Effect of virtual endoscopy simulator training on performance of upper gastrointestinal endoscopy in patients: a randomized controlled trial

Authors A. Ferlitsch, R. Schoefl, A. Puespoek, W. Miehsler, M. Schoeniger-Hekele, H. Hofer, A. Gangl, M. Homoncik

Institutions Department of Internal Medicine III, Division of Gastroenterology and Hepatology, Medical University of Vienna, Vienna, Austria

submitted 1 December 2008
accepted after revision 9 July 2010

Bibliography
DOI http://dx.doi.org/10.1055/s-0030-1255818
Published online 22 October 2010
Endoscopy 2010; 42: 1049–1056 © Georg Thieme Verlag KG Stuttgart · New York
ISSN 0013-726X

Corresponding author
A. Ferlitsch, MD
Department of Internal Medicine III
Division of Gastroenterology and Hepatology
Medical University of Vienna
Währinger Gürtel 18–20
A-1090 Vienna
Austria
Fax: +43-1-40400-4735
arnulf.ferlitsch@meduniwien.ac.at

Background: Skills in gastrointestinal endoscopy mainly depend on experience and practice. Patients upon whom trainees perform their first endoscopic examinations are likely to suffer more discomfort and prolonged procedures. Training on endoscopy simulators may reduce the time required to reach competency in patient endoscopy.

Patients and methods: Residents in internal medicine without experience of endoscopy were randomized to a group who trained on a simulator before conventional training (group S) or one that received conventional training only (group C) before starting upper gastrointestinal endoscopy in patients. After endoscopy, discomfort and pain were evaluated by patients, who were blind to the beginners' training status. Results in terms of time, technique (intubation, pyloric passage, J-maneuver), and diagnosis of pathological entities were evaluated by experts.

Results: From 2003 to 2007, 28 residents were enrolled. Comparing group S with group C in their first ten endoscopic examinations in patients, time taken to reach the duodenum (239 seconds (range 50–620) vs. 310 seconds (110–720; $P < 0.0001$) and technical accuracy ($P < 0.02$) were significantly better in group S. Diagnostic accuracy did not differ between the groups. Fourteen residents (7 simulator-trained, 7 not simulator-trained) continued endoscopy training. After 60 endoscopic examinations, investigation time was still shorter in group S. Technical and diagnostic accuracy improved during on-patient training in both groups; here differences between groups were no longer observable. There were no significant differences in discomfort and pain scores between the groups after 10 and after 60 endoscopies. Discomfort and pain were higher than for endoscopy performed by experts.

Conclusion: This randomized controlled trial shows that virtual simulator training significantly affects technical accuracy in the early and mid-term stages of endoscopic training. It helps reduce the time needed to reach technical competency, but clinically the effect is limited. Simulator training could be useful in an endoscopy training curriculum but cannot replace on-patient training.

Introduction

Skills in gastrointestinal endoscopy mainly depend on experience and practice. The patients upon whom a trainee does his or her first endoscopic examinations are likely to suffer more discomfort, longer procedures [1], and higher complication rates [2, 3]. Training on endoscopy simulators, the latest generation being electronic virtual reality devices, may decrease the time taken to reach competence in endoscopy [4–6]. Prior investigations with the GI Mentor (Simbionix, Tel Hashomer, Israel), a personal-computer-based simulator with tactile feedback, revealed its capability to distinguish between experts and beginners in endoscopy and to improve beginners' skills on the simulator [7]. The effects of simulator training on patient acceptance of trainee-performed endoscopy have not been investigated so far. We therefore evaluated, in a randomized, controlled, partially blinded trial, whether training on the GI Mentor virtual reality endoscopy simulator can improve objective performance measures of beginners in endoscopy as well as subjective measures such as patient satisfaction with gastrointestinal endoscopy.

Methods

Patients

The study was conducted in a large university hospital with a caseload of more than 5000 upper gastrointestinal endoscopies per year. Only pa-

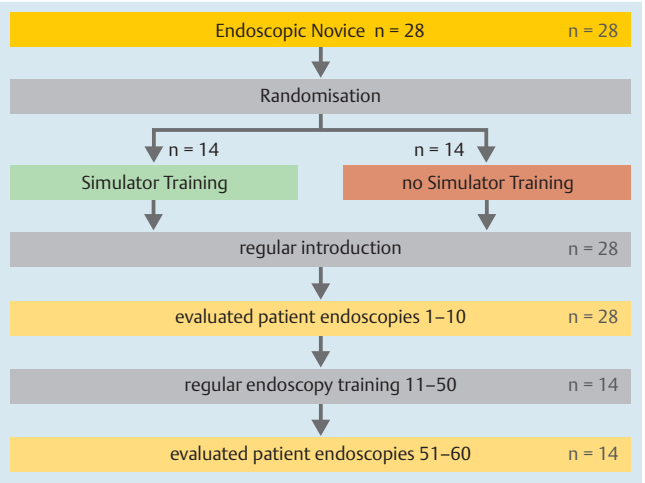

Fig. 1 Flowchart of the study protocol.

tients scheduled for diagnostic upper gastrointestinal endoscopy and unwilling to undergo sedation were selected. The indications for upper gastrointestinal endoscopy were dysphagia, reflux, abdominal pain, or a combination of these. If the patients wanted to have concomitant sedation, trainees were not allowed to perform endoscopy on them while under evaluation. No patients undergoing therapeutic interventions (polypectomy, mucosectomy, dye spraying, treatment of bleeding) were included. Patients were informed that trainees in endoscopy would perform the examination under supervision with manual assistance from an expert endoscopist. Whenever necessary, experts would take over the endoscope, but the scope was taken by the trainee again after the difficult situation had been passed, until the investigation was finished or a new insuperable obstacle was encountered, then the expert would finish the examination. All experts involved are part of the "staff team" of the endoscopy unit, each of them with experience of more than 5000 upper gastrointestinal endoscopies. The experts were informed about the training status of the endoscopic novices (i.e., which were simulator-trained), but the patients were not.

Patients gave their written informed consent to participate. The study was approved by the Vienna Medical University Ethics Committee, and was registered as a randomized, controlled trial at the Clinicaltrials.gov website (identifier NCT00576043).

Simulator training

The design of the study is outlined in ▶ **Fig. 1**.

Eligible trainees were at least 3rd-year residents in internal medicine and had to be naïve to endoscopy, including lower gastrointestinal endoscopy. They were randomized either to receive 2 hours per day of structured training [5] for a minimum of 5 hours and up to 20 hours (their choice) on the virtual endoscopy simulator (simulator group, group S) or to receive no simulator training (control/conventional training only group, group C) before starting conventional training. 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".

Training was performed on the GI Mentor virtual endoscopy simulator. With this, three-dimensional pictures are generated in real time by a computer while the endoscope is moved through the gastrointestinal mannequin. The location and movement of the scope are transmitted via sensors located at the tip and shaft of the endoscope. The oral cavity of the GI Mentor (no tongue, no teeth) leads to a plastic tube 150 cm in length into which the endoscope can be inserted. A force-feedback module simulates resistance whenever the walls of the virtual gastrointestinal tract are touched to provide realistic feeling during the examination. The endoscope used for all procedures is a modified Pentax ECS-3840F; steering and torsion of the endoscope and air inflation and suction are possible via the regular wheels and buttons of the control head of the scope. The core of the training was made up of 20 virtual gastroscopy cases as well as the haptic (targeted steering) training games "Endobasket" and "Endobubble." Trainers were present for the first 2 hours of simulator training. No colonoscopy cases were used for training.

Endoscopy

Before starting endoscopy in patients, both groups were instructed equally in upper gastrointestinal endoscopy. This common element included instruction in handling the endoscope, watching 5–10 upper gastrointestinal endoscopic examinations by the experts and withdrawing the endoscope 3–5 times from the descending duodenum in patients. Trainee residents were introduced into the specific pathological findings of the upper gastrointestinal tract, with the help of an endoscopic atlas and CD. They were trained to use a one-hand steering technique right from the beginning of their clinical track; pushing and pulling the scope is performed by the trainee with the other hand. For the first obstacle, intubation of the esophagus, trainees were allowed to try twice before the attending physician took over the scope. After returning, residents were allowed to try to perform pyloric passage twice before they were assisted by the attending; if they could not pass the pylorus with assistance after two attempts, the attending took over the scope and returned it to the trainee afterwards. Routine mucosal biopsies (gastric antrum and body) were taken by the trainee during investigation. If necessary, targeted biopsies were taken by the trainee and repeated by the expert if the trainee failed.

After the trainees had received the standard introduction to endoscopic technique, the first 10 gastroscopic examinations, performed in 10 consecutive patients who met the criteria listed above, were measured and evaluated. The parameters noted were the time between the first attempt at esophageal intubation and the end of the investigation, and the time from the first attempt at esophageal intubation until the descending part of the duodenum was reached. Technical accuracy was evaluated by recording whether the novice endoscopist was able to intubate the esophagus ("unaided"), whether manual help by the expert was needed ("expert help"), or if the expert had to take over ("expert takeover"). In the same way, pyloric passage and retroflexion of the endoscope (J-maneuver) in the gastric fundus were noted. Diagnostic accuracy was also registered, evaluated as the number of pathological entities (ulcer, erosion, polyp, hernia, diverticulum, varix, angiodysplasia) found or missed.

After the initial training period, regular endoscopy training in patients with different indications for upper gastrointestinal endoscopy continued for 14 of the 28 residents; the other 14 had to come back to the endoscopy unit later in their medical training and left the study at this point. The 14 who continued the training were not selected or randomized to continue, it was just that the hospital resident training curriculum planned a longer time in the endoscopy unit at this level of residency. For these 14, sedated patients were also selected during their continuing endos-

copy training, until they had completed 50 endoscopic examinations performed by themselves under supervision. After successfully performing 50 such supervised examinations, trainees were again assessed during 10 consecutive investigations, using the criteria listed above, in order to evaluate the medium- to long-term effect of simulator training followed by hands-on training. Fellows of the endoscopy unit served as an expert control group and were evaluated according to the same criteria while performing 10 endoscopic examinations.

A sample size calculation was performed for the outcome parameter "time to descending duodenum" and revealed that with 14 volunteers in each group a minimum difference of 20% between group S and group C could be detected with a power of 80%. These assumptions were based on an α value of 0.05, a β value of 0.2, and a standard deviation for the primary endpoint of 55 seconds.

Patient evaluation

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). Discomfort and pain were evaluated immediately after the investigation by means of a patient questionnaire that used two visual analog scales (one for discomfort, one for pain), in which marks were made with a pen across two black 10-cm lines on an otherwise blank sheet, the left end of the lines correlating with the worst (0 mm), the right end with the best possible value (100 mm) for discomfort and pain.

Statistical methods

Statistical analysis was performed using Statistica for Windows, version 6.0. All data are expressed as median and range. The study focused on the comparison between group S and group C. Follow-up data were expressed as their changes (as percentages) from baseline values. All statistical comparisons were performed using the Wilcoxon matched pairs test for post hoc comparisons. Differences between groups were assessed using the Mann-Whitney *U* test. Applying the Bonferroni correction for multiple groups and time points, a *P* value of 0.01 can be considered significant for individual post hoc tests.

Results

From 2003 to 2007, 28 residents (19-male, 9 female, mean age 31 years, range 28 – 37 years) were enrolled and 470 endoscopic examinations were evaluated. All 28 residents managed to perform at least 10 examinations in patients by themselves; 14 residents (7 of them simulator trainees) performed 60 or more examinations in patients during their 4-week training period in the endoscopy unit. Five fellows of the endoscopy unit served as the expert control group.

Simulator training

The trainees were randomized in equal proportions to receive simulator training (simulator group, group S, n = 14) or not to receive simulator training (control/conventional group, group C, n = 14). Seven male and 7 female residents were randomized to group S. The median training time was 600 minutes (range 300 – 1200).

Patient characteristics

The median age of the patients was 55 years (range 18 – 87); 55% were female. No patient received concomitant sedation. There was no significant difference between group S and group C in respect of patient sex or age ($P = 0.49$ and $P = 0.67$, respectively).

Endoscopy performance: comparison of group S and group C

Time to duodenum

After 10 endoscopic examinations, group S was significantly faster in reaching the descending duodenum [239 seconds (range 50 – 620) vs. 310 seconds (range 110 – 720), $P < 0.0001$]. At the advanced training stage (51st to 60th examinations), there was still a significant difference between the groups: the median time to the duodenum was 120 seconds in group S compared to 177 seconds in group C after 60 examinations ($P < 0.003$).

All 28 novice endoscopists managed to reach the duodenum after a median of 270 seconds (range 50 – 720). They were faster after they had completed 50 endoscopic examinations (median 125 seconds, range 20 – 400, $P < 0.0001$). Accordingly, both groups showed an improvement in the time taken to reach the descending duodenum in the comparison between the first 10 examinations and the 51st to 60th examinations ($P = 0.0001$; **Table 1**; **Fig. 2**).

For the group of experts, the median time taken to reach the duodenum was 60 seconds (range 35 – 105).

Total endoscopy time

Group S was also significantly faster in respect of the total endoscopy time during their first 10 examinations [720 seconds (range 405 – 1705) vs. 740 seconds (range 240 – 2400); $P = 0.012$] (**Table 1**; **Fig. 2**). At the advanced training stage (51st to 60th on-patient examinations), there was still a significant difference between the groups: median total endoscopy time was 495 seconds for group S and 600 seconds for group C ($P < 0.003$).

The median total endoscopy time for all beginners after their first 10 endoscopic examinations was 724 seconds (range 240 – 2400). Examinations were performed significantly faster once the trainees had performed 60 of them themselves (median 520 seconds, range 125 – 1320, $P < 0.00001$). Accordingly, both groups showed an improvement for total endoscopy time in the comparison between the first 10 examinations and the 51st to 60th examinations ($P < 0.0001$; **Table 1**; **Fig. 2**). Experts performed an upper gastrointestinal endoscopy in a median time of 220 seconds (range 125 – 310).

Technical accuracy

Group S had a better intubation ratio, needing less assistance during the first 10 examinations ($P < 0.005$) (**Fig. 3**; **Table 2**). Pyloric passage and retroflexion of the endoscope were also obviously easier for group S ($P < 0.01$ for both). These effects were no longer detectable after 60 examinations ($P = 0.55$, $P = 0.63$, and $P = 0.45$, respectively). After 60 examinations, the intubation rate and retroflexion rate had improved significantly in both group S and group C. Concerning pyloric passage, group S, which had a high success rate after 10 examinations, improved slightly ($P = 0.09$), whereas group C improved significantly ($P < 0.005$) (**Fig. 3**).

For all 28 beginners, the intubation rate (70% initial success), pyloric passage rate (78%), and rate of successful retroflection of the endoscope in the gastric fundus (74%) improved after 60 examinations (96%, 96%, and 99%, respectively).

Table 1 Effects of simulator training on time needed for endoscopic examinations. Data are shown as median (range).

	All beginners		Simulator group		Control group		*P* value, simulator vs. control group	Experts (n = 5)	
	Time, s	*P* value, examinations 1 – 10 vs. 51 – 60	Time, s	*P* value, examinations 1 – 10 vs. 51 – 60	Time, s	*P* value, examinations 1 – 10 vs. 51 – 60		Time, s	*P* value, experts vs. all beginners
Time to duodenum									
1st to 10th examinations*	270 (50 – 720)		239 (50 – 620)		310 (110 – 720)		< 0.0001	60 (35 – 105)	
51st to 60th examinations	125 (20 – 400)		120 (20 – 333)		177 (60 – 400)		< 0.01		
		< 0.00001		< 0.0001		< 0.00001			< 0.00001
Total endoscopy time									
Examinations 1 – 10	724 (240 – 2400)		720 (405 – 1705)		740 (240 – 2400)		= 0.012	220 (125 – 310)	
51st to 60th examinations	520 (125 – 1320)		495 (200 – 900)		600 (410 – 1320)		< 0.01		
		< 0.00001		< 0.001		< 0.01			< 0.00001

*In both the simulator and the control group, there were 28 trainees who performed examinations 1 – 10 but only 14 who continued training to complete examinations 51 – 60.

Table 2 Effects of simulator training on technical accuracy. Technical accuracy was evaluated by recording if the novice endoscopist was able to perform the technical tasks alone ("unaided"), if manual help by the expert was needed ("expert help") or if the expert had to take over ("expert takeover").

	Simulator group				Control group				*P* value, simulator vs. control group
	Technical accuracy, %			*P* value, examinations 1 – 10 vs. 51 – 60	Technical accuracy, %			*P* value, examinations 1 – 10 vs. 51 – 60	
	Unaided	Expert help	Expert takeover		Unaided	Expert help	Expert takeover		
Intubation									
1st to 10th examinations*	78	9	13		60	21	19		< 0.005
51st to 60th examinations	97	0	3		94	3	3		=0.55
				< 0.0001				< 0.0001	
Pyloric passage									
1st to 10th examinations	85	10	5		72	6	22		< 0.01
51st to 60th examinations	95	0	5		97	3	0		= 0.63
				=0.09				< 0.005	
Retroflexion									
1st to 10th examinations	81	12	7		66	20	14		< 0.01
51st to 60th examinations	100	0	0		98	0	0		= 0.45
				< 0.01				=0.01	

*In both the simulator and the control group, there were 28 trainees who performed examinations 1 – 10 but only 14 who continued training to complete examinations 51 – 60.

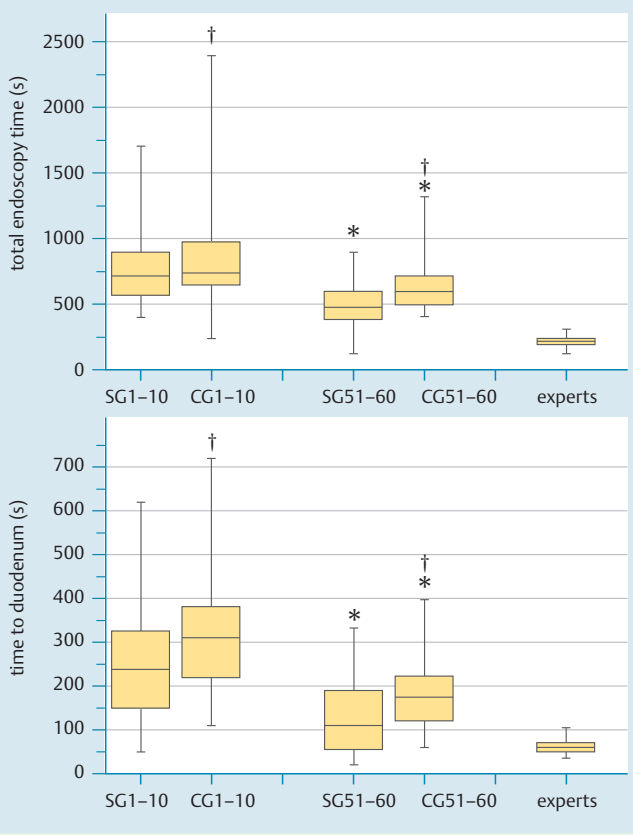

Fig. 2 Effect of simulator training on total endoscopy time and the time taken to reach the descending duodenum. *$P < 0.01$ for comparisons within the simulator-trained group (group S) and within the group who did not receive simulator training (group C) for the first 10 endoscopic examinations vs. examinations 51–60. †$P < 0.01$ for comparisons between group S and group C for the first 10 examinations (SG1–10, CG1–10) and the 51st to 60th examinations (SG51–60, CG51–60).

Diagnostic accuracy

There was no significant difference between the groups in the number of missed pathological findings after 10 endoscopic examinations (group S: 0.29 per examination vs. group C: 0.31 per examination; $P = 0.99$) and after 60 examinations ($P = 0.29$) (**Table 3**).

During the first 10 examinations the novices missed pathological entities in 29% of the examinations, while after 60 examinations only 2% of the pathological entities were missed.

Major pathological findings such as ulcers, polyps, angiodysplasia, suspected Barrett lesions, or varices were encountered in 5% of the examinations. For all beginners, there was no statistical difference in total endoscopy times for examinations in which major pathological findings were identified and any associated targeted biopsies carried out, and those in which no such findings were made ($P = 0.69$).

Patient evaluation

No patient refused to undergo endoscopy by supervised trainees.

Discomfort The evaluation scores for discomfort given by the patients on the visual analog scales after the procedures did not differ significantly between group S (median 16, range 0–98) and group C (20, 0–100), either for the first 10 examinations ($P = 0.53$) or for the 51st to 60th examinations (16, 0–82 vs. 17, 0–79, $P = 0.47$) (**Table 4**).

A slight improvement over time was observed in comparing the discomfort scores after 10 and after 60 examinations in group C ($P = 0.08$); there was no difference for group S ($P = 0.70$).

For all beginners (n = 28), the median discomfort level was 18 (range 0–100) after 10 examinations; after 60 examinations (n = 14 endoscopists) it was 16 (range 0–82), indicating a slight improvement, but no significant difference ($P = 0.16$).

Pain With regard to the occurrence of pain, neither group scored significantly differently for the first 10 examinations [median 9 (range 0–100) vs. 8 (0–100), $P = 0.24$] (**Table 4**). After 60 examinations, the scores were 9 (0–66) in group S and 5 (0–46) in group C ($P = 0.07$). After 60 examinations, the pain scores improved significantly in group C ($P = 0.008$).

For all beginners the median pain score after 10 examinations was 9 (0–100), and this was virtually unchanged after 60, at 8 (0–66, $P = 0.16$).

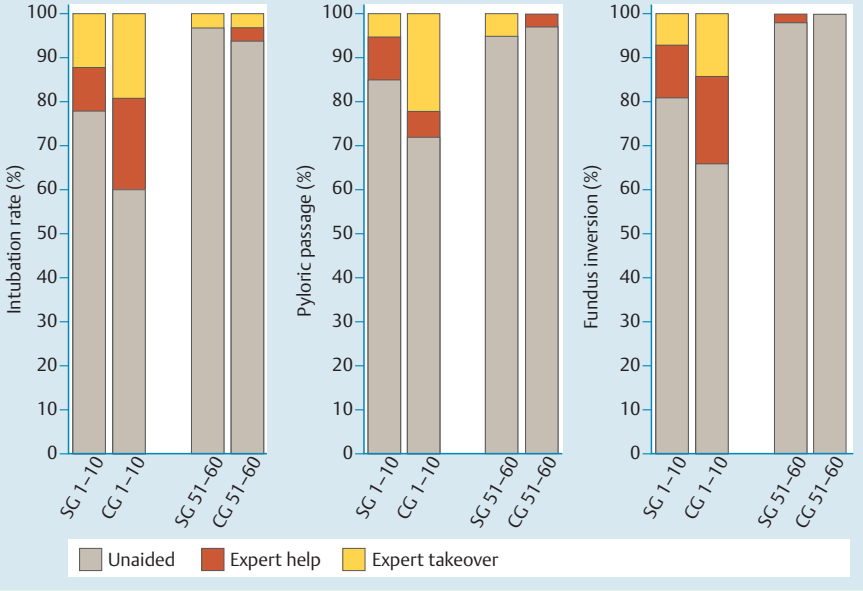

Fig. 3 Technical accuracy was evaluated by recording if the novice endoscopist was able to perform the technical tasks alone ("Unaided"), if manual help by the expert was needed ("Expert help"), or if the expert had to take over ("Expert takeover"). Technical accuracy was higher in group S; both groups improved over time, but the effect of simulator training was fading after 60 endoscopic examinations had been performed.

Table 3 Effects of simulator training on diagnostic accuracy: number of missed pathological findings per endoscopic procedure.

	Simulator group				Control group				*P* value, simulator vs. control group
	No. of missed pathological entities			*P* value, examinations 1 – 10 vs. 51 – 60	No. of missed pathological entities			*P* value, examinations 1 – 10 vs. 51 – 60	
	0	1	2+		0	1	2+		
1st to 10th examinations*	73	25	2		74	22	4		= 0.99
51st to 60th examinations	97	1.5	1.5		100	0	0		= 0.29
				< 0.001				< 0.01	

* In both the simulator and the control group, there were 28 trainees who performed examinations 1 – 10 but only 14 who continued training to complete examinations 51 – 60.

Table 4 Effects of training on patient evaluation, scored on 100-mm visual analog scales (VAS) for discomfort and pain. Scores are shown as median (range).

	All beginners		Simulator group		Control group		*P* value, simulator vs. control group	Experts (n = 5)	
	VAS score	*P* value, examinations 1 – 10 vs. 51 – 60	VAS score	*P* value, examinations 1 – 10 vs. 51 – 60	VAS score	*P* value, examinations 1 – 10 vs. 51 – 60		VAS score	*P* value, experts vs. all beginners
Discomfort									
1st to 10th examinations*	18 (0 – 100)		16 (0 – 98)		20 (0 – 100)		= 0.53	0 (0 – 98)	
51st to 60th examinations	16 (0 – 82)		16 (0 – 82)		17 (0 – 79)		= 0.47		
		= 0.16		= 0.70		= 0.08			= 0.002
Pain									
1st to 10th examinations	9 (0 – 100)		9 (0 – 100)		8 (0 – 100)		= 0.24	1 (0 – 23)	
51st to 60th examinations	8 (0 – 66)		9 (0 – 66)		5 (0 – 46)		= 0.07		
		= 0.16		= 0.89		= 0.008			= 0.005

*In both the simulator and the control group, there were 28 trainees who performed examinations 1 – 10 but only 14 who continued training to complete examinations 51 – 60.

In comparison to beginners, expert endoscopists scored 0 (0–98) for discomfort ($P = 0.002$) and 1 (0 – 23) for pain ($P = 0.005$).

Discussion

This is the first randomized, controlled and patient-blinded trial to study the effects of simulator training both on objective measures of performance in upper gastrointestinal patient endoscopy and on subjective measures such as patients' discomfort and pain scores.

Measurement of quality in endoscopy is difficult and is an intensively discussed issue [8 – 10]. While the parameters for measuring quality in colonoscopy – foremost among them cecum intubation and polyp detection rates – have been validated, parameters for measuring quality in upper gastrointestinal endoscopy are still lacking [3,11]. Patient cohorts vary, a variety of disease entities are present, and the settings for the endoscopic procedure in respect of both technical and human resources are different from one endoscopy unit to the next. Using "time" as a measurement of quality is most often misleading, as a variety of circumstances may prolong the investigation other than the technical abilities of the investigator [12]. We are fully aware that total procedure times, particularly when measured in seconds, cannot serve as a quality parameter for endoscopic procedures: quite the contrary view is supported by the results of studies in which longer endoscope withdrawal times produce better polyp detection rates [8]. For this reason we did not tell our trainees that "time" was a major outcome parameter of our study; on the contrary, we encouraged them to take the time they needed, as one would do when teaching endoscopy. Studies evaluating training in colonoscopy have demonstrated that the parameter "time to reach the cecum" is a valid tool for measuring the progress of trainees [6,8]. We therefore selected "time to reach the duodenum" as a quality indicator.

In our study simulator training improved the investigation times in comparison to those of standard (non-simulator-)trained endoscopic novices. In this first study to investigate the midterm effects of training we were able to demonstrate that, although both groups improve over time, the sustainability of the simulator training is still visible after 60 endoscopic examinations have been performed. On the other hand, the impact of clinical training as well as the need for supervised endoscopic examinations even after 60 examinations have been performed is evident. Similar study results, suggesting an even longer supervised training period of up to 160 investigations, have been demonstrated for simulator-based training in colonoscopy [6]. Experts were present during the first 2 hours of simulator training. If the experts had been present for longer, even larger differences might have resulted.

Diagnosis during the examination is mainly done during retrieval of the endoscope, and for this reason numbers of missed pathological entities and technical skills such as retroflexion of the endoscope (J-maneuver) are important indicators of quality and technical accuracy. We strongly discourage endoscopic trainees from speeding up during investigations.

Interestingly, the training group showed significantly better results for the parameter "intubation" even though the GI Mentor gives only limited training in intubation [2]. The haptic abilities gained by the trainee may have affected the outcome in this parameter. The superiority of simulator training with regard to intubation, pyloric passage, and retroflection in the fundus is evident in the early stages of on-patient endoscopy. These essential technical abilities improved over time in both groups and differences were no longer present after 60 endoscopic examinations, which can be interpreted as a major impact of on-patient endoscopy experience.

During the first endoscopic examinations in patients, the rate of missed pathological findings was high (29 % overall). Simulator training did not improve the number of pathological entities found, but after 60 examinations the skills and knowledge gained dramatically improved the rate of pathological findings during upper gastrointestinal endoscopy, which again underlines the value of on-patient endoscopy on training success.

Experts were not blinded to which group trainees were in (i.e., simulator vs. non-simulator), as the simulator was based in the endoscopy suite. This could have impacted on the reaction of the experts and might have affected outcomes. However, every trainee was supervised during their on-patient training by a variety of experts who were not involved in the simulator phase of the training and who were selected at random to minimize any possible bias as to when an expert decided to take over the procedure or might perhaps hesitate to do so with a simulator-trained trainee.

An essential part of this randomized study was the patient evaluation. Several difficulties had to be overcome. The scoring system had to be easy and reproducible, and for this reason plain 10-cm lines were chosen as a visual analog scale. We also selected patients who preferred to undergo upper gastrointestinal endoscopy unsedated in order to evaluate the immediate impression after the endoscopic procedure was over. Routinely approximately 25 % of our outpatients want to have concomitant sedation for upper gastrointestinal endoscopy. Sedation might have affected all objective outcome measures, and selecting only patients who preferred unsedated endoscopy avoided this. However, this might have led to a floor effect in scoring by selecting more insensitive patients who tolerate pain better: they might have cared less and given higher pain and discomfort scores less often. Just as for colonoscopy [13], discomfort and pain assessment were chosen. In a previous pilot study evaluating simulator training in upper gastrointestinal endoscopy, a Likert scale to describe discomfort was used to measure patients' impressions [14]. 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; however, this study was powered for the primary outcome parameter (time to duodenum) and the results on patient scores must therefore be interpreted with caution. New endoscopists should start on-patient training in sedated patients as analysis showed only a statistical trend towards improvement in both scores when comparing all beginners after 10 and after 60 endoscopic examinations. However, simulator training did not affect patient evaluation significantly. High pain values were rarely ever given; the overall tolerance of trainee-performed unsedated endoscopy was impressive. However, only the conventionally trained group improved significantly, and only in the pain assessment. These results might be statistically significant, but it must be borne in mind that the conventionally trained group had a median 5-mm line score on a 100-mm line, in comparison to 9-mm median for the simulator group, which cannot indicate a clinically relevant effect. Speculations as to whether the non-simulator-trained group might be more cautious could not be evaluated with this study setting. Patients were informed when endoscopic trainees would be performing the investi-

gation. They were blinded to which group the trainees were in (i. e., to whether or not they had undergone simulator training), but no patient refused to undergo endoscopy by a supervised trainee. Nevertheless, a bias concerning the scoring behavior cannot completely be ruled out. In a previous pilot study, four simulator-trained endoscopic novices were compared with four endoscopists trained in the regular way. Interestingly, the regular-trained group scored better in respect of patient comfort and need for sedation during upper gastrointestinal endoscopy, though no explanation for this was offered [14]. In previous trials investigating the effect of simulator training on colonoscopy, patient comfort scores were significantly better in the simulator-trained groups [6, 13].

In conclusion, this randomized controlled trial shows that virtual simulator training significantly improves and accelerates the acquisition of technical skills and reduces the time required for the endoscopic procedure in patients during the early and mid-term stages of endoscopic training. Teaching procedures are less well tolerated with reference to discomfort and pain: simulator training does not seem to have a beneficial effect on patients' subjective endoscopy tolerance scores. On-patient training improves that score, but high numbers of endoscopic procedures performed in patients are needed to achieve acceptable scores compared to endoscopy performed by experts.

The integration of simulator training into an endoscopy training curriculum, especially in institutions with a high turnover of trainees and limited resources of expert fellows to provide supervision, might be considered, since this allows technical skills to be gained earlier in the education process, making supervision necessary less often. Clinical training in endoscopy, performing supervised on-patient endoscopies, cannot, however, be reduced or replaced by simulator training.

Acknowledgments

The authors want to thank Waltraud Hoschek for her administrative contribution and Peter Will, MD, for reading the manuscript.

Competing interests: None

References

1 *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. Am J Gastroenterol 2000; 95: 3129 – 3132
2 *Bini EJ, Firoozi B, Choung RJ et al.* Systematic evaluation of complications related to endoscopy in a training setting: a prospective 30-day outcomes study. Gastrointest Endosc 2003; 57: 8 – 16
3 *Cass OW, Freeman ML, Peine CJ et al.* Objective evaluation of endoscopy skills during training. Ann Intern Med 1993; 118: 40 – 44
4 *Tuggy ML.* Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. J Am Board Fam Pract 1998; 11: 426 – 433
5 *Ahlberg G, Hultcrantz R, Jaramillo E et al.* Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? Endoscopy 2005; 37: 1198 – 1204
6 *Cohen J, Cohen SA, Vora KC et al.* Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. Gastrointest Endosc 2006; 64: 361 – 368
7 *Ferlitsch A, Glauninger P, Gupper A et al.* Virtual endoscopy simulation for training of gastrointestinal endoscopy. Endoscopy 2002; 34: 698 – 702
8 *Rex DK.* Quality in colonoscopy: cecal intubation first, then what? Am J Gastroenterol 2006; 101: 732 – 734
9 *Shah HA, Paszat LF, Saskin R et al.* Factors associated with incomplete colonoscopy: a population-based study. Gastroenterology 2007; 132: 2297 – 2303
10 *Thomas-Gibson S, Bassett P, Suzuki N et al.* Intensive training over 5 days improves colonoscopy skills long-term. Endoscopy 2007; 39: 818 – 824
11 *Cohen J, Safdi MA, Deal SE et al.; ASGE/ACG Taskforce on Quality in Endoscopy.* Quality indicators for esophagogastroduodenoscopy. Am J Gastroenterol 2006; 101: 886 – 891
12 *Staritz M, Alkier R, Krzoska B et al.* The time requirement for endoscopic diagnosis and therapy: results of a multicenter study. Z Gastroenterol 1992; 30: 509 – 518
13 *Sedlack RE, Kolars JC, Alexander JA.* Computer simulation training enhances patient comfort during endoscopy. Clin Gastroenterol Hepatol 2004; 2: 348 – 352
14 *Sedlack RE.* Validation of computer simulation training for esophagogastroduodenoscopy: pilot study. J Gastroenterol Hepatol 2007; 8: 1214 – 1219