Virtual reality simulation training in endoscopy: a Cochrane review and meta-analysis

Authors
Rishad Khan1, Joanne Plahouras2, Bradley C. Johnston3, Michael A. Scaffidi4,5, Samir C. Grover4,5, Catharine M. Walsh6,7,8

Institutions
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, Faculty of Medicine, Dalhousie University, Halifax, Canada
4 Division of Gastroenterology, St. Michael’s Hospital, University of Toronto, Toronto, Canada
5 Department of Medicine, University of Toronto, Toronto, Canada
6 Division of Gastroenterology, Hepatology, and Nutrition and the Research and Learning Institutes, Hospital for Sick Children, University of Toronto, Toronto, Canada
7 Department of Pediatrics, University of Toronto, Toronto, Canada
8 The Wilson Centre, University of Toronto, Toronto, Canada.

submitted 19.10.2018
accepted after revision 12.3.2019

ABSTRACT
Background Endoscopy programs are increasingly integrating simulation training. We conducted a systematic review to determine whether virtual reality (VR) simulation training can supplement and/or replace conventional patient-based endoscopy training for health professional trainees with limited or no prior endoscopic experience.

Methods We searched medical, educational, and computer literature databases in July 2017 for trials that compared VR simulation training with no training, conventional training, another form of simulation training, or an alternative method of VR training. We screened, abstracted data, and performed quantitative analysis and quality assessment through Cochrane methodology.

Results We included 18 trials with 3817 endoscopic procedures. VR training provided no advantage over no training or conventional training based on the primary outcome of composite score of competency. VR training was advantageous over no training based on independent procedure completion (relative risk [RR] = 1.62, 95% confidence interval [CI] 1.15 – 2.26, moderate-quality evidence), overall rating of performance (mean difference [MD] 0.45, 95%CI 0.15 – 0.75, very low-quality evidence), and mucosal visualization (MD 0.60, 95%CI 0.20 – 1.00, very low-quality evidence). Compared with conventional training, VR training resulted in fewer independent procedure completions (RR = 0.45, 95%CI 0.27 – 0.74, low-quality evidence). We found no differences between VR training and no training or conventional training for other outcomes. Based on qualitative analysis, we found no significant differences between VR training and other forms of simulation training. VR curricula based in educational theory provided benefit with respect to composite score of competency, compared with unstructured curricula.

Conclusions VR simulation training is advantageous over no training and can supplement conventional endoscopy training. There is insufficient evidence that simulation training provides benefit over conventional training.
Introduction

Endoscopy is an important diagnostic and therapeutic tool used in the evaluation and treatment of gastrointestinal disorders [1]. Novice endoscopists have traditionally acquired procedural proficiency through the apprenticeship model, whereby they learn skills under the supervision of experienced preceptors in the clinical setting [2]. The growing awareness of the need for patient comfort and safety has, however, led to the increasing use of virtual reality (VR) simulators to augment traditional endoscopy teaching [3].

Through a combination of visual and tactile interfaces, VR simulators present learners with situations that resemble reality and/or tasks that are designed to train them in a specific endoscopic skill (e.g. endoscope handling, loop reduction) [4–6]. Many such simulators employ a physical endoscope that the user navigates through a virtual lumen on a monitor or uses to perform a specific task [7]. In this learner-centered environment, trainees can acquire technical, cognitive, and non-technical skills with no risk of patient harm or discomfort [5]. Numerous studies assessing VR simulators in surgery have reported successful transfer of skills from the simulated to the clinical environment [8–11]. Additionally, endoscopic VR simulators can provide objective measures of competence, such as mucosal visualization, which can be used to assess trainees and identify gaps in skill acquisition [12].

Current American Council for Graduate Medical Education (ACGME) training guidelines mandate the incorporation of simulation training into gastroenterology fellowship programs [13]. While simulation has the potential to reduce training costs as staff endoscopists are more productive when performing procedures independently, there are costs associated with acquiring and maintaining simulators [14]. It is therefore important to ensure skills gained through simulation training positively transfer to the clinical environment.

While several systematic reviews have focused on this topic, they did not perform comprehensive searches of the educational and computer literature databases and conference proceedings [9,15–17]. Moreover, several trials have been published since the most recent reviews [18–22]. We systematically reviewed and synthesized all current evidence to determine whether VR simulation training can supplement and/or replace early conventional endoscopy training (apprenticeship model) in diagnostic esophagastroduodenoscopy (EGD), colonoscopy, and/or sigmoidoscopy for health professional trainees with limited or no prior endoscopic experience.

Methods

This report follows the Cochrane Collaboration’s methodology [23] and is an amended version of a 2018 Cochrane review [24] with a protocol published a priori [25].

Eligibility criteria

We included randomized and quasi-randomized trials comparing VR endoscopy simulation training with: (a) no training, (b) conventional patient-based training, (c) another form of endoscopy simulation training, or (d) an alternative method of VR training. We only considered trials that measured outcomes on humans in the clinical setting. Participants were health professional trainees with limited or no prior endoscopic experience, defined as previous performance of <20 cases of the procedure under study, with any level of experience in performing other endoscopic procedures not under study.

Search strategy

We searched the following health professional, educational, and computer databases from inception until 12 July 2017: The Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE, Scopus, Web of Science, BIOSIS Previews, CINAHL, Allied and Complementary Medicine Database, 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 metaRegister of controlled trials until 12 November 2017. The detailed search strategy is available in the full-length Cochrane review [24].

We hand-searched the reference lists of articles identified using the aforementioned search and abstracts of the following meetings from 2009–2017: Digestive Diseases Week, Canadian Digestive Diseases Week, British Society of Gastroenterology, United European Gastroenterology Week, The Association for Medical Education in Europe Conference, Canadian Conference on Medical Education, Research in Medical Education, American College of Surgery Clinical Congress, The Society of American Gastrointestinal and Endoscopic Surgeons Conference, and European Association for Endoscopic Surgery Congress.

Two authors (R.K. and J.P.) independently screened all titles and abstracts for inclusion, and, when needed, a third author (C.M.W.) adjudicated any disagreements. For conference abstracts which lacked sufficient detail for inclusion, we searched for corresponding full-length publications and/or contacted authors for further details. If we did not acquire sufficient information, we excluded these abstracts.

Outcomes

The primary outcome was composite score of competency in performing endoscopy. This outcome reflects an overall aggregate score derived from various workplace-based assessment tools used to assess endoscopic competence. Published validity evidence for each assessment tool is variable [12]. These tools allow for structured assessment at the “does” level of Miller’s pyramid, reflective of what an individual does during a real clinical encounter, and provide a high degree of authenticity [26].

Secondary outcomes were:

a) Independent procedure completion: number of endoscopic procedures that trainees completed without assistance from a supervisor.
b) Performance time: time required to complete a given endoscopic procedure.
c) Complication or critical flaw occurrence: procedure-related adverse events, such as bleeding and luminal perforation.
d) Patient discomfort: patient- and/or provider-rated assessment of patient discomfort during endoscopic procedures.
e) A single measure providing an overall global rating of performance in endoscopy: single rating of performance rated by an external assessor on a 5-point Likert scale.

f) Visualization of the mucosa: percentage of luminal mucosa visualized during the endoscopic procedure – higher mucosal visualization represents a more successful endoscopic procedure.

Data extraction and assessment of quality
Two authors (R.K. and J.P.) independently extracted information on participants, interventions, comparators, and outcomes, and assessed the methodological quality using the Cochrane Collaboration’s tool for assessing risk of bias [27]. We contacted study authors if the publication contained insufficient or unclear information. We evaluated the quality of evidence for each outcome that was amenable to meta-analysis using GRADE (Grading of Recommendations Assessment, Development, and Evaluation) [28] for each of the following comparisons:

- VR endoscopy simulation training versus no training
- VR endoscopy simulation training versus conventional patient-based training.

Data analysis
We performed meta-analyses using Review Manager 5.3 [29]. For dichotomous data, we calculated relative risk (RR) with 95% confidence intervals (CI). For continuous data, we used mean difference (MD) and standard deviation mean difference (SMD), both with 95% CI, when studies used the same and different outcome measures, respectively. We planned to pool data for meta-analyses if there were similarities with respect to participants, interventions, comparators, and outcome measures. We used a random-effects model because of the presence of heterogeneity and performed weighting using the Mantel-Haenszel method [23]. For studies with three or more arms [18, 19], to avoid a unit of analysis error, we excluded groups that included combination training (e.g. VR simulation training followed by patient-based training) from the meta-analyses to allow for direct comparison of VR simulation with a control.

Two authors (R.K. and J.P.) independently evaluated the included studies for clinical and methodological heterogeneity. We assessed heterogeneity using the Cochrane chi-squared test (Q test) with an alpha level of 0.10 and using the I² statistic (0–40%, low heterogeneity; 30%–60%, moderate heterogeneity; 50%–90%, substantial heterogeneity; 75%–100%, considerable heterogeneity).

To explore statistical or clinical heterogeneity, we conducted a priori subgroup analysis of trials based on the type of endoscopic procedure (EGD, colonoscopy, and sigmoidoscopy) and the level of participant experience (no prior versus limited prior endoscopic experience). We planned sensitivity analysis by excluding studies at high or unclear risk of bias and studies published in abstract form only. We also planned to examine publication bias through a funnel plot. We did not conduct sensitivity analyses or create a funnel plot because of a low number of eligible trials.

Results
We identified 1053 and 12 potentially relevant records through electronic and hand-searching, respectively. We retrieved 40 full-text articles and/or conference abstracts for further assessment. We excluded 34 references for the reasons listed in the full-length Cochrane review [24] and one study was classified as ongoing, which left five trials from this search and 13 trials from the previous version of this review, giving a total of 18 included trials. All included studies were full-text publications. ►Fig. 1 outlines the study flow diagram.

Among the 18 trials, 10 compared VR training with no intervention [22, 30–38], five compared VR training with conventional patient-based endoscopy training [18, 39–42], one compared VR training to another form of endoscopy simulation training [19], and two compared different methods of VR training [20, 21]. ►Table 1 provides an overview of the trial characteristics.

Risk of bias
We considered three trials to be at low risk of bias [20, 21, 31] and nine trials to be at high risk of bias owing to non-random sequence generation, lack of description of allocation concealment, and/or lack of blinding of outcome assessment [18, 19, 22, 33–35, 38, 39, 41]. We considered the remaining six trials to be at unclear risk of bias as the method of randomization and/or blinding of outcome assessment was unclear, and/or an assessment instrument with no validity evidence was used to assess competency [30, 32, 36, 37, 40, 42]. The risk of bias is summarized in ►Fig. 2.

Quality of evidence
We evaluated the quality of evidence using GRADE for studies evaluating VR simulation training versus no training, and studies evaluating VR simulation training versus conventional patient-based training. All outcomes were rated as moderate, low, or very low owing to risk of bias, imprecision, and/or unexplained heterogeneity (►Table 2 and ►Table 3).

Effects of interventions
All quantitative analyses are summarized in ►Fig. 3 and ►Fig. 4 and detailed in Tables 1 s and 2 s (see online-only Supplementary material). Forest plots of pooled data are shown in Figs. 1 s and 2 s. Because of clinical and methodological heterogeneity, it was not possible to perform a meta-analysis for several outcomes among our four comparisons. In addition, several trials did not appropriate central tendency (mean) and variability (standard deviation) data to allow for quantitative analysis. In the absence of data-pooling, we discuss the results of the studies qualitatively, based on outcome.

Comparison 1: Virtual reality endoscopy simulation training versus no training
One trial reported a composite score of competency and showed no statistically significant difference between the groups (MD 3.10, 95%CI =−0.16 to 6.36; n=24 procedures) [33]. Six trials reported independent procedure completion
The VR training group had a significantly higher number of independent procedure completions (RR 1.62, 95% CI 1.15 to 2.26; n = 815). Heterogeneity was statistically significant (P = 0.03; I² = 61%). Subgroup analysis showed that the VR training group had a higher number of procedure completions whether the procedure was colonoscopy (RR 1.84, 95% CI 1.35 to 2.50; I² = 11%; n = 408 procedures) or EGD (RR 1.25, 95% CI 1.13 to 1.39; n = 407), and whether participants had limited (RR 1.82, 95% CI 1.07 to 3.12; I² = 54%; n = 329) or no prior endoscopic experience (RR 1.32, 95% CI 1.09 to 1.61; I² = 13%; n = 486).

Seven trials reported performance time [22, 30, 31, 33, 35, 38]. Pooled data from two trials showed no significant difference between the groups (MD = −0.20, 95% CI −0.71 to 0.30; n = 29) [22, 38]. Heterogeneity was not statistically significant (P = 0.86; I² = 0%). Among the remaining five trials which reported this outcome, three showed a statistically significantly faster time for the VR training group [31, 34, 37], and two showed no significant difference [33, 35]. Three trials (550 procedures) reported the occurrence of complications or critical flaws, with all three reporting no complications or critical flaws in either group [30, 31, 33].
# Table 1 Description of studies included in this systematic review and meta-analysis of virtual reality (VR) simulation training in endoscopy for health professional trainees.

<table>
<thead>
<tr>
<th>Study</th>
<th>Study design</th>
<th>Trainees</th>
<th>Procedure</th>
<th>Comparator</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Type</td>
<td>Prior experience</td>
<td>n</td>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>Ahlberg et al. (2005) [31]</td>
<td>Multicenter RCT</td>
<td>Surgery residents and gastroenterology fellows</td>
<td>Limited</td>
<td>119</td>
<td>Colonoscopy</td>
</tr>
<tr>
<td>Di Giulio et al. (2004) [33]</td>
<td>Multicenter RCT</td>
<td>Gastroenterology fellows</td>
<td>None</td>
<td>420</td>
<td>EGD</td>
</tr>
<tr>
<td>Ende et al. (2012) [18]</td>
<td>Multicenter RCT</td>
<td>Surgery and medicine residents</td>
<td>None</td>
<td>39</td>
<td>EGD</td>
</tr>
<tr>
<td>Ferlitsch et al. (2010) [34]</td>
<td>Single-center RCT</td>
<td>Medicine residents</td>
<td>None</td>
<td>470</td>
<td>EGD</td>
</tr>
<tr>
<td>Gerson et al. (2003) [39]</td>
<td>Single-center RCT</td>
<td>Medicine residents</td>
<td>None</td>
<td>66</td>
<td>Sigmoidoscopy</td>
</tr>
<tr>
<td>Gomez et al. (2015) [19]</td>
<td>Single-center RCT</td>
<td>Surgery residents</td>
<td>None</td>
<td>27</td>
<td>Colonoscopy</td>
</tr>
<tr>
<td>Grover et al. (2017) [21]</td>
<td>Single-center RCT</td>
<td>Surgery and medicine residents, and gastroenterology fellows</td>
<td>Limited</td>
<td>74</td>
<td>Colonoscopy</td>
</tr>
<tr>
<td>Haycock et al. (2010) [40]</td>
<td>Multicenter RCT</td>
<td>Any health professional trainee</td>
<td>Limited</td>
<td>111</td>
<td>Colonoscopy</td>
</tr>
<tr>
<td>McIntosh et al. (2014) [22]</td>
<td>Single-center quasi-randomized trial</td>
<td>Surgery and medicine residents, and gastroenterology fellows</td>
<td>Limited</td>
<td>90</td>
<td>Colonoscopy</td>
</tr>
<tr>
<td>Sedlack and Kolars (2004) [35]</td>
<td>Single-center RCT</td>
<td>Gastroenterology fellows</td>
<td>Limited</td>
<td>120</td>
<td>Colonoscopy</td>
</tr>
<tr>
<td>Sedlack et al. (2004) [41]</td>
<td>Single-center RCT</td>
<td>Medicine residents</td>
<td>None</td>
<td>442</td>
<td>Sigmoidoscopy</td>
</tr>
<tr>
<td>Shirai et al. (2008) [42]</td>
<td>Single-center RCT</td>
<td>Residents</td>
<td>None</td>
<td>20</td>
<td>EGD</td>
</tr>
</tbody>
</table>
Seven trials reported patient discomfort [22,31,32,34,35,37,38]. Pooled data from two trials showed no significant difference between the groups (SMD 0.16, 95% CI 0.68 to 0.35; n = 145) [22,38]. Heterogeneity was not statistically significant (P = 0.13; I² = 57%). Subgroup analysis showed no significant differences between the groups whether participants had limited (SMD 0.07, 95% CI 0.35 to 0.49; n = 90) or no prior endoscopic experience (SMD -0.46, 95% CI 1.00 to 0.08; n = 45). All five remaining trials that reported this outcome found no significant difference between the groups with respect to patient discomfort [31,32,34,35,37].

Four trials reported an overall rating of performance [22,32,33,36]. Data from one trial showed statistically significantly more positive ratings in the VR training group compared with no training (MD 0.45, 95% CI 0.15 to 0.75; n = 18) [22]. Two other trials also reported statistically significantly more positive ratings in the VR training group [32,33] and one showed no significant difference between the groups [36]. Three trials reported visualization of the mucosa [35,37,38]. Data from one trial showed significantly greater visualization in the VR training group (MD 0.60, 95% CI 0.20 to 1.00; n = 55) [38]. The two other trials also showed significantly greater visualization in the VR training group [35,37].

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

One trial reported the composite score of competency and showed no statistically significant difference between the groups [40]. Two trials reported independent procedure completion [39,40]. The VR training group had a significantly lower number of independent procedure completions (RR 0.45, 95% CI 0.27 to 0.74; n = 174). Heterogeneity was not statistically significant (P = 0.47) and was low (I² = 0%). Subgroup analysis showed that the VR training group had fewer independent procedure completions in sigmoidoscopy (RR 0.41, 95% CI 0.23 to 0.72; n = 66), but not colonoscopy (RR 0.67, 95% CI 0.20 to 2.23; n = 108).

Four trials reported performance time [18,39,40,42]. Pooled data from two trials showed no significant difference between the groups (SMD 0.12, 95% CI 0.55 to 0.80; n = 34) [18,39]. Heterogeneity was not statistically significant (P = 0.73) and was low (I² = 0%). Subgroup analysis showed no significant differences between the groups in the sigmoidoscopy (SMD 0.0 minutes, 95% CI 0.99 to 0.99; n = 16) and EGD studies (SMD 0.23 minutes, 95% CI 0.69 to 1.16; n = 18). The other two trials reported no significant differences in performance time between the groups [40,42]. Three trials (72 procedures) reported the occurrence of complications or critical flaws, with all three reporting no complications or critical flaws in either group [18,39,41].

Two trials reported patient discomfort [39,41]. We did not perform a meta-analysis as neither trial had sufficient central tendency and variability data. One trial reported significantly lower patient discomfort in the VR training group [41] and the other reported no significant difference between the groups [39].

Three trials reported an overall global rating of performance [18,39,41]. Data from one trial showed statistically significantly lower positive ratings in the VR training group (MD 0.90, 95% CI -4.40 to 2.60; n = 16) [39]. Another trial showed no significant difference between the groups [41]. The third trial showed statistically significantly more positive ratings in the VR plus conventional training group compared with the VR-only training group, but no significant difference compared with the conventional-only training group [18]. Two trials reported visualization of the mucosa [18,41]. Data from one trial showed no significant difference in visualization between the groups (MD 0.0, 95% CI -6.02 to 6.02; n = 18) [18]. The other trial also showed no significant difference in visualization between the groups [41].

### Comparison 3: Virtual reality endoscopy simulation training versus training using another form of endoscopy simulation

One trial reported a composite score of competency, performance time, and visualization of the mucosa [19]. There were no significant differences between the groups for any outcome. The trial in this subgroup did not report on the other secondary outcomes.
Table 2. Summary of findings for trials comparing virtual reality (VR) simulation training versus no training.

<table>
<thead>
<tr>
<th>Patient or population: health professional trainees in gastrointestinal endoscopy</th>
<th>Setting: four single center studies from Canada, USA, and South Korea, and two multicenter European studies</th>
<th>Intervention: virtual reality endoscopy simulation training</th>
<th>Comparison: no training</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Anticipated absolute effects(^1) (MD/SMD) for risk with VR endoscopy simulation training (95 %CI)</th>
<th>Relative effect (95 %CI)</th>
<th>Number of procedures(^2) (studies)</th>
<th>Quality of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome: Composite score of competency</td>
<td>MD + 3.10 (−0.16 to + 6.36)</td>
<td>24 (1 trial)</td>
<td>⊕⊕⊕⊕ Low(^4)</td>
</tr>
<tr>
<td>Outcome: Independent procedure completion</td>
<td>RR 1.62 (1.15 to 2.26) Risk with no training, 465 per 1000 vs. risk with VR endoscopy simulation training, 754 per 1000 (535 to 1000)</td>
<td>815 (6 trials)(^5)</td>
<td>⊕⊕ Moderate(^3)</td>
</tr>
<tr>
<td>Outcome: Performance time (seven trials reported performance time, but only two provided sufficient data for quantitative analysis)</td>
<td>MD −0.20 (−0.71 to + 0.30)</td>
<td>29 (2 trials)(^6)</td>
<td>⊕⊕⊕ Very low(^4)</td>
</tr>
<tr>
<td>Outcome: Complication or critical flaw occurrence</td>
<td>All trials reporting this outcome reported an incidence of zero for complications and critical flaws in both groups</td>
<td>550 (3 trials)</td>
<td>⊕⊕ Moderate(^3)</td>
</tr>
<tr>
<td>Outcome: Patient discomfort (seven trials reported patient discomfort, but only two provided sufficient data for quantitative analysis)</td>
<td>SMD −0.16 (−0.68 to + 0.35)</td>
<td>145 (2 trials)(^6)</td>
<td>⊕⊕⊕ Very low(^4)</td>
</tr>
<tr>
<td>Outcome: Overall global rating of performance or competency (four trials reported overall global ratings, but only one with two data sets [from two types of assessor] provided sufficient data for quantitative analysis)</td>
<td>MD + 0.45 (+0.15 to + 0.75)</td>
<td>18 (1 trial)(^9)</td>
<td>⊕⊕⊕ Very low(^7)</td>
</tr>
<tr>
<td>Outcome: Visualization of the mucosa (three trials reported visualization of the mucosa, but only one provided sufficient data for quantitative analysis)</td>
<td>MD + 0.60 (+0.20 to + 1.00)</td>
<td>55 (1 trial)(^9)</td>
<td>⊕⊕⊕ Very low(^7)</td>
</tr>
</tbody>
</table>

CI, confidence interval; MD, mean difference; SMD, standardized mean difference; RR, relative risk.

\(^1\) The basis for the assumed risk is provided in the footnotes relating to quality of evidence. The risk in the intervention group (and its 95 %CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95 %CI).

\(^2\) 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 1.62-fold increased likelihood of completion of an endoscopic procedure.

\(^3\) Downgraded one level for serious risk of bias (due to unclear or inadequate methods of randomization, allocation sequence generation, and/or blinding of outcome assessment).

\(^4\) Randomized trials and two quasi-randomized trials.

\(^5\) Two quasi-randomized trials.

\(^6\) Downgraded two levels for very serious risk of bias (due to inadequate methods of randomization, allocation sequence generation, and/or blinding of outcome assessment).

\(^7\) Downgraded due to unexplained heterogeneity.

\(^8\) Quasi-randomized trial.
Comparison 4: Two different methods of virtual reality endoscopy simulation training

Two trials reported a composite score of competency [20, 21]. Both trials showed a significantly increased composite score of competency in the interventional VR training group as compared with the control VR training group. Participants in the interventional VR training group in one trial [20] received a similar curriculum as the control VR training group in the other trial [21]. Neither trial in this subgroup reported any secondary outcomes.

Discussion

We included eighteen trials with 421 participants and 3817 endoscopic procedures. Compared with no training, VR simulation training provides benefit based on independent procedure completion, procedure time, overall rating of performance, and visualization of the mucosa. The evidence is equivocal with respect to VR training compared with conventional patient-based training. VR training conferred similar advantages compared with other forms of endoscopy simulation training. With respect to comparisons of two different methods of VR training, a structured curriculum in VR endoscopy training provides an

| Table 3  Summary of findings for trials comparing virtual reality (VR) simulation training versus conventional patient-based training. |
| --- | --- | --- | --- |
| **Patient or population:** health professional trainees in gastrointestinal endoscopy  
**Setting:** one single center study from the USA and two multicenter European studies  
**Intervention:** virtual reality (VR) endoscopy simulation training  
**Comparison:** conventional patient-based training |
| Anticipated absolute effects\(^1\) (MD/SMD) for risk with VR endoscopy simulation training (95%CI) | Relative effect (95%CI) | Number of procedures\(^2\) (studies) | Quality of the evidence (GRADE) |
| **Outcome:** Composite score of competency |
| One trial reported this outcome but did not provide sufficient data for quantitative analysis | – | (0 studies) |
| **Outcome:** Independent procedure completion |
| – | RR 0.45 (0.27 to 0.74)  
Risk with no training, 337 per 1000 vs. risk with VR endoscopy simulation training, 152 per 1000 (91 to 250) | 174 (2 trials) | ⊕⊕⊖⊖ Low\(^3\) |
| **Outcome:** Performance time (four trials reported performance time but only two provided sufficient data for quantitative analysis) |
| SMD + 0.12 (−0.55 to +0.8) | – | 34 (2 trials) | ⊕⊕⊕⊕ Very low\(^3,4\) |
| **Outcome:** Complication or critical flaw occurrence |
| All trials reporting this outcome reported an incidence of zero for complications and critical flaws in both groups | – | 72 (3 trials) | ⊕⊕⊕⊕ Very low\(^3,4\) |
| **Outcome:** Patient discomfort |
| Two trials reported patient discomfort, but neither provided sufficient data for quantitative analysis | – | (0 studies) |
| **Outcome:** Overall global rating of performance or competency (three trials reported overall global ratings, but only one provided sufficient data for quantitative analysis) |
| MD −0.90 (−4.40 to +2.60) | – | 16 (1 trial) | ⊕⊕⊕⊕ Very low\(^3,4\) |
| **Outcome:** Visualization of the mucosa (two trials reported visualization of the mucosa, but only one provided sufficient data for quantitative analysis) |
| MD 0 (−6.02 to +6.02) | – | 18 (1 trial) | ⊕⊕⊕⊕ Very low\(^3,4\) |

CI, confidence interval; MD, mean difference; SMD, standardized mean difference; RR, relative risk.
\(^1\) The basis for the assumed risk is provided in the footnotes relating to quality of evidence. The risk in the intervention group (and its 95%CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95%CI).
\(^2\) 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 1.62-fold increased likelihood of completion of an endoscopic procedure.
\(^3\) Downgraded two levels for very serious risk of bias (due to inadequate methods of randomization, allocation sequence generation, and/or blinding of outcome assessment).
\(^4\) Downgraded one level for serious imprecision.
Addition-
ally, a progressive learning curriculum, whereby trainees com-
plete increasingly difficult cases [21, 43], provides an ad-
vantage over the structured curriculum.

Previously, four reviews of VR endoscopy simulation training
summarized evidence supporting the use of VR simulators to
improve learning and the performance of endoscopy [9, 15–
17]. Our study has several strengths compared with previous
reviews. First, our Cochrane methodology encompassed a
broad search of medical, computer, and educational literature
databases and used the GRADE approach to inform the quality
of our evidence. Second, we excluded observational studies
that are at high risk of bias and only considered clinical out-
comes on real patients, as opposed to outcomes on a simulator
or animal models. Finally, we have included four newer trials
published since the most recent review [15].

Our results should be interpreted with caution owing to se-
veral important limitations. The methodological quality of the
data for our outcomes of interest ranged from moderate to
very low because of risk of bias, imprecision, and unexplained
heterogeneity. The major sources of bias were inadequate ran-
domization, lack of allocation concealment or lack of reporting
with respect to allocation concealment, lack of assessor blind-
ing, and the use of outcome measures with inadequate evi-
dence of validity. Additionally, the findings are limited by small
sample sizes and variability in outcome measures across stud-
ies. There were also considerable differences among studies
with respect to VR training interventions, which made some
comparisons difficult. Finally, we focused on clinical outcomes
and did not include data on factors that may have impacted the
training environment. While this was a purposeful methodolo-
gical decision based on the Cochrane group’s focus on clinical
outcomes, we recognize that educators in endoscopy also value
data on educational outcomes, such as task difficulty, learner
comfort, cognitive load, and learning curves.

Despite these limitations, our findings support current
ACGME guidelines mandating the incorporation of simulation
training into gastroenterology fellowship programs [13]. Two

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of Number of procedures</th>
<th>Quantitative analysis</th>
<th>Effect estimate (95% CI)</th>
<th>I², %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite score of competency, mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>12 12 MD 3.10 (−0.16 to 6.36) -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of independent procedures completed</td>
<td></td>
<td></td>
<td>RR 1.62 (1.15 to 2.26) 61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>411 404 MD −0.20 (−0.71 to 0.30) 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance time, mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15 14 MD 0.45 (0.15 to 0.75) -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient discomfort, mean</td>
<td></td>
<td></td>
<td>SMD −0.16 (−0.68 to 0.35) 57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>75 70 MD 0.60 (0.20 to 1.00) -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall global rating of performance, mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10 8 MD 0.60 (0.20 to 1.00) -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualization of the mucosa, mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>25 30 MD 0.60 (0.20 to 1.00) -</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval; MD, mean difference; RR, relative risk; SMD, standardized mean difference.

Fig.3 Quantitative analysis for trials evaluating virtual reality (VR) simulation training versus no training.
studies that compared the effectiveness of different simulation curricula highlight the benefits of instructional design based on educational theory when implementing simulation-based training. Feedback is another important consideration. Simply providing trainees with access to simulators does not guarantee optimal use as extrinsic feedback is essential for the acquisition of endoscopic skills and effective simulation-based learning [20, 45–48]. Additionally, one study, which used a progressive learning approach, demonstrated the effectiveness of curricula that sequentially increase task difficulty, which has been identified as a best practice in the broader simulation literature [45, 47].

One area that remains unclear, however, is the length and intensity of training required for optimal learning. For example, one report only demonstrated a benefit after 6–10 hours of training and not after 5 hours [37]. Finally, it is unclear whether the use of simulators leads to a reduction in the median number of procedures required to achieve technical and cognitive competence. Studies to date have yet to demonstrate a reduction in the learning curve of more than 25%, a threshold the American Society for Gastrointestinal Endoscopy has proposed for widespread adoption of simulators during training [49].

Future research must consist of comparative effectiveness trials that adhere to strict quality standards, such as adequate randomization and allocation concealment, and which evaluate the impact of different educational theory-based endoscopy simulation curricula or compare VR training with lower cost simulation approaches. Future studies should also evaluate the characteristics of instruction, feedback, and duration of training required to optimize skill transfer to the clinical setting. In conclusion, novice endoscopists who receive VR simulation training have an advantage over their untrained peers during initial clinical endoscopic procedures. Thus, learning during VR simulation training transfers to patient care, suggesting this modality can effectively supplement conventional endoscopy training. Conversely, simulation training provides equivocal results compared with conventional endoscopy training. We therefore cannot advise for or against the use of VR simulation training as a replacement for early conventional endoscopy training. Results from our review support the implementation of VR simulation-based training prior to clinical procedures on live patients in endoscopy training programs. This strategy provides opportunities to acquire technical, cognitive, and non-technical skills in a low risk environment with no risk to patients. Future research should compare the effectiveness of different theory-based endoscopy simulation curricula using rigorous methodology.

### Acknowledgments

The authors wish to thank Ms. Thomasin Adams-Webber (Manager, Library and Archives Services, Learning Institute, the Hospital for Sick Children) for assisting with the electronic search strategy and the Cochrane Colorectal Cancer Editorial office and editors for their support throughout the review process.
Competing interests

Rishad Khan was an author on a study included in this review [21]. Michael A. Scaffidi was an author on two studies included in this review [20, 21]. Samir C. Grover was the first author on two studies included in this review [20, 21]. Catharine M. Walsh was the senior author on two studies included in this review [20, 21]. The remaining authors have no conflicts of interest to declare. The authors have no financial relationships relevant to this article to disclose.

References


Gerson L, van Dam J. A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. Endoscopy 2003; 35: 569 – 575


Brydges R, Carnahan H, Rose D et al. Coordinating progressive levels of simulation fidelity to maximize educational benefit. Acad Med 2010; 85: 806 – 812


Issenberg SB, McGaghie WC, Petrusa ER et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005; 27: 10 – 28


