

# Polypectomy skills of gastroenterology fellows: can we improve them?

## Authors

Sascha C. van Doorn<sup>1</sup>, Barbara A. J. Bastiaansen<sup>1</sup>, Siwan Thomas-Gibson<sup>2</sup>, Paul Fockens<sup>1</sup>, Evelien Dekker<sup>1</sup>

## Institutions

<sup>1</sup> Department of Gastroenterology and Hepatology, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands

<sup>2</sup> Wolfson Unit for Endoscopy, St. Mark's Hospital and Imperial College London, London, United Kingdom

## submitted

10. September 2015

## accepted after revision

27. October 2015

## Bibliography

**DOI** <http://dx.doi.org/10.1055/s-0041-109086>

Published online: 8.1.2016

Endoscopy International Open

2016; 04: E182–E189

© Georg Thieme Verlag KG

Stuttgart · New York

E-ISSN 2196-9736

## Corresponding author

**Sascha van Doorn, MD**

Department of

Gastroenterology and

Hepatology

Academic Medical Centre

Meibergdreef 9

1105 AZ Amsterdam

The Netherlands

Fax: +31-20-691-7033

s.c.vandoorn@amc.uva.nl

**Background and aims:** Currently, most training programs for gastroenterology (GI) fellows lack systematic training in polypectomy. Systematic education and direct feedback with the direct observational polypectomy skills (DOPyS) method is a simple and inexpensive way to train GI fellows in practical endoscopy. Our primary aim was to evaluate whether a lecture-based training course could improve the polypectomy skills of GI fellows. As a secondary aim, the interobserver agreement among the three assessors was evaluated.

**Participants and methods:** We invited GI fellows to record five polypectomies, after which they attended a training course consisting of three lectures on polyps and polypectomy methods given by expert endoscopists. After training, the fellows recorded five polypectomies again. All videos were blindly assessed by three expert endoscopists, who used the DOPyS method.

## Introduction

Over the past decade, an increasing number of Western countries have started colorectal cancer (CRC) screening programs, generally with fecal immunochemical testing (FIT) or primary colonoscopy [1]. These screening programs can reduce CRC-related mortality by detecting CRC in an early and treatable stage, and can also reduce the incidence of CRC by detecting and removing precursor lesions [2,3]. As a result, the safe and radical removal of polyps has become the focus of increased attention. Recently, the direct observation of polypectomy skills (DOPyS) method was developed for the purpose of formal training in and assessment of basic and advanced polypectomy skills [4,5]. Gupta et al. demonstrated that the DOPyS method can help to differentiate reliably among endoscopists of varying degrees of skill [5].

**Results:** Eight GI fellows participated in this study. There was no significant difference in the median overall competency scores before and after training; before training, 25% (10/40) of the polypectomies were scored as “pass,” compared with 37.5% (15/40) after training ( $P=0.56$ ). The interobserver agreement among the experts was fair (intraclass correlation coefficient [ICC] 0.34, 95% confidence interval [CI] 0.14–0.52).

**Conclusions:** Our lecture-based training course did not result in an improvement in overall competency scores for the polypectomy skills of GI fellows. Besides, the overall quality of the polypectomy techniques of the fellows was considered low. To optimize polypectomy training and competency, we believe that direct feedback in the endoscopy suite and hands-on training by dedicated teachers are essential.

In 2014, a nationwide FIT-based screening program for CRC was started in the Netherlands. Because of recognized variability in colonoscopy skill [6–8], endoscopists who are to perform colonoscopies in FIT-positive screenees are obliged to undergo a formal accreditation process. Strict requirements were developed, including an assessment of their polypectomy skills with the DOPyS method [9].

During the training of GI fellows, systematic education in polypectomy is important. However, structured training programs are currently not available in most countries [4,5]. In order to structure endoscopy training methods, several simulator models have been developed over the years, the first dating from 1997 (Erlangen model) [10–13]. The Welsh Institute for Minimal Access Therapy (WIMAT) colonoscopy suitcase was recently developed as an ex vivo porcine simulator for polypectomy and has shown content validity for training in polypectomy skills [14–16]. However,

## License terms



among the disadvantages of this simulated polypectomy training simulator are the limited number of techniques that can be practiced, its scarce capacity, and its high cost [15,17]. Systematic education and direct feedback in daily practice with the DOPyS method comprise a simpler and less expensive method for training, providing both feedback and practical endoscopy practice. To our knowledge, no studies have evaluated the effect of lecture-based education on polypectomy skills. We hypothesized that increasing basic knowledge on assessing and removing polyps would improve endoscopic skills.

The aim of this study was to evaluate whether a short, lecture-based training course given by three expert endoscopists would improve the polypectomy skills of GI fellows, as measured with the validated DOPyS method before and after the training.

## Methods

### Study design

This was a prospective study involving GI fellows trained at the Academic Medical Center (AMC) in Amsterdam, the Netherlands. Review by the institutional review board was not required, as in agreement with the Medical Research Involving Human Subjects Act. In the Netherlands, the gastroenterology fellowship lasts 6 years, and from the fourth year, fellows start performing endoscopic procedures. In our center, the first 100 colonoscopies are completely supervised by a consultant who is present in the endoscopy suite. “On-demand” supervision follows, which means that a consultant is present whenever the fellow asks for supervision (e.g., if the fellow has difficulty reaching the cecum) and always when a polypectomy is performed. The polyp is shown to the consultant, and therapeutic options are discussed as well as anticipated possible difficulties in removing the polyp as a consequence of its location, size, or morphology. The consultant offers advice or intervenes when the fellow experiences difficulties during the polypectomy. This supervision is provided by most members of the academic staff of our department, all of whom are experienced in colonoscopy and polypectomy and some of whom are experts in advanced polypectomy.

Only GI fellows who had performed at least 100 colonoscopies and at least 20 polypectomies were invited to participate in this study. For the purpose of the study, the supervisor present during each polypectomy did not intervene or advise the fellow. In cases in which the supervisor felt it was in the patient’s best interest to intervene, the case was not used for analysis.

The fellows were asked to record by video five polypectomy procedures performed on polyps smaller than 15 mm at baseline. Before the start of the study, all fellows were informed about the DOPyS assessment forms, so they knew which aspects of the polypectomies would be scored. After all participating fellows had obtained their videos, they attended a training course based on three lectures. To participate in the study, they were required to attend all training lectures. After the course, they recorded another five polypectomy procedures. The videos were arranged in random order and assessed by three expert endoscopists. The experts independently scored all videos with the DOPyS method.

### Polypectomy videos

All fellows recorded five consecutive polypectomies before and five after the lectures. Only polyps smaller than 15 mm without any suspicion of malignancy were recorded for the study. During colonoscopy, all quality indicators as well as polyp features were

recorded in a standardized colonoscopy reporting system [18]. Polyp features were the following: morphology (pedunculated, sessile, flat, or depressed according to the Paris classification [19]); estimated size; optical diagnosis; and polypectomy technique. All videos were edited to include an endoscopic view of the entire procedure, from detection to polypectomy and polyp retrieval. The segment of the colon was not included because it was impossible to assess this from the endoscopic view.

### Training

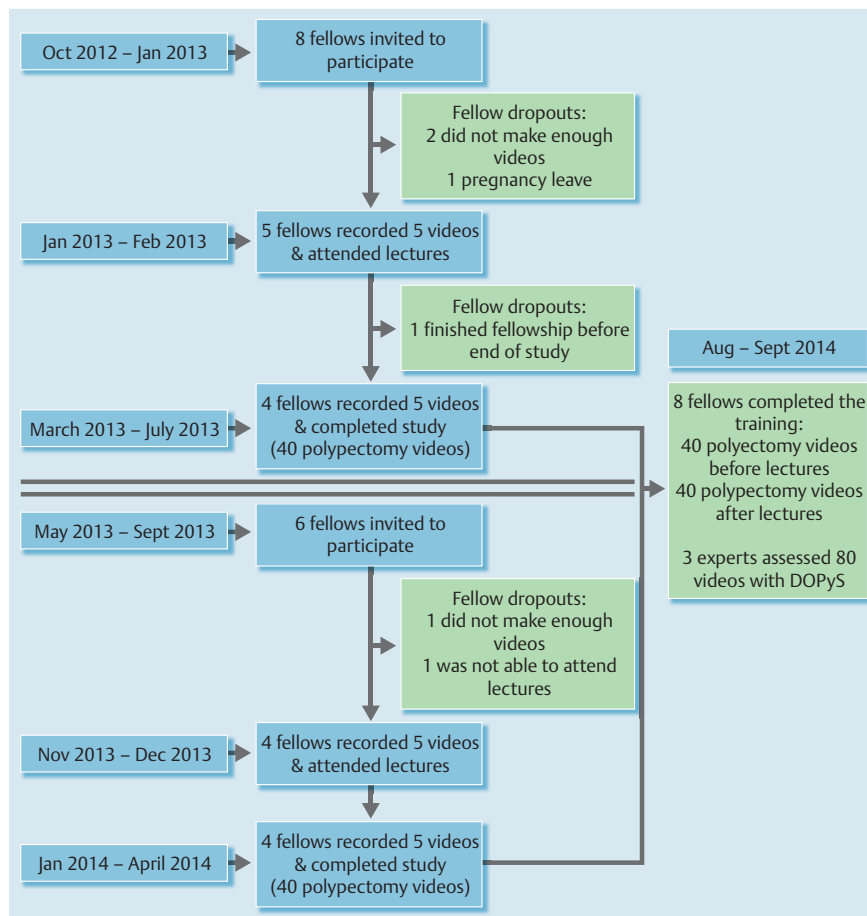
After recording five polypectomies, the fellows attended a training course of three lectures presented by three expert endoscopists specializing in colonoscopy and advanced polypectomy (E.D., B.B., and P.F.). To stimulate interaction, the lectures were given for a small group of fellows only. The lectures were created for the purpose of the study, and the fellows had not attended any similar lectures before. The first lecture was on the optical diagnosis of polyps. All aspects of polyp characterization and observation were discussed: location, size, morphology, Kudo pit pattern, endoscopic assessment of histology, signs of invasive growth, and on-site decision making regarding therapy and surveillance intervals. The second lecture was on the effective and safe removal of polyps. Different polypectomy techniques were discussed, and video examples with correct and incorrect techniques were shown. The final lecture discussed the most common complications of colonoscopy and polypectomy. The fellows were taught how to recognize symptoms of complications and how to treat and prevent them.

### Direct observational polypectomy skills assessment

After all videos had been recorded and edited, the same expert endoscopists independently assessed the videos, which were placed in a file in random order. The assessors were blinded to the endoscopist’s identity and the timing of the polypectomy (before or after training). Each video was scored with the DOPyS method. Before the start of the study, the experts had had experience with the DOPyS method but had not discussed the assessment criteria or assessed a selection of videos together. Each of the 32 parameters of the DOPyS method, as well as overall competency, was scored as 1 (standards not met), 2 (some uncorrected errors), 3 (competent and safe performance), or 4 (highly skilled performance). Scores below 3 implied that part of the procedure had not been adequately performed and thus that the endoscopist had “failed.” Only parameters that could be scored with a video were assessed (excluding all underlined items on the original DOPyS assessment form) [4,5].

### Study outcomes and statistical analysis

The primary study aim was to evaluate whether the polypectomy skills of GI fellows had improved after the training course. Three expert endoscopists assessed the polypectomies with the DOPyS method, and the median score of the three assessments was used for further analysis. Comparisons were made for the overall competency score per polypectomy and for each individual parameter of the DOPyS method. Because the data were not parametric, we used Wilcoxon’s signed rank test to evaluate the differences in these parameters on a scale of 1 to 4 (1, standards not met; 4, highly skilled performance) before and after the training course, with the median of the five polypectomy scores per fellow per occasion used as the unit of analysis. The differences in overall competency and in each individual DOPyS parameter before and after training were also calculated across the pass/fail divide, and the



**Fig. 1** Study flowchart – inclusion of participating fellows. DOPyS, direct observational polypectomy skills.

scores were separated into two groups: 1 s and 2 s (i.e., suboptimal performance/fail) and 3 s and 4 s (i.e., competent performance/pass). A linear mixed model was used to evaluate this difference with the separate scores per polypectomy used as unit of analysis, where a random intercept on the fellow level was entered to account for the correlation between polyp scores from the same fellow.

Secondary aims were to evaluate the differences in overall competency among fellows and the interobserver agreement among the three assessors. The chi-squared test was used to compare the scores among the fellows. The interobserver agreement among the three experts was calculated with an intraclass correlation coefficient (ICC) (ordinal scale), and a Fleiss kappa (pass/fail divide, more than two observers). Kappa values were interpreted according to Landis and Koch [20,21].

## Results

### Fellows

Initially, eight GI fellows were invited to participate in the study in October 2012 (Fig. 1). Four of the eight fellows dropped out. Reasons for failing to finish the study were inability to record enough polypectomy videos before the scheduled lectures (2), completion of fellowship before enough procedures were recorded after the lectures (1), and pregnancy leave (1). In the first group, the average time between recording the videos before and after the course was 3 1/2 months, and in those months the lectures were attended. In May 2013, a second group consisting of six fellows was invited to participate. Of these fellows, four

completed the study. The average time between recording the polypectomy videos before and after the lectures in the second group was longer, 8 months. A total of eight fellows completed the entire study, resulting in 40 pre- and 40 post-training videos. The duration of the final videos ranged from 30 seconds to 1.40 minutes, with an average of 1.10 minutes.

### Polyp characteristics

Table 1 shows the characteristics of the polyps that were removed before and after training, as described by the fellows in the colonoscopy reports. The mean polyp size before training was 5.8 mm (range 2–15 mm) and after training 5.0 mm (range 2–10 mm, not significant). Polyp morphology was mostly sessile, also not significantly different before and after training.

### Polypectomy technique

Before training, most of the polyps (23/40, 58%) were removed by lifting with saline and snare diathermy (Table 2). However, after training, significantly more polyps were removed by the cold snare technique (28/40, 70%;  $P=0.032$ ).

### Overall direct observational polypectomy skills scores

Table 3 shows the overall scores (scale of 1–4) for polypectomy before and after training of the eight GI fellows who completed the study. There was no significant difference between the median overall competency scores before and after training; before training 25% (10/40) of the polypectomies were scored as “pass,” compared with 37.5% (15/40) after training ( $P=0.56$ , Table 4). The mixed model assessing the difference between passing and failing on the polypectomy level showed an odds ra-

	Before training (N=40)	After training (N=40)	P value (chi-squared test)
Size, mm			0.788
1–5	25	27	
6–9	9	9	
15	6	4	
Morphology <sup>1</sup>			0.676
Flat	6	9	
Sessile	29	27	
Pedunculated	5	4	

<sup>1</sup> According to the Paris classification of superficial neoplasia [19]

**Table 1** Characteristics of the polyps removed by the gastroenterology fellows participating in the study.

	Before training (N=40)	After training (N=40)	P value (chi-squared test)
Polypectomy technique			0.032
Cold snare alone	11	14	
Lift and cold snare	6	14	
Snare diathermy alone	0	1	
Lift and snare diathermy	23	11	

**Table 2** Polypectomy methods used by the study participants.

tio of 1.91 (95%CI 0.69–5.31) for a “pass” score after training when compared with before training. The estimated number of “passed” polypectomies, accounting for the correlation of polypectomy scores from the same fellow, were 22% before training vs. 36% after training, which approximates the observed numbers. The increase in “passed” polypectomies was not statistically significant ( $P=0.21$ ).

### Individual direct observation polypectomy skills parameter scores

The median scores given for each individual DOPyS parameter (on the scale of 1–4 with Wilcoxon’s signed rank test) that could be scored with a video per fellow were compared before and after training (Supplementary Table e5, available online only). There was no significant difference between any of the parameter scores before and after training except parameter 30, “identifies and appropriately treats residual polyp,” for which the median score was lower after training than before (2.3 vs. 1.5,  $P=0.01$ ). We used the mixed models method to assess the differences in passed and failed polypectomies for the different parameters (odds ratios per parameter shown in Supplementary Table e5). The differences between “passed” individual polypectomy parameters before and after training were not statistically significant.

### Differences in overall scores among fellows

The differences in overall competency scores among the fellows were evaluated. Before the training, the overall competency scores across the pass/fail levels differed significantly (Table 6,  $P<0.05$ ). After training, the differences were no longer significant.

### Interobserver agreement among three experts

Analysis of the interobserver agreement of the three expert endoscopists showed a fair agreement for overall DOPyS competency (ICC 0.34, 95%CI 0.14–0.52). Sensitivity analyses showed that this moderate agreement was caused by a deviation in assessment by expert 3; the agreement between expert 1 and expert 2 was also fair (kappa 0.32); however, between observer 1 and 3 and between observer 2 and observer 3, there was only a slight agreement (kappa 0.05 and kappa 0.15, respectively). The interobserver agreement among the overall competency scores

across pass/fail divide was also fair, with a Fleiss kappa of 0.28 (95%CI 0.15–0.40). For most individual DOPyS parameters, the interobserver agreement was slight, fair, or moderate (data not shown).

### Discussion

In most countries, GI fellows acquire endoscopic skills through experiential learning, and fellowship programs lack structured training in polypectomy. Although endoscopy simulators can be helpful for training in simple endoscopic interventions (e.g., polypectomy), they are not ideal and not available in many teaching hospitals. In this study, we aimed to improve polypectomy skills with a training interaction consisting of three dedicated lectures on polyps and polypectomy, a simple and inexpensive method that could easily be applied in any training center. To evaluate its effect, we assessed the polypectomy skills of eight GI fellows with the DOPyS method before and after they attended this training. Regrettably, our study did not show an improvement in the overall competency scores of the fellows, or in the scores of any of the individual parameters of the DOPyS. Besides, the overall quality of the polypectomy techniques of the fellows was considered poor. The percentage of polypectomies that were scored as “pass” increased from 25.0% before to 37.5% after training, but the increase did not reach statistical significance ( $P=0.56$ ). The interobserver agreement of overall competency scores among the three expert endoscopists who assessed the videos was only fair (ICC 0.34).

Based on our results, we conclude that a short, lecture-based curriculum is insufficient to improve the polypectomy skills of GI fellows. After training, only a small, insignificant increase in the percentage of “passed” polypectomies, from 25.0% to 37.5%, was observed. All the fellows in our study independently performed polypectomies, but the levels of experience differed among them. Interestingly, the two fellows whose skills did improve after the course recorded their first set of polypectomy videos very early in their training. Because the learning curve is steepest at the beginning of training, this could explain their improvement over time [22–24]. The difference in overall competency scores among the fellows was evaluated, and before training, the overall competency scores differed significantly ( $P<0.05$ ). After

**Table 3** Direct observational polypectomy skills overall competency scores of the eight gastroenterology fellows before and after training.

	Overall competency score, median of 3 observers	Polypectomies before training, n	Polypectomies after training, n
Fellow 1	1		
	2	2	5
	3	3	
	4		
	Fail (1 or 2)	2	5
	Pass (3 or 4)	3	
Fellow 2	1	1	1
	2	4	2
	3		2
	4		
	Fail (1 or 2)	5	3
	Pass (3 or 4)		2
Fellow 3	1		
	2	5	3
	3		2
	4		
	Fail (1 or 2)	5	3
	Pass (3 or 4)		2
Fellow 4	1	1	1
	2	4	3
	3		1
	4		
	Fail (1 or 2)	5	4
	Pass (3 or 4)		1
Fellow 5	1		1
	2	5	4
	3		
	4		
	Fail (1 or 2)	5	5
	Pass (3 or 4)		
Fellow 6	1		
	2	3	2
	3	2	3
	4		
	Fail (1 or 2)	3	2
	Pass (3 or 4)	2	3
Fellow 7	1	1	1
	2	2	
	3	2	4
	4		
	Fail (1 or 2)	3	1
	Pass (3 or 4)	2	4
Fellow 8	1		1
	2	2	1
	3	3	2
	4		1
	Fail (1 or 2)	2	2
	Pass (3 or 4)	3	3

training, however, the difference was no longer significant. Therefore, it seems that the overall skills of the fellows were more alike after the training course than before. However, this could also be a coincidence because of the small number of fellows included in the study.

We also compared the techniques for polypectomy that the fellows used before and after training. After training, significantly more polyps were removed with the cold snare technique (70%) than before training, when most (58%) were removed with diathermy snare. Surprisingly, this did not result in different scores for the DOPyS parameter “appropriate technique.” Although a dif-

**Table 4** Differences in overall competency before and after training.

	Overall competency		Wilcoxon's signed rank test
	Before training, median score	After training, median score	$P = 0.56$
Fellow 1	3	2	
Fellow 2	2	2	
Fellow 3	2	2	
Fellow 4	2	2	
Fellow 5	2	2	
Fellow 6	2	3	
Fellow 7	2	3	
Fellow 8	3	3	

**Table 6** Differences among fellows in overall competency scores across the pass/fail divide.

	Fail	Pass	$P$ value
<i>Before training</i>			0.045
Fellow 1	2	3	
Fellow 2	5	0	
Fellow 3	5	0	
Fellow 4	5	0	
Fellow 5	5	0	
Fellow 6	3	2	
Fellow 7	3	2	
Fellow 8	2	3	
<i>After training</i>			0.08
Fellow 1	5	0	
Fellow 2	3	2	
Fellow 3	3	2	
Fellow 4	4	1	
Fellow 5	5	0	
Fellow 6	2	3	
Fellow 7	1	4	
Fellow 8	2	3	

ferent method was applied, this was not scored higher or lower by the observers. We think the difference in applied techniques is the result of increasing experience of the participating fellows, but it may partially have been caused by the training course.

The initial study describing the development of the DOPyS method was designed to test its content validity, not to assess the competency of individual endoscopists [4]. Videos of polypectomies performed by accredited screening endoscopists within the British screening program were assessed by seven individuals, and the authors concluded that the DOPyS method can be used as training tool for all endoscopists [4]. In a second study from the same group, the DOPyS method was further validated. This study concluded that the DOPyS method can reliably differentiate among endoscopists of varying skills; however, the authors also concluded that untrained examiners might not make reliable assessments. The authors also advised that further large-scale validation studies are required to strengthen the evidence base for this novel assessment tool. In both the British and Dutch screening programs, the DOPyS method is already used in the accreditation process for endoscopists.

This study has several limitations. We included a limited study group of eight fellows. Initially, more fellows started the training course but dropped out for reasons not related to their skills; in most cases, the reasons were logistical. The fellows, both those



who finished and those who dropped out, were in various stages of training. We therefore do not believe that this resulted in a bias in our study. Besides, our study was not powered to detect a difference in DOPyS scores, nor did we perform any sample size calculations. We chose to evaluate the effect of three lectures because this was logistically feasible within one teaching hospital. However, if a larger group of fellows had been included, preferably from several hospitals, a larger number of videos could have been assessed. The time between the lectures and the second set of videos was relatively long (6 months). We observed a small, insignificant increase in “passed” polypectomies. This improvement could also have been the result of the fellows’ expected experiential learning and receipt of feedback from their supervisors in the endoscopy suite over time, and thus independent of the training lectures.

The study was designed to test our hypothesis that increasing fellows’ basic knowledge of assessing and removing polyps could improve their endoscopic skills. Although the polypectomy skills of the fellows did not significantly improve after the training, their knowledge actually might have improved, but this was not measured. Maybe the lack of benefit from the lectures did not result from the inability of the lectures to improve technical skills but from the key messages of the lectures not being retained. However, the major conclusion of our study would still be valid, as the lessons did not improve the skills.

Although the assessors were familiar with the DOPyS method, they were not actually trained in using it, and the scores of the observers for the GI fellows varied widely. Observer 3 was more severe than the other two, resulting in a median score per fellow of “fail” for all fellows by this observer, whereas observer 1 scored the majority of the polypectomies as “pass.” The assessors in our study may have been too severe, resulting in many failed polypectomies. The expectations may have been too high and our assessors overly critical, as if these expert endoscopists were looking for expert fellows instead of GI fellows using competent and safe techniques.

In line with the data of Gupta et al., in our study the interobserver agreement among assessors who were not trained in use of the DOPyS method was low [5]. Several ways to optimize the actual scoring in such a study should be considered. Training observers in using the DOPyS method might have resulted in more equal scores, or feedback to the scorers during the course of the study might have acted as a benchmark and balanced the “hawk-dove” phenomenon (a harsh observer vs. a generous observer) [25,26]. Before the start of the study, all participating fellows were informed about the DOPyS assessment form, so they knew what would be tested. However, they did not receive personal feedback on their pre-training videos, which could have resulted in higher scores after the training. We did not provide them with feedback because we intended to evaluate only the effect of the lectures, so the assessors assessed all videos in random order. Finally, the fellows independently performed polypectomies with a supervisor present in the room (according to local protocol). Although the research fellow was usually also present to make a video of the polypectomy, this could have resulted in the generation of verbal or nonverbal instructions not visible in the endoscopy video. However, because the DOPyS scores in our study were still remarkably low, we do not think this caused an important bias in our study.

Our study suggests that the fellows in our academic center lack competence to perform therapeutic colonoscopy. We believe that there are two potential explanations for these disappointing

results. Firstly, the competency scores of our fellows could have been low because the assessors were very harsh. However, these disappointing scores could also have been the result of an insufficient training program in our center, requiring improvement. We have learned that it is important to provide fellows with more direct feedback in the endoscopy-suite. To improve polypectomy skills, hands-on training by trained trainers should be provided [24,27,28]. Therefore, we should also assess the teaching skills of our staff endoscopists and give them tools to optimize their competence as teachers – for example with a “teach-the-teacher” program. Such dedicated training courses aiming to develop teaching skills have been developed and are available in the United Kingdom (Joint Advisory Group on Gastrointestinal Endoscopy [JAG] Endoscopy Training System, or JETS) to improve the effectiveness of clinical training [22, 29–31]. In this study, the DOPyS method was used as a form of summative assessment. However, the DOPyS might have been better used as a formative assessment tool to structure teaching and providing immediate feedback to the trainee. A future study is planned to address this way of teaching, using the DOPyS method as a platform to aid training and define a learning curve for polypectomy.

Worldwide, training in colonoscopy and polypectomy varies widely, reflecting uncertainties regarding the optimal design of such training programs [27]. For a standardized polypectomy training program, several elements are important. First, as only limited guidelines are currently available, evidence-based polypectomy guidelines should be developed. Second, dedicated gastroenterologists who provide training to fellows should be advised about their teaching skills in a “teach-the-teacher” program. Last but not least, a standardized program for colonoscopy training, including regular assessments with DOPS (direct observation of procedure or skills) and DOPyS, should be designed. Future studies could then evaluate the effect of the implementation of standardized colonoscopy and polypectomy training programs on actual polypectomy skills.

**Competing interests:** None

## References

- 1 Zavoral M. Colorectal cancer screening in Europe. *World J Gastroenterol* 2009; 15: 5907–5915
- 2 Libby G, Brewster DH, McClements PL et al. The impact of population-based faecal occult blood test screening on colorectal cancer mortality: a matched cohort study. *Br J Cancer* 2012; 107: 255–259
- 3 Zauber AG, Winawer SJ, O'Brien MJ et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 2012; 67: 687–696
- 4 Gupta S, Anderson J, Bhandari P et al. Development and validation of a novel method for assessing competency in polypectomy: direct observation of polypectomy skills. *Gastrointest Endosc* 2011; 73: 1232–1239.e2
- 5 Gupta S, Bassett P, Man R et al. Validation of a novel method for assessing competency in polypectomy. *Gastrointest Endosc* 2012; 3: 568–575
- 6 de Jonge V, Sint Nicolaas J, Cahen DL et al. Quality evaluation of colonoscopy reporting and colonoscopy performance in daily clinical practice. *Gastrointest Endosc* 2012; 75: 98–106
- 7 Piette C, Manfredi S. Variations between endoscopists in rates of detection of colorectal neoplasia and their impact on a regional screening program based on colonoscopy after fecal occult blood testing. *Gastrointest Endosc* 2010; 71: 335–341
- 8 Chen SC, Rex DK. Endoscopist can be more powerful than age and male gender in predicting adenoma detection at colonoscopy. *Am J Gastroenterol* 2007; 17: 856–861
- 9 National Institute for Public Health and the Environment. Protocol for the authorization and auditing of colonoscopy centres and endoscopists. April 30, 2014. Available from: <http://www.rivm.nl/en/Docu>

- ments\_and\_publications/Professional\_Serviceable/Protocols/Disease\_Prevention\_and\_Healthcare/Protocol\_for\_the\_authorization\_and\_auditing\_of\_colonoscopy\_centres\_and\_endoscopists (Accessed 29 October 2015)
- 10 Hochberger J, Matthes K, Maiss J et al. Training with the compactEASIE biologic endoscopy simulator significantly improves hemostatic technical skill of gastroenterology fellows: a randomized controlled comparison with clinical endoscopy training alone. *Gastrointest Endosc* 2005; 61: 204–215
  - 11 Matthes K, Cohen J, Kochman ML et al. Efficacy and costs of a one-day hands-on EASIE endoscopy simulator train-the-trainer workshop. *Gastrointest Endosc* 2005; 62: 921–927
  - 12 Maiss J, Millermann L, Heinemann K et al. The compactEASIE is a feasible training model for endoscopic novices: a prospective randomised trial. *Dig Liver Dis* 2007; 39: 70–78; discussion 79–80
  - 13 Hochberger J, Euler K, Naegel A et al. 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. *Scand J Gastroenterol* 2004; 39: 895–902
  - 14 Ansell J, Arnaoutakis K, Goddard S et al. The WIMAT colonoscopy suitcase model: a novel porcine polypectomy trainer. *Colorectal Dis* 2013; 15: 217–223; discussion 223
  - 15 Ansell J, Hurley JJ, Horwood J et al. The Welsh Institute for Minimal Access Therapy colonoscopy suitcase has construct and concurrent validity for colonoscopic polypectomy skills training: a prospective, cross-sectional study. *Gastrointest Endosc* 2014; 79: 490–497
  - 16 Ansell J, Hurley JJ, Horwood J et al. Can endoscopists accurately self-assess performance during simulated colonoscopic polypectomy? A prospective, cross-sectional study *Am J Surg* 2014; 207: 32–38
  - 17 Haycock AV, Youd P, Bassett P et al. Simulator training improves practical skills in therapeutic GI endoscopy: results from a randomized, blinded, controlled study. *Gastrointest Endosc* 2009; 70: 835–845
  - 18 van Doorn SC, van Vliet J, Fockens P et al. A novel colonoscopy reporting system enabling quality assurance. *Endoscopy* 2014; 46: 181–187
  - 19 Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy* 2005; 37: 570–578
  - 20 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174
  - 21 Landis JR, Koch GG. The measurement of observer agreement for categorical data. International Biometric Society. Available from: [http://www.dentalage.co.uk/wp-content/uploads/2014/09/landis\\_jr\\_koch\\_gg\\_1977\\_kappa\\_and\\_observer\\_agreement.pdf](http://www.dentalage.co.uk/wp-content/uploads/2014/09/landis_jr_koch_gg_1977_kappa_and_observer_agreement.pdf) (Accessed 30 October 2015)
  - 22 Ward ST, Mohammed MA, Walt R et al. An analysis of the learning curve to achieve competency at colonoscopy using the JETS database. *Gut* 2014; 63: 1746–1754
  - 23 Shahidi N, Ou G, Telford J et al. Establishing the learning curve for achieving competency in performing colonoscopy: a systematic review. *Gastrointest Endosc* 2014; 80: 410–416
  - 24 Sedlack RE. Training to competency in colonoscopy: assessing and defining competency standards. *Gastrointest Endosc* 2011; 74: 355–366.e1–2
  - 25 Barton JR, Corbett S, van der Vleuten CP. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. *Gastrointest Endosc* 2012; 75: 591–597
  - 26 Wilkinson JR, Crossley JGM, Wragg A et al. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ* 2008; 42: 364–373
  - 27 Ekkelenkamp VE, Koch AD, de Man RA et al. Training and competence assessment in GI endoscopy: a systematic review. *Gut* 30. 01 2015: [Epub ahead of print]. DOI 10.1136/gutjnl-2014-307173
  - 28 Sedlack RE. The Mayo Colonoscopy Skills Assessment Tool: validation of a unique instrument to assess colonoscopy skills in trainees. *Gastrointest Endosc* 2010; 72: 1125–1133.e1–3
  - 29 Joint Advisory Group on GI Endoscopy. Royal College of Physicians . JAG Endoscopy Training System. Available from: <http://www.jets.nhs.uk> Accessed 27 October 2015
  - 30 Norton ID, Bell C. Delegate perceptions of a national train-the-colonoscopy trainer program: an essential component of quality colonoscopy training. *Gastrointest Endosc* 2010; 71: AB215
  - 31 Steinert Y, Mann K, Centeno A et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach* 2006; 28: 497–526

**Supplementary Table e5** Comparison of individual direct observational polypectomy score (DOPyS) parameters for polypectomy across the pass/fail divide.

DOPyS parameter	Parameter description	Polypectomies before training, median [IQR]; n (% pass)	Polypectomies after training, median [IQR]; n (% pass)	Wilcoxon's signed rank test / linear mixed model
1	Optimizes polyp position	2.5 [2;3]; 19/40 (47.5%)	2 [2;3]; 20/40 (50%)	0.66, OR 1.1, <i>P</i> =0.82
2	Optimizes view	2 [2;2.8]; 17/40 (42.5%)	2 [2]; 14/40 (35%)	0.32, OR 0.72, <i>P</i> =0.49
3	Visualizes full extent of polyp	2 [2]; 12/40 (30%)	2 [2]; 11/40 (27.5%)	1.00, OR 0.88, <i>P</i> =0.80
4	Appropriate technique	2.5 [2;3]; 21/40 (52.5%)	3 [3]; 23/40 (57.5%)	0.32, OR 1.2, <i>P</i> =0.65
11	Prophylactic hemostatic measures	3 [2.4;3]; 6/8 (75%)	2 [1;2]; 1/4 (25%)	1.10, OR 0.10, <i>P</i> =0.16
12	Appropriate snare size	2.75 [2.4;3]; 5/8 (62.5%)	2.5 [2;2.5]; 2/4 (50%)	1.00, OR 0.6, <i>P</i> =0.69
13	Directs snare over head	2.25 [2;2.7]; 2/8 (25%)	1.5 [1;1.5]; 2/4 (50%)	0.41, OR 11.5, <i>P</i> =0.27
14	Stalked polyps: Selects en bloc or piecemeal removal	3 [2.8;3]; 7/8 (87.5%)	2 [2]; 2/4 (50%)	0.32, OR 0.07, <i>P</i> =0.26
15	Snare to stalk	2 [1.9;2.5]; 1/8 (12.5%)	1.5 [1;1.5]; 2/4 (50%)	0.41, OR 9.1, <i>P</i> =0.23
16	Snare position on stalk	2.25 [1.4;2.6]; 3/8 (37.5%)	1.5 [1;1.5]; 2/4 (50%)	1.00, OR 3.5, <i>P</i> =0.49
17	Mobilizes polyp to ensure appropriate amount of tissue is trapped within snare	2.4 [1.9;2.9]; 3/8 (37.5%)	1.5 [1;1.5]; 2/4 (50%)	0.59, OR 1.7, <i>P</i> =0.69
18	Applies appropriate diathermy	3 [2;3]; 6/8 (75%)	1 [1]; 2/4 (50%)	0.18, OR 0.15, <i>P</i> =0.36
19	Adequate submucosal injection	2 [1.6;2.2]; 6/29 (20.7%)	2 [1.1;3.4]; 6/27 (22.2%)	0.40, OR 1.1, <i>P</i> =0.89
20	Proceeds only if lesion lifts adequately	2.5 [2.1;3]; 16/29 (55.2%)	2.8 [2.5;3]; 14/25 (56%)	0.34, OR 1.1, <i>P</i> =0.89
21	Appropriate snare size	3 [3]; 28/38 (73.7%)	3 [3]; 34/39 (87.2%)	0.32, OR 2.4, <i>P</i> =0.15
22	Directs snare over lesion	2.3 [2;3]; 14/38 (36.8%)	2.5 [2;3]; 18/39 (46.2%)	0.32, OR 1.5, <i>P</i> =0.41
23	Small sessile polyps: Selects en bloc or piecemeal removal	3 [3]; 35/38 (92.1%)	3 [3]; 37/39 (94.9%)	0.89, OR 1.4, <i>P</i> =0.71
24	Positions snare over lesion as snare is closed	2 [2;2.4]; 8/38 (21.1%)	2 [2;2.8]; 11/39 (28.2%)	1.00, OR 1.6, <i>P</i> =0.44
25	Traps appropriate amount of tissue	2 [2]; 9/38 (23.7%)	2 [2]; 7/39 (17.9%)	0.71, OR 0.69, <i>P</i> =0.52
26	Tents lesion away from mucosa	2.5 [2;3]; 18/38 (47.4%)	2.8 [2;3]; 16/39 (41%)	1.00, OR 0.78, <i>P</i> =0.58
27	Chooses diathermy or cold snare	2.8 [2;3; 18/38 (47.4%)]	3 [2.4;3]; 25/39 (64.1%)	0.39, OR 2.0, <i>P</i> =0.15
28	Ensures adequate hemostasis before further resection	3 [2;3]; 5/7 (71.4%)	2.5 [2, 3]; 3/6 (50%)	0.32, OR 0.35, <i>P</i> =0.43
29	Examines remnant stalk/polyp base	2 [2,3]; 18/40 (45%)	2 [2]; 13/40 (32.5%)	0.58, OR 0.59, <i>P</i> =0.26
30	<b>Identifies and treats residual polyp</b>	<b>2.3 [2;3]; 14/30 (46.7%)</b>	<b>1.5 [1;1.9]; 4/30 (13.3%)</b>	<b>0.01, OR 0.18, <i>P</i>=0.01</b>
31	Identifies and treats bleeding	2.6 [2;3.4]; 14/29 (48.3%)	2.5 [2;2.5]; 10/23 (43.5%)	0.40, OR 0.82, <i>P</i> =0.73
32	Retrieves or attempts retrieval of polyp	2 [1.1;2]; 12/39 (30.8%)	2 [1.1;2.4]; 14/40 (35%)	0.68, OR 1.2, <i>P</i> =0.69

IQR, interquartile range; OR, odds ratio.