(Continued)*

Details of patients used for live assessment: Asymptomatic patients referred for routine colorectal cancer screening via flexible sigmoidoscopy.

Outcome measures:

1. Independent completion (yes/no).
2. Examination duration (time).
3. Required assistance (yes/no).
4. Flexure recognition (yes/no).
5. Completion of retroflexion (yes/no).
6. Ability to recognise pathology (yes/no).
7. Expert global rating (rated by attending, 1-to-5 Likert scale: 1 = unable to clear the rectum; 2 = unable to clear the rectosigmoid junction; 3 = unable to pass 1 turn without assistance; 4 = able to perform independently, but more than 20 min required; 5 = independent examination less than 20 min in duration).
8. Level of patient comfort/discomfort (rated by patient, 1-to-5 Likert scale: 1 = strongly agree; 2 = agree; 3 = not sure; 4 = disagree; 5 = strongly disagree).
9. Patient satisfaction (rated by patient, 1-to-5 Likert scale: 1 = strongly agree; 2 = agree; 3 = not sure; 4 = disagree; 5 = strongly disagree).
10. Technical competence (rated by patient, 1-to-5 Likert scale: 1 = strongly agree; 2 = agree; 3 = not sure; 4 = disagree; 5 = strongly disagree).

Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Inadequate: Sequential allocation. Quote: "Residents were assigned in a sequential fashion by one of the investigators to a simulator-trained group or a traditional teaching group."
Allocation concealment (selection bias)	High risk	Inadequate: Not concealed. Quote: “Neither the investigators nor participating residents were blinded to the group assignment.”
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	High risk	Inadequate: Assessing physicians were not blinded; participating patients were blinded to resident's training method. Quote: "The attending physicians grading the test cases were not blinded to the mode of training." Quote: "Participating patients were blinded to the residents training method."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.

Gerson 2003 *(Continued)*

Other bias	Unclear risk	Adequate: No sample size calculation (outcome not likely to be influenced by lack of sample size calculation).

Gomez 2015

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2012 to 2013 **Generation of allocation sequence:** Consecutive allocation of participants rotating through the study setting. **Allocation concealment:** No. **Blinding of assessors:** Adequate (physician assessor blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None.
Participants	**Country:** USA. **Year(s) participants randomised:** 2012 to 2013. **Inclusion criteria:** Trainees in first year of the general surgery program. **Exclusion criteria:** None stated. **Health profession:** Medical trainees (general surgery residents). **Level of training:** First-year residents. **Endoscopy experience:** None. **Sex:** 1. VR simulator training group: 6 males, 3 females. 2. Another method of VR simulator training group: 5 males, 4 females. 3. Another form of endoscopy simulation group: 6 males, 3 females. **Age (median):** 1. VR simulator training group: 29. 2. Another method of VR simulator training group: 28. 3. Another form of endoscopy simulation group: 29.
Interventions	**Learning theory:** None stated. Participants were randomised to 1 of 3 groups. Each participant performed a baseline colonoscopy on a real patient, then completed 3 online modules. Module 1 familiarised residents with endoscopic equipment. Module 2 described fundamental concepts of endoscopy practice. Module 3 described use of the 2 training platforms available at the simulation centre. Each participant then completed 1 of 3 flexible endoscopy courses based on their group.

Gomez 2015 *(Continued)*

GROUP 1: VR simulator training in addition to another form of endoscopy simulation training (n = 9)

1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA Corp., Cleveland, OH, USA).
2. **Non-VR simulator:** Kyoto Kagaku colonoscopy physical model simulator (Kyoto Kagaku Co. Ltd., Kyoto, Japan).
3. **Duration of training and/or training endpoint:** 3 weeks.
4. **Description of intervention:** On the GI Mentor II simulator, participants were required to complete 2 practice exercises and at least 1 of the 10 available simulated colonoscopy cases. On the Kyoto Kagaku simulator, participants were required to complete at least 1 of the 6 available colonoscopy modules.
5. **Observation, instruction, and feedback:** Not observed and no external instruction provided. Performance quality parameters were provided to participants by the GI Mentor endoscopy simulator: time to reach the caecum, percentage of time with a clear view of the lumen.

GROUP 2: VR simulator training only (n = 9)

1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA Corp., Cleveland, OH, USA).
2. **Duration of training and/or training endpoint:** 3 weeks.
3. **Description of intervention:** Participants were required to complete 2 practice exercises and at least 1 of the 10 available simulated colonoscopy cases.
4. **Observation, instruction, and feedback:** Not observed and no external instruction provided. Performance quality parameters were provided to participants by the GI Mentor endoscopy simulator: time to reach the caecum, percentage of time with a clear view of the lumen.

GROUP 3: Another form of endoscopy simulation only (n = 9)

1. **Non-VR simulator:** Kyoto Kagaku colonoscopy physical model simulator (Kyoto Kagaku Co. Ltd., Kyoto, Japan).
2. **Duration of training and/or training endpoint:** 3 weeks.
3. **Description of intervention:** Participants were required to complete at least 1 of the 6 available colonoscopy modules.
4. **Observation, instruction, and feedback:** None.

Outcomes

Time to assessment: Assessment took place immediately after completion of the 3-week course.

Assessment model: 1 patient-based colonoscopy under the guidance of an expert endoscopist.

Details of patients used for live assessment: Patients were included only if they were older than 18 years, scheduled for an elective screening colonoscopy, and had no prior history of any major intestinal or abdominal operations.

Outcome measures:

1. Procedural proficiency (rated by an expert endoscopist using the Global Assessment of Gastrointestinal Endoscopic Skills - Colonoscopy tool) (Vassiliou 2010).
2. Total procedure time (min).
3. Time to reach the caecum (min).
4. Time with a clear view of the lumen (min).
5. Number of times a faculty took full control of the colonoscope (n).
6. Need for endoscopic instrumentation (n).

Notes

Funding: None stated.

Declarations of conflicts of interest for primary investigators: None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement

Gomez 2015 *(Continued)*

Random sequence generation (selection bias)	High risk	Inadequate: Group assignment was not completely random. Quote: "...each resident was randomly assigned to 1 of 3 training conditions based on equipment availability at our simulation centre"
Allocation concealment (selection bias)	High risk	Inadequate: No allocation concealment (through direct contact with authors).
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Expert faculty were blinded to the training status of participants. Quote: "It should be noted that the expert faculty scoring both the GAGES-C performance and colonoscopy conditions was blinded regarding the training condition"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Grover 2015

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2011 to 2012 **Generation of allocation sequence:** Blinded random draw of numbers contained within sealed envelopes. **Allocation concealment:** Adequate (sealed envelope). **Blinding of assessors:** Adequate (physician assessors blinded). **Inclusion of all randomised participants:** 33/34 (97%). **Sample size calculation:** Yes.
Participants	**Country:** Canada. **Year(s) participants randomised:** 2011 to 2012. **Inclusion criteria:** Postgraduate trainees from adult gastroenterology, general surgery, and internal medicine residency training programs at the University of Toronto.

Grover 2015 *(Continued)*

Exclusion criteria: Performance of > 20 EGDs and/or colonoscopies in the clinical and/or simulated setting.

Health profession: Medical trainees (internal medicine and general surgery residents, gastroenterology fellows).

Level of training: Postgraduate years 2 to 4.

Endoscopy experience (average number of procedures):

1. VR simulator training group: 0.6 independent colonoscopies, 2.7 assisted colonoscopies.
2. Another method of VR simulator training group: 0.7 independent colonoscopies, 0.8 assisted colonoscopies.

Sex:

1. VR simulator training group: 13 males, 4 females.
2. Another method of VR simulator training group: 7 males, 9 females (no significant difference between groups).

Age (mean ± SD):

1. VR simulator training group: 29.7 ± 3.8.
2. Another method of VR simulator training group: 28.4 ± 1.3 (no significant difference between groups).

Interventions

Learning theory: A structured comprehensive curriculum that incorporates teaching of technical, cognitive, and integrative competencies related to colonoscopy (Palter 2013), and self regulated learning, whereby trainees direct their own acquisition of knowledge and skills (Brydges 2015; Murad 2010).

All participants performed a baseline procedure on the VR simulator which simulated a screening colonoscopy. Both groups received 8 hours of simulation-based training with a prespecified list of cases.

GROUP 1: VR simulator training (n = 16)

1. **VR simulator:** EndoVR VR endoscopy simulator (CAE Healthcare Canada, Montreal, Quebec, Canada).
2. **Duration of training and/or training endpoint:** 6 hours of lectures and 8 hours of endoscopy VR simulation-based training.
3. **Description of intervention:** Participants received 6 hours of interactive small-group lectures and 8 hours of supervised 1-on-1 endoscopy VR simulation-based training led by experienced endoscopists. Didactic sessions were led by faculty gastroenterologists and covered the theory of colonoscopy and mechanics of performance of colonoscopic procedures. Simulation-based training consisted of a prespecified list of cases.
4. **Observation, instruction, and feedback:** During simulation-based training, an experienced endoscopist demonstrated procedural elements of colonoscopy, answered questions, and provided direct verbal feedback to the participant. At the end of each case, participants had the opportunity to review simulator-generated metrics of their performance (specific metrics not stated).

GROUP 2: Another method of VR simulator training (n = 17)

1. **VR simulator:** EndoVR VR endoscopy simulator (CAE Healthcare Canada, Montreal, Quebec, Canada).
2. **Duration of training and/or training endpoint:** 8 hours of endoscopy VR simulation-based training.
3. **Description of intervention:** Participants received 8 hours of VR simulation-based training. They were provided with a list of desired objectives and proceeded through the same prespecified list of cases as Group 1. Participants were also provided with a link to a website with the set of lecture content, which was accessible during training.
4. **Observation, instruction, and feedback:** Experienced endoscopists only provided information regarding the technical use of the simulator, and provided no feedback on performance. At the end of each case, participants had the opportunity to review simulator-generated metrics of their performance (specific metrics not stated).

Grover 2015 *(Continued)*

Outcomes	**Time to assessment:** Assessment took place immediately and 4 to 6 weeks after training. Patient-based colonoscopies were only performed at the 4- to 6-week mark. **Assessment model:** 2 patient-based colonoscopies under the guidance of an expert endoscopist. **Details of patients used for live assessment:** Patients were excluded if they had a history of colonic or pelvic surgery or difficult colonoscopy. **Outcome measures:** 1. Procedural proficiency on 2 patient-based colonoscopies (rated by an expert endoscopist using the UK Joint Advisory Group colonoscopy Director Observation of Procedural Skills (JAG DOPS) assessment form) (4 to 6 weeks' post-training) (JAG Central Office 2010). 2. Procedural knowledge assessed by multiple-choice tests (immediately post-training). 3. Procedural proficiency, communication skills, and global performance on simulated colonoscopies (immediately post-training and 4 to 6 weeks' post-training) (rated by an expert endoscopist using the JAG DOPS, the Integrated Scenario Communication Rating Form (LeBlanc 2009), and the Integrated Scenario Global Rating Form (Hodges 2003), respectively).
Notes	**Funding:** Yes (peer-reviewed research grant). **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Blinded random draw of numbers contained within sealed envelopes. Quote: “Participants were randomised using a sealed envelope technique”
Allocation concealment (selection bias)	Low risk	Adequate: Sealed envelopes. Quote: “Participants were randomised using a sealed envelope technique...”
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians were blinded to the training status of participants. Quote: “The raters were blinded to group assignment”
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: Accounted for missing outcome data. Quote: “Thirty-four participants were randomised, with 33 completing the study. One participant was recruited and randomised but could not participate because of a scheduling conflict.”
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No intention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

Grover 2017

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2013 to 2014 **Generation of allocation sequence:** Blinded random draw of numbers contained within sealed envelopes. **Allocation concealment:** Adequate (sealed envelope). **Blinding of assessors:** Adequate (physician assessors blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** Yes.
Participants	**Country:** Canada. **Year(s) participants randomised:** 2013 to 2014. **Inclusion criteria:** Postgraduate trainees from adult gastroenterology, general surgery, and internal medicine residency training programs at the University of Toronto. **Exclusion criteria:** Performance of > 20 EGDs and/or colonoscopies in the clinical and/or simulated setting. **Health profession:** Medical trainees (internal medicine and general surgery residents, gastroenterology fellows). **Level of training:** Postgraduate years 2 to 4. **Endoscopy experience (average number of procedures):** 1. VR simulator training group: 0.8 independent colonoscopies, 5.5 assisted colonoscopies, 1.8 independent EGDs, and 6.0 assisted EGDs. 2. Another method of VR simulator training group: 0.2 independent colonoscopies, 3.6 assisted colonoscopies, 0.9 independent EGDs, and 3.2 assisted EGDs. **Sex:** 1. VR simulator training group: 13 males, 5 females. 2. Another method of VR simulator training group: 10 males, 9 females. **Age (mean ± SD):** 1. VR simulator training group: 28.1 ± 3.0. 2. Another method of VR simulator training group: 28.1 ± 2.0
Interventions	**Learning theory:** Progressive learning, in which trainees transition from tasks of low complexity to high complexity (Brydges 2010; Guadagnoli 2012). All participants performed a baseline procedure on the VR simulator that simulated a screening colonoscopy. Both groups received 4 hours of didactic small-group lectures and 6 hours of 1-on-1 simulation-based training. Lectures were led by faculty gastroenterologists and covered the theory of colonoscopy and mechanics of performance of colonoscopic procedures. **GROUP 1: VR simulator training (n = 18)** 1. **VR simulator:** EndoVR VR endoscopy simulator (CAE Healthcare Canada, Montreal, Quebec, Canada).

Grover 2017 *(Continued)*

2. **Non-VR simulator:** Bench-top endoscopy simulator (Walsh 2008).
3. **Duration of training and/or training endpoint:** 4 hours of lectures and 6 hours of endoscopy VR simulation-based training.
4. **Description of intervention:** Participants spent 1 hour on a bench-top simulator and 5 hours on the VR simulator in addition to receiving 4 hours of didactic sessions. They performed simulated cases in order of increasing difficulty.
5. **Observation, instruction, and feedback:** During simulation-based training, an experienced endoscopist demonstrated procedural elements of colonoscopy, answered questions, and provided direct verbal feedback to the participant. At the end of each case, participants had the opportunity to review simulator-generated metrics of their performance (specific metrics not stated).

GROUP 2: Another method of VR simulator training (n = 19)

1. **VR simulator:** EndoVR VR endoscopy simulator (CAE Healthcare Canada, Montreal, Quebec, Canada).
2. **Duration of training and/or training endpoint:** 4 hours of lectures and 6 hours of endoscopy VR simulation-based training.
3. **Description of intervention:** Participants spent 6 hours on the bench-top simulator in addition to receiving 4 hours of didactic sessions. They completed a prespecified list of cases with a random order of task difficulty.
4. **Observation, instruction, and feedback:** During simulation-based training, an experienced endoscopist demonstrated procedural elements of colonoscopy, answered questions, and provided direct verbal feedback to the participant. At the end of each case, participants had the opportunity to review simulator-generated metrics of their performance (specific metrics not stated).

Outcomes	**Time to assessment:** Assessment took place immediately and 4 to 6 weeks after training. Patient-based colonoscopies were only performed at the 4- to 6-week mark. **Assessment model:** 2 patient-based colonoscopies under the guidance of an expert endoscopist. **Details of patients used for live assessment:** Patients were excluded if they had a history of colonic or pelvic surgery or difficult colonoscopy. **Outcome measures:** 1. Procedural proficiency on 2 patient-based colonoscopies (rated by an expert endoscopist using the UK Joint Advisory Group colonoscopy Director Observation of Procedural Skills (JAG DOPS) assessment form) (4 to 6 weeks' post-training) (JAG Central Office 2010). 2. Procedural knowledge assessed by multiple-choice tests (immediately post-training). 3. Procedural proficiency, communication skills, and global performance on simulated colonoscopies (immediately post-training and 4 to 6 weeks' post-training) (rated by an expert endoscopist using the JAG DOPS, the Integrated Scenario Communication Rating Form (LeBlanc 2009), and the Integrated Scenario Global Rating Form (Hodges 2003), respectively).
Notes	**Funding:** Yes (peer-reviewed research grant). **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Blinded random draw of numbers contained within sealed envelopes. Quote: “Participants were randomised by using a sealed envelope technique”
Allocation concealment (selection bias)	Low risk	Adequate: Sealed envelopes.

Grover 2017 *(Continued)*

		Quote: "The random allocation sequence was generated by another author (J.Y), and this sequence was concealed from participants and from other study staff until assignment of intervention."
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians were blinded to the training status of participants. Quote: "Assessors were blinded to group allocation for evaluation of the primary outcome measure"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No intention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

Haycock 2010

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy **Language of publication:** English. **Number of centres:** Multicentre (4). **Year(s) of conduct of trial:** Not stated. **Generation of the allocation sequence:** Computer-generated, block randomisation protocol (8 per block, enrolled by subinvestigator and randomised to simulator vs traditional patient-based bedside training). A**llocation concealment:** No. **Blinding of assessors:** Adequate (physician assessors blinded). **Inclusion of all randomised participants:** (36/40) 90%. **Sample size calculation:** Yes. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** United Kingdom, Netherlands, Italy. **Year(s) participants randomised:** Not stated. **Number:** 40 enrolled and 36 analysed. **Inclusion criteria:** Any medical background (physicians, surgeons, nurses) or position recognised by the training institution as appropriate for training in colonoscopy.

Haycock 2010 *(Continued)*

Exclusion criteria: Performance of > 25 previous colonoscopies or flexible sigmoidoscopies; previous participation in an intensive colonoscopy training course, colonoscopy training or simulator training study; performance of > 10 laparoscopic surgical procedures.

Health profession: Any health profession background (medical trainees (general trainee, specialist in training), nurses, etc.).

Level of training: Not stated.

Endoscopy experience (average number of procedures):

1. VR simulator training group: 15 observed colonoscopies, 0 assisted colonoscopies.
2. Conventional endoscopy training group: 45 observed colonoscopies, 1 assisted colonoscopy.

Sex:

1. VR simulator training group: 6 males, 13 females.
2. Conventional endoscopy group: 10 males, 8 females (no significant difference between groups).

Age (mean (range)):

1. VR simulator training group: 31 (26 to 33).
2. Conventional endoscopy group: 28 (26 to 30) (no significant difference between groups).

Interventions	**Learning theory:** None stated. Prior to undergoing the training task, all participants received a standardised tutorial on the fundamentals of colonoscopy. All participants then performed 3 validated pre-test simulator cases to assess baseline performance. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 18)** 1. **VR simulator:** Endo TS-1 Olympus colonoscopy simulator (Olympus KeyMed, Southend, UK). 2. **Duration of training and/or training endpoint:** 16 hours. 3. **Description of intervention:** 16 hours of standardised simulator training. The training package included knowledge and skill-based learning with formative assessments in a multimedia environment and incorporated a simulated 3-dimensional (3-D) image viewer. It was structured in a sequential fashion to introduce the skills and knowledge needed to progress from rectum to caecum. 4. **Observation, instruction, and feedback:** Trainers expected to provide minimal tutoring and feedback. **GROUP 2: Conventional patient-based endoscopy training (n = 18)** 1. **Duration of training and/or training endpoint:** 16 hours (minimum 8 colonoscopies). 2. **Description of intervention:** 16 hours of patient-based training (4 half-day sessions) by an expert trainer using a ScopeGuide 3-D endoscopic imager. Participants performed a minimum of 8 colonoscopies under 1:1 supervision. Recommendations made for topics to be covered aiming to standardise training. All trainees taught to use single-handed, 1-person technique for colonoscopy, but instructor otherwise told to provide "usual" training for a novice colonoscopist**.** 3. **Observation, instruction, and feedback:**Use of ScopeGuide imager. Instructor told to teach single-handed, 1-person technique, but instructor otherwise told to provide "usual" training for a novice colonoscopist. Details of instruction and feedback not stated.
Outcomes	**Time to assessment:** Not stated. **Assessment model:** 3 patient-based colonoscopies were completed, under the supervision and evaluation of an expert assessor. Assessors were asked not to provide any assistance (verbal, practical) unless there were safety concerns. A ScopeGuide 3-D endoscopic imager view used for all colonoscopies performed. Procedures terminated at 20 minutes or earlier if caecal intubation achieved (confirmed by visualisation of 2 of 3 landmarks (ileocaecal valve, appendix orifice, triradiate fold) and imager view

Haycock 2010 *(Continued)*

compatible with tip of endoscope in caecum). An assessment was repeated if a procedure was terminated due to patient factors (e.g. poor prep, poor patient tolerance).

Details of patients used for live assessment: < 75 years old, no history of pelvic or colonic surgery or difficult colonoscopy.

Outcome measures:

1. Procedural proficiency (rated by attending using an abbreviated version of the UK Joint Advisory Group colonoscopy Direct Observation of Procedural Skills assessment form (JAG Central Office 2010), which rated 9 domains of "endoscopic skills during insertion and withdrawal" on a 1-to-4-point scale).
2. Global score (rated by attending using Global Performance Score assessment form (Park 2007), which rates 7 domains on a 1-to-5 Likert scale: atraumatic technique, colonoscope advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure, overall performance).
3. Time to completion.Depth of insertion (cm and anatomical position).

Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Computer-generated, block randomisation. Quote: "...randomised into subjects (simulator training) and controls (patient-based training) by the lead investigator, by using a computer-generated, block randomisation protocol with 8 per block."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding). Quote: "Participants, sub investigators, and trainers in each institution were not blinded to the group allocation."
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians were blinded. Quote: "An expert assessor blinded to the group allocation of the trainee was present during all assessments."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: Accounted for missing outcome data. Quote: "Forty trainees were randomised, with 36 completing the study. Two trainees did not start because of limitations in availability of endoscopy sessions, 1 trainee completed the simulator pre-training assessment but had to leave for personal reasons before commencing the training, and 1 trainee completed the training and simulator assessments but did not complete all 3 patient-based assessment cases."
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Unclear risk	Unclear: Use of an assessment instrument with no evidence of validity (there is insufficient evidence to suggest that this would have introduced bias). No in-

Haycock 2010 *(Continued)*

	tention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

McIntosh 2014

Methods	**Study design:** Quasi-randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2009 to 2011. **Generation of allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Adequate (physician assessors blinded, nurse assessors blinded, patients not stated). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** Yes.
Participants	**Country:** Canada. **Year(s) participants randomised:** 2009 to 2011. **Inclusion criteria:** Enrolment in internal medicine, gastroenterology, or general surgery subspecialties at Western University between postgraduate years 2 and 4. **Exclusion criteria:** Performance of > 10 EGDs, sigmoidoscopies, and/or colonoscopies. **Health profession:** Medical trainees (internal medicine and general surgery residents, gastroenterology fellows). **Level of training:** Postgraduate years 2 to 4. **Endoscopy experience (average number of procedures):** 1. VR simulator training group: 0.8 independent colonoscopies, 5.5 assisted colonoscopies, 1.8 independent EGDs, and 6.0 assisted EGDs. 2. No-intervention group: 0.2 independent colonoscopies, 3.6 assisted colonoscopies, 0.9 independent EGDs, and 3.2 assisted EGDs. **Sex:** 1. VR simulator training group: 9 males, 1 female. 2. No-intervention group: 8 males, 0 females. **Age (mean):** 1. VR simulator training group: 29. 2. No-training group: 29.
Interventions	**Learning theory:** None stated. Residents were assigned (non-randomly) to a simulator-training group or a control group. Gastroenterology residents were assigned to the simulator group. General surgery and internal medicine resi-

McIntosh 2014 *(Continued)*

	dents with an interest in endoscopy and gastroenterology, and gastroenterology residents who could not complete simulator training before starting their fellowship were assigned to the control group. **GROUP 1: VR simulator training (n = 10)** 1. **VR simulator:** GI Mentor II simulator (Simbionix USA, Cincinnati, OH, USA). 2. **Duration of training and/or training endpoint:** 10 to 20 hours of the simulator over 4 weeks. 3. **Description of intervention:** Residents performed 10 to 20 hours of training on the simulator over 4 weeks before patient-based colonoscopies. They were free to complete 1 to 10 modules of upper endoscopy and 1 to 10 modules of lower endoscopy at their discretion. 4. **Observation, instruction, and feedback:** None. **GROUP 2: No intervention (n = 8)** 1. **Description of intervention:** No intervention. 2. **Observation, instruction, and feedback:** None.
Outcomes	**Time to assessment:** Assessment took place immediately after the 4-week training period for participants in the VR simulator training group, and immediately after the start of their gastroenterology rotation for participants in the no-intervention group. **Assessment model:** 5 patient-based colonoscopies under the guidance of an expert endoscopist. **Details of patients used for live assessment:** Patients were included if they gave informed consent, were undergoing a screening or surveillance colonoscopy, were between 18 and 75 years of age, and had previously undergone colonoscopy without reported difficulty. Patients were excluded if they failed to give consent or were not willing to complete the post-endoscopy questionnaire. **Outcome measures:** 1. Mean number of proctor assists required per colonoscopy (n). 2. Procedure time (min). 3. Median depth of insertion (1 = rectum, 2 = sigmoid, 3 = descending colon, 4 = splenic flexure, 5 = transverse colon, 6 = hepatic flexure, 7 = ascending colon, 8 = caecum). 4. Proportion of cases in which the caecum was successfully intubated (%).
Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Inadequate: Non-random allocation. Quote: “Residents in the gastroenterology program at the start of their fellowship or residents selected to be in the gastroenterology program were assigned to the simulator training group. Similarly matched controls were selected from internal medicine residents interested in gastroenterology, general surgery residents with interest in endoscopy and gastroenterology residents who could not complete the simulator training before starting their fellowship"
Allocation concealment (selection bias)	High risk	Inadequate: Non-random allocation. See quote above.
Blinding of participants and personnel (performance bias)	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).

McIntosh 2014 *(Continued)*
All outcomes

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate. Quote: "Preceptors were blinded as to who had received simulator training."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Unclear risk	Unclear: Use of an assessment instrument with no evidence of validity (there is insufficient evidence to suggest that this would have introduced bias). No intention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

Park 2007

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** Not stated. **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Adequate (physician assessors blinded). **Inclusion of all randomised participants:** (24/28) 85.71%. **Sample size calculation:** Yes. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** Canada. **Year(s) participants randomised:** Not stated. **Number:** 28 enrolled and 24 analysed. **Inclusion criteria:** Internal medicine and surgery residents. **Exclusion criteria:** Experience in endoscopy defined as the primary endoscopist for 3 procedures of any type. **Health profession:** Medical trainees (internal medicine and surgery residents). **Level of training:** Postgraduate years 1 to 3. **Endoscopy experience:** < 3 endoscopic procedures (of any kind) performed. **Sex:** Details not stated (no significant difference between groups).

Park 2007 *(Continued)*

	Age: Details not stated (no significant difference between groups).
Interventions	**Learning theory:** None stated. Prior to undergoing the training task, all participants viewed an introduction to colonoscopy video and were given the opportunity to familiarise themselves with the components and handling of a colonoscope. No formal instruction was given at this time. All participants then performed 1 pre-test simulator sequence to assess baseline performance. Between the VR simulator pre-test and the test in the clinical setting, participants in both groups were allowed to attend and view colonoscopies performed by faculty endoscopists as per their normal experience during a clinical rotation. They did not receive specific teaching regarding the technical aspects of endoscopy or perform any procedures prior to their clinical test. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 12)** 1. **VR simulator:** AccuTouch VR endoscopy simulator version 1.2 (Immersion Medical, Inc., Gaithersburg, MD, USA). 2. **Duration of training and/or training endpoint:** 2 to 3 hours. 3. **Description of intervention:** Participants practiced independently for 2 to 3 hours (average time (mean ± SEM): 125 ± 37 minutes) on the simulator, during which time they had access to the range of 6 available simulator cases. 4. **Observation, instruction, and feedback:** Participants were not observed, and no external instruction was provided. Simulator training included the use of all simulator-based resources (e.g. computer-generated anatomical views). 14 performance quality parameters were provided to participants by the simulator after each procedure, including: procedure time, insertion length, degree of air insufflation, percentage of mucosa visualised, time in red-out, patient discomfort, recognition of pathology, occurrence of perforation, performance of retroflection. **GROUP 2: No intervention (n = 12)** 1. **Description of intervention:** No intervention. 2. **Observation, instruction, and feedback:** None.
Outcomes	**Time to assessment:** Within 2 weeks (range 2 to 14 days) of participants' simulator pre-test and training. **Assessment model:** 1 colonoscopy (insertion only, maximum 30 minutes) was completed under the supervision and evaluation of 1 of 3 blinded attending endoscopists (different from the pre-test examiner) who allowed the participants as much independence as possible while ensuring patient safety, and could provide verbal instruction if necessary. If, in the opinion of the attending, the resident was not making progress, the attending was permitted to take control of the colonoscope and navigate through the difficult section before returning it to the resident. If the test procedure was terminated due to patient factors (e.g. extensive diverticulosis), the resident was given the opportunity to repeat the procedure on a second suitable patient. **Details of patients used for live assessment:** Patients between the ages of 40 and 75 years with no previous colon or rectal resection, no history of difficult colonoscopy (secondary to anatomy or patient compliance), and no history of inflammatory bowel disease. **Outcome measures:** 1. Global performance score (rated by attending, 1-to-5 Likert scale of 7 domains: atraumatic technique, colonoscope advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure, overall performance). 2. Ability to independently reach the caecum (yes/no). 3. Number of critical flaws (perforation or significant bleeding) during the procedure (n).
Notes	**Funding:** Yes (peer-reviewed research grant).

Park 2007 *(Continued)*

Declarations of conflicts of interest for primary investigators: None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear: Method of sequence generation not specified. Quote: "...residents were randomly assigned to 1 of 2 groups."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians were blinded. Quote: "...under the supervision of 1 of 3 faculty endoscopist evaluators (different from the pre-test examiner) blinded to the residents training group."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: Accounted for missing outcome data. Quotes: “4 residents (2 in each group) were unable to complete the clinical phase because of scheduling difficulties, and their data were excluded from analyses.” and “Procedures were terminated on 1 occasion in each group because of patient-related factors (difficulty anatomy). Each of these residents performed a colonoscopy on a second suitable patient, and only evaluations from the second procedure were included in the analysis.”
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Unclear risk	Unclear: Use of an assessment instrument with no evidence of validity (there is insufficient evidence to suggest that this would have introduced bias). No intention-to-treat analysis (outcome not likely to be influenced by lack of intention-to-treat analysis).

Sedlack 2004

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** Not stated. **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Inadequate (physician assessors not blinded, patients not stated). **Inclusion of all randomised participants:** 100%.

Sedlack 2004 *(Continued)*

	Sample size calculation: None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** USA. **Year(s) participants randomised:** Not stated. **Number:** 8 randomised and analysed. **Inclusion criteria:** First-year gastroenterology fellows who had completed 2 months of EGD training. **Exclusion criteria:** Prior colonoscopy training or simulator experience. **Health profession:** Medical trainees (gastroenterology fellows). **Level of training:** First-year fellows. **Endoscopy experience:** 2 months of EGD training, no prior colonoscopy training or simulator experience. **Sex:** 5 males, 3 females. **Age:** Not stated.
Interventions	**Learning theory:** None stated. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 4)** 1. **VR simulator:** AccuTouch VR endoscopy simulator version 1.1 (Immersion Medical, Inc., Gaithersburg, MD, USA). 2. **Duration of training and/or training endpoint:** 6 hours (over 2 days). 3. **Description of intervention:** 6 hours of simulator training over a 2-day period, comprising a brief multimedia tutorial followed by the performance of 10 to 25 simulated colonoscopies (average 21, range 19 to 26). 6 colonoscopy scenarios of varying complexity were used. Simulator curriculum previously validated (Sedlack 2002). 4. **Observation, instruction, and feedback:** Not stated. It was not stated whether participants had access to the performance quality parameters generated by the simulator during practice. **GROUP 2: No intervention (n = 4)** 1. **Description of intervention:** No intervention (see 'Notes' section below). 2. **Observation, instruction, and feedback:** None.
Outcomes	**Time to assessment:** Not stated. **Assessment model:** 4 to 8 weeks of patient-based colonoscopy training during which participants were supervised and evaluated by 1 of 38 faculty gastroenterologists during one half-day (i.e. 4 hour) assignment intervals. Outcomes were compared between groups for procedures 1 to 15, 16 to 30, 31 to 45, and 46 to 60. **Details of patients used for live assessment:** Not specified. **Outcome measures:** 1. Time to reach maximum insertion (min). 2. Depth of unassisted insertion (1 = rectum, 2 = sigmoid, 3 = splenic flexure, 4 = hepatic flexure, 5 = caecum, 6 = terminal ileum). 3. Independent procedure completion (yes/no, defined as independently reaching the caecum or terminal ileum).

Sedlack 2004 *(Continued)*

4. Ability to identify endoscopic landmarks (rated by attending, 1-to-7 Likert scale, 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
5. Ability to insert in a safe manner (rated by attending, 1-to-7 Likert scale, 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
6. Ability to adequately visualise mucosa on withdrawal (rated by attending, 1-to-7 Likert scale, 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
7. Ability to respond appropriately to patient discomfort (rated by attending, 1-to-7 Likert scale, 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
8. Patient discomfort (rated by patient, 10-point scale: 1 = minimal or no pain, 10 = worst pain of life).
9. Faculty productivity during the training phase (number of procedures completed).
10. Faculty productivity during the assessment phase (number of procedures completed).

Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated. The authors state “the remaining 4 fellows served as a control group and underwent traditional colonoscopy training consisting of staff-supervised patient-based colonoscopy.” However, the performance of participants in both groups was evaluated (and compared) in the clinical setting from the first procedure they completed, therefore Group 2 was considered to have ‘no intervention’ prior to evaluation.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear: Method of sequence generation not specified. Quote: "8 fellows were randomly assigned to 1 of 2 different colonoscopy training curricula."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	High risk	Inadequate: Assessing physicians were not blinded to the training status of participants. It was not stated whether the assessing patients were blinded. Quote: “...evaluating staff were not blinded to the type of training curriculum that the fellow underwent...”
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Sedlack 2004a

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Flexible sigmoidoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2001 to 2002. **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Inadequate (physician assessors not blinded, patients not stated). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** USA. **Year(s) participants randomised:** Not stated. **Number:** 38 randomised and analysed. **Inclusion criteria:** Second-year internal medicine residents. **Exclusion criteria:** Prior endoscopy experience. **Health profession:** Medical trainees (internal medicine residents). **Level of training:** Second-year residents. **Endoscopy experience:** None. **Sex:** Not stated. **Age:** Not stated.
Interventions	**Learning theory:** None stated. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training followed by conventional patient-based endoscopy training (n = 19)** 1. **VR simulator:** AccuTouch VR endoscopy simulator version 1.1.1 (Immersion Medical, Inc., Gaithersburg, MD, USA). 2. **Duration of training and/or training endpoint:** 3 hours of simulator-based training followed by 6 hours (over 2 days) patient-based endoscopy training. 3. **Description of intervention:** 3 hours of simulator-based training under the supervision of a senior gastroenterology fellow, comprised of a brief multimedia tutorial followed by the performance of 8 to 10 simulated sigmoidoscopies (average 9, range 6 to 11). 6 sigmoidoscopy scenarios of varying complexity were used. Simulator training was followed by 2 additional afternoons (3 hours per day) of staff-supervised patient-based endoscopy training. 4. **Observation, instruction, and feedback:** a. Simulated setting: "Under the supervision of a senior gastroenterology fellow." It was not stated whether participants had access to the performance quality parameters generated by the simulator during practice. b. Clinical setting: "Staff-supervised."

Sedlack 2004a *(Continued)*

	GROUP 2: Conventional patient-based endoscopy training (n = 19) 1. **Duration of training and/or training endpoint:** 9 hours (over 3 days) patient-based endoscopy training. 2. **Description of intervention:** 3 afternoons (3 hours per day) of staff-supervised patient-based endoscopy training. 3. **Observation, instruction, and feedback:** "Staff-supervised."
Outcomes	**Time to assessment:** Not specified. **Assessment model:** 1 afternoon (3 hours) of staff-supervised patient-based endoscopy. **Details of patients used for live assessment:** Not specified. **Outcome measures:** 1. Patient discomfort (rated by patient, 1-to-10 Likert scale: 1 = no pain, 10 = worst pain of life). 2. Resident's ability to perform flexible sigmoidoscopy independently (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 3. Resident's ability to identify pathology (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 4. Resident's ability to identify landmarks (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 5. Resident's ability to respond to patient discomfort (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 6. Resident's ability to insert scope safely (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 7. Resident's ability to adequately visualise mucosa on withdrawal. 8. Resident's ability to routinely reach 40 cm (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 9. Resident's ability to perform biopsies (rated by attending and self rated, 1-to-10 Likert scale: 1 = strongly agree, 5 = neutral, 10 = strongly disagree). 10. Faculty productivity during training (number of procedures completed).
Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear: Method of sequence generation not specified. Quote: "19 subjects were randomly assigned to complete independently a 3-hour simulator-based training curriculum and the other 19 residents underwent staff-supervised patient-based training."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	High risk	Inadequate: Assessing physicians were not blinded to the training status of participants. It was not stated whether the assessing patients were blinded.

Sedlack 2004a *(Continued)*

		Quote: "...the evaluating staff was not blinded to the training curriculum undertaken by the residents..."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Sedlack 2007

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** EGD. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** Not stated. **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Adequate (physician assessors blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** USA. **Year(s) participants randomised:** Not stated. **Number:** 8 randomised and analysed. **Inclusion criteria:** First-year gastroenterology fellows. **Exclusion criteria:** Prior endoscopy or simulator experience. **Health profession:** Medical trainees (gastroenterology fellows). **Level of training:** First-year fellows. **Endoscopy experience:** None. **Sex:** Not stated. **Age:** Not stated.
Interventions	**Learning theory:** None stated. Participants were randomly assigned to 2 groups:

Sedlack 2007 *(Continued)*

GROUP 1: VR simulator training (n = 4)

1. **VR simulator:** GI Mentor II simulator (Simbionix USA, Cincinnati, OH, USA).
2. **Duration of training and/or training endpoint:** 6 hours (over 2 days).
3. **Description of intervention:** 6 hours of simulation training in EGD over 2 consecutive afternoons immediately prior to beginning patient-based training. Simulation training was comprised of a 15-minute introduction to the use of the simulator by a supervising staff member, followed by self directed, sequential progression through a curriculum consisting of 20 EGD simulation scenarios (2 modules consisting of 10 cases each). For the first case and every fourth case thereafter, the participant completed a standardised scenario (module 1, case 3) to allow tracking of learning curves during simulation training.Participants were required to complete at least 21 cases (average 22 cases, range 21 to 25).
4. **Observation, instruction, and feedback:** 15-minute introduction to the use of the simulator by a supervising staff member followed by self directed simulator use. It was not stated whether participants had access to the performance quality parameters generated by the simulator during practice.

GROUP 2: No intervention (n = 4)

1. **Description of intervention:** No intervention.
2. **Observation, instruction, and feedback:** None.

Outcomes

Time to assessment: Assessment began the day following simulation-based training and continued for 4 weeks.

Assessment model: The initial 4 weeks of staff-supervised patient-based EGD training. Each participant's performance was rated by the supervising staff member at the end of each training day, based on observation of the fellow's performance. Outcomes were compared between groups for procedures performed on days 1 to 5, 6 to 10, and 11 to 15.

Details of patients used for live assessment: Not specified.

Outcome measures:

1. Intubates safely (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
2. Reaches the second portion of the duodenum expediently (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
3. Completes the procedure without hands-on assistance (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
4. Uses sedation appropriately (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
5. Recognises and responds to patient discomfort (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree).
6. Is competent to perform EGD independently (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree).

Notes

Funding: Yes (research grant).

Declarations of conflicts of interest for primary investigators: None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear: Method of sequence generation not specified. Quote: "...carried out in a randomised, controlled trial, where each of the eight first-year fellows was randomly assigned to one of two possible EGD training curricula."

Sedlack 2007 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear: Participants instructed not to disclose their training status, but blinding was not confirmed. Quote: "Fellows were instructed not to reveal their arm of training to the evaluating staff but no other steps were specifically taken to ensure that evaluations were completed only by blinded staff members." and "although fellows were instructed not to disclose to their teaching staff the training arm to which they were assigned, specific blinding was not queried for individual evaluators."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Shirai 2008

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** EGD. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** October 2004 to March 2006. **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Adequate (physician assessors blinded). **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** Japan. **Year(s) participants randomised:** Not stated. **Number:** 20 randomised and analysed. **Inclusion criteria:** Residents rotating through gastroenterology.

Shirai 2008 *(Continued)*

	Exclusion criteria: Prior experience in performing endoscopy. **Health profession:** Medical trainees (internal medicine residents). **Level of training:** Not stated. **Endoscopy experience:** None. **Sex:** 1. VR simulator training group: 5 males, 5 females. 2. Conventional endoscopy training group: 6 males, 4 females. **Age (mean ± SD):** 1. VR simulator training group: 26 ± 0.77. 2. Conventional endoscopy training group: 27 ± 1.91.
Interventions	**Learning theory:** None stated. All participants received a 3-hour explanation regarding manipulation of an endoscope, endoscopic observation, and endoscopic diagnosis of common diseases. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training followed by conventional patient-based endoscopy training (n = 10)** 1. **VR simulator:** GI Mentor endoscopy simulator (Simbionix USA Corp., Cleveland, OH, USA). 2. **Duration of training and/or training endpoint:** 5, 1-hour simulator training sessions within 2 weeks followed by 15 hours bedside teaching. 3. **Description of intervention:** 5, 1-hour sessions of simulator training within 2 weeks. First, the level-1 EndoBubble and EndoBasket tasks were performed 3 times each, and then EGD training modules were completed. Case 1-1 was performed in each session, and the remaining time was used for other cases of the EGD module. Participants also received 15 hours of bedside training during which they could observe EGD performed by experienced doctors and work as an assistant, but were not allowed to perform EGD on patients. 4. **Observation, instruction, and feedback:** a. Simulated setting: "The residents were not supervised or instructed during the simulator training." It was not stated whether participants had access to performance quality parameters generated by the simulator during practice. b. Clinical setting: Staff-supervised, otherwise not specified. **GROUP 2: Conventional patient-based endoscopy training (n = 19)** 1. **Description of intervention:** 15 hours of bedside training during which participants could observe EGD performed by experienced doctors and work as an assistant, but were not allowed to perform EGD on patients. 2. **Observation, instruction, and feedback:** Staff-supervised, otherwise not specified.
Outcomes	**Time to assessment:** Not stated ("after completion of training schedules"). **Assessment model:** 2 EGD procedures carried out (within 1 week of each other) on volunteer patients without sedation, under the supervision and evaluation of 2 attending physicians who simultaneously assessed the procedures independently of each other. After the first evaluation, the supervisors gave the resident some advice (provided orally) to improve their skills. The time limit for each item assessed (see below), aside from insertion into the oesophagus and insertion in to the third part of the duodenum, was set at 2 min. Up to 3 attempts were allowed for insertion into the oesophagus, crossing the oesophagogastric junction, passing through the pyloric ring, and insertion into the third part of the duodenum. Instructions were provided when the supervisor considered the manoeuvre risky or when the endoscope re-

Shirai 2008 *(Continued)*

mained at the same site for 2 minutes or greater. A manoeuvre was defined as risky when there was a possibility of mucosal injury or perforation due to insertion of the endoscope without any confirmation of the position of the lumen. When the response to the instructions was inadequate, a supervisor assumed direct charge of the procedure until the next item at which time the participant resumed.

Details of patients used for live assessment:

Volunteers who were doctors and residents in the department. There was no significant difference in age or sex between the volunteers used within each group. Some of the volunteers had duodenal ulcer scars, hiatus hernia, or reflux oesophagitis, but the authors commented that these findings were not considered to influence the difficulty of performing EGD.

Outcome measures:

1. Total procedure time (min).
2. The following outcomes rated by 2 attendings (mean score used for analysis) using a 1-to-5 Likert scale: 1 = direct assistance by the supervisor was required; 2 = instructions were required; 3 = the resident could performed the manoeuvre without receiving instructions from the supervisor; 4 = skill was good, but not as good as that of the supervising physician; 5 = the resident could perform the manoeuvre as well as the supervising physician.
   a. Insertion into the oesophagus.
   b. Crossing the oesophagogastric junction (EGJ).
   c. Passing from the EGJ into the gastric antrum.
   d. Passing through the pyloric ring.
   e. Examination of the duodenal bulb.
   f. Insertion into the third part of the duodenum.
   g. Examination of the gastric antrum.
   h. Examination of the gastric angle.
   i. Manipulation for retroflexion.
   j. Looking down the gastric body.
   k. Viewing the fornix.

Notes	**Funding:** Yes (research grant). **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Adequate: Blinded random draw of numbers contained within sealed envelopes. Quote: "...a series of envelopes in a numbered sequence and with every second designated to training. Envelopes were drawn in a blinded fashion when each trainee was randomised." (personal correspondence) Quote: “10 residents were each randomised to simulator and non-simulator groups by envelopes.”
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified. Quote: “10 residents were each randomised to simulator and non-simulator groups by envelopes.”
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).

Shirai 2008 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Adequate: Assessing physicians and participating patients were blinded to the training status of participants. Quote: "The supervising physicians.... were unaware of whether the residents belonged to the simulator or non-simulator group." and "The volunteers did not know whether the residents were in the simulator group or not."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Tuggy 1998

Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Flexible sigmoidoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** Not stated. **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Not stated. **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** USA. **Year(s) participants randomised:** Not stated. **Number:** 10 randomised and analysed. **Inclusion criteria:** Family medicine residents. **Exclusion criteria:** Prior flexible sigmoidoscopy experience. **Health profession:** Family medicine residents. **Level of training:** Not stated. **Endoscopy experience:** None. **Sex:** Not stated. **Age:** Not stated.

Tuggy 1998 *(Continued)*

Interventions	**Learning theory:** None stated. Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 5)** 1. **VR simulator:** Gastro-Sim flexible sigmoidoscopy simulator (Interact Medical). 2. **Duration of training and/or training endpoint:** 10 hours total (5 prior to first live patient examination). 3. **Description of intervention:** 5 hours of simulation training prior to the first live patient examination and up to an additional 5 hours after the first live patient examination and prior to the second live patient examination. 4. **Observation, instruction, and feedback:** No guidance or training on the skills required for sigmoidoscopy other than what was encountered during the simulation. It was not stated whether participants had access to the performance quality parameters generated by the simulator during practice. **GROUP 2: No intervention (n = 5)** 1. **Description of intervention:** No intervention received prior to the first live patient. After the first live patient examination (and before the second), this group of residents was allowed to access the simulator to complete 5 hours of training. 2. **Observation, instruction, and feedback:** None.
Outcomes	**Time to assessment:** Not stated **Assessment model:** Residents were placed in matched pairs, consisting of 1 resident from Group 1 and 1 resident from Group 2. For the first examination, the 2 residents in each matched pair sequentially performed a flexible sigmoidoscopy procedure on the same patient to reduce the risk of encountering a different colon structure, which could affect performance. Residents were monitored by an experienced sigmoidoscopist who inserted and retracted the sigmoidoscope at the command of the resident. The trainee performed all steering and torque manoeuvres. Examinations were videotaped. For the second examination, the 2 residents in each matched pair once again sequentially performed a flexible sigmoidoscopy procedure on the same patient. During this second examination, the paired residents performed the procedure on the volunteer patient that they had not previously examined. **Details of patients used for live assessment:** 2 live patient volunteers who were healthy men aged 25 to 35 years who were compensated for their participation in the study. **Outcome measures:** 1. Time to reach 30 cm, 40 cm, and maximal insertion (seconds). 2. Total examination time (seconds). 3. Total time in red-out (seconds). 4. Quality of visualisation of the colon walls (rated by attending, 1-to-3 Likert scale: 1 = organised, 2 = adequate, 3 = haphazard). 5. Estimated percentage of the colon visualised (rated from the videotape, %). 6. Directional errors defined as the inability of the examiner to direct the sigmoidoscopy correctly toward the lumen when it was visualised (n). 7. Pain (rated by patient). 8. Perceived confidence of the examiner (rated by patient). 9. Duration of examination (rated by patient).
Notes	**Funding:** None stated. **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement

Tuggy 1998 *(Continued)*

Random sequence generation (selection bias)	Unclear risk	Unclear: Method of sequence generation not specified. Quote: "The volunteers were randomly assigned to an experimental (n = 5) and a matched control (n = 5 group)."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear: Participating patients were blinded to the experience and training status of participants; however, it is unclear whether the assessing physicians were blinded. Quote: “The patient was blinded to the experience of the examiner and to which arm of the study the trainee was assigned.” and “...before the examinations the residents read a prepared script.... requesting that they not reveal to which arm of the study they were assigned.”
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

Yi 2008

Methods	**Study design:** Quasi-randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** October 2006 to February 2007 **Generation of the allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Not stated. **Inclusion of all randomised participants:** 100%. **Sample size calculation:** None. **Intention-to-treat analysis:** Not stated.
Participants	**Country:** South Korea. **Year participants randomised:** Not stated.

Yi 2008 *(Continued)*

Number: 11 assigned to 2 groups and analysed.

Inclusion criteria: Not stated.

Exclusion criteria: Not stated.

Health profession: Medicine (fellows and residents).

Level of training: Not stated (fellows and residents).

Endoscopy experience: None.

Sex: 2 males, 9 females.

Age: Not stated.

Interventions	**Learning theory:** None stated. All participants received basic instruction for the operation of the colonoscope and colonoscopy. Participants were assigned (non-randomly) to 2 groups: **GROUP 1: VR simulator training (n = 5)** 1. **VR simulator:** KAIST-Ewha Colonoscopy Simulator II. 2. **Duration of training and/or training endpoint:** Until achievement of established training goals (scoring system based on performance criteria derived from experts' profiles). 3. **Description of intervention:** Participants practiced the targeted skills of colonoscopy using 2 training scenarios with different colon flexures and degrees of difficulty. Training scenario A was designed to teach practical skills to navigate the colon applying torque and up-down angulations. Scenario B was designed to teach skills to manage a loop formed in the sigmoid colon. Participants were required to practice until they reached all established training goals (scoring system based on performance criteria derived from experts' profiles). The average training time was 229.4 (range 82 to 377) minutes for scenario A (53.4 (range 26 to 100) procedures) and 232 (range 141 to 414) minutes for scenario B (68.2 (range 33 to 105) procedures). 4. **Observation, instruction, and feedback:** Not stated. It was not stated whether participants had access to performance quality parameters generated by the simulator during practice. **GROUP 2: No intervention (n = 6)** 1. **Description of intervention:** No intervention. 2. **Observation, instruction, and feedback:** None.
Outcomes	**Time to assessment:** Not stated. **Assessment model:** 5 colonoscopies under the supervision of experts. **Details of patients used for live assessment:** Average age was 49.6 (range 24 to 71) for the VR simulator training group and 53.5 (range 25 to 79) for the no-intervention group. **Outcome measures:** 1. Insertion time (min). 2. Success rate. 3. Number of red-outs. 4. Number of air inflations. 5. Number of loop formations. 6. Number of abdominal pressure applications. 7. Number of changes in patient posture. 8. Mucosal visualisation (rated by attending, 1-to-5 Likert scale: 1 = poor; 5 = excellent). 9. Overall performance accuracy (rated by attending, 1-to-5 Likert scale: 1 = poor; 5 = excellent). 10. Extent of abdominal pain (rated by patient, 1-to-5 Likert scale: 1 = no pain; 5 = worst pain).

Yi 2008 *(Continued)*

	11.Extent of abdominal inflation (rated by patient, 1-to-5 Likert scale: 1 = no pain; 5 = worst pain). 12.Extent of anus discomfort (rated by patient, 1-to-5 Likert scale: 1 = no pain; 5 = worst pain).
Notes	**Funding:** Yes (research grant). **Declarations of conflicts of interest for primary investigators:** None stated.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Inadequate: Non-random allocation. Quote: "The fellows and residents were divided in two groups."
Allocation concealment (selection bias)	Unclear risk	Unclear: Not specified.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Adequate: Unable to blind participants or personnel due to nature of intervention (outcome not likely to be influenced by lack of blinding).
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear: Not specified.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Adequate: No missing outcome data. Analysis was performed on all participants randomised.
Selective reporting (reporting bias)	Low risk	Adequate: Analysis and results are in accordance with the predefined study protocol.
Other bias	Low risk	Adequate: No sample size calculation and no intention-to-treat analysis (outcome not likely to be influenced by lack of sample size calculation and no intention-to-treat analysis).

EGD: oesophagogastroduodenoscopy
SD: standard deviation
SEM: standard error of the mean
VR: virtual reality

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Ahad 2011	Outcome in the simulated setting
Ahad 2013	Outcome in the simulated setting
Ahn 2016	A realism-validation study, not a randomised trial
Ansell 2013	A realism-validation study, not a randomised trial
Bai 2011	Outcome in the simulated setting

Study	Reason for exclusion
Bai 2012	Written in Chinese. We contacted the authors to request a translation but did not receive a reply.
Carot 2015	Aim was to determine the rate of detection of various colonic lesions by different colon screening techniques; was not related to virtual reality simulation training of trainees.
Carot 2016	Aim was to determine the rate of detection of various colonic lesions by different colon screening techniques; was not related to virtual reality simulation training of trainees.
Castells 2014	Aim was to determine the rate of detection of various colonic lesions by different colon screening techniques; was not related to virtual reality simulation training of trainees.
Ekkelenkamp 2016	Review
Elvevi 2012	Assessment validation study, not a randomised trial
Grover 2016	Outcome in the simulated setting
Hritz 2013	Assessment of endoscopic retrograde cholangiopancreatography skills
Jirapinyo 2014	Abstract from a scientific conference for which no published report of this trial was identified. We contacted the authors to request further information but did not receive a reply.
Jirapinyo 2015	Abstract from a scientific conference for which no published report of this trial was identified. We contacted the authors to request further information but did not receive a reply.
Jun 2013	Outcomes not directly compared between groups.
Kaltenbach 2011	Simulator used is not a virtual reality simulator.
Koch 2015	Outcomes not directly compared between groups.
Li 2012	Written in Chinese. We contacted the authors to request a translation but did not receive a reply.
Liao 2013	Assessment of endoscopic retrograde cholangiopancreatography skills
Lim 2011	Assessment of endoscopic retrograde cholangiopancreatography skills
Meng 2016	Assessment of endoscopic retrograde cholangiopancreatography skills
NCT01405443	Trial identified from a trial registry that was classified as 'awaiting assessment' in the previous version of this review. No corresponding published report. We contacted the authors to request further information but did not receive a reply.
Nehme 2013	Assessment of natural orifice transluminal endoscopic surgery skills
Plooy 2016	Simulator used is not a virtual reality simulator.
Qiao 2014	Systematic review
Santos 2017	Outcome in the simulated setting
Scaffidi 2018	Outcome in the simulated setting
Seshadri 2014	Outcome in the simulated setting

Study	Reason for exclusion
Singh 2014	Systematic review and meta-analysis
Snyder 2011	Outcome in an animal model
Strosberg 2017	Design and validation study, not a randomised trial
Van Sickle 2011	Outcome in the simulated setting
Williams 2015	Retrospective observational study, not a randomised trial

Characteristics of ongoing studies *[ordered by study ID]*

Grover 2017a

Trial name or title	A virtual reality curriculum in non-technical skills improves performance in simulated colonoscopy: a randomized trial
Methods	**Study design:** Prospective, randomised clinical trial. **Endoscopic procedure:** Colonoscopy. **Language of publication:** English. **Number of centres:** Single centre. **Year(s) of conduct of trial:** 2015 to 2016 **Generation of allocation sequence:** Not stated. **Allocation concealment:** Not stated. **Blinding of assessors:** Not stated. **Inclusion of all randomised participants:** Not stated. **Sample size calculation:** Not stated.
Participants	**Country:** Canada. **Year(s) participants randomised:** 2015 to 2016. **Inclusion criteria:** Postgraduate trainees from adult gastroenterology, general surgery, and internal medicine residency training programmes at the University of Toronto. **Exclusion criteria:** Performance of > 25 oesophagogastroduodenoscopy and/or colonoscopies in the clinical and/or simulated setting. **Health profession:** Medical trainees (internal medicine and general surgery residents, gastroenterology fellows). **Level of training:** Postgraduate years 2 to 4. **Endoscopy experience (average number of procedures):** Not stated. **Sex:** Not stated. **Age (mean ± SD):** Not stated.
Interventions	**Learning theory:** Not stated.

Grover 2017a *(Continued)*

	Participants were randomly assigned to 2 groups: **GROUP 1: VR simulator training (n = 21)** **- VR simulator:** EndoVR VR endoscopy simulator (CAE Healthcare Canada, Montreal, Quebec, Canada). - **Non-VR simulator:** Bench-top endoscopy simulator (Walsh 2008). **- Duration of training and/or training endpoint:** 7 hours of lectures and 6 hours of endoscopy VR simulation-based training. **- Description of intervention:** Participants spent 1 hour on a bench-top simulator and 5 hours on the VR simulator in addition to 7 hours of didactic sessions. 1 hour of the didactic teaching was dedicated to non-technical skills. They performed simulated cases in order of increasing difficulty. Participants also reviewed a checklist of tasks relevant to non-technical skills concepts prior to each integrated scenario case and were provided with dedicated feedback on their non-technical skills performance during the integrated scenario practice. **- Observation, instruction, and feedback:** Not stated. **GROUP 2: Another method of VR simulator training (n = 21)** **- VR simulator:** EndoVR VR endoscopy simulator (CAE Healthcare Canada, Montreal, Quebec, Canada). - **Non-VR simulator:** Bench-top endoscopy simulator (Walsh 2008). **- Duration of training and/or training endpoint:** 6 hours of lectures and 6 hours of endoscopy VR simulation-based training. **- Description of intervention:** Participants spent 1 hour on a bench-top simulator and 5 hours on the VR simulator in addition to 6 hours of didactic sessions. They performed simulated cases in order of increasing difficulty. **- Observation, instruction, and feedback:** Not stated.
Outcomes	**Time to assessment:** Assessment took place immediately and 4 to 6 weeks after training. Patient-based colonoscopies were only performed at the 4- to 6-week mark. **Assessment model:** 2 patient-based colonoscopies under the guidance of an expert endoscopist. **Details of patients used for live assessment:** Not stated. **Outcome measures:** (1) Procedural proficiency on 2 patient-based colonoscopies (rated by an expert endoscopist using the UK Joint Advisory Group colonoscopy Director Observation of Procedural Skills (JAG DOPS) assessment form) (4 to 6 weeks post-training) (JAG Central Office 2010). (2) Procedural knowledge assessed by multiple-choice tests (immediately post-training). (3) Procedural proficiency, communication skills, and global performance on simulated colonoscopies (immediately post-training and 4 to 6 weeks post-training) (rated by an expert endoscopist using the JAG DOPS, the Integrated Scenario Communication Rating Form (LeBlanc 2009), and the Integrated Scenario Global Rating Form (Hodges 2003), respectively). (4) Patient comfort during clinical colonoscopies as assessed by the Nurse-Assessed Patient Comfort Score (Rostom 2013). (5) Non-technical performance on 2 patient-based colonoscopies (rated by an expert endoscopist using the Modified Objective Structured Assessment of Nontechnical Skills (MOSANTS) (Dedy 2015).

Grover 2017a *(Continued)*	(6) Participant self efficacy (immediately post-training) as assessed by the General Self-Efficacy Scale (Chen 2001). (7) Practice case length on the simulator.
Starting date	June 2015
Contact information	Corresponding author: Dr Samir C Grover Address: 16-036 Cardinal Carter Wing, 30 Bond Street, St. Michael's Hospital, Toronto, Canada, ON M5B 1W8 Phone: 416-864-5628 Fax: 416-864-5882 Email: samir.grover@utoronto.ca
Notes	We contacted study authors for full details. All participants have completed the study. Data collection from videotaped performances of clinical endoscopic procedures is ongoing.

SD: standard deviation
VR: virtual reality

DATA AND ANALYSES

Comparison 1. Virtual reality endoscopy simulation training versus no training

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Composite score of competency	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2 Independent procedure completion: type of endoscopic procedure under study	6	815	Risk Ratio (M-H, Random, 95% CI)	1.62 [1.15, 2.26]
2.1 Colonoscopy	5	408	Risk Ratio (M-H, Random, 95% CI)	1.84 [1.35, 2.50]
2.2 Oesophagogastroduodenoscopy	1	407	Risk Ratio (M-H, Random, 95% CI)	1.25 [1.13, 1.39]
3 Independent procedure completion: level of participant endoscopy experience	6	815	Risk Ratio (M-H, Random, 95% CI)	1.62 [1.15, 2.26]
3.1 Limited prior training in endoscopy	3	329	Risk Ratio (M-H, Random, 95% CI)	1.82 [1.07, 3.12]
3.2 No prior training in endoscopy	3	486	Risk Ratio (M-H, Random, 95% CI)	1.32 [1.09, 1.61]
4 Performance time	2	29	Mean Difference (IV, Random, 95% CI)	-0.20 [-0.71, 0.30]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5 Patient discomfort: level of participant endoscopy experience	2	145	Std. Mean Difference (IV, Random, 95% CI)	-0.16 [-0.68, 0.35]
5.1 Limited prior training in endoscopy	1	90	Std. Mean Difference (IV, Random, 95% CI)	0.07 [-0.35, 0.49]
5.2 No prior training in endoscopy	1	55	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [1.00, 0.08]
6 Overall global rating of performance or competency	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
7 Visualisation of mucosa	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Analysis 1.1. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 1 Composite score of competency.

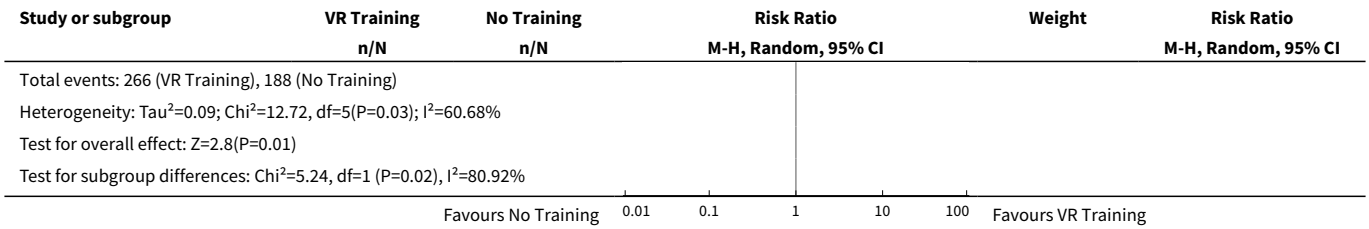
Study or subgroup	VR Training N	VR Training Mean(SD)	No Training N	No Training Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
Park 2007	12	17.9 (5.2)	12	14.8 (2.5)		3.1[-0.16,6.36]

Favours No Training -100 -50 0 50 100 Favours VR Training

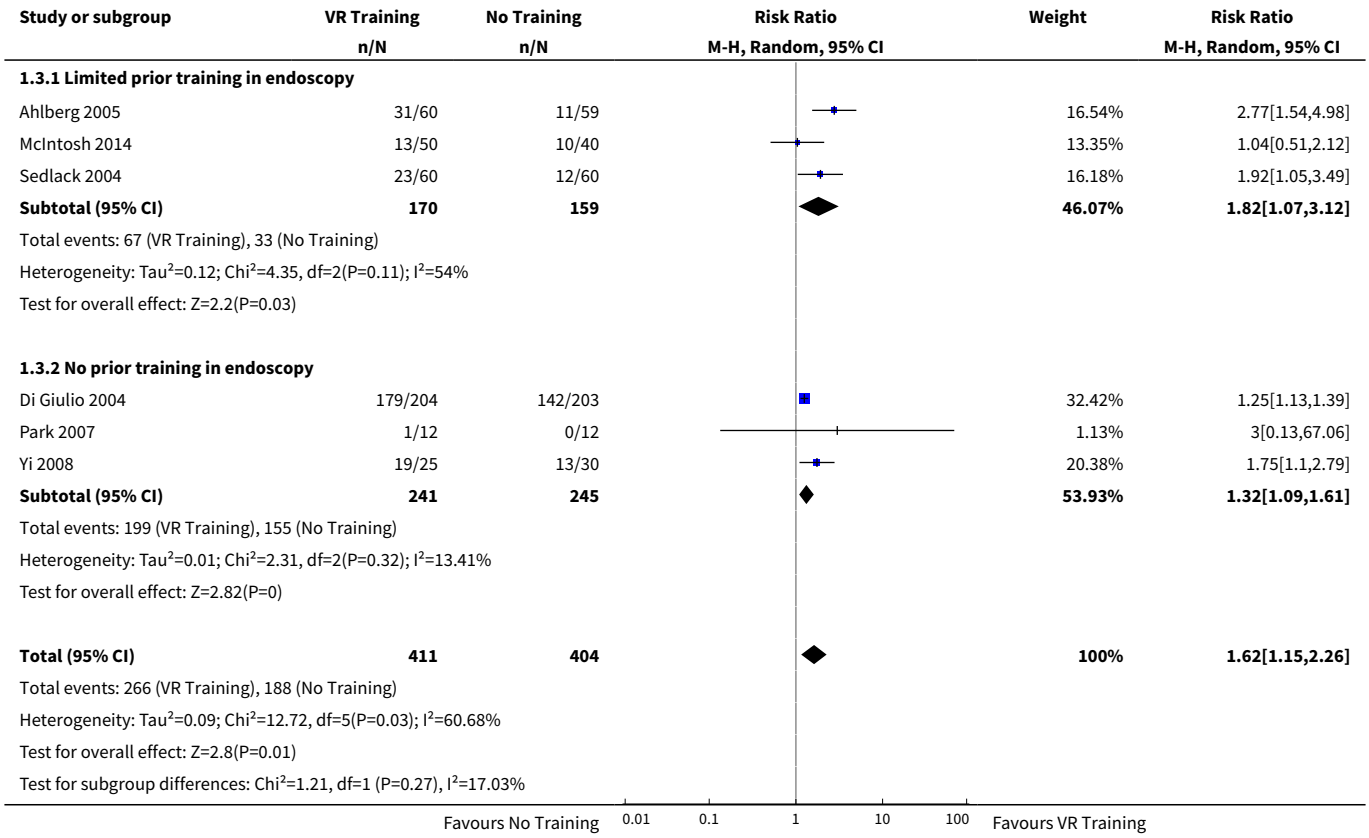
Analysis 1.2. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 2 Independent procedure completion: type of endoscopic procedure under study.

Study or subgroup	VR Training n/N	No Training n/N	Risk Ratio M-H, Random, 95% CI	Weight	Risk Ratio M-H, Random, 95% CI
1.2.1 Colonoscopy					
Ahlberg 2005	31/60	11/59		16.54%	2.77[1.54,4.98]
McIntosh 2014	13/50	10/40		13.35%	1.04[0.51,2.12]
Park 2007	1/12	0/12		1.13%	3[0.13,67.06]
Sedlack 2004	23/60	12/60		16.18%	1.92[1.05,3.49]
Yi 2008	19/25	13/30		20.38%	1.75[1.1,2.79]
Subtotal (95% CI)	**207**	**201**		**67.58%**	**1.84[1.35,2.5]**
Total events: 87 (VR Training), 46 (No Training)					
Heterogeneity: Tau^2=0.01; Chi^2=4.52, df=4(P=0.34); I^2=11.43%					
Test for overall effect: Z=3.85(P=0)					
1.2.2 Oesophagogastroduodenoscopy					
Di Giulio 2004	179/204	142/203		32.42%	1.25[1.13,1.39]
Subtotal (95% CI)	**204**	**203**		**32.42%**	**1.25[1.13,1.39]**
Total events: 179 (VR Training), 142 (No Training)					
Heterogeneity: Not applicable					
Test for overall effect: Z=4.28(P<0.0001)					
Total (95% CI)	**411**	**404**		**100%**	**1.62[1.15,2.26]**

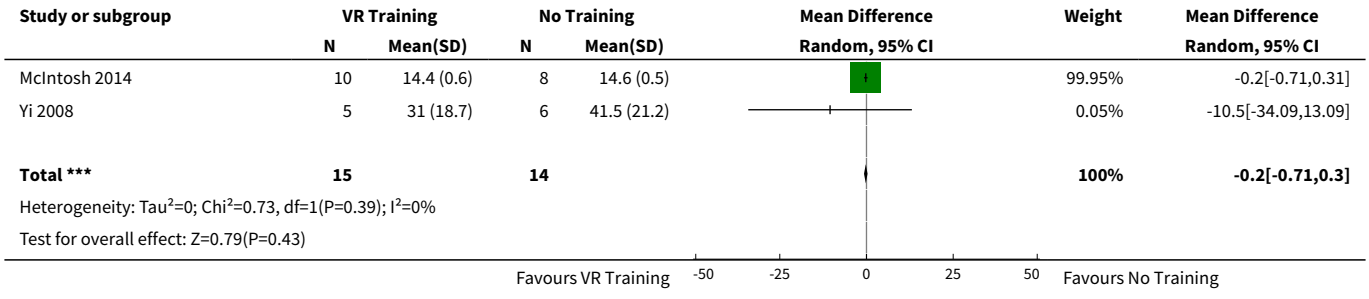
Favours No Training 0.01 0.1 1 10 100 Favours VR Training



Analysis 1.3. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 3 Independent procedure completion: level of participant endoscopy experience.

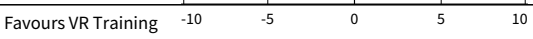


Analysis 1.4. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 4 Performance time.



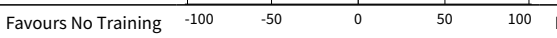
Analysis 1.5. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 5 Patient discomfort: level of participant endoscopy experience.

Study or subgroup	VR Training N	VR Training Mean(SD)	No Training N	No Training Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
1.5.1 Limited prior training in endoscopy							
McIntosh 2014	50	2 (0.5)	40	2 (0.3)		55.45%	0.07[-0.35,0.49]
Subtotal ***	**50**		**40**			**55.45%**	**0.07[-0.35,0.49]**
Heterogeneity: Not applicable							
Test for overall effect: Z=0.33(P=0.74)							
1.5.2 No prior training in endoscopy							
Yi 2008	25	2.9 (0.7)	30	3.3 (1)		44.55%	-0.46[-1,0.08]
Subtotal ***	**25**		**30**			**44.55%**	**-0.46[-1,0.08]**
Heterogeneity: Tau^2=0; Chi^2=0, df=0(P<0.0001); I^2=100%							
Test for overall effect: Z=1.66(P=0.1)							
Total ***	**75**		**70**			**100%**	**-0.16[-0.68,0.35]**
Heterogeneity: Tau^2=0.08; Chi^2=2.31, df=1(P=0.13); I^2=56.78%							
Test for overall effect: Z=0.63(P=0.53)							
Test for subgroup differences: Chi^2=2.31, df=1 (P=0.13), I^2=56.78%							



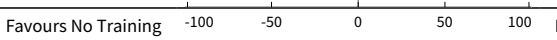
Analysis 1.6. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 6 Overall global rating of performance or competency.

Study or subgroup	VR Training N	VR Training Mean(SD)	No Training N	No Training Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
McIntosh 2014	10	2.4 (0.2)	8	2 (0.4)		0.45[0.15,0.75]



Analysis 1.7. Comparison 1 Virtual reality endoscopy simulation training versus no training, Outcome 7 Visualisation of mucosa.

Study or subgroup	VR Training N	VR Training Mean(SD)	No Training N	No Training Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
Yi 2008	25	3.5 (0.8)	30	2.9 (0.7)		0.6[0.2,1]

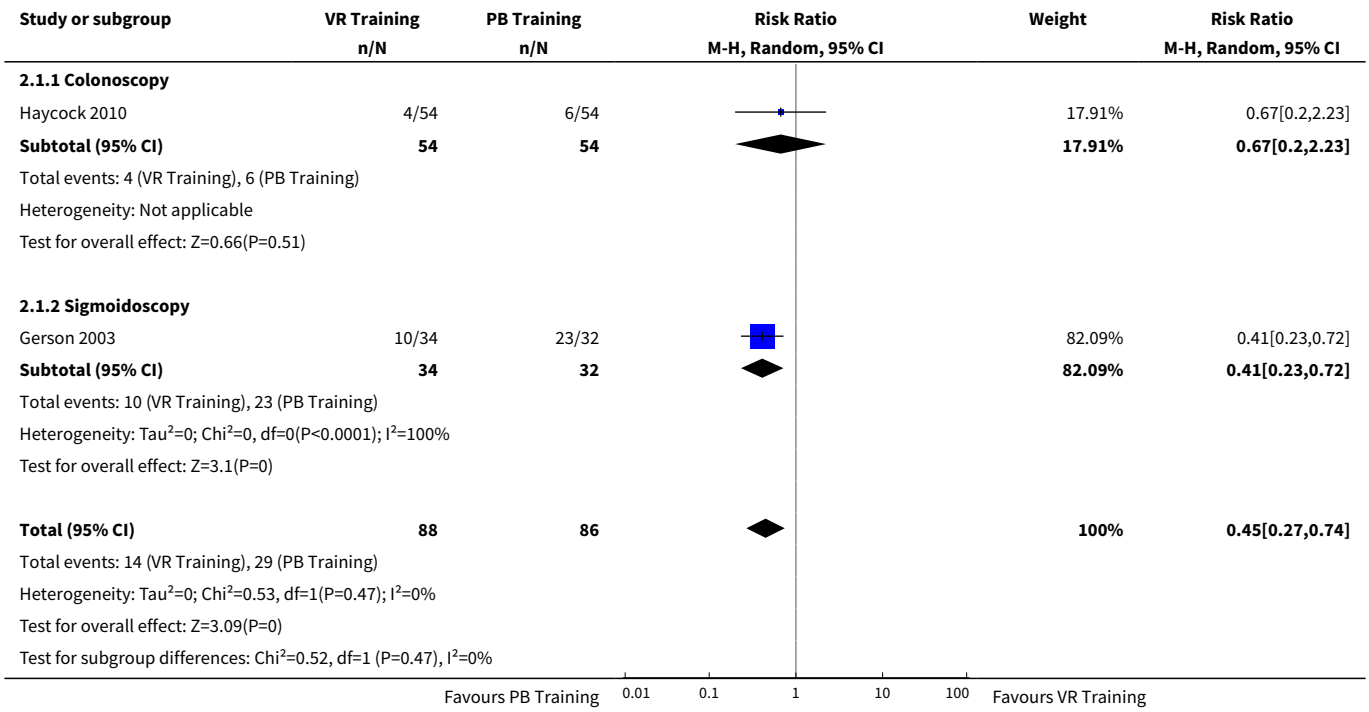


Comparison 2. Virtual reality endoscopy simulation training versus conventional patient-based training

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Independent procedure completion	2	174	Risk Ratio (M-H, Random, 95% CI)	0.45 [0.27, 0.74]

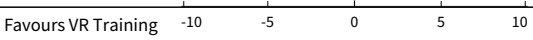
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Colonoscopy	1	108	Risk Ratio (M-H, Random, 95% CI)	0.67 [0.20, 2.23]
1.2 Sigmoidoscopy	1	66	Risk Ratio (M-H, Random, 95% CI)	0.41 [0.23, 0.72]
2 Performance time	2	34	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.55, 0.80]
2.1 Sigmoidoscopy	1	16	Std. Mean Difference (IV, Random, 95% CI)	0.0 [-0.99, 0.99]
2.2 Oesophagogastroduodenoscopy	1	18	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.69, 1.16]
3 Overall global rating of performance or competency	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4 Visualisation of mucosa	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Analysis 2.1. Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 1 Independent procedure completion.



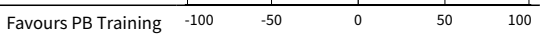
Analysis 2.2. Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 2 Performance time.

Study or subgroup	VR Training N	VR Training Mean(SD)	PB Training N	PB Training Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
2.2.1 Sigmoidoscopy							
Gerson 2003	9	24 (1)	7	24 (1.1)		46.88%	0[-0.99,0.99]
Subtotal ***	**9**		**7**			**46.88%**	**0[-0.99,0.99]**
Heterogeneity: Not applicable							
Test for overall effect: Not applicable							
2.2.2 Oesophagogastroduodenoscopy							
Ende 2012	9	16.1 (2.3)	9	15.4 (3.3)		53.12%	0.23[-0.69,1.16]
Subtotal ***	**9**		**9**			**53.12%**	**0.23[-0.69,1.16]**
Heterogeneity: Not applicable							
Test for overall effect: Z=0.5(P=0.62)							
Total ***	**18**		**16**			**100%**	**0.12[-0.55,0.8]**
Heterogeneity: Tau^2=0; Chi^2=0.11, df=1(P=0.73); I^2=0%							
Test for overall effect: Z=0.36(P=0.72)							
Test for subgroup differences: Chi^2=0.11, df=1 (P=0.73), I^2=0%							



Analysis 2.3. Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 3 Overall global rating of performance or competency.

Study or subgroup	VR Training N	VR Training Mean(SD)	PB Training N	PB Training Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
Gerson 2003	9	2.9 (4)	7	3.8 (3.1)		-0.9[-4.4,2.6]



Analysis 2.4. Comparison 2 Virtual reality endoscopy simulation training versus conventional patient-based training, Outcome 4 Visualisation of mucosa.

Study or subgroup	VR Training N	VR Training Mean(SD)	PB Training N	PB Training Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
Ende 2012	9	92 (6)	9	92 (7)		0[-6.02,6.02]

ADDITIONAL TABLES

Table 1. Details of training and assessment

Study	Simulator	Procedure	Training endpoint for VR simulator training group	Comparison group	Assessment in the clinical setting	Assessment scoring	Validity evidence of assessment
Ahlberg 2005	AccuTouch VR endoscopy simulator	Colonoscopy	Attainment of predefined expert level of performance on a VR examination case (1 to 2 hours VR training over at least 4 days, median total time = 20 hours)	No intervention	10 patient-based colonoscopies	**Objective:** (1) Time (time to reach caecum or total procedure time in unsuccessful cases)(2) Completed procedure rate (3) Segment of colon where procedure stopped(4) Analgesic drugs given (5) Complications **Rater-based (rated by *blinded assessor*):** (1) Reason for stopping procedure (if applicable) **Rater-based (rated by *blinded patient*):** (1) Maximum discomfort	Not stated
Cohen 2006	GI Mentor VR endoscopy simulator	Colonoscopy	10 hours VR training (5, 2-hour sessions over a maximum of 8 weeks)	No intervention	200 patient-based colonoscopies (or number performed prior to study completion). Outcomes were compared for every group of 20 cases (i.e. procedures 0 to 20, 21 to 40, 41 to 60, etc.).	**Objective:** (1) Objective competence defined as (a) ability to reach transverse colon and caecum without assistance and (b) ability to correctly recognise and identify abnormalities (2) Median number cases required to reach 90% competence **Rater-based (rated by blinded *assessor*):** (1) Overall rating of competency (2) Patient discomfort	Authors report evaluation form (rating ability to reach transverse colon and caecum, ability to correctly recognise and identify abnormalities, and overall competency) used in previous study (Cass 1996). Other measures not stated.

Table 1. Details of training and assessment *(Continued)*

						Rater-based (*self rated*): (1) Usefulness of simulation training	
Di Giulio 2004	GI Mentor VR endoscopy simulator	EGD	10 hours VR training (over 3 to 5 sessions)	No intervention	20 consecutive patient-based EGDs	**Objective:** (1) Number of times manual assistance required (2) Number of times verbal assistance required (3) Number of identified or missed lesions (4) Number of complications (5) Failure to effect oesophageal intubation (6) Number of attempts at oesophageal intubation **Rater-based (rated by *non-blinded assessor*):** (1) Completeness of procedure (2) Overall judgement of performance	Not stated
Ende 2012	GI Mentor VR endoscopy simulator	EGD	18 to 20 hours over 9 to 10 sessions and conventional patient-based training over 4 months (average of 29 ± 21 EGDs)	Comparison group 1: conventional patient-based training over 4 months (average of 19 ± 18 EGDs) Comparison group 2: VR simulator training only (18 to 20 hours over 9 to 10 sessions)	3 patient-based EGDs	**Objective:** (1) Time to reach descending duodenum (2) Procedure times (for oesophageal intubation, to pass the pylorus, to reach the descending duodenum, overall procedure time) (3) Percentage of estimated visualised mucosal surface (4) Incidence of complications **Rater-based (rated by *1 blinded and 1 non-blinded assessor*):** (1) Endoscopic skill	Not stated

Table 1. Details of training and assessment *(Continued)*

Ferlitsch 2010	GI Mentor VR endoscopy simulator	EGD	2 hours VR training per day for 5 to 20 hours total (range 5 to 20 hours, median 10 hours)	No intervention	10 consecutive patient-based EGDs. Outcomes were compared for procedures 1 to 10 and 51 to 60.	**Objective:** (1) Total time (2) Time to reach descending duodenum (3) Diagnostic accuracy **Rater-based (rated by *non-blinded assessor*):** (1) Intubation of oesophagus completed "unaided", with "expert help", or "expert takeover" (2) Pyloric passage completed "unaided", with "expert help", or "expert takeover" (3) Retroflexion in gastric fundus completed "unaided", with "expert help", or "expert takeover" **Rater-based (rated by *blinded patient*):** (1) Discomfort	The authors stated that the "parameters chosen in our evaluation were suitable for discriminating endoscopic examinations performed by experts from those performed by beginners, documenting the validity of the method."
Gerson 2003	AccuTouch VR endoscopy simulator	Sigmoidoscopy	2 weeks unlimited VR training (average time (mean ± SEM): 138 ± 28 minutes)	Conventional patient-based training (10 sigmoidoscopies in clinical setting over 2 weeks)	5 patient-based sigmoidoscopies	**Objective:** (1) Independent completion (2) Examination duration (3) Requirement for assistance (4) Flexure recognition (5) Completion of retroflexion (6) Ability to recognise pathology **Rater-based (rated by *non-blinded assessor*):** (1) Expert global rating **Rater-based (rated by *blinded patient*):** (1) Level of patient comfort/discomfort	Not stated

Table 1. Details of training and assessment *(Continued)*

						(2) Patient satisfaction (3) Technical competence	
Gomez 2015	GI Mentor VR endoscopy simulator	Colonoscopy	3 weeks unlimited VR (required to complete 2 practice tests and 1 of 10 simulated colonoscopy cases) and non-VR simulator training (required to complete 1 of 6 simulated colonoscopy cases)	Comparison group 1: 3 weeks unlimited VR training (required to complete 2 practice tests and 1 of 10 simulated colonoscopy cases) Comparison group 2: 3 weeks unlimited non-VR simulator training (required to complete 1 of 6 simulated colonoscopy cases)	1 patient-based colonoscopy	**Objective:** (1) Total time (2) Time to reach the caecum (3) Time with a clear view of the lumen (4) Number of times faculty took control of the colonoscope (5) Number of times there was a need for endoscopic instrumentation **Rater-based (rated by *blinded assessor*):** (1) Global Assessment of Gastrointestinal Endoscopic Skills - Colonoscopy (GAGES-C) tool (Vassiliou 2010)	Validity evidence of the GAGES-C tool has been assessed (Vassiliou 2010). Other measures not stated.
Grover 2015	AccuTouch VR endoscopy simulator	Colonoscopy	6 hours of lectures and 8 hours VR training	8 hours of VR training	2 patient-based colonoscopies	**Objective:** (1) Knowledge of endoscopy **Rater-based (rated by *blinded assessor*):** (1) UK JAG Colonoscopy DOPS assessment form on clinical colonoscopy (JAG Central Office 2010) (2) JAG DOPS assessment form on simulated colonoscopy (3) ISCRF on simulated colonoscopy (LeBlanc 2009) (4) Integrated Scenario Global Rating Form on simulated colonoscopy (Hodges 2003)	Validity evidence of the UK JAG DOPS has been assessed (Barton 2008; Barton 2012). Validity evidence of the ISCRF has been studied in other settings (Hodges 2003), but not in endoscopy. Other measures not stated.

Table 1. Details of training and assessment *(Continued)*

Grover 2017	AccuTouch VR endoscopy simulator	Colonoscopy	4 hours of lectures and 6 hours VR training	4 hours of lectures and 6 hours VR training	2 patient-based colonoscopies	**Objective:** (1) Knowledge of endoscopy **Rater-based (rated by *blinded assessor*):** (1) UK JAG Colonoscopy DOPS assessment form on clinical colonoscopy (JAG Central Office 2010) (2) JAG DOPS assessment form on simulated colonoscopy (3) ISCRF on simulated colonoscopy (LeBlanc 2009) (4) Integrated Scenario Global Rating Form on simulated colonoscopy (Hodges 2003)	Validity evidence of the UK JAG DOPS has been assessed (Barton 2008; Barton 2012). Validity evidence of the ISCRF has been studied in other settings (Hodges 2003), but not in endoscopy. Other measures not stated.
Haycock 2010	Endo TS-1 Olympus colonoscopy simulator	Colonoscopy	16 hours VR training	Conventional patient-based training (16 hours, minimum 8 colonoscopies)	3 patient-based colonoscopies	**Objective:** (1) Time to completion (2) Depth of insertion **Rater-based (rated by *blinded assessor*):** (1) Modified JAG DOPS assessment form (JAG Central Office 2010) (2) Global Performance Score (Park 2007)	Validity evidence of the UK JAG DOPS has been assessed for the assessment tool as a whole (Barton 2008; Barton 2012); however, validity evidence of the abbreviated version utilised in this study has not been studied. The authors report the Global Performance Score is "validated"; however, no details of validity evidence were provided in reference source (Park 2007).

Table 1. Details of training and assessment *(Continued)*

							Other measures not stated.
McIntosh 2014	GI Mentor VR endoscopy simulator	Colonoscopy	10 to 20 hours VR training over 4 weeks	No intervention	5 patient-based colonoscopies	**Objective:** (1) Number of proctor assists required (2) Total time (3) Depth of insertion (4) Caecal intubation rate	Not stated
Park 2007	AccuTouch VR endoscopy simulator	Colonoscopy	2 to 3 hours VR training	No intervention	1 patient-based colonoscopy	**Objective:** (1) Ability to independently reach the caecum (2) Number of critical flaws **Rater-based (rated by *blinded assessor*):** (1) Global Performance Score	Authors report Global Performance Score is "validated"; however, no reference or details of validity evidence were provided. Other measures not stated.
Sedlack 2004	AccuTouch VR endoscopy simulator	Colonoscopy	6 hours VR training over 2 days. Previously validated curriculum (Sedlack 2002)	No intervention	4 to 8 weeks of patient-based colonoscopy training. Outcomes were compared between groups for procedures 1 to 15, 16 to 30, 31 to 45, and 46 to 60.	**Objective:** (1) Time to reach maximum insertion (2) Depth of insertion (3) Independent procedure completion (4) Faculty productivity **Rater-based (rated by *non-blinded assessor*):** (1) Ability to identify endoscopic landmarks (2) Ability to insert in a safe manner (3) Ability to adequately visualise mucosa on withdrawal (4) Ability to respond appropriately to patient discomfort	Not stated

Table 1. Details of training and assessment *(Continued)*

						Rater-based (rated by *patient, unclear if blinded*): (1) Patient discomfort	
Sedlack 2004a	AccuTouch VR endoscopy simulator	Sigmoidoscopy	3 hours VR training followed by 6 hours (over 2 days) patient-based endoscopy training	Conventional patient-based training (9 hours over 3 days)	3 hours of patient-based flexible sigmoidoscopy	**Objective:** (1) Faculty productivity **Rater-based (rated by *non-blinded assessor* and*self rated*):** (1) Resident's ability to respond to patient discomfort (2) Resident's ability to perform flexible sigmoidoscopy independently (3) Resident's ability to identify pathology (4) Resident's ability to identify landmarks (5) Resident's ability to insert scope safely (6) Resident's ability to adequately visualise mucosa on withdrawal (7) Resident's ability to routinely reach 40 cm (8) Resident's ability to perform biopsies **Rater-based (rated by *patient, unclear if blinded*):** (1) Patient discomfort	Not stated
Sedlack 2007	GI Mentor VR endoscopy simulator	EGD	6 hours VR training (over 2 days)	No intervention	4 weeks patient-based EGD training. Outcomes were compared between groups for procedures performed on days 1 to 5, 6	**Objective:** None **Rater-based (rated by *assessor, unclear if blinded*):** (1) Intubates safely(2) Reaches the second portion of the duodenum expediently (3) Completes the procedure without hands-on assistance	Not stated

Table 1. Details of training and assessment *(Continued)*

					to 10, and 11 to 15.	(4) Uses sedation appropriately (5) Recognises and responds to patient discomfort(6) Is competent to perform EGD independently	
Shirai 2008	GI Mentor VR endoscopy simulator	EGD	5, 1-hour VR training sessions over 2 weeks plus 15 hours patient-based training (observed or assisted)	Conventional patient-based training (15 hours, observed or assisted)	2 patient-based EGDs	**Objective:** (1) Total procedure time **Rater-based (rated by *blinded assessor*)**: (1) Insertion into the oesophagus (2) Crossing the oesophagogastric junction (3) Passing from the oesophagogastric junction into the gastric antrum (4) Passing through the pyloric ring (5) Examination of the duodenal bulb (6) Insertion into the third part of the duodenum (7) Examination of the gastric antrum (8) Examination of the gastric angle (9) Manipulation for retroflexion (10) Looking down the gastric body (11) Viewing the fornix	Not stated
Tuggy 1998	Gastro-Sim VR endoscopy simulator	Sigmoidoscopy	5 hours VR training	No intervention	1 patient-based flexible sigmoidoscopy	**Objective:** (1) Time to reach 30 cm, 40 cm, and maximal insertion (2) Total examination time (3) Total time in red-out **Rater-based (rated by *assessor, unclear if blinded*):**	Not stated

Table 1. Details of training and assessment *(Continued)*

						(1) Estimated percentage of colon visualised (2) Number of directional errors (3) Quality of visualisation of colon walls **Rater-based (rated by *blinded patient*):** (1) Pain (2) Perceived confidence of the examiner (3) Duration of examination	
Yi 2008	KAIST-Ewha Colonoscopy Simulator	Colonoscopy	Attainment of predefined expert level of performance on VR simulator (2 practice scenarios, mean practice time 229.4 (53.4 procedures) and 232 minutes (68.2 procedures) for scenario A and B)	No intervention	5 patient-based colonoscopies	**Objective:** (1) Insertion time (2) Success rate (3) Number of red-outs (4) Number of air inflations (5) Number of loop formations (6) Number of abdominal pressure applications (7) Number of changes in patient posture **Rater-based (rated by *assessor, unclear if blinded*):** (1) Mucosal visualisation (rated by attending) (2) Overall performance accuracy (3) Extent of abdominal pain (4) Extent of abdominal inflation (5) Extent of anus discomfort	Not stated

DOPS: Direct Observation of Procedural Skills
EGD: oesophagogastroduodenoscopy

ISCRF: Integrated Scenario Communication Rating Form
JAG: Joint Advisory Group
SEM: standard error of the mean
VR: virtual reality

Table 2. Summary of outcomes - composite score of competency in performing endoscopy

Study	Procedure	Comparison group(s)	Method	VR versus no training	VR versus conventional endoscopy training	VR versus another form of endoscopy simulation	VR versus another method of VR training
Gomez 2015	Colonoscopy	Comparison group 1: VR training only Comparison group 2: another form of endoscopy simulation only	Procedural proficiency (rated by an expert endoscopist using the GAGES-C tool, which rated 5 domains (scope navigation; strategies for scope advancement; clear field; instrumentation (when performed); and overall quality) on 1-to-5-point scales)	-	-	No significant differences in GAGES-C scores (P value of ANOVA not reported). Numerical GAGES-C values not reported.	No significant differences in GAGES-C scores (P value of ANOVA not reported). Numerical GAGES-C values not reported.
Grover 2015	Colonoscopy	Another method of VR training	Procedural proficiency (rated by 2 expert endoscopists using JAG DOPS colonoscopy assessment form, which rated 20 items based on 4 domains (assessment, consent, communication; safety and sedation; endoscopic skills during insertion and withdrawal; and diagnostic and therapeutic ability) on 1-to-4-point scales)	-	-	-	Mean JAG DOPS scores for procedure 1: 72.2 (SD 10.9) and procedure 2: 71.9 (SD 16.7) in intervention group, and for procedure 1: 31.8 (SD 14.8) and procedure 2: 32.3 (SD 18.3) in control group. Intervention group had significantly higher scores, P < 0.001.
Grover 2017	Colonoscopy	Another method of VR training	Procedural proficiency (rated by 2 expert endoscopists using JAG DOPS colonoscopy assessment form, which rated 20 items based on 4 domains (assessment, consent, communication; safety and sedation; endoscopic skills during insertion and withdrawal; and diagnostic and therapeutic ability) on a 1-to-4 point scale)	-	-	-	**Assessor 1**: Mean JAG DOPS scores for procedure 1: 72.2 (SD 12.1) and procedure 2: 72.3 (SD 11.1) in intervention group, and for procedure 1: 58.3 (SD 8.3) and procedure 2: 58.2 (SD 13.4) in control group. Intervention group had

Table 2. Summary of outcomes - composite score of competency in performing endoscopy *(Continued)*

							significantly higher scores, P < 0.001. **Assessor 2**: Mean JAG DOPS scores for procedure 1: 64.3 (SD 4.1) and procedure 2: 64.3 (SD 3.3) in intervention group, and for procedure 1: 59.7 (SD 7.1) and procedure 2: 60.0 (SD 5.4) in control group. Intervention group had significantly higher scores, P = 0.006.
Haycock 2010	Colonoscopy	Conventional patient-based training	1) Procedural proficiency (rated by attending using abbreviated version of UK JAG DOPS colonoscopy assessment form, which rated 9 domains of "endoscopic skills during insertion and withdrawal" on 1-to-4 point scales) 2) Global performance score (rated by attending, 7 domains rated on a 1-to-5 Likert scale: atraumatic technique, colonoscope advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure, overall performance)	-	1) Procedural proficiency (JAG DOPS). Median score 16 (IQR (14 to 22) for VR group versus 18 (IQR 14 to 21) for control group. No significant difference between groups, P = 0.92 2) Global performance. Median score 18 (IQR 14 to 19) for VR group versus 17 (IQR 14 to 19) for control group.	-	-

Table 2. Summary of outcomes - composite score of competency in performing endoscopy *(Continued)*

				No significant difference between groups, P = 0.35			
Park 2007	Colonoscopy	No training	Global performance score (rated by attending, 7 domains rated on a 1-to-5 Likert scale: atraumatic technique, colonoscope advancement, use of instrument controls, flow of procedure, use of assistants, knowledge of specific procedure, overall performance)	Mean score 17.9 (SD 5.2) for VR group versus 14.8 (SD 2.5) for control group. SMD 0.73 (-0.10, 1.57) VR trained group had significantly higher scores, P = 0.04.	-	-	-

ANOVA: analysis of variance
DOPS: Direct Observation of Procedural Skills
GAGES-C: Global Assessment of Gastrointestinal Endoscopic Skills - Colonoscopy
IQR: interquartile range
JAG: Joint Advisory Group
SD: standard deviation
SMD: standardised mean difference
VR: virtual reality

Table 3. Summary of outcomes - independent procedure completion

Study	Procedure	Comprison group	Method	VR versus no training	VR versus conventional endoscopy training
Ahlberg 2005	Colonoscopy	No training	Completed procedure rate (intubation of caecum within given time limit)	RR 2.77 (1.54, 4.98) VR trained group completed significantly more procedures independently, P = 0.0011.	-
Di Giulio 2004	EGD	No training	Number of complete procedures (completeness of procedure (rated by attending, "complete" = oesophageal intubation achieved, participant identified, within 20 minutes, all anatomical landmarks (oesophagogastric mucosal junction, gastric angulus, pylorus) and performed certain basic manoeuvres (aspiration of gastric juice, pylorus intubation in no more than 3 attempts, duodenal bulb exploration, intubation of the second part of the duodenum and retroflexion) without verbal direction)	RR 1.25 (1.13, 1.39) VR trained group completed significantly more procedures independently, P < 0.001.	-
Gerson 2003	Sigmoidoscopy	Conventional patient-based training	Independent completion (yes/no)	-	RR 0.41 (0.23, 0.72) VR trained group completed significantly fewer procedures, P = 0.02.
Haycock 2010	Colonoscopy	Conventional patient-based training	Completion of case – insertion to caecum independently (yes/no)	-	RR 0.67 (0.20, 2.23) No significant difference between groups, P = 0.51
McIntosh 2014	Colonoscopy	No training	Completed procedure rate (intubation of caecum within given time limit)	RR 1.04 (0.51, 2.12) No significant difference between groups, P = 0.06	-
Park 2007	Colonoscopy	No training	Ability to independently reach the caecum (yes/no)	RR 3.00 (0.13, 67.06) No significant difference between groups, P > 0.05	-

Table 3. Summary of outcomes - independent procedure completion *(Continued)*

Sedlack 2004	Colonoscopy	No training	Independent procedure completion defined as independently reaching the caecum or terminal ileum (yes/no)	RR 1.92 (1.05, 3.49) VR trained group completed significantly more procedures, P = 0.027 (procedures 1 to 15).	-
Yi 2008	Colonoscopy	No training	Success rate	RR 1.75 (1.10, 2.79) VR trained group completed significantly more procedures, P = 0.006.	-

EGD: oesophagogastroduodenoscopy
RR: risk ratio
VR: virtual reality

Table 4. Summary of outcomes - performance time

Study	Procedure	Comparison group	Method	VR versus no training	VR versus conventional endoscopy training	VR versus another form of endoscopy simulation
Ahlberg 2005	Colonoscopy	No training	Time to reach caecum in successful cases (min)	Median 30 min (IQR 17 to 38) for VR group versus 40 min (IQR 25 to 45) for control group. VR trained group significantly faster, P = 0.008	-	-
Di Giulio 2004	EGD	No training	Duration of procedure (defined as the length of time the light source was switched on)	Mean 10.5 min for VR group versus 12.4 min for control group. No significant difference between groups, P > 0.05	-	-
Ende 2012	EGD	Comparison group 1: conventional patient-based training; Comparison group 2: VR training only	Time to reach the descending part of the duodenum	-	Mean 822 s (SD 163) for VR plus conventional training group versus 922 s (SD 186) for conventional training-only group versus 968 s (SD 139) for VR training-only group. No significant difference between groups, P = 0.201	-
Ferlitsch 2010	EGD	No training	Time between the first attempt at oesophageal	Mean 239 s (range 50 to 620) for VR group versus 310 s	-	-

Table 4. Summary of outcomes - performance time *(Continued)*

			intubation until the descending part of duodenum was reached (measured after 10 endoscopic examinations)	(range 110 to 720) for control group. VR trained group significantly faster, P < 0.001 (procedures 1 to 10)		
Gerson 2003	Sigmoidoscopy	Conventional patient-based training	Examination duration (min)	-	Mean 24 min (SEM 1.0) for VR group versus 24 min (SEM 1.1) for control group SMD 0.00 (-0.99, 0.99) No significant difference between groups, P > 0.05	-
Gomez 2015	Colonoscopy	Comparison group 1: VR training only;Comparison group 2: another form of endoscopy simulation only	Time to reach caecum in successful cases (min)	-	-	Median 23.7 min for VR plus another endoscopy simulator group versus 23.9 for VR training-only group versus 28.2 for another endoscopy simulator-only group. No significant difference between groups, P = 0.084
Haycock 2010	Colonoscopy	Conventional patient-based training	Time to completion in complete cases	-	Median 20 min (IQR 20 to 20) for VR group versus 20 min (IQR 19 to 20) for control group. No significant difference between groups, P = 0.11	-
McIntosh 2014	Colonscopy	No training	Total insertion time (min)	Mean 14.4 min (SD 0.6) for VR group versus 14.6 (SD 0.5) for control group.	-	-

Table 4. Summary of outcomes - performance time *(Continued)*

				No significant difference between groups, P = 0.37		
Sedlack 2004	Colonoscopy	No training	Time to reach maximum insertion (min)	Median 23 min (IQR 19 to 30) for VR group versus 23 min (IQR 20 to 30) for control group. No significant difference between groups, P = 0.16 (procedures 1 to 15)	-	-
Shirai 2008	EGD	Conventional patient-based training	Total procedure time (min)	-	14:40 min (12:15 to 16:07) for VR group versus 14:05 min (13:30 to 16:00) for control group. No significant difference between groups, P > 0.05	-
Tuggy 1998	Sigmoidoscopy	No training	Total examination time (seconds)	**5 hours VR training:** Mean 530 s for VR group after 5 hours training versus 654 s for control group. No significant difference between groups, P = 0.31 **6 to 10 hours VR training:** Mean 323 s for VR group after 6 to 10 hours training versus 654 s for control group. VR group significantly faster, P = 0.01	-	-
Yi 2008	Colonoscopy	No training	Total insertion time (min)	Mean 31 min (SD 18.7) for VR group versus 41.5 min (SD 21.2) for control group SMD -0.48 (-1.69, 0.74) VR trained group significantly faster, P = 0.028	-	-

EGD: oesophagogastroduodenoscopy
IQR: interquartile range
SD: standard deviation
SEM: standard error of the mean
SMD: standardised mean difference
VR: virtual reality

Table 5. Summary of outcomes - complication or critical flaw occurrence

Study	Procedure	Comparison group	Method	VR versus no training	VR versus conventional endoscopy training
Ahlberg 2005	Colonoscopy	No training	Complications (n)	No complications in either group. No significant difference between groups, P > 0.05	-
Di Giulio 2004	EGD	No training	Complications (n)	No complications in either group. No significant difference between groups, P > 0.05	-
Ende 2012	EGD	Comparison group 1: conventional patient-based training; Comparison group 2: VR training only	Complications (n)	-	No complications in any of the 3 groups. No significant difference between groups, P > 0.05
Gerson 2003	Sigmoidoscopy	Conventional patient-based training	Adverse events (n)	-	No adverse events occurred in either group. No significant difference between groups, P > 0.05
Park 2007	Colonoscopy	No training	Number of critical flaws (perforation or bleeding) during the procedure (n)	No complications in either group. No significant difference between groups, P > 0.05	-
Sedlack 2004a	Sigmoidoscopy	Conventional patient-based training	Number of adverse events (n)	-	No complications in either group. No significant difference between groups, P > 0.05

EGD: oesophagogastroduodenoscopy

Table 6. Summary of outcomes - patient discomfort

Study	Procedure	Comparison group	Method	VR versus no training	VR versus conventional endoscopy training
Ahlberg 2005	Colonoscopy	No training	Maximum discomfort (rated by patient, visual analogue scale)	Median 4 (IQR 2.5 to 6) for VR group versus 5 (IQR 4 to 7) for control group. Significantly less pain in VR trained group, P = 0.02	-

Table 6. Summary of outcomes - patient discomfort *(Continued)*

Cohen 2006	Colonoscopy	No training	Patient discomfort level (rated by attending, 1-to-5 Likert scale: 1 = very comfortable to 5 = severe pain)	Mean 25.7 for VR group versus 31.4 for control group. No significant difference between groups, P = 0.42 (procedures 1 to 20)	-
Ferlitsch 2010	EGD	No training	Pain and discomfort (rated by patient, 2 separate 10-centimetre visual analogue scales for pain and discomfort)	**Discomfort:** Median discomfort for 1st 10 procedures was 16 (range 0 to 98) for VR group versus 20 (range 9 to 100) for control group. No significant difference in discomfort between groups, P = 0.53 (procedures 1 to 10) **Pain:** Median pain for 1st 10 procedures was 9 (range 0 to 100) for VR group versus 8 (1 to 100) for control group. No significant difference in pain between groups, P = 0.24 (procedures 1 to 10)	-
Gerson 2003	Sigmoidoscopy	Conventional patient-based training	Level of patient pain and discomfort (rated by patient, 1-to-5 Likert scale: 1 = strongly agree; 2 = agree; 3 = not sure; 4 = disagree; 5 = strongly disagree)	-	53% patients in the VR group versus 42% in the control group agreed they "had a lot of pain." 43% patients in the VR group versus 31% in the control group agreed the procedure "caused great discomfort." No significant difference between groups, P > 0.05
McIntosh 2014	Colonoscopy	No training	Level of patient pain (rated by patient, 0-to-5 Likert scale: 0 = no pain, 5 = extreme pain)	Mean patient-rated pain 1.98 (SD 0.48) for VR group versus 1.95 (SD 0.33) for control group. No significant difference in pain between groups, P = 0.9	-

Table 6. Summary of outcomes - patient discomfort *(Continued)*

Sedlack 2004	Colonoscopy	No training	Patient discomfort (rated by patient, 10-point scale: 1 = minimal or no pain, 10 = worst pain of life)	Median patient-rated discomfort 2 (IQR 1 to 4) for VR group versus 4 (IQR 1.5 to 5) for control group. Statistically significantly less pain in VR trained group, P = 0.019 (procedures 1 to 15)	-
Sedlack 2004a	Sigmoidoscopy	Conventional patient-based training	Patient discomfort (rated by patient, 1-to-10 Likert scale: 1 = no pain, 10 = worst pain of life)	-	Median patient-rated discomfort 3 (IQR 2 to 5) for VR group versus 4 (IQR 2 to 6) for control group. Statistically significantly less pain in VR trained group, P < 0.01
Tuggy 1998	Sigmoidoscopy	No training	Pain scale (rated by patient)	No significant difference between groups, P > 0.05	-
Yi 2008	Colonoscopy	No training	Extent of abdominal pain and anus discomfort (rated by patient. 1-to-5 Likert scale: 1 = no pain; 5 = worst pain)	**Abdominal pain:** Mean patient-rated abdominal pain 3.1 (SD 0.8) for VR group and 3.2 (SD 1.1) for the control group SMD -0.10 (-0.63, 0.43) **Anus discomfort:** Mean patient-rated anus discomfort 2.7 (SD 0.8) for the VR group and 3.4 (SD 0.9) for the control group SMD -0.81 (-1.36, -0.25) **Pooled discomfort:** Mean pooled patient-rated discomfort 2.9 (SD 0.8) for the VR group and 3.3 (SD 1.0) for the control group SMD -0.46 (-1.00, 0.08)	-

EGD: oesophagogastroduodenoscopy
IQR: interquartile range
SD: standard deviation
SMD: standardised mean difference
VR: virtual reality

Table 7. Summary of outcomes - overall global rating of performance or competency

Study	Procedure	Comparison group	Method	VR versus no training	VR versus conventional endoscopy training
Cohen 2006	Colonoscopy	No training	Overall objective rating of competency (ability to reach the transverse colon and the caecum without assistance, and the ability to correctly recognise and identify abnormalities) Overall subjective rating of competency (rated by attending, 1-to-5 Likert scale: 1 = totally unskilled, 5 = competent and expedient)	**Objective competency:** Mean score 50.4 for VR group versus 40.9 for control group. Statistically significantly more positive ratings in VR trained group, P = 0.06 (procedures 1 to 20) **Subjective competency:** Mean score 47.6 for VR group versus 36.6 for control group. Statistically significantly more positive ratings in VR trained group, P = 0.08 (procedures 1 to 20)	-
Di Giulio 2004	EGD	No training	Expert global rating of performance based on "completeness" of the examination, the need for assistance, and the presumed difficulty of the procedure (rated by attending, 0-to-10 Likert scale with a procedure receiving a score of 5 or less being classified as "negative" and a procedure receiving a score of 6 or more as "positive": 0 = bad; 10 = good)	86.8% positive scores for VR group versus 56.7% for control group. Statistically significantly more positive ratings in VR trained group, P < 0.001	-
Ende 2012	EGD	Comparison group 1: conventional patient-based training; Comparison group 2: VR training only	Endoscopic skills rated using a 10-point visual analogue scale (rated by expert endoscopist, 1 = worst performance; 10 = optimal performance).	-	**Blinded tutor:** Median score 6.6 (IQR 6.0 to 7.75) for VR plus conventional training group versus 5.5 (IQR 4.75 to 7.0) for conventional training-only group versus 5.1 (IQR 4.0 to 6.0) for VR training-only group. Statistically significantly

Table 7. Summary of outcomes - overall global rating of performance or competency *(Continued)*

					more positive ratings for VR plus conventional training group versus VR training-only group, P = 0.035. Other comparisons not significant, P > 0.05. **Unblinded tutor:** Median score 7.7 (IQR 7.0 to 8.0) for VR plus conventional training group versus 6.3 (IQR 4.75 to 7.25) for conventional training-only group versus 4.7 (IQR 3.0 to 6.0) for VR training-only group. Statistically significantly more positive ratings for VR plus conventional training group versus VR training-only group, P = 0.004. Other comparisons not significant, P > 0.05.
Gerson 2003	Sigmoidoscopy	Conventional patient-based training	Expert global rating (rated by attending, 1-to-5 Likert scale: 1 = unable to clear the rectum; 2 = unable to clear the rectosigmoid junction; 3 = unable to pass 1 turn without assistance; 4 = able to perform independently, but more than 20 min required; 5 = independent examination less than 20 min duration)	-	Mean score 2.9 (SEM 0.2) for VR group versus 3.8 (SEM 0.2) for control group SMD -0.23 (-1.22, 0.76) Statistically significantly more negative score in the VR group, P < 0.001
McIntosh 2014	Colonoscopy	No training	1) Overall skill and technique (rated by expert endoscopist, 1-to-5 Likert scale: 1 = poor technique; 3 = competent; 5 = expert) 2) Overall skill and technique (rated by nurse, 1-to-5 Likert scale: 1 = poor technique; 3 = competent; 5 = expert)	**Expert endoscopist:** Mean score 2.28 (SD 0.21) for VR group versus 1.88 (SD 0.45) for control group. Statistically significantly more positive ratings in VR group, P = 0.02 **Nurse:** Mean score 2.56 (SD 0.26) for VR group versus 2.05	-

Table 7. Summary of outcomes - overall global rating of performance or competency *(Continued)*

				(SD 0.28) for control group. Statistically significantly more positive ratings in VR group, P = 0.001 **Pooled:** Mean score: 2.42 (SD 0.24) for VR group versus 1.97 (SD 0.37) for control group. Statistically significantly more positive ratings in VR group, P = 0.009	
Sedlack 2004a	Sigmoidoscopy	Conventional patient-based training	Expert global rating of competence to perform endoscopy independently (rated by attending, 1-to-10 Likert scale: 1 = strongly agree; 5 = neutral; 10 = strongly disagree)	-	Median score 8 (IQR 7 to 9) for VR group versus 8 (IQR 7 to 9) for control group. No significant difference between groups, P = 0.893
Sedlack 2007	EGD	No training	Expert global rating of competence to perform EGD independently (rated by attending, 1-to-7 Likert scale: 1 = strongly disagree; 4 = neutral; 7 = strongly agree)	-	No significant difference between groups, P > 0.05 (procedure days 1 to 5)

EGD: oesophagogastroduodenoscopy
IQR: interquartile range
SD: standard deviation
SEM: standard error of the mean
SMD: standardised mean difference
VR: virtual reality

Table 8. Summary of outcomes - visualisation of mucosa

Study	Procedure	Comparison group	Method	VR versus no training	VR versus conventional endoscopy training	VR versus another form of endoscopy simulation
Ende 2012	EGD	Comparison group 1: conventional patient-based training; Comparison group 2: VR training only	% of mucosa visualised	-	Mean 94% (SD 4) for VR plus conventional training group versus 92% (SD 7) for conventional training-only group versus 92% (SD 6) for	-

Table 8. Summary of outcomes - visualisation of mucosa *(Continued)*

					VR training-only group. No significant difference between groups, P = 0.211	
Gomez 2015	Colonoscopy	Comparison group 1: VR training only; Comparison group 2: another form of endoscopy simulation only	Time with a clear view of the lumen	-	-	Median 23.7 min for VR plus another endoscopy simulator group versus 23.9 for VR training-only group versus 28.2 for another endoscopy simulator-only group. No significant difference between groups, P = 0.084
Sedlack 2004	Colonoscopy	No training	Adequacy of mucosal visualisation on withdrawal (1 = strongly disagree, 4 = neutral, 7 = strongly agree)	Median 6.0 (IQR 6.0 to 7.0) for VR group versus 6.0 (IQR 5.0 to 7.0) for control group. Significantly greater visualisation in VR trained group, P = 0.009 (procedures 1 to 15)	-	-
Sedlack 2004a	Sigmoidoscopy	Conventional patient-based training	Adequacy of mucosal visualisation on withdrawal (1 = strongly agree, 5 = neutral, 10 = strongly disagree)	-	Median 7 (IQR 3 to 8) for VR group versus 5 (IQR 4 to 7) for control group. No significant difference between groups, P = 0.33	-
Tuggy 1998	Sigmoidoscopy	No training	% of colon visualised (assessed from videotapes of procedures)	**5 hours VR training:** Mean 55% in VR group versus 45% in control group. No significant difference between groups, P = 0.60 **6 to 10 hours VR training:** Mean 79% in VR group versus 45% in control group.	-	-

Table 8. Summary of outcomes - visualisation of mucosa *(Continued)*

				Significantly greater visualisation in VR trained group, P = 0.02		
Yi 2008	Colonoscopy	No training	Mucosal visualisation (1 = poor, 5 = excellent)	Mean 3.5 (SD 0.8) in VR trained group versus 2.9 (SD 0.7) in control group. Significantly greater visualisation in VR trained group, P = 0.002	-	-

EGD: oesophagogastroduodenoscopy
IQR: interquartile range
SD: standard deviation
VR: virtual reality

APPENDICES

Appendix 1. Search strategies for identification of studies

Database	Period	Search strategy used
The Cochrane Central Register of Controlled Trials (OVID)	2017, Issue 6 (Searched 12 July 2017)	#1 (endoscop* or colonoscop* or sigmoidoscop* or duodenoscop* or gastroscop* or proctoscop* or esophagoscop* or eosphagoscop* or oesphagoscop* or esophagoduodenoscop* or eosophagoduodenoscop* or oesophagoduodenoscop* or esophagogastroduodenoscop* or eosophagogastroduodenoscop* oesophagogastroduodenoscop*OR rectoscop*).mp. #2 (virtual realit* or simulat*).mp. #3 (#1 AND #2)
MEDLINE (Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R))	1946 - 12 July 2017	#1 endoscopy, digestive system/ or endoscopy, gastrointestinal/ or colonoscopy/ or sigmoidoscopy/ or duodenoscopy/ or esophagoscopy/ or gastroscopy/ or proctoscopy/ #2 ((gastrointestinal adj2 endoscop*) or (intestin* adj2 endoscop*) or colonoscop* or duodenoscop* or eosophagoduodenoscop* or eosophagogastroduodenoscop* or eosphagoscop* or esophagoduodenoscop* or esophagogastroduodenoscop* or esophagoscop* or gastroscop* or oesophagoduodenoscop* or oesophagogastroduodenoscop* or oesophagoscop* or proctoscop* or rectoscop* or sigmoidoscop* or (upper adj2 endoscop*)).tw,kf. #3 (#1 OR #2) #4 programmed instruction as topic/ or computer-assisted instruction/ or simulation training/ or high fidelity simulation training/ or patient simulation/ #5 diagnosis, computer-assisted/ or surgery, computer-assisted/ #6 Video-Assisted Surgery/ #7 computer simulation/ #8 user-computer interface/ or video games/ #9 ((virtual adj2 realit*) or (virtual adj realis*) or VR or simulat*).tw,kf.

(Continued)

		#10 (OR/#4-9) #11 (#3 AND #10) #12 clinical trial/ or clinical trial, phase i/ or clinical trial, phase ii/ or clinical trial, phase iii/ or clinical trial, phase iv/ or controlled clinical trial/ or randomized controlled trial/ or pragmatic clinical trial/ or comparative study/ or meta-analysis/ or multicenter study/ or validation studies/ #13 controlled clinical trials as topic/ or randomized controlled trials as topic/ or pragmatic clinical trials as topic/ or double-blind method/ or random allocation/ or single-blind method/ #14 (rct or rcts or random* or placebo* or cct or ccts or (control* adj2 trial*)).tw,kf. #15 ((singl* or doubl* or tripl* or trebl*) adj2 (mask* or blind*)).tw,kf. #16 (OR/#12-15) #17 (#11 AND #16)
Embase **(OVID)**	1947 - 12 July 2017	)#1 digestive tract endoscopy/ or esophagogastroduodenoscopy/ or esophagoscopy/ #2 gastrointestinal endoscopy/ or gastroscopy/ #3 intestine endoscopy/ or colonoscopy/ or duodenoscopy/ or rectoscopy/ or sigmoidoscopy/ #4 ((gastrointestinal adj2 endoscop*) or (intestin* adj2 endoscop*) or colonoscop* or duodenoscop* or eosophagoduodenoscop* or eosophagogastroduodenoscop* or eosphagoscop* or esophagoduodenoscop* or esophagogastroduodenoscop* or esophagoscop* or gastroscop* or oesophagoduodenoscop* or oesophagogastroduodenoscop* or oesophagoscop* or proctoscop* or rectoscop* or sigmoidoscop* or (upper adj2 endoscop*)).tw,kw. #5 (OR/#1-4) #6 computer assisted diagnosis/ #7 simulation/ or computer simulation/ or disease simulation/ or vignette/ #8 simulation training/ or high fidelity simulation training/ #9 educational technology/ #10 teaching/ #11 computer assisted surgery/ #12 virtual reality/ #13 (((computer* or video*) adj5 assist* adj5 (instruct* or teach* or educat*)) or ((virtual adj2 realit*) or (virtual adj realis*) or VR or simulat*) or (video* adj5 game*)).tw,kw. #14 (OR/#6-13) #15 (#5 AND #14) #16 comparative study/ or intermethod comparison/ #17 clinical trial/ or multicenter study/ or phase 1 clinical trial/ or phase 2 clinical trial/ or phase 3 clinical trial/ or phase 4 clinical trial/

(Continued)

		#18 controlled clinical trial/ or randomized controlled trial/ #19 controlled study/ #20 double blind procedure/ or single blind procedure/ or triple blind procedure/ #21 randomization/ #22 "clinical trial (topic)"/ or exp "controlled clinical trial (topic)"/ or "multicenter study (topic)"/ or "phase 1 clinical trial (topic)"/ or "phase 2 clinical trial (topic)"/ or "phase 3 clinical trial (topic)"/ or "phase 4 clinical trial (topic)"/ #23 (rct or rcts or random* or placebo* or cct or ccts or (control* adj2 trial*) or ((singl* or doubl* or tripl* or trebl*) adj2 (mask* or blind*))).tw,kw. or ct.fs. #24 (OR/#16-23) #25 (#15 AND #24)
Scopus	1960 - 12 July 2017	#1 TITLE-ABS-KEY ("gastrointestinal endoscop*" OR "intestinal endoscop*") #2 TITLE-ABS-KEY (colonoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR eosophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR "upper endoscop*" OR rectoscop*) #3 TITLE-ABS-KEY (simulat* OR vr OR "virtual realit*" OR cai OR "computer assisted instruct*" OR "computer assisted diagnos*" OR "computer assisted surger*") #4 TITLE-ABS-KEY (trial OR trials OR randomization OR randomization OR random OR randomised) #5 ((#1 OR #2) AND #3 AND #4)
Web of Science (includes (a) Science Citation Index Expanded; (b) Social Sciences Citation Index; (c) Arts & Humanities Citation Index; (d) Conference Proceedings Citation Index - Science and (e) Conference Proceedings Citation Index - Social Science	Science Citation Index Expanded (1900 - 12 July 2017) Social Sciences Citation Index (1956 - 12 July 2017) Arts & Humanities Citation Index (1975 - 12 July 2017) Conference Proceedings Citation Index - Science (1990 - 12 July 2017) Conference Proceedings Citation Index - Social Science (1990 - 12 July 2017)	#1 TS=("gastrointestinal endoscop*" OR "intestinal endoscop*" OR colonoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesphagoscop* OR esophagoduodenoscop* OR eosophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR "upper endoscop*" OR rectoscop*) #2 TS=(simulat* OR vr OR “virtual realit*” OR cai OR "computer assisted instruct*" OR "computer assisted diagnos*" OR "computer assisted surger*") #3 (#1 AND #2) #4 TS=(trial OR trials OR randomization OR randomisation OR random OR randomized) #5 (#3 AND #4)
Biosis Previews (OVID)	1980 - 12 July 2017	#1 TS=("gastrointestinal endoscop*" OR "intestinal endoscop*" OR colonoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesphagoscop* OR esophagoduodenoscop* OR eosophagoduodenoscop* OR oesophagoduodenoscop* OR

(Continued)

		esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR "upper endoscop*" OR rectoscop*) #2 TS=(simulat* OR vr OR "virtual realit*" OR cai OR "computer assisted instruct*" OR "computer assisted diagnos*" OR "computer assisted surger*") #3 (#1 AND #2) #4 TS=(trial OR trials OR randomization OR randomisation OR random OR randomized) #5 (#3 AND #4)
CINAHL **(EBSCO)**	1981 - 12 July 2017	#1 (MH "Endoscopy") OR (MH "Endoscopy, Digestive System") OR (MH "Endoscopy, Gastrointestinal") OR (MH "Colonoscopy") OR (MH "Sigmoidoscopy") OR (MH "Gastroscopy") OR (MH "Proctoscopy") OR (MH "Esophagoscopy") #2 TI (duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosophagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR eosophagoduodenoscop* OR oesophagoduodenoscop*OR esophagogastroduodenocop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*) #3 AB (duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosophagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR eosophagoduodenoscop* OR oesophagoduodenoscop*OR esophagogastroduodenocop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*) #4 (#1 OR #2 OR #3) #5 (MH "Diagnosis, Computer Assisted") #6 (MH "Simulations") OR (MH "Computer Simulation") OR (MH "Patient Simulation") OR (MH "Vignettes") OR (MH "Programmed Instruction") OR (MH "Computer Assisted Instruction") #7 (MH "Computerized Clinical Simulation Testing") #8 TI (virtual* OR VR OR simulat* OR cai OR "computer assisted") OR AB (virtual* OR VR OR simulat* OR cai OR "computer assisted") #9 (#5 OR #6 OR #7 OR #8) #10 (#4 AND #9) #11 (MH "Clinical Trials+") #12 TI (rct OR rcts OR random* OR placebo* OR cct OR ccts OR "controlled trial*") OR AB (rct OR rcts OR random* OR placebo* OR cct OR ccts OR "controlled trial*") #13 (#11 OR #12) #14 (#10 AND #13)
Allied and Complementary Medicine Database **(OVID)**	1985 - 12 July 2017	#1 endoscopy/ #2 (endoscop* or colonoscop* or sigmoidoscop* or duodenoscop* or gastroscop* or proctoscop* or esophagoscop* or eosphagoscop* or oesphagoscop*or esophagoduodenoscop* or eosophagoduodenoscop* or oesophagoduodenoscop* or esophagogastroduodenoscop* or eosophagogastroduodenoscop* or oesophagogastroduodenoscop* or rectoscop*).mp.

(Continued)

		#3 (#1 AND #2) #4 virtual reality/ #5 computer assisted instruction/ or computer simulation/ #6 (simulat* or vr or (virtual adj2 realit*) or (virtual adj2 realis*) or cai or computer assisted instruct* or computer assisted diagnos* or (computer adj2 (assisted adj2 surger*))).mp. #7 (#4 OR #5 OR #6) #8 (#3 AND #7)
ERIC **(ProQuest)**	1966 - 12 July 2017	#1 ti((((gastrointestinal or intesin*) NEAR/2 endoscop*) or colonoscop* or endoscop* or sigmoidoscop* or duodenoscop* or gastroscop* or proctoscop* or esophagoscop* or eosphagoscop* or oesophagoscop* or esophagoduodenoscop* or eosophagoduodenoscop* or oesophagoduodenoscop* or (upper NEAR/2 endoscop*) or rectoscop* or esophagogastroduodenoscop* or eosophagogastroduodenoscop* or oesophagogastroduodenoscop*)) OR ab((((gastrointestinal or intesin*) NEAR/2 endoscop*) or colonoscop* or endoscop* or sigmoidoscop* or duodenoscop* or gastroscop* or proctoscop* or esophagoscop* or eosphagoscop* or oesophagoscop* or esophagoduodenoscop* or eosophagoduodenoscop* or oesophagoduodenoscop* or (upper NEAR/2 endoscop*) or rectoscop* or esophagogastroduodenoscop* or eosophagogastroduodenoscop* or oesophagogastroduodenoscop*)) OR su((((gastrointestinal or intesin*) NEAR/2 endoscop*) or colonoscop* or endoscop* or sigmoidoscop* or duodenoscop* or gastroscop* or proctoscop* or esophagoscop* or eosphagoscop* or oesophagoscop* or esophagoduodenoscop* or eosophagoduodenoscop* or oesophagoduodenoscop* or (upper NEAR/2 endoscop*) or rectoscop* or esophagogastroduodenoscop* or eosophagogastroduodenoscop* or oesophagogastroduodenoscop*))
Education Full Text **(EBSCOHost)**	1969 - 12 July 2017	#1 TI (colonoscop* OR endoscop* OR sigmoidoscop* OR duodenoscop*OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR esophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*) #2 AB (colonoscop* OR endoscop* OR sigmoidoscop* OR duodenoscop*OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR esophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*) #3 SU (colonoscop* OR endoscop* OR sigmoidoscop* OR duodenoscop*OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR esophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*) #4 (#1 OR #2 OR #3 OR #4) #5 TI (virtual* OR VR OR simulat* OR cai OR “computer assisted”) OR AB (virtual* OR VR OR simulat* OR cai OR “computer assisted”) #6 (#4 AND #5)
CBCA Education **(ProQuest)**	1933 - 12 July 2017	#1 ab((colonoscop* OR endoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR esophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosopha-

(Continued)

		gogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*)) OR ti((colonoscop* OR endoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR esophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*)) OR su((colonoscop* OR endoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR esophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*))
ACM Digital Library **(ACM Portal)**	1948 - 12 July 2017	#1 (+endoscopy +simulat*) (+endoscopy +virtual)
IEEE Xplore	1950 - 12 July 2017	#1 (duodenoscopy OR gastroscopy OR proctoscopy OR esophagoscopy OR eosophagoscopy OR oesophagoscopy OR esophagoduodenoscopy OR eosophagoduodenoscopy OR oesophagoduodenoscopy OR esophagogastroduodenoscopy OR eosophagogastroduodenoscopy OR oesophagoduodenoscopy OR rectoscopy) AND (virtual OR cai OR 'computer assisted' OR 'computer based' OR simulation OR simulated OR simulations)
Abstracts in New Technologies and Engineering **(ProQuest)**	1981 - 12 July 2017	#1 (ALL(endoscop* OR colonoscop* OR sigmoidoscop*) OR ALL(duodenoscop* OR gastroscop* OR proctoscop*) OR ALL (esophagoscop* OR eosophagoscop* OR oesophagoscop*) OR ALL(esophagoduodenoscop* OR eosophagoduodenoscop* OR oeosophagoduodenoscop*) OR ALL(esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop*) OR ALL(rectoscop*)) AND (ALL(simulat* OR VR OR ("virtual realit*")) OR ALL(cai OR ("computer based train*") OR ("computer assist*"))) AND (ALL(Random* NEAR/3 trial*) OR ALL(random* OR trial*))
Computer & Information Systems Abstracts **(ProQuest)**	1981 - 12 July 2017	#1 (ALL(endoscop* OR colonoscop* OR sigmoidoscop*) OR ALL(duodenoscop* OR gastroscop* OR proctoscop*) OR ALL (esophagoscop* OR eosophagoscop* OR oesophagoscop*) OR ALL(esophagoduodenoscop* OR eosophagoduodenoscop* OR oeosophagoduodenoscop*) OR ALL(esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop*) OR ALL(rectoscop*)) AND (ALL(simulat* OR VR OR ("virtual realit*")) OR ALL(cai OR ("computer based train*") OR ("computer assist*"))) AND (ALL(Random* NEAR/3 trial*) OR ALL(random* OR trial*))
***meta*Register of controlled trials** **(active registers:** www.controlled-trials.com/mrct/ **and archived registers:** www.controlled-trials.com/mrct/archived**)**	12 November 2017	#1 (virtual realit* OR VR OR simulat* OR cai OR computer assisted instruct* OR computer based train* OR computer assisted train*) AND (endoscop* OR colonoscop* OR sigmoidoscop* OR duodenoscop* OR gastroscop* OR proctoscop* OR esophagoscop* OR eosphagoscop* OR oesophagoscop* OR esophagoduodenoscop* OR eosophagoduodenoscop* OR oesophagoduodenoscop* OR esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop* OR rectoscop*)
Dissertations & Theses **(ProQuest)**	1997 - 12 July 2017	#1 (ALL(endoscop* OR colonoscop* OR sigmoidoscop*) OR ALL(duodenoscop* OR gastroscop* OR proctoscop*) OR ALL (esophagoscop* OR eosophagoscop* OR oesophagoscop*) OR ALL(esophagoduodenoscop* OR eosophagoduodenoscop* OR oeosophagoduodenoscop*) OR ALL(esophagogastroduodenoscop* OR eosophagogastroduodenoscop* OR oesophagogastroduodenoscop*) OR ALL(rectoscop*)) AND (ALL(simulat* OR VR OR ("virtual realit*")) OR ALL(cai OR ("computer based train*") OR ("computer assist*"))) AND (ALL(Random* NEAR/3 trial*) OR ALL(random* OR trial*))

Appendix 2. Criteria for judging risk of bias in the 'Risk of bias' assessment tool

RANDOM SEQUENCE GENERATION Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence	
Criteria for a judgement of 'low risk' of bias	The investigators describe a random component in the sequence generation process such as: • referring to a random number table; • using a computer random number generator; • coin tossing; • shuffling cards or envelopes; • throwing dice; • drawing of lots; • minimisation.* *Minimisation may be implemented without a random element, and this is considered to be equivalent to being random.
Criteria for the judgement of 'high risk' of bias	The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example: • sequence generated by odd or even date of birth; • sequence generated by some rule based on date (or day) of admission; • sequence generated by some rule based on hospital or clinic record number. Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgement or some method of nonrandom categorisation of participants, for example: • allocation by judgement of the clinician; • allocation by preference of the participant; • allocation based on the results of a laboratory test or a series of tests; • allocation by availability of the intervention.
Criteria for a judgement of 'unclear risk' of bias	Insufficient information about the sequence generation process to permit judgement of 'low risk' or 'high risk'
ALLOCATION CONCEALMENT Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment	
Criteria for the judgement of 'low risk' of bias	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • central allocation (including telephone, web-based and pharmacy-controlled randomisation); • sequentially numbered drug containers of identical appearance; • sequentially numbered, opaque, sealed envelopes.
Criteria for a judgement of 'high risk' of bias	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: • using an open random allocation schedule (e.g. a list of random numbers); • assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not

(Continued)

	sequentially numbered); • alternation or rotation; • date of birth; • case record number; • any other explicitly unconcealed procedure.
Criteria for the judgement of 'unclear risk' of bias	Insufficient information to permit judgement of 'low risk' or 'high risk'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement - for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.
BLINDING OF PARTICIPANTS AND PERSONNEL Performance bias due to knowledge of the allocated interventions by participants and personnel during the study	
Criteria for the judgement of 'low risk' of bias	Any one of the following: • no blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding; • blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
Criteria for a judgement of 'high risk' of bias	Any one of the following: • no blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; • blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.
Criteria for the judgement of 'unclear risk' of bias	Any one of the following: • insufficient information to permit judgement of 'low risk' or 'high risk'; • the study did not address this outcome.
BLINDING OF OUTCOME ASSESSMENT Detection bias due to knowledge of the allocated interventions by outcome assessors	
Criteria for the judgement of 'low risk' of bias	Any one of the following: • no blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; • blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.
Criteria for a judgement of 'high risk' of bias	Any one of the following: • no blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding; • blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
Criteria for the judgement of 'unclear risk' of bias	Any one of the following: • insufficient information to permit judgement of 'low risk' or 'high risk'; • the study did not address this outcome.
INCOMPLETE OUTCOME DATA Attrition bias due to amount, nature, or handling of incomplete outcome data	

(Continued)

Criteria for the judgement of 'low risk' of bias	Any one of the following: • no missing outcome data; • reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); • missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; • for dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • for continuous outcome data, plausible effect size (difference in means or standardised difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • missing data have been imputed using appropriate methods.
Criteria for a judgement of 'high risk' of bias	Any one of the following: • reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • for dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; • for continuous outcome data, plausible effect size (difference in means or standardised difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size; • 'as-treated' analysis done with substantial departure of the intervention received from that assigned at randomisation; • potentially inappropriate application of simple imputation.
Criteria for the judgement of 'unclear risk' of bias	Any one of the following: • insufficient reporting of attrition/exclusions to permit judgement of 'low risk' or 'high risk' (e.g. number randomised not stated, no reasons for missing data provided); • the study did not address this outcome.
SELECTIVE REPORTING Reporting bias due to selective outcome reporting	
Criteria for the judgement of 'low risk' of bias	Any of the following: • the study protocol is available and all of the study's prespecified (primary and secondary) outcomes that are of interest in the review have been reported in the prespecified way; • the study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were prespecified (convincing text of this nature may be uncommon).
Criteria for a judgement of 'high risk' of bias	Any one of the following: • not all of the study's prespecified primary outcomes have been reported; • one or more primary outcomes is reported using measurements, analysis methods, or subsets of the data (e.g. subscales) that were not prespecified; • one or more reported primary outcomes were not prespecified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • one or more outcomes of interest in the review are reported

(Continued)

	incompletely so that they cannot be entered in a meta-analysis; • the study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'unclear risk' of bias	Insufficient information to permit judgement of 'low risk' or 'high risk'. It is likely that the majority of studies will fall into this category.
OTHER BIAS Bias due to problems not covered elsewhere in the table	
Criteria for the judgement of 'low risk' of bias	The study appears to be free of other sources of bias.
Criteria for a judgement of 'high risk' of bias	There is at least one important risk of bias. For example, the study: • had a potential source of bias related to the specific study design used; • has been claimed to have been fraudulent; or • had some other problem.
Criteria for the judgement of 'unclear risk' of bias	There may be a risk of bias, but there is either: • insufficient information to assess whether an important risk of bias exists; or • insufficient rationale or evidence that an identified problem will introduce bias.

WHAT'S NEW

Date	Event	Description
20 December 2017	New citation required and conclusions have changed	Substantively updated review with new conclusions. Author by-line changed.
12 July 2017	New search has been performed	New literature search was performed to update the review. New studies added.

CONTRIBUTIONS OF AUTHORS

Rishad Khan and Joanne Plahouras independently assessed the eligibility of article abstracts for inclusion in the review. Both review authors were responsible for data extraction and analysis. Rishad Khan and Catharine M Walsh were responsible for the writing of the final review manuscript. Coauthors Bradley Johnston, Michael A Scaffidi, and Samir C Grover provided supervisory support and content expert advice.

DECLARATIONS OF INTEREST

Rishad Khan was an author on a study included in this review (Grover 2017). He has received research funding from AbbVie and Ferring Pharmaceuticals outside the submitted work.

Joanne Plahouras has no conflicts of interest to declare.

Bradley C Johnston has no conflicts of interest to declare.

Michael A Scaffidi was an author on two studies included in this review (Grover 2015; Grover 2017).

Samir C Grover was the first author on two studies included in this review (Grover 2015; Grover 2017). He has received research funding from AbbVie and Ferring Pharmaceuticals, payments for consulting and speaking from AbbVie and Takeda, and has stock in Volo Healthcare outside the submitted work.

Catharine M Walsh was the senior author on two studies included in this review (Grover 2015; Grover 2017).

SOURCES OF SUPPORT

Internal sources

- New source of support, Other.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

This updated review has been performed according to the required Methodological Expectations of Cochrane Intervention Reviews (MECIR).

In this update, we modified one participant inclusion criterion from the 2012 version of this review. Specifically, limited endoscopic experience is defined here as previous performance of no greater than 20 cases of the procedure under study in the clinical or simulated setting or both, while previously it was defined as previous performance of no greater than 10 cases. This change reflects a changing definition of limited endoscopic experience in the literature, as evidenced by inclusion criteria in several new endoscopy simulation trials (Grover 2015; Grover 2017). In addition, we removed two secondary outcome measures, as the GRADE 'Summary of findings' table limits the total number of outcomes to seven. We removed insertion depth and error rate, as we perceived these outcomes to be of the least value from an educational standpoint with regard to acquisition of endoscopic competence (Walsh 2016).

INDEX TERMS

Medical Subject Headings (MeSH)

*Clinical Competence; *Virtual Reality; Endoscopy, Gastrointestinal [*education]; Health Personnel [*education]; Randomized Controlled Trials as Topic; Simulation Training [*methods]

MeSH check words

Humans