Vivian E. Ekkelenkamp

Competence & quality assessment

The future of training in GI endoscopy

Competence and quality assessment - the future of training in GI endoscopy

Vivian E. Ekkelenkamp

COLOFON

Cover design: Dennis Happé
Layout and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

ISBN: 978-94-6169-532-1

The work presented in this thesis was conducted at the department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, the Netherlands.

Financial support for printing this thesis was kindly given by:
Zambon Nederland B.V.; Olympus Nederland B.V.; Dr. Falk Pharma B.V.; Ferring B.V.; Bayer B.V.; Vifor Pharma Nederland B.V.; Norgine B.V.; Vrest Medical B.V.; ERBE Nederland B.V.; Gilead Sciences Netherlands B.V.; Astellas Pharma B.V.; Nederlandse Vereniging voor Gastroenterologie; Department of Gastroenterology and Hepatology, Erasmus University Medical Center Rotterdam; Erasmus University Rotterdam.

Competence and quality assessment –
the future of training in GI endoscopy

Competentie- en kwaliteitsevaluatie –
de toekomst van opleiding in
gastro-intestinale endoscopie

PROEFSCHRIFT
ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 5 september 2014 om 13:30 uur

door

Vivian Elske Ekkelenkamp

geboren te Norg

PROMOTIECOMMISSIE

Promotor: Prof.dr. E.J. Kuipers
Copromotor: Dr. R.A. de Man

Overige leden: Prof.dr. M.J. Bruno
Prof.dr. A.A.M. Masclee
Prof.dr. J.L.C.M. van Saase

Paranimfen: drs. A.F. Klijnsma
dr. E.C.M. Sikkens

Voor mijn ouders

TABLE OF CONTENTS

Section I

Introduction

Chapter 1

Training and competence assessment in gastrointestinal endoscopy: a systematic review

Vivian E. Ekkelenkamp*, Arjun D. Koch*, Robert A. de Man, Ernst J. Kuipers

** Authors contributed equally to this work*

Submitted

ABSTRACT

Introduction: Training procedural skills in gastrointestinal endoscopy once focused on threshold numbers. However, as threshold numbers poorly reflect individual competence, the focus gradually shifts towards a more individual approach. Tools to assess and document individual learning progress are being developed and incorporated in dedicated training curricula. However, there is a lack of consensus and training guidelines differ worldwide, which reflects uncertainties on optimal set-up of a training program.

Aims: The primary aim of this systematic review was to evaluate the currently available literature for the use of training and assessment methods in GI endoscopy. Secondly, we aimed to identify the role of simulator-based training as well as the value of continuous competence assessment in patient-based training. Thirdly, we aimed to propose a structured training curriculum based on the presented evidence.

Methods: A literature search was carried out in the available medical and educational literature databases. The results were systematically reviewed and studies were included using a predefined protocol with independent assessment by two reviewers and a final consensus round.

Results: The literature search yielded 5846 studies. Ninety-four relevant studies on simulators, assessment methods, learning curves and training programs for gastrointestinal endoscopy met the inclusion criteria. Twenty-seven studies on simulator validation were included. Good validity was demonstrated for four simulators. Twenty-three studies reported on simulator training and learning curves, including 17 RCT's. Increased performance on a virtual reality simulator was shown in all studies. Improved performance in patient-based assessment was demonstrated in 14 studies. Four studies reported on the use of simulators for assessment of competence levels. Simulator-based performance did not reflect competence in patient-based endoscopy. Eight out of fourteen studies on colonoscopy, ERCP and EUS reported on learning curves in patient-based endoscopy and proved the value of this approach for measuring performance. Ten studies explored the numbers needed to gain competence, but the proposed thresholds varied widely between them. Five out of nine studies describing the development and evaluation of assessment tools for gastrointestinal endoscopy provided insight in performance of endoscopists. Five out of seven studies proved that intense training programs result in good performance.

Conclusions: The use of validated virtual reality simulators in the early training setting accelerates learning of practical skills. Learning curves are valuable for continuous assessment of performance and are more relevant than threshold numbers. Future research will strengthen these conclusions by evaluating simulation-based as well as patient-based training in gastrointestinal endoscopy. A complete curriculum with assessment of competence throughout training needs to be developed for all gastrointestinal endoscopy procedures.

INTRODUCTION

The focus on training in procedural skills in gastrointestinal (GI) endoscopy is shifting from threshold numbers towards an individual approach. This illustrates the awareness that the classic master-apprentice model may not reflect all necessary aspects of training. Moreover, the old adage 'see one, do one' seems no longer appropriate for educating health professionals to perform complex technical procedures, such as flexible endoscopy.[1] Virtual reality (VR) simulators may be of benefit in the education of gastroenterology trainees. However, a substantial part of training still has to be patient-based. The assessment of a trainee's competence is not clearly defined and competence benchmarks for trainees are sparse. The use of threshold numbers is nowadays considered a poor surrogate marker for competence. Keeping track of one's performance by measuring skill development seems preferable. However, training guidelines differ worldwide and there is no consensus on the skills a trainee has to possess at the end of education. On top of that, for most procedural skills in flexible endoscopy, the proper assessment tools to measure these skills are lacking.

The aim of this systematic review was therefore to evaluate the available literature on different training and assessment methods in gastrointestinal endoscopy. Secondly, we aimed to identify the role of simulator training and competence development in patient-based training, specifically for procedures that normally will be learnt during residency. Thirdly, we aimed to propose a structured training curriculum based on the presented evidence.

METHODS

Literature search strategy

A systematic literature search was carried out in July 2013 in seven different medical and educational literature databases: Embase, Medline OvidSP, Web of Science, Cochrane central, Google Scholar, Research and Development Resource Base (RDRB) and Education Recourse Information Center (ERIC). There was no restriction regarding time of publication or language.

In- and exclusion criteria

All studies pertaining to training and assessment in gastrointestinal endoscopy (colonoscopy, endoscopic retrograde cholangiopancreatography (ERCP), endosonography (EUS), and upper gastrointestinal endoscopy) were included in this review. The studies were to report outcome measures with respect to learning curves, assessment methods or tools and training programs including simulators. Two reviewers independently examined all retrieved studies. When disagreement existed over studies to be in- or excluded, these were discussed until consensus was reached. Reviews, systematic reviews, meta-analyses and abstracts were excluded, as well as studies on tools to improve completion of colonoscopy. However, reference lists of potential relevant systematic reviews and meta-analyses were checked for any missed papers.

Data extraction and analysis

For each study, the methods, way of assessment, and endpoints were recorded according to a predefined protocol. Two reviewers extracted all data. The quality of the studies was appraised and the reviewers assigned a level of evidence to each study using a tool developed by the Oxford Centre for Evidence-based Medicine (CEBM) as shown in Table 1.[2] A grade of

Table 1. Levels of evidence and grades of recommendation (CEBM).

Level of evidence	
1a	Systematic reviews (meta-analysis) containing at least some trials of Level Ib evidence, in which results of separate, independently conducted trials are consistent
1b	RCT of good quality and of adequate sample size (power calculation), cohort study with good follow up
2a	RCT of reasonable quality and/or if inadequate sample size
2b	Nonrandomized trials, comparative research parallel cohort
3	Nonrandomized, noncomparative trials, descriptive research
4/5	Case series, expert opinions, including the opinion of work group members
Grades of recommendation	
A	Consistent level 1 studies
B	Consistent level 2 or 3 studies or extrapolations form level 1 studies
C	Level 4 studies or extrapolations from level 2 or 3 studies
D	Level 5 evidence or troubingly inconsistent or inconclusive studies of any level

recommendation was given for each subgroup of studies included in this systematic review using the same tool provided by the CEBM (Table 1). The validation method and type of each simulator study was designated according to the consensus guidelines for validation of virtual reality simulators as described by Carter et al.[3] Validation of simulators is in most cases performed by demonstrating different types of validity. *Validity* in itself is defined as the extent to which an assessment tool, in this case a simulator, measures what it is supposed to measure. One of the simplest forms of validity is *face validity*. This is demonstrated by questioning a defined group of subjects, to judge the simulator on realism between the simulator and the real activity. Usually, a group of experts is questioned. This is why the term *expert validity* is also used. *Construct validity* describes the extent to which the simulator can distinguish between different levels of expertise. The most used method of establishing *construct validity* is that the simulator can distinguish beginners from more experienced endoscopists and experts by the simulators performance parameters. *Reliability* of the simulator relates to the power of the simulator to provide consistent results. The most commonly used test is the test-retest reproducibility. It predicts to what extent a subject can 'beat the test' by repeated assessment. The most powerful evidence of validity is *concurrent validity*. This refers to the level of which performance on the simulator correlate to the real activity, in this case patient-based endoscopy.

Since we aimed to provide a complete overview of the available literature on training and assessment in GI endoscopy, the included studies were fairly heterogeneous. Therefore, it was judged that statistical pooling of the data was not suitable.

RESULTS

Inclusion

Figure 1 shows the flowchart of the selection process of the included studies. Ninety-four studies investigating simulators, assessment methods, learning curves and training programs for gastrointestinal endoscopy were included in this review. These are summarised in Table 2. In order to provide a systematic overview of these studies, they were divided into different categories (simulator training, learning curves, numbers needed to gain competence, assessment of performance and evaluation of (patient-based) training models). In the more detailed discussion, we will focus on studies providing level 1 and 2 evidence.

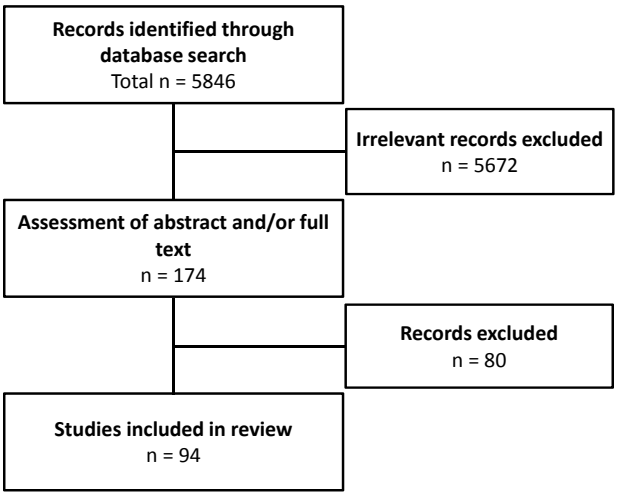


Figure 1. Flow diagram of the studies included.

Table 2. Overview of included studies per subject.

Subject	Number of studies included
Simulator validation	27
Colonoscopy	11
Flexible sigmoidoscopy	2
EGD	1
ERCP	7
EUS	2
Basic flexible endoscopy in general	3
Dexterity exercises	1
Simulator-based training and learning curves	23
Colonoscopy	14
Flexible sigmoidoscopy	3
EGD	3
ERCP	1

Table 2. Overview of included studies per subject. (continued)

Subject	Number of studies included
EUS	1
Training upper GI bleeds	1
Simulator competence assessment studies	4
Colonoscopy	2
Flexible sigmoidoscopy	1
EGD and flexible sigmoidoscopy	1
Learning curves and threshold numbers	24
Colonoscopy	13
EGD and colonoscopy	2
Flexible sigmoidoscopy	1
ERCP	6
EUS	2
Assessment tools	9
Colonoscopy	6
Flexible sigmoidoscopy	1
Colonoscopy and sigmoidoscopy	1
EGD and colonoscopy	1
Training models	7
Colonoscopy	4
Flexible sigmoidoscopy	1
EGD	2

Simulator validation studies

Twenty-seven simulator validation studies were retrieved.[4-29] We included eleven studies on colonoscopy, two on flexible sigmoidoscopy, three studies on basic flexible endoscopy in general, one study on EGD, seven studies on ERCP, two on EUS, and one study on dexterity exercises in forward viewing endoscopy. All studies are shown in Table 3. Besides the ERCP and EUS studies, all five other categories of studies focused on conventional, forward viewing flexible endoscopy with a large overlap in outcome parameters. Procedures like ERCP and EUS show profound differences compared to basic forward viewing flexible endoscopy, not only because of combination with radiological or ultrasonographical imaging, but also because of a complete different perception by the endoscopist in side viewing endoscopy. We have therefore analysed them separately from the larger group that we refer to as forward viewing flexible endoscopy procedures. Eight validation studies on flexible endoscopy tasks were performed using the Simbionix GI Mentor VR computer simulator.[8-10,12,14,15,22,25] Two studies reported on face validity. The largest study included 35 experts and demonstrated good face validity for colonoscopy.[15] A smaller study reported low level of realism as judged by six experts on all modules of the simulator.[14] All studies reported consistent results and good construct for performance metrics on procedure times of the GI Mentor. These procedural times varied from time to cecal intubation, time spent with clear view and time spent with

Table 3. Studies on validation of simulators for all gastrointestinal endoscopy procedures.

Study	Setting and participants	Simulator	Procedure	Methods	Validation method	Results	LOE
Bhutani (1998)	Feasibility study for EUS	Live porcine model	EUS	Comparing EUS images to autopsy findings	Feasibility only, no validation	Visualization of most anatomical landmarks in a swine model seem comparable to human anatomy	4
Bhutani (2000)	Feasiblity study for interventional EUS	Live porcine model	EUS	Creation, location and puncture of pseudo-lesions	Feasibility only, no validation	Demonstrated feasibility to perform FNA of a pseudo-lymph node, submucosal lesion, pancreatic tissue and celiac neurolysis	4
Neumann (2000)	Feasibility study for ERCP	Erlangen Endo-Trainer	ERCP	ERCP cannulation, sphincterotomy and stent placement	Feasibility only, no validation	Demonstrated feasibility to perform ERCP with deep cannulation, sphincterotomy and stent placement	4
Datta (2002)	15 Novices, 15 intermediate and 15 experts	AccuTouch	Sigmoidoscopy	3 cases performed by all after a familiarization tour. Recorded parameters were mucosa visualized, time taken, scope pathlength, time in red-out and efficiency ratio.	Construct validity	Significant difference of mucosa visualized and efficiency ratio between all groups, difference in time and scope pathlength for novices compared to experienced	2b
Ferlitsch (2002)	13 Novices, 11 experts	GI Mentor	Flexible endoscopy	VR simulator tasks on the GI Mentor included EndoBubble, gastroscopy and colonoscopy. Additionally, 7 novices performed unsupervised training: 2 hours, 5 days a week for 3 weeks	Construct validity and early learning curve	Baseline assessment revealed significant differences in all tasks in favor of experts proving construct validity. After 3 weeks of training, differences were no longer seen whereas these differences persisted for the no-training group	2b
MacDonald (2003)	10 Novices, 19 experienced and 5 experts	AccuTouch	Sigmoidoscopy	3 VR sigmoidoscopies. Performance is measured by the simulator	Construct validity	Significant differences on all parameters between the untrained and trained endoscopists. Experts perform better than senior residents but with no significant differences.	2b

Table 3. Studies on validation of simulators for all gastrointestinal endoscopy procedures. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Validation method	Results	LOE
Mahmood (2003)	11 Novices (<10 procedures), 7 intermediates (11-100 procs), 7 experienced endoscopists (>100 procs)	AccuTouch	Colonoscopy	2 VR colonoscopies on the simulator. Performance is measured by the simulator	Construct validity	Significantly improved performance with increased experience on all parameters measured except path-length	2b
Ritter (2003)	9 Novices and 5 intermediate experienced endoscopists (100-250 procs)	GI Mentor	Dexterity	3 VR tasks EndoBubble level 2 on the GI Mentor II simulator. Performance parameters were recorded for each attempt and compared to previous attempts.	Construct validity and reliability	Intermediates performed better in each trial without improvement of performance. Novices improved significantly in completion time. Consistency and reliability were greater than 0.8 for both groups on all measures except completion time in the novice group	2b
Sedlack (2003)	6 Novices, 6 trainees (88-202 colonoscopies) and 10 experts	AccuTouch	Colonoscopy	2 VR colonoscopies on the simulator. Performance is measured by the simulator. A survey on realism by experts	Face and construct validity	Favourable degree of realism as judged by experts but substantially easier. Only procedure time, insertion time and time in red-out significantly different between groups	2b
Sedlack (2003)	10 Trainees and 10 trainers	Live porcine model, Erlangen Endo-Trainer, GI Mentor	ERCP	Procedures in live porcine models, the Erlangen Endo-Trainer and the GI Mentor. Survey on realism and performance. Ranking of results	Face validity	Overall results for realism and educational value highest for the Erlangen model and lowest for the GI Mentor. GI Mentor was easiest incorporated in a training program	2b
Moorthy (2004)	11 Novices, 11 intermediates (10-50 procs), 10 experienced (>200 procs)	GI Mentor	EGD	2x2 VR tasks by all participants. Performance measured by simulator and video assessment by blinded experts using standardized assessment form	Construct validity and reliability	Significant differences on all simulator output parameters between the 3 groups. Significant differences in global scores between groups, only percentage of mucosa visualized correlated with global scores	2b

Table 3. Studies on validation of simulators for all gastrointestinal endoscopy procedures. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Validation method	Results	LOE
Felsher (2005)	39 Novices and 36 experienced endoscopists	GI Mentor	Colonoscopy	Colonoscopy task on the GI Mentor II including insertion, withdrawal and polypectomy	Construct validity	Experienced ednoscopists performed the tasks more efficient, visualized more mucosa, achieved greater polypectomy rates and spent more time in clear view of the lumen than novices ($p<0.05$)	2b
Grantcharov (2005)	10 Novices, 10 intermediate and 8 experienced endoscopists	GI Mentor	Colonoscopy	Single VR colonoscopy on the GI Mentor II simulator	Construct validity	Significant differences in performance by experts versus intermediates and novices on all recorded parameters. Differences between intermediates and novices not significant	2b
Leung (2007)	6 Unexperienced (<100 ERCPs), 9 experienced (>100 ERCPs), 8 experts	ERCP Mechanical Simulator - EMS	ERCP	Performance times are measured for ERCP total procedure times and simulated fluoroscopy times. Participants completed a survey on realism	Face and construct validity	Significant shorter procedure times for more experienced endoscopists on all aspects. Very high credibility scores	2b
Sedlack (2007)	13 Novices, 13 trainees (100-150 colonoscopies) and 13 experts	Ex vivo bovine colon model	Colonoscopy	Single colonoscopy on a bovine ex vivo colon model. Standardized performance measures were maximum insertion depth and time, scope length, withdrawal time, % of mucosa visualized and quality of mucosal exam. Performance was compared to known patient-based performance for each endoscopist	Face, construct and concurrent validity	Favourable degree of realism as judged by experts, significantly greater performance compared to less experienced groups. High correlation to patient-based cecal intubation times ($r=0.764$)	2b
Frimberger (2008)	26 Trainees and 7 trainers	X-Vision ERCP Training System	ERCP	ERCP cannulation, sphincterotomy and stent placement	Face validity	Favourable results in all categories assessed	4

Table 3. Studies on validation of simulators for all gastrointestinal endoscopy procedures. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Validation method	Results	LOE
Koch (2008)	26 Novices, 23 experts	Endo TS-1	Colonoscopy	VR tasks included a hand-eye coordination game and a colonoscopy. Experts completed a survey on realism	Face and construct validity	Experts completed tasks significantly faster and rated face validity as good.	2b
Koch (2008)	35 Novices, 15 intermediates, 20 experienced and 35 experts	GI Mentor	Colonoscopy	VR tasks included a hand-eye coordination game and 2 colonoscopies. Experts completed a survey on realism	Face and construct validity	Novices performed significantly worse than the other groups. The experienced groups performed equal. Expert validity was rated as good	2b
Haycock (2009)	10 Novices, 13 intermediates and 11 experts	Endo TS-1	Colonoscopy	3 VR colonoscopy cases on the Endo TS-1 with increasing difficulty	Face and construct validity	Experts performed significantly better compared to intermediates and novices. Face validity was rated as good	2b
von Delius et al. (2009)	6 Advanced, 10 intermediate, 12 novice endoscopists	X-Vision ERCP Training System	ERCP		Construct validity, Expert validity, didactic value	Significant differences for construct validity between experts, experienced and novice endoscopists	2b
Bittner (2010)	6 Novices and 6 experts	GI Mentor	ERCP	2 ERCP tasks performed on the simulator	Face and construct validity	Significant differences between novices and experts performing simulated ERCP cases demonstrating both face and construct validity	2b
Fayez (2010)	12 Novices (<5 procedures), 8 experienced (>50 procedures)	GI Mentor	Colonoscopy	3 VR colonoscopy cases on the GI Mentor II wit increasing difficulty	Construct validity	Differences in performance between novices and experienced on the simulator are greater in the more complex scenarios. During easy tasks only time to cecum differed. In difficult tasks ther difference in experience resulted in significantly better performance for the experienced group	2b

Table 3. Studies on validation of simulators for all gastrointestinal endoscopy procedures. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Validation method	Results	LOE
Kim (2010)	5 Novices, 6 experts	GI Mentor	Flexible endoscopy	Novices perform 50 supervised cases of which 6 predetermined cases are used for evaluation, experts perform 6 cases. Evaluation form on realism by experts	Face and construct validity	No significant overall differences between 2 groups on performance parameters. Only differences on time parameters of the procedures. No realistic representation	2b
Leung (2010	22 Experienced endoscopists	EMS and ex vivo porcine model	ERCP	Same ERCP tasks; scope insertion, selective bile duct cannulation and biliary stent placement, were performed on the EMS and the PSM. Survey on credibility was performed before and after traning	Face validity	Before hands-on practice, both EMS and PSM received high scores. After practice, there was a significantly greater increase in confidence score for EMS than PSM ($p<0.003$). Participants found EMS more useful for training ($p=0.017$)	2b
McConnell (2012)	5 Novices and 6 experts	AccuTouch	Flexible endoscopy	18 VR colonoscopies and 6 gastroscopies performed by all participants. Survey on realism by all	Face and construct validity	No significant overall differences between 2 groups on 38/57 performance parameters. Only differences in 19/57 parameters of which 8/19 on time parameters of the procedures. No realistic representation as rated by experts	2b
Plooy (2012)	18 Novices and 21 experienced endoscopists	Kyoto Kagaku Colonoscope Training Model	Colonoscopy	2x4 standard cases on the simulator by all participants. Performance measures included completion rate, time to cecum and peak force	Construct validity	Experts had significantly higher completion rates and shorter times to cecum for all cases. Peak force was significantly lower in 2 cases	2b
Sugden (2012)	30 Novices, 10 intermediates (100-500 colonoscopies) and 10 experienced (>500 procs)	Endo TS-1	Colonoscopy	VR tasks included 3 abstract tasks, 9 part-procedural and 1 whole procedure task. Each set was performed 2 times by intermediates and experienced and 10 times by novices. Benchmarks were established and learning curves for novices to reach these	Construct validity	Construct validity was demonstrated for 3 abstract, 4 part-procedural and the whole procedure task. Significant differences were demonstrated between novices and both experienced groups but not between intermediates and experienced. Learning curves plateaued consistently at the 9th or more attempt	2b

endoscope loops. Although these types of parameters, measuring a time aspect, are usually considered surrogate markers for competence, it seems to be the most consistent and therefore the most reliable parameter to distinguish between competence levels. There is a fairly large heterogeneity on other outcome parameters. Five studies reported on the AccuTouch Immersion Medical computer simulator.[4,19-21,27] Two studies reported on face validity with conflicting results. Again, as for the GI Mentor, realism was judged as valid by experts for the colonoscopy module but not for the complete set of modules on the simulator as a whole. The AccuTouch simulator seemed to have the same construct validity profile as the GI Mentor. That is, construct validity was consistently reported as good for performance measures related to procedural times in all published studies. Three validation studies reported on the Olympus Endo TS-1 VR computer simulator for colonoscopy.[13,16,28] Face validity was rated as good by two studies and all three demonstrated good construct validity on all studied procedures. One study reported good construct validity of the Kyoto Kagaku Colonoscope Training Model. Face validity was not studied.[24] The last study demonstrated good face, construct and concurrent validity in a bovine explant colon model.[26]

Six validation studies were performed on ERCP. Two studies were feasibility studies and no formal validation was done. These two were both in mechanical models.[11,23] Only one validation study was performed using a VR computer based simulator. This study demonstrated both face and construct validity for the ERCP modules in the Simbionix GI Mentor II simulator.[7] A similar study was done for the X-Vision ERCP Training System, a mechanical simulator, showing both face and construct validity.[29] Two studies on the same mechanical ERCP training simulator were performed by the same research group.[17,18] The ERCP Mechanical Simulator (EMS) demonstrated a good construct validity and excellent face validity. In a direct comparison to an ex vivo porcine stomach model, the EMS was rated more realistic and useful. Another study compared live porcine models versus the Erlangen Endo-Trainer versus the Simbionix GI Mentor VR simulator for ERCP.[30] The Erlangen model scored highest on realism and educational value. The GI Mentor scored lowest. However, it was felt that the GI Mentor was more easily incorporated in a training program. Although the validation studies for ERCP simulation comprised a fairly heterogeneous group of simulators, the strongest evidence was provided for the mechanical simulators. For EUS, only two studies by the same author reported on feasibility to perform EUS and FNA in a porcine model.[5,6] No attempt at validation has been published to date.

Simulator training and learning curve studies

Twenty-three studies reported on simulator training and learning curves.[31-53] Twenty studies reported on forward viewing flexible endoscopy (3 EGD, 3 sigmoidoscopy, 14 colonoscopy), one study reported on EUS, one on ERCP and one on training haemostasis in upper GI bleeds. The studies are shown in Table 4. Eleven studies were performed using the AccuTouch Immersion Medical VR computer simulator for training, ten with a level 2 evidence and one study with a level 1 evidence.[31,33,34,40,44,45,47,49-52] All studies on flexible sigmoidoscopy and colonoscopy had a randomized design and compared simulator-based training groups versus controls. Acquired competence was evaluated using the same simulator and in six studies also during patient-based assessment. The most consistent outcome parameters demonstrating improved performance were on procedural times, cecal intubation rates, and times in red-out, meaning that luminal view was lost. Patient comfort scores were measured in two

Table 4. Simulator-based training and learning curve studies.

Study	Setting and participants	Simulator	Procedure	Methods	Assessment method	Results	LOE
Tuggy (1998)	Single center, randomized, 5 novice endoscopists in the simulator group, 5 controls	AccuTouch	Sigmoidoscopy	5 And 10 hours VR sigmoidoscopy training on the simulator versus no training	Improvement of performance during patient based sigmoidoscopy	Significantly shorter procedure times, no difference in times in red-out or completeness of mucosal inspection	2b
Gerson (2003)	Single center, randomized, 9 novices in simulator-trained group, 7 in control group	AccuTouch	Sigmoidoscopy	Unsupervised training on the simulator versus patient-based training	5 Patient-based sigmoidoscopies per trainee evaluated by supervisor and patients	Patient-based trained group completed significantly more cases, completed more cases independently, and performed more successful retroflexion. There was no difference in procedure times and patients satisfaction or discomfort	2a
Eversbusch (2004)	Single center, randomized, 2 groups; group I: 8 experienced endoscopists (>200 procedures), 10 intermediate (<50 procedures), 10 novices. Group II: 10 novices psychomotor training versus 10 novices control.	GI Mentor	Dexterity and colonoscopy	Group I: learning curve psychomotor training using EndoBubble task 10 times. Group II: VR colonoscopy followed by psychomotor training versus no training and again VR colonoscopy	Improved performance during VR colonoscopy	No learning curve during psychomotor training for experts, short learning curve for intermediate experienced and a longer learning curve for novices. Statistically significant improved performance during VR colonoscopy on all parameters after psychomotor training versus control.	2b
Giulio (2004)	Multicenter, randomized, 22 novices	GI Mentor	EGD	10 hours unsupervised training on the simulator versus no training	20 Patient-based gastroscopies per trainee evaluated by non-blinded supervisors	Significantly more complete examinations, less need for assistance and more positive scores by supervisors in the simulator trained group	2a

Table 4. Simulator-based training and learning curve studies. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Assessment method	Results	LOE
Mahmood (2004)	Single center, 26 Novices performing VR colonoscopies on the AccuTouch Immersion Medical simulator	AccuTouch	Colonoscopy	38 trials of 5 consecutive VR colonoscopies. Performance measures are recorded by the simulator	Improvement of performance measures on the simulator in the absence of feedback	In the absence of feedback there was no learning effect measured on the simulator after repetitive training	2b
Sedlack (2004)	Single center, randomized, 4 novices to VR colonoscopy training versus 4 controls	AccuTouch	Colonoscopy	VR colonoscopy training for 6 hours	Patient-based assessment of colonoscopy performance using standardized assessment forms by experts and patient survey on discomfort	Study group outperformed controls on all parameters except insertion time during first 15 procedures and on 3 parameters (depth of insertion, independent completion and ability to identify landmarks) after 30 colonoscopies ($p<0.05$)	2a
Sedlack (2004)	Single center, randomized, 19 novices versus 19 controls	AccuTouch	Sigmoidos-copy	19 Novices receiving 3 hours of simulator-based training prior to patient-based sigmoidoscopy versus 19 controls	1 Week of patient-based sigmoidoscopy. Assessment of performance using a standardized assessment form by supervisors and patients survey on discomfort	Median patient discomfort scores were significantly lower in the study group ($p<0.01$). No differences in procedural performance as judged by supervisors	2a
Ahlberg (2005)	Single center, randomized, 12 novice colonoscopists with only gastroscopy experience	GI Mentor	Colonoscopy	VR colonoscopy training to a predefined expert level on the simulator versus no training	Patient-based colonoscopy	Significantly improved performance in the simulator-trained group ($p=0.0011$) and 4.53 times more likely to reach the cecum	2a

Table 4. Simulator-based training and learning curve studies. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Assessment method	Results	LOE
Cohen (2006)	Multicenter, randomized, 45 novices	GI Mentor	Colonoscopy	Randomized to no simulator training versus 10 hours of unsupervised simulator training	Assessment of technical and cognitive performance as well as patient discomfort during the first 200 consecutive patient-based colonoscopies	Simulator group had significantly higher objective competency rates during the first 80 cases. The overall number needed to reach a 90% competency level was 160 cases in both groups	1b
Barthet (2007)	17 novice trainees performing EUS	Live porcine model	EUS	Visualizing anatomical landmarks, perform FNA and celiac neurolysis	Pre- and post-test assessment of performance	Increased performance after training in both diagnostic (visualization) and therapeutic procedures (time to complete and precision, not the rate of technical errors)	3
Buzink (2007)	Single center, 35 novices and 5 expert endoscopists	GI Mentor	Dexterity and colonoscopy	All participants performing a predefined protocol colonoscopy and hand-eye coordination tasks	Assessment by simulator construct repeating predefined tasks on the simulator itself	Novices significantly improve performance in the early learning curve by training on the GI Mentor II, whereas experts do not	2b
Maiss (2007)	27 Novices randomized to intensive simulator training versus controls	Compac-tEASIE Simulator	Haemostasis	14 Novices are trained 12 times in a 7 months period in endoscopic skills, injection therapy, clipping and band ligation	Final assessment by blinded experts	Study group performed significantly better on all tasks as judged by blinded experts. Results were comparable to similar studies performed in New York and France.	2a
Park (2007)	Single center, randomized, 12 novices receiving simulator-based training versus 12 controls	AccuTouch	Colonoscopy	All participants performed baseline VR colonoscopy. SG received 2-3 hours of training on the simulator	Patient-based assessment by blinded experts using GRS. Scores were compared to simulator performance metrics	SG had a significantly higher GRS. Only 2/8 simulator performance metrics correlated with GRS (procedure time and time in red-out)	2a

Table 4. Simulator-based training and learning curve studies. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Assessment method	Results	LOE
Ferlitsch (2010)	Single center, randomized, 28 novices, 14 received simulator training prior to patient-based gastroscopy training	GI Mentor	EGD	5 to 20 hours of VR simulator training in upper GI endoscopy and training games with the first 2 hours supervised	Patient-based gastroscopy with standardized assessment of performance by the unblinded supervising expert on time taken, technical and diagnostic accuracy and blinded patients evaluation on discomfort and pain.	Time and technical accuracy was significantly better in the simulator-trained group ($p<0.05$) in the early learning curve, after 60 procedures time was still significantly different. Diagnostic accuracy and patient discomfort or pain was not different in any point of time.	2a
Haycock (2010)	Multicenter, randomized, blinded, 36 novice colonoscopists	Endo TS-1	Colonoscopy	16 hours VR colonoscopy training versus patient-based training	Patient-based and simulator-based assessment	No differences in performance between subjects and controls during patient-based assessment. Subjects performed significantly better in simulator-based assessment.	1b
Kruglikova (2010)	Single center, randomized, 22 novices	Accutouch	Colonoscopy	Simulator training with concurrent feedback versus no feedback	Post-training simulator-based assessment	Subjects reached proficiency levels significantly faster than controls. This effect diminished during the retention/ transfer tests	2a
Kruglikova (2010)	Single center, 30 female novices, 10 nurses, 10 endoscopy nurses, 10 residents	AccuTouch	Colonoscopy	10 repetitions of same VR task on the simulator	Comparing first and last colonoscopy task to each group and expert reference	Residents performed the tasks faster with a similar learning curve pattern, other parameters were equal.	2b
Snyder (2010)	Single center, randomized, 8 proctored versus 5 unproctored trainees	AccuTouch	Colonoscopy	Both groups performed VR colonoscopy assessment: baseline directly after training & after 4.5 months	A combined proficiency score of 10 (best performance on all parameters) was the maximum score and considered proficient	All participants improved to proficient level post-training. Retention scores after 4.5 months were unchanged for all participants	2a

Table 4. Simulator-based training and learning curve studies. (continued)

Study	Setting and participants	Simulator	Procedure	Methods	Assessment method	Results	LOE
Kaltenbach (2011)	Single center, pilot study, 3 trainees	Mechanical colon model	Colonoscopy	Half-day one-on-one training using a colon model with ScopeGuide imaging creating different loops during the trainees colonoscopy learning curves	Colonoscopy performance was appraised (non-blinded) before and after the training	Performance scores and CIRs improved significantly	4
Lim (2011)	Multicenter, randomized, 16 trainees in ERCP	ERCP Mechanical Simulator - EMS	ERCP	2 Sessions of hands-on training in selective bile duct cannulation on the EMS for the study group	Patient-based ERCP with cannulation success rates, cannulation times and blinded supervisor assessment	Significant higher cannulation success rates in less time in the study group. Same competency scores for both groups	1b
Van Sickle (2011)	Multicenter, prospective, 41 novice endoscopists	GI Mentor	Dexterity	Novice endoscopists training EndoBubble on the simulator	Pre- and posttraining assessment using Colonoscopy case module I outcome parameters and GAGES assessment form	Performance on the simulator, as measured by the simulator parameters, improved significantly after performing EndoBubble tasks ranging 13 +/-10 for EndoBubble level 1 and 23 +/- 16 for level 2	2b
Ende (2012)	Multicenter, randomized, 28 novices, EGD	Various	EGD	3 groups; clinical+simulator training, clinical training only, simulator training only. Different simulators used.	Assessment by simulator and blinded experts using standardized forms	Clinical trained groups outperformed the simulator trained group significantly. Results in the combination group were better than clinical training alone	1b
Ahad (2013)	Single center, randomized, 32 novice endoscopists	AccuTouch	Colonoscopy	Low versus high-fidelity model simulator training	Pre- and posttraining assessment using colonoscopy tasks on the Immersion AccuTouch simulator	Colonoscopy skills acquisition of basic endoscopy skills on a low-fidelity model is just as effective as on a high-fidelity simulator	1b

studies.[40,50] One study favoured simulator training versus no simulator training prior to starting patient endoscopies, the second study showed no difference between groups.

Six studies were carried out using the Simbionix GI Mentor VR simulator for training and learning curves.[32,34,36-38,41] Five studies provided level 2 evidence, one study level 1 evidence. Four studies were on colonoscopy tasks, two on EGD. All studies demonstrated that simulator training improved performance of novices. There were no learning effects for experienced endoscopists. Due to the heterogeneity of these studies, improved performance could not be expressed in terms of exact numbers. Performance was assessed by means of the simulator construct in three studies. Two studies used patient-based assessment for evaluation of the simulator-based learning effect. The competence parameters that consistently improved significantly were; (I) procedure time, (II) cecal intubation rate (CIR); a direct comparison of simulator-training versus controls showed a 4.5-fold increased cecal intubation rate in the simulator-training group in the early learning curve[34], (III) time with clear view, (IV) time of endoscope looping, and (V) objective performance scores, as judged by expert supervisors during patient-based endoscopy assessment. Improved performance in the simulator-trained groups versus controls was observed in up to 60 patient-based assessed EGDs and 80 procedures in colonoscopy training. Only one study used the Olympus Endo TS-1 colonoscopy simulator for training.[42] This multicenter, randomized study compared simulator-based training versus patient-based training. Blinded experts assessed performance during patient-based endoscopy. Both groups showed equal performance. (LOE 1b) One multicenter, randomized study was performed using all kinds of simulators.[53] The study showed that patient-based training with complementary simulator training was superior to patient- or simulator-based training alone. (LOE 1b) One study was done on ERCP.[46] This study had a multicenter, randomized design. It demonstrated significantly higher cannulation success rates in less time in the study group after training on the ERCP Mechanical Simulator. (LOE 1b) One study was performed evaluating the CompactEASIE simulator, a mechanical simulator with an ex vivo porcine stomach.[48] Significant improvement in skills in endoscopic haemostatic therapy was demonstrated with a sufficient level of evidence. No previous formal validation of the model was carried out. Only one study was performed on the subject of learning diagnostic and therapeutic EUS.[35] Only a description of improved performance on live porcine models before and after a hands-on training course was provided. No formal statistical calculation was carried out. The model had not been previously validated.

Simulator competence assessment studies

Four studies reported on the use of simulators for assessment of competence.[54-57] Two studies focussed on colonoscopy, one on sigmoidoscopy and one on both EGD and sigmoidoscopy. The studies are summarized in Table 5. Only two studies reached a 2b LOE.[55,57] In both studies performance parameters derived from the simulators did not correlate to performance scores given by blinded experts. It seems that current simulators lack the discriminative power to assess performance and determine competence levels in patient-based endoscopy.

Learning curves

Fifteen studies reported on learning curves for colonoscopy (n=8), ERCP (n=5) and EUS (n=2).[58-72] These are shown in Table 6 (A,B,C, respectively).

For colonoscopy, four studies reached a sufficient evidence level 1 or 2.[59,61,62,65] These studies had a prospective design and evaluated 8 to 41 trainees with procedure numbers varying

Table 5. Simulator competence assessment studies.

Study	Setting and participants	Simulator	Procedure	Assessment method	Results	LOE
Moorthy (2004)	Single center, 7 novices, 7 intermediates (20-80 procs), 6 experts (>200 procs) performing VR sigmoidoscopy on the AccuTouch Immersion Medical simulator	AccuTouch	Sigmoidoscopy	Dual video assessment by blinded experts using standardized assessment forms and simulator performance parameters	Significant discrimination of skills levels by the global scores. Only "time in red-out" correlated with global scores and not for time, depth of insertion and mucosa visualized	2b
Phitayakorn (2009)	23 Experts perform a single colonoscopy task on the GI mentor II	GI Mentor	Colonoscopy	Performance parameters results with their range are demonstrated as an attempt to set a benchmark	Considerable variety in performance on the simulator among experts	4
Sarker (2010)	37 Trainees (<100 procs) and 18 consultants (>500 procs) performing both patient-based and simulator-based EGD and sigmoidoscopy, GI Mentor	GI Mentor	EGD and sigmoidoscopy	Assessment of patient-based endoscopy by blinded experts using video recordings, simulator-based assessment only by composed outcome measures of the simulator itself: mucosal score and time score	Construct validity was good for the patient-based assessment (p=0.000-0.002) but not for the simulator based outcome parameters (p=0.263-0.701)	2b
Elvevi (2012)	Single center, 12 novices in colonoscopy, GI Mentor	GI Mentor	Colonoscopy	Competency level assessment at the start of training and after 60 patient-based colonoscopies using 2 GI Mentor colonoscopy modules	Only time to cecal intubation improved significantly during simulator assessment. The simulator is considered not useful for competency assessment	3b

from 2887 to 4351. However, outcome measures and use of competence standards were fairly heterogeneous. The studies reported on cecal intubation rate (CIR) or completion rate, time to cecum, or a combination of those outcomes. One group described the learning curve by means of scoring different aspects of the procedure on a newly developed assessment tool (Mayo Colonoscopy Skills Assessment Tool), but also described learning curves for outcomes such as CIR.[65] The number of colonoscopies that trainees needed to perform in order to achieve a CIR of > 85-90% varied from 150 to 280 procedures.

From the five studies focusing on ERCP, only two reached a sufficient evidence level.[60,71] These described a prospective evaluation of respectively 17 and 20 trainees, with the following outcome measures: subjective score regarding performance (overall and per part of the procedure) on a 6-point scale where a score of 1,2 or 3 was considered competent, and success of selective cannulation of the common bile duct (CBD) or pancreatic duct (PD). One study concluded that an overall sufficient score was reached after 137 (probability of success = 0.8) or 185 ERCPs (probability of success = 0.9).[60] A different group reported that an 85% selective cannulation rate was reached after 70 procedures for the PD and after >100 ERCPs for the CBD.[71]

The two studies on EUS described the performance per anatomic station of the procedure.[70,72] There was a large variability in achieving overall competence, with acceptable performance after a range of 255 to >400 EUS procedures.[70] One study did not report on overall competence, but stated that 78 procedures were necessary for competence in duodenal examination.[72]

Threshold numbers needed to gain competence

Nine studies reported numbers needed to gain competence in different procedures in gastrointestinal endoscopy.[72-81] These studies are shown in Table 6 as well. Two studies handled both EGD and colonoscopy[73,80], whereas most of the studies pertained to colonoscopy alone.[74,75,77-79] There were two single studies on sigmoidoscopy and ERCP.[76,81] The level of evidence was moderate for most studies due to the designs and numbers of procedures evaluated. Only three groups performed studies (regarding EGD and colonoscopy) with a prospective design and a considerable amount of trainees evaluated, resulting in LOE 2.[73,74,80] These will be discussed in further detail.

For EGD, competence was measured in two ways: intubation of the oesophagus and reaching a sufficient score on the Global Assessment of Gastrointestinal Endoscopic Skills (GAGES). One group demonstrated an 80% success rate of oesophageal intubation after 100 procedures, whereas another study concluded a plateau in the GAGES score after 50 procedures.[73,80] Concerning colonoscopy, competence was measured trough CIR and scores on the GAGES form as well. Two studies concluded that 100 colonoscopies was insufficient for reaching a >90% CIR[73,74], whereas the GAGES score displayed a plateau score at n=75 procedures.[80] All studies confirmed that performance of trainees increased with experience.

Assessment and grading of performance

Nine studies described the development and evaluation of assessment tools for colonoscopy (n=6), sigmoidoscopy (n=1), both (n=1) and both colonoscopy and EGD (n=1).[82-90] These are shown in Table 7. All evidence level 2 studies focused on colonoscopy, flexible sigmoidoscopy or both, had a prospective design, and reported on 18 to 162 participants.[82,86-89] The British Direct Observation of Procedural Skills (DOPS) appears effective for evaluation of competence

Table 6A. Learning curves and threshold numbers as a measure for EGD, sigmoidoscopy and colonoscopy.

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Sedlack (2011)	Single center, prospective, trainees at different levels of experience	41	Learning curves through assessment (MCSAT) with scores from 1-4: competency when score >3.5. CIR & time to cecum.	4103 of 6635 colonoscopies (62%) assessed (mean n per trainee: 399; 95% CI 365-433). Competency after mean 275 procedures. CIR of 85% after 250-275 colonoscopies, time to cecum <16 min after 275.	1b
Koch et al (2012)	Single center, prospective, trainees at different levels of experience	19	Learning curves through self-assessment (Rotterdam Assessment Form for Colonoscopy). CIR and time to cecum.	2887 colonoscopies (152 per trainee, range 91-347). Baseline CIR: 65% improved to 78% after 100 and 85% after 200 colonoscopies ($p<0.001$). After 280 ≥90% CIR. Time to cecum decreased from baseline 13:10 min to 8:30 min after 200.	1b
Chung et al (2009)	Single center, prospective, GI fellows with no colonoscopy experience	12	Learning curves per 50 procedures for adjusted completion rate (>90%) and time to cecum (<20 min).	3243 colonoscopies (n per trainee not mentioned). First 50 completion was 37% and improved to 94% after 250. Mean time improved from 12.9 min to 8.0 min. After 200 procedures, >90% completion rate within 20 min was reached.	2b
Lee et al (2008)	Multicenter, prospective, first-year GI fellows with no colonoscopy experience	24	Outcome: adjusted CIR (>90%) and time to cecum (<20 min). Learning curves per 50 colonoscopies.	4351 colonoscopies (n per trainee not mentioned). CIR per 50 consecutive procedures: 71.5% - 82.6% - 91.3% - 94.4% - 98.4% - 98.7% ($p<0.05$). Time to cecum decreased from 11.2 min to 6.6 min. Competency after 150 procedures.	2b
Tassios et al (1999)	Single center, prospective, trainees with no colonoscopy experience	8	Learning curves for completion rate. Outcome: completion of colonoscopy (CIR) per training year and per number of procedures performed.	N = 978 colonoscopies performed by trainees (median n 1st year: 43, range 31-69 and 2nd year 91 (79-143). Completion rates in %: first-year 33 (range 15-42); second-year 60 (range 55-79) and third-year 75 (range 58-94) ($p<0.001$). Per number: after 60 52.6 (95% CI 45.2-59.9); after 100 67.1 (95% CI 59.1-75.2) and after 180 76.8 (95% CI 66.1-87.6).	3

Table 6A. Learning curves and threshold numbers as a measure for EGD, sigmoidoscopy and colonoscopy. (continued)

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Marshall et al (1995)	Multicenter, prospective, first and second-year trainees with different levels of experience.*	9	Frequency of reaching cecum <30 min in last 7 months of 1st and 2nd year training.	N colonoscopies: 186 (n per trainee not mentioned) by first-year and 203 by second-year fellows. CIR 1st year: 53.7%, 2nd year 85.8%. Time to cecum 28, 19 and 9 min, respectively. Difference in performance by groups (p<0.001).	3
Selvasekar et al (2012)	Single center, prospective, fellows with different levels of experience in colonoscopy	6	Learning curves by completion and time to cecum. Cusum analysis. Competency defined as significant reduction in time and 80% CIR within 35 min.	1498 out of 2904 colonoscopies performed by fellows (mean n per fellow 249). Mean procedure time 30.2 +-15 min; decreased significantly during first 120 procs (p<0.001). 80% completion within 35 min after 114 colonoscopies.	3
Parry et al (1990)	Single center, retrospective, trainee with no experience in colonoscopy	1	Learning curve by cusum analysis for completion rate.	305 of 334 consecutive colonoscopies analyzed. First 100 procedures completion rate 67%, 101-200 was 88% and 201-305 91%. After 200 procedures, steady >90% completion.	4
Cass et al (1993)**	Single center, prospective cross-sectional design, surgical and GI trainees with different levels of endoscopy experience. EGD	12	Intubation esophagus	Median 113 UGI endoscopies per trainee (range 54-162). Esophagus intubation 90% after 50 procedures, but declined and reached 80% after 100 procedures.	2b
Cass et al (1993)**	Colonoscopy	12	Intubation cecum	Median 49 colonoscopies per trainee (range 39-127). CIR of 80% after 50 procedures but only 85% at 100 procs.	2b

Table 6A. Learning curves and threshold numbers as a measure for EGD, sigmoidoscopy and colonoscopy. (continued)

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Chak et al (1996)	Single center, prospective, GI fellows at various stages of endoscopy training, attendings	12 fellows, 17 attendings	Cecal intubation rate, time to cecum per training year, assistance needed. Adequacy of threshold of 100 colonoscopies	496 colonoscopies performed (79 by 5 first-years, 102 by 7 second-years, 315 by attendings). Second-years performed mean of 123 colonoscopies prior to study. First-years required assistance in 92% of procs vs 36% for second-years. CIR second-years vs attendings: 84 vs 94%, p<0.05). Time to cecum second-years vs attendings: 14.5 vs 10.5 min (p<0.01). 90% success rate within 15 min not reached after 100.	2b
Vassiliou et al (2010)**	Multicenter, prospective design, surgical and GI trainees with different levels of endoscopy experience. Single observations. UGI endoscopy	86	Score on GAGES for UGI, 3 different groups: <35 cases (1), 35-130 cases (2), >130 cases (3)	86 evaluations, group (1) n=35, (2) n=22, (3) n=29. Mean +-SD score: (1) 14.4 +- 3.7; (2) 17.8 +- 1.8; (3) 19.1 +- 1.1. Difference between groups significant (p<0.05) but not for (2) and (3). Plateau score at n=50.	2b
Vassiliou et al (2010)**	Colonoscopy	57	Scores on GAGES colonoscopy, 2 groups: cutoff <50 cases (1) and <140 cases (2); novices and experienced	57 evaluations, cutoff (1) 29 novices vs 28 experienced. Mean +- SD score: (1) novices 11.8 +-3.8 vs. experienced 18.8 +- 1.3 (p<0.001). Cutoff (2) 32 vs. 25. Novices 12.4 +- 4.2 vs. experienced 18.8 +- 1.3. No difference between cutoff groups. Plateau score at n=75.	2b
Spier et al (2010)	Single center, surgical trainees who participated in 2-month endoscopy rotation, survey	21	Numbers performed, CIR, perception of colonoscopy training	100% response rate, 15 2nd year residents, 6 4th year. Mean 80 +- 35 colonoscopies during rotation. Average CIR 47% (range 9-78%). Adequacy of training: 67% of 2nd year felt it was adequate vs 100% 4th years.	3

Table 6A. Learning curves and threshold numbers as a measure for EGD, sigmoidoscopy and colonoscopy. (continued)

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Spier et al (2010)	Single center, retrospective, GI fellows at various stages of endoscopy training	11	Total colonoscopy time, time to cecal intubation, independent completion rates. Adequacy of threshold of 140 colonoscopies	770 colonoscopies performed (369 by 9 first-years, 158 by 4 second-years, 243 by 5 third-years). Improvements from first- to third-year: mean colonoscopy time from 48 to 33 min, time to cecal intubation 19 to 11 min, completion rate 63 to 92% (all p<0.001). No independent completion of >90% after 140 colonoscopies.	3
Leyden et al (2011)	Single center, retrospective, GI and surgical trainees with >2 years endoscopic experience	13	Completion rates, polyp detection rates, withdrawal time in subset of patients, comparison of GI and surgical trainees	1998 and 1081 colonoscopies performed by GI and surgical trainees, respectively. Crude completion rate for GI vs surgical: 84 vs 78% (p<0.0001). PDR was 21 vs. 14% (p<0.001); withdrawal time 5 vs 2.5 min (p=0.003).	3
Church et al (2002)	Single center, prospective, trainees with different endoscopy experience	18	Performance during first 125 colonoscopies, completion rate defined as cecum reached by trainee as a percentage of completion by staff. Time to cecum	Completion improved from first 25 cases to fifth 25 cases: 43.1% to 75.1%. Time to cecum from 18.7 to 17.1 min.	4
Hawes et al (1986)	Single center, prospective, residents rotating on GI ward performing sigmoidoscopy	25	Assessment of overall skill on 6-point competence scale (1-3 not competent, 4-6 competent), accuracy of diagnosis	495 of 662 sigmoidoscopies were graded. Initial 10 examinations largely graded 3 or less. Examinations 10 to 25: largely grade 3 or 4, and >25, 82% was graded 4 or above. With increasing competence score, more correct diagnoses.	3

*2 testing periods. Period 1: 4 first-year fellows, 2 second-year. Period 2: 3 first-year fellows, 4 second-year.
** These studies reported on both UGI endoscopy and colonoscopy, the study population is the same.

Table 6B. Learning curves and threshold numbers as a measure of performance in ERCP.

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Jowell et al (1996)	Single center, prospective, fellows with different levels of ERCP experience	17	Outcome: grading of performance on 6-point scale. Competence: score 1,2,3; overall and per part of procedure (i.e. CBD cannulation). Learning curve by probability of success.	1450 of 1796 ERCPs were evaluated. Median n per trainee=132 (range 57-186). Overall score: 0.8 after 137 ERCPs and 0.9 after 185 procedures. Score and numbers for CBD cannulation: 0.65 after 180 ERCPs.	2b
Watkins et al (1996)	Multicenter, prospective, trainees with no previous ERCP experience	20	Outcome: selective cannulation of CBD and PD. Analysis per 10 consecutive procedures.	641 ERCPs performed by trainees (mean 31, range 10-96). First 10 ERCPs (n=199); PD cannulation 46%, CBD cannulation 32%. Rapid increase during first 40 procedures; 85% selective cannulation after 70 ERCPs for PD and >100 for CBD.	2b
Waller et al (2009)	Single center, prospective, trainee without ERCP experience	1	Outcome: successful cannulation. Learning curves through cumulative failure charting and (cusum) analysis (unacceptable failure rate 0.35, acceptable failure rate 0.20).	290 ERCPs. Cumulative failure charting: after 60 ERCPs no unacceptable performance. Acceptable performance after 100 procedures with cusum analysis.	3
Biau et al (2008)	Single center, retrospective, endoscopist without ERCP experience	1	Learning curves through cusum analysis, success rates for selective cannulation (acceptable failure rate 0.10).	529 ERCPs, selective cannulation in 479 (90.5%). success rate after 100 ERCPs was 82%, 200 ERCPs 88%, 300 ERCPs 90% and 400 95%. Competence at n=79.	3
Verma et al (2007)	Single center, retrospective, endoscopist without ERCP experience	1	Outcome: cannulation success of CBD in patients with native papillary anatomy. Learning curve per 50 procedures.	1097 ERCPs (697 as trainee, 400 as independent endoscopist). Success rates: baseline 42% and ≥80% after 350-400 procedures. Posttraining, the following 300 ERCPs >96% success.	4
Vitale et al (2005)	Single center, retrospective, fellows with different experience in ERCP (training period 6 to 14 months)	13	Type of ERCP, success of cannulation of intended duct per 3-month period, faculty success where fellows failed	2008 ERCPs performed by fellows (median 135, range 53-351). First period, 95% diagnostic ERCPs, last period 95% therapeutic. Cannulation success improved from 77.3% in first period to 84.4% in the 4th period. After 7.1 months and 102 ERCPs, cannulation of >85%.	4

Table 6C. Learning curves and threshold numbers as a measure of performance in EUS.

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Wani et al (2013)	Multicenter, prospective, trainees without prior EUS experience	5	Outcome: performance per anatomic station. Learning curves through cumulative sum (cusum) analysis. Acceptable and unacceptable failure rates: 0.10 and 0.20.	n = 1412 EUS examinations (median per trainee 295, range 175-402). Acceptable overall performance based on scores per anatomic station for 2 trainees after EUS 255 and 295. 2 trainees trend to competency (n=255 and 196), 1 needed ongoing observation after n=402.	2b
Meenan et al (2003)	Single center, prospective, GI trainees and nurse endoscopist with different experience in EUS	4 & 1	Assessment of the ability to reproduce set views from esophagus, stomach and duodenum, points per structure (18, 8 and 11, respectively), 5 examinations each	Previous EUS experience: observed ranged from 55 to 170 procedures, performed from 25 to 124. Mean overall score per trainee: 37.4; 32; 29.7; 23.3. Nurse: 12.5 out of 18 (esophagus alone). Correlation between n performed and score for duodenal views ($p<0.01$). Competence after 25 procs in esophagus, 35 in the stomach, 78 in duodenum.	3

for already registered endoscopists.[82] The Mayo Colonoscopy Skills Assessment Tool (MCSAT) was more effective in discriminating different experience levels, and therefore applicable in training settings.[87] Two studies reported on some sort of video assessment of endoscopic skills.[88,89] The tri-split video recording assessment tool proved to be valid, but reliability was lacking.[88] The other study on video assessment described the development of an assessment tool for sigmoidoscopy withdrawals in a series of five experiments.[89] They concluded that the sequential assessment of five withdrawals led to the highest agreement. However, all procedures included in this video study were performed by experienced endoscopists. Some assessment tools were applicable in training situations, while others were only evaluated in a setting with experienced endoscopists. This difference makes it therefore difficult to compare the assessment tools.

Training models

Finally, seven studies reported on different kinds of training models for colonoscopy (n=4), sigmoidoscopy (n=1) and EGD (n=2).[91-97] Table 8 provides an overview of these studies. Two groups described the evaluation of the accelerated colonoscopy training course (ACTC) as it is carried out in the UK.[96,97] Both concluded that performance in knowledge, colonoscopy performance, and Direct Observation of Procedural Skills (DOPS) scores improved significantly after the training week. Thomas-Gibson et al. added an evaluation at a median follow-up of 9 months. There were however no differences between post-training assessment and follow-up. A different training model was the 'gastroenterological education – training endoscopy' (GATE) model.[91] This training model showed improvement in post-test results and simulator performance. A German group tried to identify predictors for performance in a 1-week training course by psychological and psychomotor tests.[94] The training week resulted in improved performance, but only one specific (double labyrinth) test was identified as a predictor for improvement in performance.

One RCT evaluated the impact of systematic feedback on colonoscopy performance.[92] Although only four trainees were evaluated, there was a significant improvement in CIR performance in the feedback group, while the control group showed no improvement.

DISCUSSION

Forward viewing flexible endoscopy procedures

Gastrointestinal endoscopic procedures are fairly complex. The sole use of the classic master-apprentice model for teaching endoscopy is nowadays less accepted. The use of simulators in the early training phase is gaining acceptance and several VR endoscopy simulators have been validated (Table 2). The GI Mentor, AccuTouch, and Endo TS-1 were shown to have good validity.[4,8-10,12-16,19-22,25,27,28] These can thus be considered as realistic devices that have discriminative abilities for distinguishing dexterity and competence levels in flexible endoscopy. Based on these LOE 2b studies, the grade of recommendation for the validity of the mentioned mechanical simulators is B.

Following validation, the impact of simulator training on learning curves needs to be assessed. A VR simulator with good validity, but not improving performance after repeated exercise, is not suitable for implementing in a training program. Three studies with LOE 1b provided evidence for the positive effect of simulator training in novices in flexible endoscopy, measured

Table 7. Assessment tools for colonoscopy, sigmoidoscopy and EGD.

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Barton et al (2012)	Candidates for bowel cancer screening program, prospective	162	Evaluation of colonoscopy by experts through DOPS. Grading performance through DOPS over 2 consecutive colonoscopies, global expert assessment, reliability and validity	193 assessments, 46 excluded, 147 candidates remained. Reliability of 0.81 (high). Validity: 72.6% of candidates and 92.9% of assessors experienced DOPS as valid. Global expert assessment and DOPS grade were similar in 97%. Variance with candidates.	2b
Shah et al (2002)	Prospective, single center, endoscopists with different endoscopy experience	18	Tri-split video of colonoscopy (endoscopists hands, endoscopy view, magnetic endoscopic imaging) during insertion. Scoring system for instrument controls, insertion tube, depth of insertion	22 colonoscopies scored by each of 3 observers. 4 endoscopists with <100 colonoscopies experience, 3 with 250, 2 with 500 and 9 with >1000 experience. Significant differences between scorers ($p<0.001$). Good validity (differences for competence categories, $p<0.001$). Good interobserver agreement and correlation between individual scores and global assessment of competence ($p<0.001$). Reliability is lacking.	2b
Sedlack (2010)	Single center, prospective, trainees	41	Development and evaluation of colonoscopy assessment tool, assessment of colonoscopy skills, validity testing, correlation between item scores and overall	3936 MCSAT completed (62%) by 58 staff members. Correlation between average and overall cognitive and motor scores (0.79 and 0.88, respectively, $p<0.01$). Difference in scores related to experience ($p<0.01$)	2b
Sarker et al (2008)	Multicenter, prospective, trainees and consultants	21	Assessment of skills by Likert scale, hierarchical task analysis. Assessment of flexible sigmoidoscopy and colonoscopy, generic technical skills, two assessors, validity and reliability	135 endoscopies (75 flex sig, 60 colonoscopies) assessed, 9 consultants and 12 trainees. Cronbach's alpha for flex sig (generic vs. specific): 0.81 and 0.79 ($p<0.05$). For colonoscopy generic vs. specific 0.85 and 0.80 ($p<0.05$). Construct validity for experience in flex sig and colonoscopy: generic vs. specific $p=0.005$ and 0.003 and $p=0.012$ and 0.004.	2b

Table 7. Assessment tools for colonoscopy, sigmoidoscopy and EGD. (continued)

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Thomas-Gibson et al (2006)	UK flexible sigmoidoscopy screening trial, experienced endoscopists	43	Development of scoring system for assessment of accuracy for flexible sigmoidoscopy, series of 5 experiments for video scoring of extubation, up to 6 scorers per experiment. First two experiments scores on VAS, last three on 5-point Likert scale. Different items of extubation assessed	Overall ICC for first 2 experiments: 0.10. In experiment 3, all scores varied significantly between scorers ($p<0.05$). In experiment 4, overall parameters scored first with higher ICC (quality improved from 0.45 to 0.72). Experiment 5: sequential assessment of videos vs. random. Agreement was low for individual extubations (ICC 0.13) but high for series of 5 extubations (ICC 0.89).	2b
Boyle et al (2012)	Multicenter, prospective, trainees and consultants	27	Development and evaluation of colonoscopy assessment tool. Assessment of elective colonoscopies, dichotomous checklist and global assessment	81 procedures (24 by 8 consultant and 57 by 19 trainees) assessed. Significant differences in overall score for novice, intermediate and expert endoscopists ($p<0.001$). Checklist: 3 items distinguished between two or more groups, global assessment: 6 items distinguished between three levels	3
Vassiliou et al (2010)	Multi center, prospective, novice and experienced endoscopists	139	Development and evaluation of GAGES for EGD and colonoscopy, reliability and validity. Assessment by operator, attending and observer. Novices and experienced endoscopists	Data on attending and observer evaluations: 18 EGD and 13 colonoscopies. Data on attending and participants: 77 EGD and 57 colonoscopies. ICC for attendings and observers for both tools were 0.96 and 0.97. ICC for attending and participants: 0.78 and 0.89. Validity for both tools ($p<0.001$).	3
Hope et al (2013)	Single center, prospective, surgical residents	100	Assessment of colonoscopy performance by two assessments tools (8 objective criteria vs 10 generic and specific endoscopic skills). Performance per PGY level	100 colonoscopies performed by residents, 89 assessed (72% PGY-3). Tool 1 showed on some items significant differences for PGY-level ($p<0.05$) but not for all. Tool 2 as well ($p<0.05$). Both tools not for all categories significantly different; they show improvement with experience but not for all items.	3
Mohamed et al (2011)	Single center, prospective, first year GI fellows and third year surgery fellows	7	Assessment of DOPS reliability for trainees for colonoscopy. Comparison of 7-step model for training (GI fellows) colonoscopy with master-apprentice model (surgical fellows)	4 GI fellows with experience of 30 colonoscopies, 3 surgery fellows with 5 colonoscopies experience. Scores for GI fellows and surgical were comparable. Test-retest of DOPS wide range of correlations, no reliability.	4

Table 8. Training models for colonoscopy, sigmoidoscopy and EGD.

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Götzberger et al (2011)	Multicenter, training courses, trainees with different levels of endoscopy experience	98	Evaluation of GATE model (basic and therapeutic course), pre- and postcourse knowledge, acceptance, simulator assessment, colonoscopy	78 trainees provided complete data sets. Acceptance for both courses: 88% would advise participation. 77% underlined realism. Improvement in pre-and post-test results ($p<0.001$). Simulator assessment: reduction in time needed for procedure ($p<0.01$).	2b
Harewood et al (2008)	Single center, RCT, trainees with experience of 500-600 colonoscopies	4	Evaluation of effect of systematic feedback on performance (1-on-1, monthly) in colonoscopy. Evaluation of CIR and PDR.	581 colonoscopies performed, 296 in feedback group, 285 in control group. After feedback, 364 procs performed (211 feedback, 153 control). In feedback group, CIR improved from 72.9 to 83.4% ($p=0.04$). In control group CIR from 78 to 71.9% ($p=0.2$). PDR not significantly different ($p=0.2$ vs $p=0.5$).	2b
Neumann et al (2005)	Multicenter, prospective, trainees without prior endoscopic experience	58	Evaluation of psychological, psychomotor, cognitive tests and subjective assessment of trainer before and after 1-week training course in EGD	58 trainees from 12 centers evaluated. All showed significant improvement in performance ($p<0.001$). Only double labyrinth test was associated with improvement (OR 11.5, $p=0.035$), expert assessment of at least moderate improvement (OR 41.5, $p=0.018$)	2b
Suzuki et al (2006)	Multicenter, prospective, trainees with different endoscopic experience	50	Assessment of accelerated colonoscopy training course: multiple choice questions (MCQ), simulator and hands-on training, evaluation DOPS	50 trainees attended. MCQ score improved from 57 to 66% posttraining ($p<0.001$). Median of 15 live colonoscopies performed: DOPS showed improvement on all aspects ($p=0.007$). Global score improved from 1.7 to 2.0 ($p<0.001$).	2b
Thomas-Gibson et al (2007)	Multicenter, prospective, trainees with different endoscopic experience	21	Assessment of accelerated colonoscopy training course, MCQ, simulator sessions, hands on training, live case assessment, trisplit video assessment. Follow up: MCQ, simulator, live case assessment, trisplit video assessment	13 trainees with <200 colonoscopies, 5 with 200-500 colonoscopies, 3 with 500-1000 colonoscopies; 16 follow-up. Significant improvement in MCQ scores ($p<0.001$), simulator test times (diagnostic vs therapeutic $p=0.02$ vs $p=0.003$). DOPS scores improved significantly on all aspects expect general approach. Trisplit video: 1 scorer gave higher grades ($p=0.008$), 1 did not ($p=0.11$). No difference between post-training week scores and at follow up, except simulator performance.	2b

Table 8. Training models for colonoscopy, sigmoidoscopy and EGD. (continued)

Study	Setting and participants	n participants	Assessment method – outcome measures	Results	LOE
Neumann et al (2003)	Multi center, prospective, trainees with different endoscopic experience.	56	Evaluation of training course for EGD with Erlangen Endo-Trainer, endoscopy Score Cards, tests on each day, self-assessment, video assessments	Assessment of day 1 and 5 training significantly different ($p<0.001$) for both specialist and self-assessment. During entire training, differences in scores of self-assessment and expert assessment on 24 items ($p<0.05$).	3
Proctor et al (1998)	Single center, prospective evaluation, trainees without previous endoscopic experience	10	Evaluation of teaching model for flexible sigmoidoscopy, 8 components with two levels of competency and 1 level non-competent. Comparison with independent observer method	120 of 128 sigmoidoscopies evaluated using teaching model method; 73 (60.8%) competent, 47 incompetently performed. 1 component excluded due to lack of variation; 6 of 7 remaining components were associated with competency in teaching model. 50 of 120 procedures Overall Competence Score; correlation between this and teaching model was 0.71 ($p<0.001$).	3

in terms of both virtual reality as well as live endoscopy.[37,42,53] Two of these were well-designed randomized multicenter trials comparing the combination of simulator- and bedside training vs bedside training alone for colonoscopy training of novices. These studies demonstrated that simulator training is effective.[37,42] Several studies on simulator learning curves for EGD, sigmoidoscopy and colonoscopy gained a LOE of 2a or 2b.[31,32,34,36,38-41,43-45,47,49-52] Based on this evidence, one can conclude that simulator training is complementary to patient-based learning and is useful in the early training phase, resulting in a grade of recommendation B.

The four studies that reported on the use of a simulator as a competence assessment tool showed diverging results.[54-57] Therefore, no grade of recommendation is given.

Elaborating further on the learning curve, the next step is (continuous) assessment of a trainee's performance during patient-based training. The currently available recommendations and guidelines focus mainly on minimum numbers as a threshold for competence.[73-80] However, outcomes and proposed minimum numbers for flexible endoscopic procedures vary widely. Nowadays there is a tendency to define more objective criteria for competence. Two large prospective single-center studies with LOE 1b provided evidence for the use of an assessment form as a measure of competence, respectively the Mayo Colonoscopy Skills Assessment Tool (MCSAT) and the Rotterdam Assessment Form for Colonoscopy (RAF-C).[61,65] The learning curves obtained in these studies were similar. Both forms are good methods to continuously assess performance, resulting in a grade of recommendation A. The DOPS on the other hand is more appropriate for assessment of 'end-stage' competence.[82]

Overall, some high-quality studies have been performed for each individual step in training, providing valuable information on the effect of simulator training, learning curves and assessment methods. The most and best evidence for all these stages regarding basic flexible endoscopy is available for colonoscopy. However, one can imagine that some results can be extrapolated to other basic GI endoscopy procedures as well, since the techniques are comparable.

ERCP

One of the most challenging procedures with high complication rates in GI endoscopy is ERCP. It takes a great deal of training and a large number of procedures to reach competence. However, little is known about the learning curve for trainees in ERCP. A number of questions remain unanswered when it comes to the shape of the learning curve, the number of procedures needed to gain competence, and the definition of competence itself. The six studies on learning curves in ERCP varied widely in design, number of trainees and procedures included, as well as outcome, resulting in a large heterogeneity among them. Successful cannulation in >85% of the patients was seen after a number of 100 to 185 ERCPs. Due to the heterogeneity, a grade of recommendation is not provided. There would be a great benefit if part of the learning curve for endoscopists could be accomplished by training on simulator models. In reality the number of available simulators for training in ERCP is limited. Seven validation studies have been performed in an attempt to validate 6 different ERCP simulator models. A 2b level of evidence was reached in five studies, two studies were merely feasibility studies. The GI Mentor is the only validated VR simulator for ERCP.[7,30] The face and construct validity was demonstrated in these two studies and although it received lower scores than the ex vivo or live porcine model in a head-to-head comparison, it was considered the easiest of all ERCP simulator models to incorporate in a training curriculum.[7,30] The live porcine model was validated only once in comparison to the ex vivo model and the GI Mentor in the same study.[30] The ex vivo simulators and purely mechanical simulators are highly comparable

among each other and achieve similar results. All of these models require a real endoscope to be introduced to reach a papilla which is either a synthetic or an ex vivo papilla located in a mechanical tube representing the duodenum or an ex vivo duodenum. Overall, these ERCP simulator models receive the highest scores on realism. In total four studies were performed reaching a 2b level of evidence with a fairly good concordance resulting in a grade B recommendation. Only one learning curve study was performed, demonstrating higher success in the simulator group.[46] Since this was the only learning curve study on simulator-based ERCP in the literature, no grade of recommendation can be provided. There were no studies found on validated competence assessment tools to objectify performance in ERCP. The most common performance parameter is cannulation success rate. This only partly reflects the extent, therapeutic intent and diversity of a therapeutic procedure like an ERCP.

Endosonography

Endosonography or endoscopic ultrasound (EUS) is widely practiced with an increasing number of therapeutic possibilities since the first reports of transgastric drainage of pseudocysts by Grimm et al.[98] This makes EUS more complex. Especially the therapeutic procedures have a marked overlap with ERCP and demand a great deal of experience. There are only a few reports on simulator-based training in EUS.[5,6] Training diagnostic and interventional EUS seems logical and feasible in a live porcine model but no formal attempt at validation has been made. No grade of recommendation can be given based on these studies. A learning effect by repeated exercise and improvement of performance during EUS procedures in the live porcine model itself was documented in one study.[35] There is a lack of scientific evidence of transfer of competence to a patient-based setting. There is an even greater scarcity of evidence on learning curves and numbers to reach competence in EUS. Two studies were performed that both included five trainees. The first study included only radial EUS.[72] They reported no additional effect of observing large numbers of procedures; the largest benefit was achieved during hands-on training. There is only one LOE 2b study performed.[70] The learning curves differed considerably among the five trainees. These studies demonstrated the substantial need for much more training than the 150 procedures recommended by the American Society for Gastrointestinal Endoscopy (ASGE) in order to reach proficiency. It leads to a grade C recommendation.

Limitations

The heterogeneity of the studies regarding forward viewing endoscopy limits the conclusions that can be drawn. This systematic review covers a broad range of studies regarding training and assessment in GI endoscopy. This broad approach automatically results in a large variety of methodology, devices used and endpoints measured. This hampers head-to-head comparison of individual studies. Another limitation concerns the fact that all studies focused on specific aspects of the endoscopic procedure, instead of on overall performance, which is both overall competence assessment from novice to experienced, certified endoscopy, as well as expert levels for specific procedures.

The evidence in the literature on learning curves and competence measures for ERCP is highly heterogeneous. This makes it impossible to provide a level of recommendation. Also, cannulation success rates do relate to improved performance but do not entirely reflect the diversity of a complex procedure like ERCP. No solid data are currently available on other aspects of therapeutic interventions related to learning curves and benchmarks in ERCP. As of

yet, no validated competence assessment tools have been developed for ERCP. This should be a prerequisite before attempts to define learning goals and benchmarks are made.

Future research

Future research, based on the presented evidence in this review, should therefore include a complete training program. We propose a pre-patient curriculum using simulator training. The transfer of simulation skills to patient-based procedures needs to be further explored. Simulation training needs to be followed by continuous assessment of patient-based endoscopies to provide individual and group learning curves and after a period of time, (repeated) overall assessments of performance by an expert. Therefore, the development of validated assessment tools is necessary and the effect of expert assessments on daily practice needs to be measured.

With respect to ERCP, there is a rationale to start training using simulators. There is however no evidence yet as to what extent or performance level simulator-based training has to be carried out. The next step would be to investigate the transfer of skills to patient-based training. These research objectives seem to be clear goals for future research. There is a need for the development of validated objective assessment tools in ERCP to document progress in training and finally proficiency. Benchmarks can be set using the same assessment tools in ERCP performed by experts.

The evidence on training and competence assessment in EUS is extremely scarce. Although training in a live porcine model seems logical, in the current era of evidence-based medicine, validation studies should be carried out to establish the degree of realism and training potential. Current threshold numbers for training appear to be inadequate, but the available data are sparse. We seem to be far away from establishing benchmarks for competence in EUS and validated assessment tools are lacking.

General conclusions and recommendations

Based on the presented evidence, we propose implementation of simulator training in gastrointestinal endoscopic training curricula. Regarding basic flexible endoscopy (EGD, sigmoidoscopy and colonoscopy), simulator-based training has proven its value and it is justifiable to start a pre-patient training course using a validated simulator. The extent to which simulator-based training should be carried out is still a matter of debate. Furthermore, objective outcome parameters should be measured continuously in patient-based training. This provides insight in the learning curve and is preferable to threshold numbers. The MCSAT, RAF-C and DOPS assessment forms seem to be the best forms to document progress or proficiency levels. Regarding ERCP training, we would recommend a pre-patient training curriculum using a validated simulator as well. Evidence for evaluation of learning curves and continuous assessment in ERCP is scarce. This makes competency based training difficult. The available data support prolonged training, at least to a larger extent than current upheld threshold numbers in most countries. The results so far may hopefully stimulate further research. The evidence on endosonography training and competence is yet the least investigated. A pre-patient training curriculum is logical and attractive. However, the evidence is too scarce to give recommendations at this moment.

Chapter 2

Aims and outline of the thesis

AIMS AND OUTLINE OF THE THESIS

Aims of this thesis

In the introduction, **chapter 1**, we aimed to provide a complete overview of the available literature on assessment of competency, skill development and quality of gastrointestinal endoscopy. However, after reviewing the literature, some topics regarding training and assessment of GI endoscopy still need clarification. This led to the following research questions in this thesis.

Section II of this thesis aims to investigate the learning curves of trainees in colonoscopy. In order to get more grip on the learning process, we firstly explore the opinion of experts regarding endoscopist-related factors essential for a high-quality colonoscopy. Moreover, we want to identify factors that might influence this learning curve, and finally develop a complete training and evaluation program for a solid and reproducible quality assessment of colonoscopy performance.

In section III, we intend to provide a practical tool to evaluate ERCP performance in both trainees and experienced endoscopists. Repeated self-assessment allows to develop a learning curve of trainees in ERCP. Insight in procedural quality of ERCP is merited, and finally, we aim to recognize predictors of procedural failure in ERCP in general practice.

Outline of this thesis

In **chapter 3**, we explore the opinion of experts in the field of gastrointestinal endoscopy on factors important for high-quality colonoscopy. This study is carried out through an anonymous Delphi survey. We try to provide supervisors some support regarding competency assessment in colonoscopy.

As pointed out in the introduction, many studies have been published on virtual reality (VR) simulation in endoscopy. However, data on the transfer of skills from VR training to patient-based colonoscopy are lacking. In **chapter 4**, we evaluate the effect of a prolonged virtual reality training program on patient-based performance in colonoscopy.

In **chapter 5**, we assess the learning curve for colonoscopy of trainees from different hospitals in the Netherlands. We try to identify specific factors influencing their speed of learning the many skills needed for this procedure. Furthermore, we evaluate a newly developed assessment program and aim to evaluate the value of video-assessment, by comparing scores given by a supervising expert endoscopist, video-assessor and through self-assessment by the trainee.

Nowadays, it is common practice to keep track of performance indicators in colonoscopy services, especially since the widespread implementation of organized, nationwide bowel cancer screening programs. However, it is important to realize that a quality indicator such as the cecal intubation rate (CIR) can be a surrogate marker, and that an endoscopist must not pursuit a high CIR ‘at all costs’. Therefore, we explore the relationship of patient comfort with different established quality indicators for colonoscopy in **chapter 6**.

ERCP is one of the most challenging procedures in gastrointestinal endoscopic practice. There is an ongoing discussion on the use of threshold numbers to reach predefined competency levels. Solid data on this topic are scarce and controversial. In **chapter 7**, we aim to provide insight in the learning curve for ERCP using a self-assessment method (Rotterdam Assessment Form for ERCP – RAF-E).

The ERCP practice of experienced endoscopists in our department is prospectively evaluated through the RAF-E as well. We investigate individual as well as group performance regarding ERCP in **chapter 8**.

Following this single-center study, a nationwide prospective quality evaluation of ERCP is carried out. The RAF-E is implemented in an electronic portfolio and all ERCP-performing endoscopists in the Netherlands are invited to register their procedures. In **chapter 9**, we evaluate the participation rate, nationwide quality defined by pre-determined treatment aims, and explore the relationship between determinants and procedural outcome.

Finally, in **chapter 10**, the main findings and conclusions of the studies presented in this thesis are being discussed and put into perspective.

Section II

Colonoscopy learning curves and assessment of performance

Chapter 3

Endoscopist-related factors contributing to high-quality colonoscopy: results of a Delphi survey

Vivian E. Ekkelenkamp, Arjun D. Koch, Jelle Haringsma, Ernst J. Kuipers, Robert A. de Man

Perspectives on Medical Education 2014;3:31-40.

ABSTRACT

Introduction: Education and competency assessment in gastrointestinal endoscopy is important. Concerning colonoscopy, it is not completely clear what the best way is to learn this procedure, what defines competency in colonoscopy, and which factors define a high-quality colonoscopy.

Aims and methods: The aim of this study was to determine the endoscopist-related factors that define a high-quality colonoscopy. A three-round Delphi survey among expert endoscopists was carried out. In round 1, the panel was invited to identify factors essential for a good colonoscopy. The listed factors were to be ranked during the second round. In the third round, a 5-point Likert scale was added. A reference panel was invited to assess the items as well.

Results: Fourteen expert endoscopists from the Netherlands were invited, of whom 8 participated (57%). A list of 30 items important for colonoscopy was formulated. After the following rounds, consensus was reached on 16 items. Validation was conducted among 8 trainees and 8 experienced endoscopists (response 100%). The groups agreed on the importance of all but one factor (p=0.001).

Conclusions: This Delphi survey has made explicit the endoscopist-related factors that are important for optimal colonoscopy. This might provide trainers more support regarding concrete competency assessment of trainees in endoscopy.

INTRODUCTION

Nowadays, colonoscopy is the most commonly performed gastrointestinal endoscopic procedure in Western countries. The implementation of colorectal cancer screening programmes all over the world has largely increased colonoscopy demand, and put major emphasis on quality assessment.[99,100] Indicators such as cecal intubation rate (CIR) and adenoma detection rate (ADR) are now recognized worldwide as outcome parameters on quality.[101] However, these outcome parameters primarily indicate quality over larger numbers of procedures and do not address the quality of a single colonoscopy. Quality is narrowly linked with competency, which is endoscopist-dependent.

The ultimate goal of education and training in colonoscopy is to deliver competent endoscopists, but concrete measures to define competency in colonoscopy are sparse. Procedural knowledge and skill development are both important domains in endoscopic education and determining competency should be based on at least parts of those domains. Skills acquisition in colonoscopy is a topic that has increasingly received attention.[61,63,102,103] There are nonetheless no standards available for structured assessment of trainees and it is not clear which domains deserve the most attention. Assessment forms that are commonly used in Dutch practice are the Objective Structured Assessment of Technical Skills (OSATS) form, similar to the Direct Observation of Procedural Skills (DOPS) form developed in the UK and the Rotterdam Assessment Form for colonoscopy (RAF-C).[61,104] The RAF-C is primarily developed for self-assessment of colonoscopy and therefore not appropriate for expert assessment of the trainee's competency. The OSATS is being used in different specialties and directly derived from the surgical variant. Even for surgical procedures, the value of OSATS for measuring progress as well as defining competency is not clear[105], let alone for colonoscopy. Therefore, colonoscopy trainers as well as trainees feel that the OSATS and RAF-C do not reflect all aspects of colonoscopy training and should be optimized.

There is thus a need to identify specific factors that can be used in the assessment of colonoscopy quality and in skills assessment of trainees in colonoscopy. There are different methods to explore the opinion of experts on this topic. A Delphi survey is a well-recognized method to reach expert consensus through an anonymous group process.[106-110] The aim of this study was to visualize an optimal colonoscopy through the eyes of expert endoscopists, with explicit identification of important factors that define a high-quality colonoscopy.

METHODS

A three-round Delphi survey among expert endoscopists in the Netherlands was conducted. Endoscopists were selected based on their experience in colonoscopy and reputation in the field. Invitations to participate in the survey were sent to the selected endoscopists by e-mail and regular mail. Round one of the survey was sent together with the invitation. This was an open round. The panel was invited to openly list endoscopist-related factors that were in their opinion essential for a high-quality colonoscopy. The factors identified in the first round were collated and compiled into a new questionnaire.

This list was distributed as the second round of the Delphi study. Again, the questionnaire was spread by e-mail and regular mail. In round two, the listed factors had to be ranked by the expert panel. The most important item was ranked as no. 1, the second most important

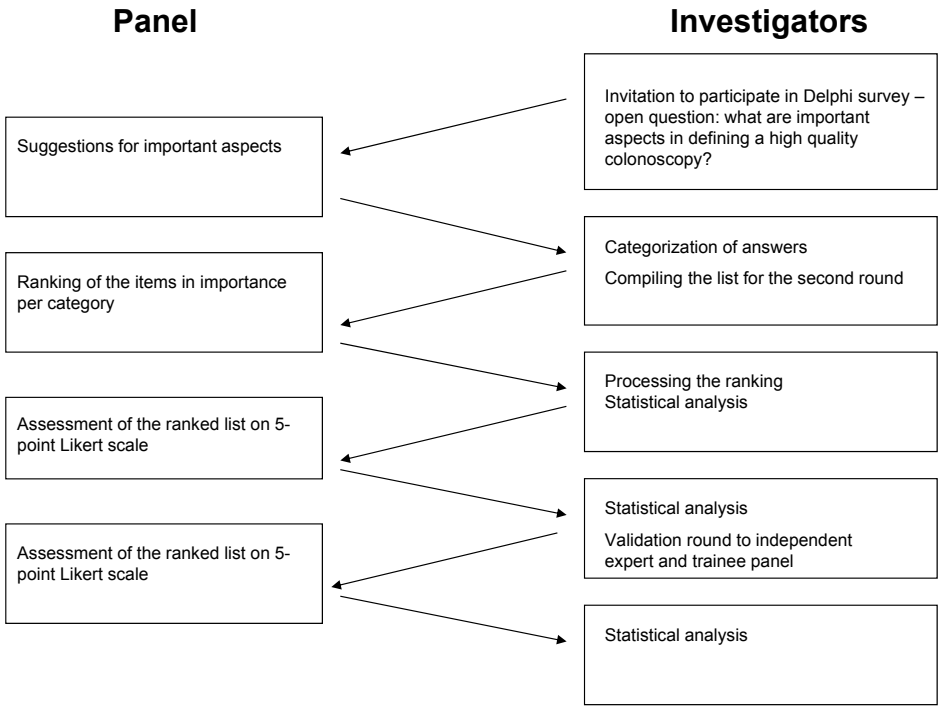


Figure 1. Flowchart of the Delphi process.

item as no. 2, and so forth. The outcome of the ranking in round two was sent back to the panellists as feedback.

The third round consisted of a consensus round. Now, a five-point Likert scale with values ranging from '1 = not important' to '5 = very important' was added to each item. After the three rounds of this Delphi survey were completed, two reference panels of trainees and certified endoscopists were put together. They were invited to assess the items that resulted from this Delphi study on the same five-point Likert scale. The purpose of this last evaluation was to assess whether the results could be extrapolated into clinical practice as items for assessing competency.

Fig. 1 shows the flowchart of the Delphi process as it was carried out in this study.

Data analysis

Mean scores per item were calculated after the second round and mean Likert scores were calculated after the third and validation round. Cronbach's alpha was used to quantify the reliability of the summation of entities, in this case the members of the Delphi panel. Where the responses of the members are highly correlated, they are considered to be internally consistent or homogeneous. A value of >0.8 was considered significant for consensus. Kendall's tau-c was used to analyze the scores of the validation panels. A p-value of <0.05 was considered significant. Analysis was carried out using IBM SPSS Statistics 20.

RESULTS

Expert panel

Initially, 14 experienced endoscopists were identified and invited to participate in the Delphi study. They were selected based on their reputation in the field and they had to have a colonoscopy experience of >1000 procedures. In the first round, 10 out of 14 endoscopists responded (71%). Those 10 experts were invited for the second round; 9 out of 10 responded (90%). Eight out of 9 experts (89%) responded to the third round. The overall response rate was 57%.

Round 1

The first questionnaire resulted in an overall list of 30 items: 10 in the first category representing efficacy and endpoints and 20 in the category representing safety and behaviour. Table 1 lists the factors provided by the panel. Some items were mentioned by different panel members, but these were only entered once in Table 1.

Table 1. Results of the first and second round of the Delphi survey.

Efficacy/endpoints – mean score		Safety/side effects/behaviour – mean score	
Knowledge			
Adequate identification of endoscopic image	3.6	Knowledge of own boundaries	2.4
Basic colonoscopy technique	4.2	Knowledge of material and options for polypectomy	4.2
Knowledge of complications and registration	4.9	Knowledge of the burden for patients	4.4
Use of ADR as marker of performance	5.8	Understanding and solving loops	4.4
Knowledge of anatomy	8	Knowledge of calmly withdrawing the scope	6.6
		Use of CO2 insufflation	7.9
Skills			
Cecal intubation rate	4.2	Skills and hand-eye coordination	3.1
Polyp detection and removal	7.8	Rotation and straightening of the scope	4.7
Competency in intervention techniques	8	Patience and precision	6.1
Duration of the procedure and withdrawal time	8.6	Endoscopy with clear view	6.1
Proper assessment of mucosa	9	Small, gentle movements	6.1
		Anticipation and tip control	7.1
		Minimizing insufflation	7.9
		Proper position for intervention	8.8
		Feeling of equipment	8.9
		Adequate and ergonomic placing of equipment	9.6
		Localizing optimal pressure points and effects	9.6
		Scope positioning by changing patients position	9.9
		Experience of patients	10
		Use of opioids	10

Round 2

In the second round, the random list with factors was sent to the expert panel members who had completed the first round. They were asked to give a number to each item in order to rank the importance. Items with number 1 were considered to be the most important and items with the lowest numbers the least significant by the expert panel. Cronbach's alpha was 0.94, indicating a high level of agreement between the panellists. After analyzing the ranking of the different items, the list was segmented in order to provide a clear overview of the items. Two subclassifications were created for both categories, i.e. 'knowledge' and 'skills'. Items were divided in those classifications by the investigators. The results are also shown in Table 1.

Round 3

In the third round, the list as described in Table 1 was presented to the expert panel with an additional Likert scale. The mean score per item on the Likert scale, with a minimum of 1 and a maximum of 5, varied from 3.1 to 4.9. Cronbach's alpha was 0.60 calculated over all items.

After this analysis, a list of items with the highest scores on the Likert scale per category was created. For each score of Likert 4 or 5, items received 1 point. These points were added up and the top three items with the highest scores per category were selected. When more than three items had the same amount of points, all of them were selected. A new list of factors important for a high-quality colonoscopy was created. All items in this selection, except for 'assessment of mucosa' (n=1), received a score of 4 or 5, representing the values important and very important.

Validation

Two validation panels were created. One panel consisted of eight experienced endoscopists working in our endoscopy department; the other panel was composed of eight gastroenterologists in training. We created these validation panels in order to evaluate the feasibility of using these items for assessing competency in clinical practice. All sixteen selected panel members returned the survey.

Cronbach's alpha was calculated for all three panels together. This was 0.86, which means that a high level of consensus was reached. To explore the differences in scores of the three groups in more detail, Kendall's tau-c was calculated. There were no significant differences in scores between the three groups for all items, except for 'proper position for intervention' (Kendall's tau-c=-.41; p=0.001). Further evaluation showed that trainees scored this item significantly lower than the expert panel (p=0.001). Table 2 shows the final result of this Delphi survey.

DISCUSSION

This Delphi survey has made explicit the endoscopist-related factors that play an important role in defining an optimal colonoscopy. Experts reached consensus on this topic in a three round survey. A list of 16 items was identified during the process. A validation panel of endoscopists and trainees agreed on the importance of nearly all the factors. This may provide trainers more support regarding concrete competency assessment of trainees in endoscopy.

Table 2. Final results of Delphi survey.

Efficacy/endpoints	Safety/side effects/behaviour
Knowledge	
Adequate identification of endoscopic image	Knowledge of own boundaries
Basic colonoscopy technique	Knowledge of material and options for polypectomy
Knowledge of complications and registration	Understanding and solving loops
Skills	
Cecal intubation rate	Skills and hand-eye coordination
Polyp detection and removal	Patience and precision
Competency in intervention techniques	Small, gentle movements
Assessment of mucosa	Minimizing insufflation
	Proper position for intervention
	Feeling of equipment

It is important to assure quality in endoscopy training, but to be able to do that, a standard needs to be set.[111] Up until now, there is no universal method to assess a trainee's ability and capacity. There have been several studies published on skill development in colonoscopy in general, not focusing on assessment. These studies mainly addressed motor or technical skills, such as cecal intubation rate.[61,65] In one study, a learning curve for CIR was created through self-assessment, which was a novel method to gain insight into progression of learning.[61] However, objective assessment by a supervisor did not play a role in this study. In order to provide a more complete picture of competency, objective assessment of other factors next to CIR is necessary. When taking a closer look at the DOPS (Direct Observation of Procedural Skills) evaluation method[112], the focus again lies mainly on technical skills. The Mayo Colonoscopy Skills Assessment Tool (MCSAT) is an assessment device as well, developed to assess gastroenterologists in training on their colonoscopy performance.[87] There are quite some similarities between the results of this Delphi study and the DOPS and MCSAT assessment forms. Basic colonoscopy technique, suction and loop recognition and reduction can be found in each of the tools. The same goes for adequate visualization of the mucosa and identification of landmarks and pathology, as well as completion of the procedure (CIR) and applying the correct intervention.

However, there were definitely some new aspects identified through our Delphi study. Where MCSAT and DOPS really focus on technical skills, as mentioned before, our experts valued factors such as knowledge of own boundaries, patience and precision and knowledge of complications very highly. The MCSAT mentions cognitive skills only once and does not specify them. Knowledge of material and equipment was also considered important by our experts; this is not mentioned at all in both the existing assessment forms. In summary, the existing assessment instruments are imperfect with comparable aspects regarding technical skills, but the Delphi study added an important area for assessment with items on knowledge, safety and behaviour.

The MCSAT was primarily based on literature and guidelines. A focus group of experienced endoscopists from one center (Mayo Clinic) reviewed the blueprint, which eventually resulted in the final form. In this process, there is a risk that endoscopists with strong opinions have great influence on the final composition of the assessment form. A Delphi process seems

more likely to result in honest answers and therefore a better representation of the complete panel's opinion. Nonetheless, there are similarities between the forms despite the different methods of developing them, i.e. pre-procedural assessment, safe advancement of the scope and adequate visualization of the mucosa. The Canadian guidelines on endoscopic quality indicators were put together through a Delphi approach as well.[113]

In this study, a Delphi survey was used to determine specific factors in defining a high-quality colonoscopy. The Delphi process is a well-recognized method to achieve consensus in a group on a given topic. It is a commonly used approach and the method utilizes the information from experience and knowledge of the panel members, mainly experts. Delphi has an anonymous nature. This could be an advantage because there is no place for dominance of specific panel members. On the other hand, discussions and hearing other's arguments might lead to a higher level of consensus, because learning about different perspectives could influence the opinion of panel members.

The final list of items identified by the expert panel was sent to two validation panels, consisting of certified endoscopists and trainees in gastroenterology. Those panels agreed on the importance of almost the entire list; only one item was not assessed as important by the trainees as by the expert panel. The outcome of this internal validation makes it more likely that the factors can be extrapolated to clinical practice and are indeed important for assessment.

One limitation of this study is that the number of participants in the expert panel was relatively small (overall, eight experts completed the survey). There is, however, no strict guideline for the number of panel members when carrying out a Delphi study. The loss of two panellists during the second and third round is another limitation. The responses of those two members could have influenced the outcome and agreement, especially when taking into account that the overall number of panellists was relatively small.

This Delphi study was carried out in order to gain insight into the thoughts of experts on competency in colonoscopy and eventually to create a clear list with items on which competency can be based. The first round was open and no suggestions were given. This resulted in an unbiased response of the panel members, since they had to define for themselves what factors they considered important. After the second round, a high degree of agreement was reached; after three rounds, Cronbach's alpha was slightly lower. In other words, there was a high agreement on the overall set of items, but importance scores on the Likert scale per separate item varied. The first evaluation of the items (after round two) was carried out through a simple ranking; in the second assessment, a Likert scale was added. After evaluating the results of the scores on this Likert scale in the last round, a selection was made of the items with the highest mean scores. After this selection, expert agreement on this last group of factors was again higher. It can therefore be concluded that the final record of factors important in defining competency in colonoscopy is correct, given that Cronbach's alpha had increased again, compared with the previous analysis. After evaluation of this list by the validation panels, consensus was reached on all items but one. Compared with the expert panel and the certified endoscopists, the trainees considered 'proper position for intervention' less important. One possible explanation for this is that trainees have less experience in intervention techniques and still lack knowledge in this area.

The next step in this research process will include the development of a new assessment form with the factors derived from this Delphi study and to test and implement it in clinical practice.

Conclusion

This Delphi study provides valuable insight into the opinion of experts regarding competency in colonoscopy. Endoscopist-related factors for a high-quality colonoscopy have been made explicit through this survey. Taking the outcome into account, we believe that it is justified to implement the factors that resulted from this study in an assessment device for trainees in colonoscopy. This might provide endoscopy trainers more support in concrete competency assessment.

Acknowledgements

We would like to thank the following persons for their contribution and participation in the expert panel: M. Van Leerdam, E. Mathus-Vliegen, S. Van Deventer, J. Dees, C.J. Van der Woude, J. Monkelbaan, A. Geraedts, F. ter Borg and W. Lesterhuis.

Chapter 4

Prolonged colonoscopy simulator training leads to improved performance during patient-based assessment

Arjun D. Koch, Vivian E. Ekkelenkamp, Jelle Haringsma, Erik J. Schoon, Robert A. de Man, Ernst J. Kuipers

Submitted

ABSTRACT

Introduction: Virtual reality endoscopy simulators are increasingly used in the training of novice endoscopists. There are however insufficient data regarding the effect of simulator training on the early learning curve of novice endoscopists. The aim of this study therefore was to assess the clinical performance of novice endoscopists during colonoscopy after intensive and prolonged training on a virtual reality endoscopy simulator.

Methods: Eighteen trainees without any endoscopic experience were included in the study. They were divided into two groups. The simulator-training program consisted of either 50 (group I) or 100 virtual-reality colonoscopies. After 10, 30 and 50 (group II) (group I), and after 20, 60 and 100 (group II) virtual colonoscopies, trainees underwent both simulator-based (SBA) and patient-based (PBA) assessment.

Results: Eighteen novices participated in the study. All completed virtual training and assessments. The mean cecal intubation time on the SBA decreased from baseline 9.50 min. to 2.20 min. at completion of the training ($p=0.002$). Colonic insertion depth during PBA improved from 29.4 to 63.7 cm ($p<0.001$). The learning effect of simulator training ceased after 60 colonoscopies.

Conclusions: Virtual reality training by means of a colonoscopy simulator leads to a significant improvement of performance on the simulator itself and, more importantly, to significantly improved performances during patient-based colonoscopy. This study demonstrates the rationale for intensive simulator training in the early learning curve of novices performing colonoscopy.

INTRODUCTION

Training in procedural skills in gastrointestinal endoscopy is gradually changing from the use of threshold numbers towards a more competency based training approach. Virtual reality (VR) endoscopy simulators may be of benefit in the education of gastroenterology trainees, especially in the early learning curve. However, the major part of training remains patient-based. VR endoscopy simulators are increasingly used in the training of novice endoscopists.[32-34,53,114,115] This is among others due to the continued further upgrading of simulators to high levels of virtual reality[15,16,34,46], the introduction of competency models in medical training[116], and demands from health authorities and the public regarding physician training in general.[50,117] Training the basic endoscope navigational skills in patients only is losing acceptance.

The old adagium “See one, do one” seems no longer appropriate for educating health professionals to perform complex procedures. A recent paper defined a pre-patient training for technical skills for complex medical procedures in a virtual reality surrounding.[118] This is complementary to the ancient master-apprentice model with graduated independence and ‘on the job’ training. A recent Cochrane review showed that endoscopy simulators accelerate the early learning curve of novice endoscopists. There is however no convincing evidence to support superiority for either simulator-based training or patient-based endoscopy training.[119]

A variety of simulators have been developed for virtual reality simulation of endoscopic procedures and interventions. These simulators provide the opportunity to familiarize endoscopists with new procedures and to repeatedly train complex procedures in order to reach a higher level of experience before performing the same procedures in patients. Numerous simulator validation studies have been performed[9,15,16,21,27,28,115,120,121], as well as studies to demonstrate increased performance of novice endoscopists after simulator training.[36,47,50,115] Remaining questions are on the optimal extension of VR training, and whether there is a point at which prolonged simulator training does not further improve performance? What is the optimal use of VR endoscopy simulator training prior to transfer to a clinical setting in a master-apprentice situation?

The aim of this study therefore was to assess the performance of novice endoscopists during patient-based colonoscopy after intensive and prolonged training on a virtual reality endoscopy simulator and to identify the point where continuation of training on the simulator ceases to have additional value on skills acquisition.

PATIENTS AND METHODS

We performed a prospective, single-center evaluation for training colonoscopy at the Erasmus MC - University Medical Center in Rotterdam, the Netherlands. The study was approved by the institutional review board.

All participants were young physicians at the start of their training in gastroenterology. Trainees were excluded if they had any form of previous simulator or patient-based endoscopic experience. All trainees provided informed consent and completed a questionnaire providing data on demographics, previous endoscopic experience, and simulator experience.

Simulator

The simulator used in this study was the Simbionix GI Mentor II (Simbionix Ltd. Israel, software version 2.7.4). The endoscope used is a customized Pentax ECS-3840F endoscope. The simulator is equipped with a training program for 20 virtual colonoscopies, two modules with each ten colonoscopies. All colonoscopy cases were randomly used for training in order to avoid bias by training only one patient scenario. Case number three of the first module was repeated after each training session for assessment of performance. This case was chosen because of its discriminative value in measuring performance as demonstrated in earlier validation studies.[15,36] It is a straightforward colonoscopy without any abnormalities such as polyps, tumors, or inflammation; however, the relatively winding sigmoid and a built in loop in the descending and transverse colon make it a fairly complicated case.

Training program

All participants received a standardized tutorial on the fundamentals of colonoscopy to ensure minimum background knowledge with respect to the basic concepts of colonoscopy and colonoscope handling. An instructor (ADK or VEE) was present during the entire simulator training program. Two cohorts of participants were formed. The training program contained 20 or 10 sessions of five consecutive colonoscopies; each session ended with case three from the first module for colonoscopy. To avoid bias in the performance scores, the participants were not notified about the repetitive nature of the last VR-colonoscopy in each session. Two sessions were performed each week. This five or ten week schedule was based on the idea that distributed learning is more effective than massed training.[122] After 10, 30 and 50 (group I), or 20, 60 and 100 (group II) virtual colonoscopies, participants performed two patient-based colonoscopies (Figure 1). All participants were randomly allocated to the first or second group. This elaborate division in two cohorts was chosen to minimize the learning effect of each patient-based colonoscopy performed by the trainees. As our primary aim was to assess the effect of simulator-based training on patient-based performance, we tried to limit the

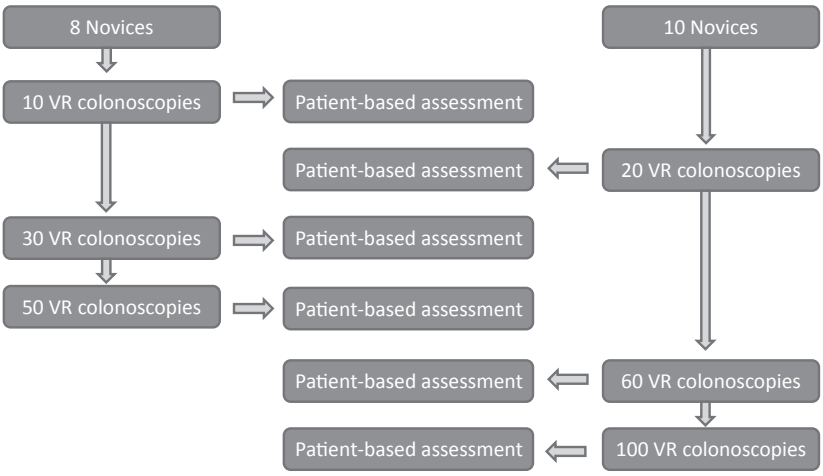


Figure 1. Training program and patient-based assessment.

number of bedside colonoscopies as much as possible. The division into two cohorts led to a reduction of 50% on the total number of patient-based colonoscopies, from a total of 12 to 6 procedures per participant. The simulator-based learning curve was analyzed for the entire group. This means that 18 participants, both group I and II, performed 50 VR colonoscopies, while 10 of them performed an additional 50 VR colonoscopies.

Patient-based colonoscopies

Patient-based colonoscopies were performed after 10, 20, 30, 50, 60 and 100 simulator colonoscopies. To minimize the impact of learning acquired by performing the patient-based colonoscopies, two actions were taken: (1) the number of patient-based colonoscopies was limited to two colonoscopies only, and (2) participants were allocated to two groups, both performing three of the total of six patient-based assessments (Figure 1). The number of two colonoscopies was consciously chosen to correct for any difficult colonoscopy, for example with an inadequately cleaned colon, fixed sigmoid because of diverticulosis or previous abdominal surgery. Although ideal for statistical purposes, a greater number of real-life colonoscopies was considered to have too much impact on the simulator-derived learning curve of novice endoscopists. The mean results in terms of the number of cecal intubations and the maximum insertion depth of these two clinical colonoscopies were used for analysis.

For the same reasons, to minimize the impact on the simulator-derived learning curve, the trainees were allocated to two groups with the same training program but patient-based assessments at different intervals. Colonoscopies were performed using an Olympus CF-Q160DL colonoscope. A ScopeGuide™ 3-D magnetic endoscope imager view was utilised for all PBA procedures. Participants had access to the information on the endoscope position and loop formation, provided by the ScopeGuide™ imager. The patient-based assessments were carried out in randomly selected patients who were already scheduled for routine colonoscopy. Exclusion criteria were (i) previous colonic resections, and (ii) documented previous colonoscopies where the cecum was not reached by an expert endoscopist. An expert endoscopist was present at all times during the patient-based assessment. The expert was unaware of the number of VR colonoscopies performed by the trainee. The expert was instructed to take over the procedure (i) if he was concerned at any time about the safety of the procedure or the patient's well-being, (ii) after a fixed 20 minutes time limit, or (iii) when the trainee had reached the cecum. Patients provided informed consent for inclusion in the study. In case the trainee did not reach the cecum, the maximum insertion depth from the anal verge was measured after straightening the endoscope using the endoscopic and the ScopeGuide™ view. In case a loop was present, the endoscope was straightened while maintaining the tip of the endoscope at the point of maximum insertion as displayed by the ScopeGuide™ imager.

Statistical analysis

Analyses were carried out in SAS® version 9.2. The two cohorts, as described, were combined and considered as one integral group for the analysis. This means that the learning curves for both groups were fused and considered as a single, group learning curve. Linear mixed models were used for analysis of the patient-based assessments of the integrated group. These models included a random intercept for trainees, in order to take the repeated measurements of the individual trainees into account. A Wilcoxon signed rank test was used for analysis of group simulator performance. Two-sided P-values less than 0.05 were considered significant. Differences between patient-based assessments at the predetermined

intervals were calculated using two-tailed, paired and unpaired t-tests. Graphs and trend lines were created using standard software.

RESULTS

Eighteen trainees were included, five male and thirteen female. Eight participants were allocated to the first group, ten to the second. Their mean age was 27 years. All 18 trainees were trainees at the start of their training in gastroenterology. None of them had any previous endoscopic experience. All completed the training program and patient-based assessments (PBAs).

Simulator performance

The mean cecal intubation time on the simulator for the entire group improved from baseline 9.50 min. to 2.20 min. at completion of the training (p=0.002). The learning curve is displayed in Figure 2. Eighteen participants performed 50 VR colonoscopies and only ten participants, allocated to the second group proceeded to perform 50 more VR colonoscopies to a total of one hundred. The learning curve is displayed as a mean group learning curve for the entire group. The results show a rapid improvement of performance on the simulator during the first 50 VR colonoscopies. From 60 VR colonoscopies the learning effect of prolonged training on the simulator seems to diminish.

Patient-based assessments

Tables 1 and 2 show the data on the patient-based assessments for each group. The insertion depth is reported with the range and the 95% confidence interval. Also the number of cecal intubations that occurred at each session is reported. Figure 3 shows the learning curve for

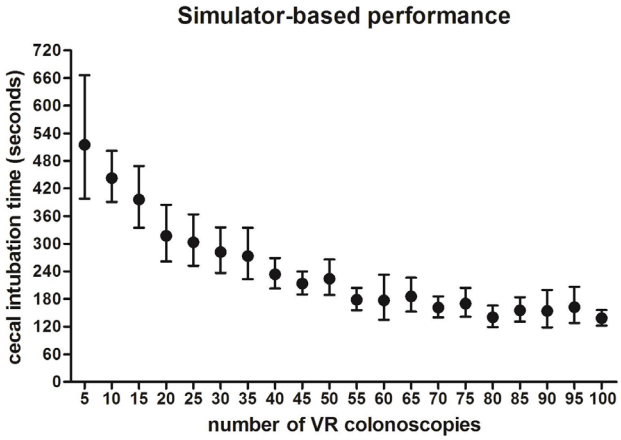


Figure 2. Simulator-based performance curve.

insertion depth during patient-based assessment related to the number of VR colonoscopies. This performance curve is displayed as a mean for the entire group.

Analysis of the patient-based assessments was carried out using a linear mixed model, with insertion depth as dependent variable, simulator session as predictor and a random intercept for trainees. Insertion depth increased from 29.4 cm at the first PBA to 63.7 cm during the last PBA. This increment in insertion depth over the amount of training was significant ($p<0.001$). There was no significant difference in performance between trainees ($p=0.29$). Differences between separate PBAs were calculated. The results are shown in Table 3. After 50 VR colonoscopies, a statistically significant increment in performance could no longer be demonstrated ($p=0.361$). A visual estimate of the patient-based performance curve demonstrates a plateau

Table 1. Performance during patient-based assessment (PBA). Group I (n=8).

Number of VR-colonoscopies	Number of PBA colonoscopies	Insertion depth (cm)	Range (cm)	95% CI (cm)	Number of cecal intubations
10	16	29.4	12-48	22.6-36.2	0
30	16	38.6	18-70	31.4-46.1	1
50	16	58.5	23-85	50.9-66.1	1

Table 2. Performance during patient-based assessment (PBA). Group II (n=10).

Number of VR-colonoscopies	Number of PBA colonoscopies	Insertion depth (cm)	Range (cm)	95% CI (cm)	Number of cecal intubations
20	20	36.5	18-60	30.3-42.7	0
60	20	60.5	32-95	51.8-69.1	1
100	20	63.7	25-110	55.1-72.2	3

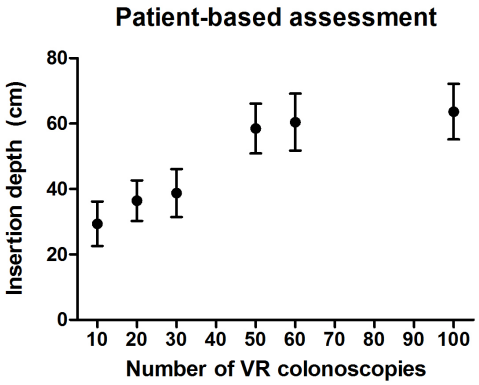


Figure 3. Patient-based assessment.

Table 3. Differences in insertion depths.

PBA session	Mean difference in insertion	p-value*
10 vs. 100	34.3	<0.001
20 vs. 100	27.2	<0.001
30 vs. 100	25.1	<0.001
50 vs. 100	5.2	0.361
60 vs. 100	3.2	0.561

*Differences were tested with two-tailed paired t-tests within the same group and unpaired t-tests between groups.

phase with no further improvement after approximately 60 VR colonoscopies, similar to the simulator-based performance curve.

DISCUSSION

Numerous studies have been performed to demonstrate improved performance of novice endoscopists after simulator-based training. Most of these studies show increased performance either on the simulator itself, and some measure improved performance and transfer to patient-based endoscopy.[34,36,38,39,49] A limited number of studies have been carried out to determine the extent of simulator training versus no training. A large multicenter trial demonstrated improved competency during the first 80 patient-based colonoscopies after simulator training.[37] Another multicenter trial compared simulator-based training with traditional bedside training and demonstrated equal acquisition of skills and performance in patient-based colonoscopy.[42] A recent Cochrane review confirmed the additional value of simulator-based endoscopy training in the early learning curve of novice trainees. No superiority could be demonstrated for either simulator-based or traditional patient-based training.[119]

In this study we have chosen a different approach and investigated how simulator-based training affects clinical performance and how much training is useful to still generate an increment in performance. We have demonstrated the rationale for prolonged and intensive training on a colonoscopy simulator. Continued training shows improvement of performance of novice endoscopists on the simulator itself, which is accompanied by significant improvement of performances during actual patient based colonoscopy. Prolonged simulator training provides novices with a head start for clinical endoscopy. This is thought to contribute to improvement of patient safety and diminish patient risk.

We have used both a statistical and a visual estimate to identify the point where there is no more improvement with continued training on the VR simulator. In our study, there was no longer a statistically significant increase in performance measured after 50 VR colonoscopies. However, at this point there is still a slight rise in the learning curve, indicating that there is a potential benefit for additional VR training. Visually, this point is reached after approximately 60 VR colonoscopies. From this point on there is no more improvement in simulator-based performance as well as patient-based performance. This practical approach simplifies preclinical training: it seems that when the acquisition of skills on the simulator reaches a plateau phase, the same is true for the transfer of these skills to patient-based colonoscopy. This can be translated into a training curriculum where trainees start to train their colonoscopy

skills using VR simulators. When the learning curve on the simulator itself seems to reach the plateau phase, there is no need to continue VR training and the trainee can progress to “on the job” training in patient-based colonoscopy. This method replaces a threshold number of procedures by a more competence-based approach.

Another advantage of simulator training program is that it potentially diminishes instructor time. This was not measured in this study but previously reported.[42] A possible pitfall in leaving novice endoscopists completely alone during their simulator training is that ‘bad habits’ in handling the endoscope are being developed which have to be ‘unlearned’ during patient-based colonoscopy. This problem was avoided in this study by providing all participants a standardized introduction on how to handle the instrument. Also during the training an instructor (ADK or VEE) was present at intervals to observe the trainees and give instructions if necessary.

Ideally, for statistical purposes, a larger number of colonoscopies had been performed per session. More patient-based assessments could have resulted in a more fitted learning curve. However, having the trainees perform more patient-based colonoscopies during ongoing simulator training would inevitably have affected the learning curve and would have biased the effect of the simulator training. In order to avoid this bias, the trainees were divided into two separate groups. Each group performed their patient-based assessments at different points in time in their training program.

Based on the findings in this study, in the Erasmus MC - University Medical Center, a ‘pre-patient’ training curriculum has been implemented, containing a simulator training course until a plateau phase is reached. In our study, this occurred around 60 VR colonoscopies performed on the GI Mentor.

Conclusions

Virtual reality training on a simulator in a ‘pre-patient’ training curriculum leads to a significant improvement of performance on the simulator itself and, more importantly, to significantly improved performances during patient-based colonoscopy. This study demonstrates the rationale for intensive simulator training in the early learning curve of novices performing colonoscopy and provides a practical approach to define the extent of VR training that is useful.

Chapter 5

Evaluating the learning curve in colonoscopy – the value of a repetitive assessment and feedback program

Vivian E. Ekkelenkamp, Arjun D. Koch, Wilco Lesterhuis, Robert A. de Man, Ernst J. Kuipers

Submitted

ABSTRACT

Introduction: Colonoscopy is a practical skill requiring extensive training. There has been little attempt to comprehensively assess both generic and specific technical skills in colonoscopy. Standardized self-assessment during colonoscopy training provides insight in performance and enables specific intervention to improve skills. Implementation of a further intensified assessment and feedback program with close trainee observation at regular intervals in training, provides the opportunity for such interventions. Video-assessment may complement this. The aim of this study was to assess the value of a standardized assessment program on top of self-assessment of performance and to evaluate the value of video-assessment.

Methods: We developed a standardized skills training and assessment program for colonoscopy. All procedures were self-assessed. Trainees were systematically evaluated after every 50 colonoscopies performed during their formal colonoscopy training. For this evaluation, a single expert closely observed the trainee during two subsequent colonoscopies and gave feedback on the trainees' performance. In order to assess the value of this assessment program on top of the self-assessment program, we used a historical control group of trainees as a reference. The historical group only participated in the self-assessment program. The primary outcome parameter was the incremental change in cecal intubation rate (CIR) expressed in a learning curve. Video-assessments and supervisor-assessments were compared and correlations were calculated.

Results: The study group consisted of sixteen trainees from different centers. The reference group consisted of 19 trainees, all from one center. A total number of 3857 colonoscopies were performed in the study group against 2887 in the reference group. Ninety patient-based colonoscopies (range 3-10 per trainee) were carried out during the assessments. Baseline CIR after 20 procedures was 43.6% versus 64.1% in the reference group ($p<0.001$). The slopes of the two learning curves differed significantly between the study and control group (β 0.11 vs. 0.07, $p<0.001$). Evaluation of the program showed that trainees experienced the assessment program as positive and useful (mean score of 4 on a Likert scale of 1-5). Scores on different aspects of the colonoscopy, given by the supervisor and the video-expert, correlated highly.

Conclusions: The addition of a standardized assessment and feedback program on top of self-assessment in colonoscopy training is useful and results in a steeper learning curve. Trainees experience this program as valuable. Video-assessment is a promising addition to colonoscopy training programs.

INTRODUCTION

Colonoscopy is a complex medical procedure with various important aspects, such as cecal intubation and visualization of the entire colon, patient burden, adverse events, complications and performing additional techniques such as biopsy sampling, polypectomy and hemostasis. Achieving competence in colonoscopy requires extensive training. The concept of training and competency assessment in colonoscopy is gradually shifting from training and certification based on threshold numbers to a more competency-based approach. The use of learning curves and the development of specialized assessment tools have made such an approach possible.[61,118]

Cecal intubation is a prerequisite for total colonoscopy and therefore one of the most obvious primary parameters to evaluate the learning process of novices in colonoscopy. There is a rationale to start the learning process with achieving navigational control, focusing on insertion and loop-management leading to cecal intubation.

Nowadays, the first basic endoscopic skills are often being trained on a virtual reality (VR) simulator before moving on to patient-based training.[12,13] Starting patient-based colonoscopy, continuous assessment of performance of the trainee can be carried out by using the Rotterdam Assessment Form for Colonoscopy (RAF-C).[61] This has proven its value for providing insight in learning curves of groups as well as individuals through self-assessment. This method can act as an ongoing monitoring process in order to evaluate performance.[123] A standardized assessment program at fixed intervals may potentially speed up the learning curve even further. In the UK, where an expert supervisor always attends during colonoscopy lists, trainees have a steep learning curve.[124] Traditionally, especially in the Netherlands, when trainees have reached a certain competency level, there is a shift from standard supervision to feedback on request. The yield of unrequested feedback to improve on minor points resulting in gradual perfection is thereby probably lost. The use of video assessment on specific moments in training with expert feedback can be valuable as well.[88] The so-called Tri-split Video

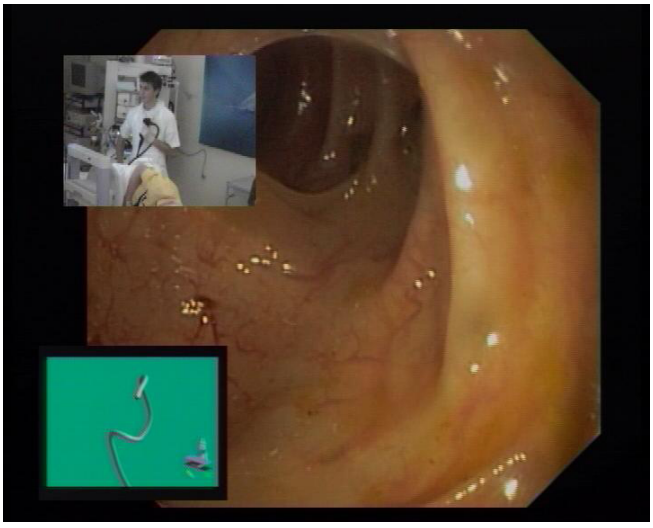


Figure 1. Tri-split video recording.

technique (TsV), which is a combination of endoscopic luminal view, ScopeGuide (SG) image and instrument handling, enables observation of technical competence and can be used for assessment of trainees (Figure 1).

The aim of this study was to evaluate the value of a repetitive standardized assessment program with structured feedback on top of continuous self-assessment in colonoscopy training. Secondly, we aimed to objectively assess progression during training using two different methods: (I) direct observation and assessment, and (II) blinded video-assessment.

METHODS

From 2011 to 2014 we performed a prospective, multicenter evaluation of performance of novices in colonoscopy. After two years of basic training in internal medicine, trainees start with their gastroenterology residency. All trainees perform endoscopy during this period. Trainees from tertiary as well as regional training centers were included. The internal review board approved this study. Figure 2 shows a flowchart of the study course.

Participants and inclusion

Participants were trainees in gastroenterology with limited previous flexible endoscopy experience at inclusion, recruited during a basic flexible endoscopy course in the study center. This is a two-day introductory course on basic navigational skills in endoscopy, compulsory for residents who start their endoscopy training. The course provides in a theoretical background followed by hands-on training on VR-simulators. Participants filled out two questionnaires.

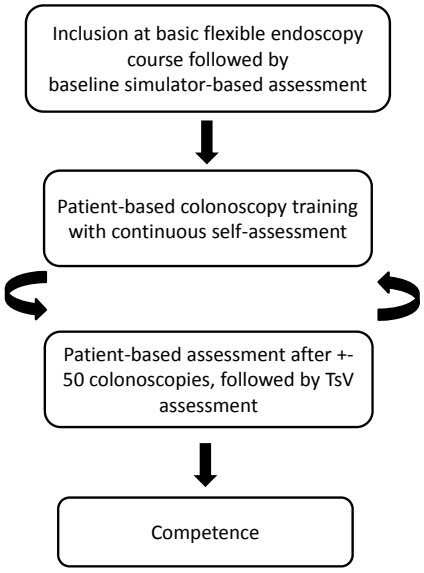


Figure 2. Flowchart of the study.

Data collection

The baseline questionnaire contained questions on demographic data and experience in endoscopy. Each participant received a personal log with a set of self-assessment forms (Rotterdam Assessment Form for colonoscopy or RAF-C), which are based on previously validated assessment forms.[52,125] This form was filled out by the trainee after each colonoscopy and included objective data on CIR, procedural time, as well as subjective information on handling of the endoscope, loop reduction techniques, mucosal visualization and overall performance, graded on a visual analogue scale (VAS). These data were entered in a personal electronic portfolio, used by all trainees in gastroenterology to collect all performed endoscopic procedures, facilitating summative and formative assessment. At predefined intervals (after approximately every fifty colonoscopies), trainees were invited to undergo a standardized assessment and feedback session, which took place at the endoscopy unit of the Erasmus MC - University Medical Center in Rotterdam. All patient-based procedures were recorded for assessment on DVD using the TsV technique.

Assessment program

The assessment program consisted of a patient-based assessment and a patient-based TsV assessment. The patient-based assessment consisted of two colonoscopies, supported by ScopeGuide, with an expert observing both procedures. The number of two colonoscopies was chosen to correct for any aborted colonoscopy due to procedural problems. A twenty-minute time limit was set for intubation of the cecum. The supervisor took over when this time limit was exceeded, or at any point where he felt that the patient's safety was compromised. CIR, the time needed for cecal intubation as well as the patient's comfort was recorded. The assisting endoscopy nurse assessed the comfort levels of the patient. This has proven to be a valuable method.[126] Trainees self-assessed their performance by filling out the RAF-C. Directly after the assessment, the trainees received verbal and written feedback, RAF-forms with suggestions for improvement given by the supervisor and the DVD of the colonoscopies for their portfolio.

Finally, trainees completed an evaluation form on their perception of security in this training environment, the usefulness of the feedback provided by the supervisor, and whether they would recommend this assessment session to a fellow trainee. All items were scored on a Likert scale from 1-5.

The TsV recording was assessed by an expert endoscopist from the study center or the trainer from the local center, using the same RAF-C forms. We aimed to track the trainees for approximately 300 patient-based colonoscopies, since at that point most trainees were expected to have reached a CIR of $\geq 90\%$.[61,65]

Patients

Patients were routinely scheduled from the waiting list for colonoscopy. Informed consent was obtained before the procedure. No in- or exclusion criteria were upheld. Colonoscopies took place under conscious sedation or without medication, according to the patient's preferences.

Reference group

For evaluation of the learning curve, we used a historical cohort as a comparison.[61] This reference group consisted of 19 trainees, who participated in a similar endoscopy training program in a single-center training setting. All trainees in the reference group had performed

upper GI endoscopy and sigmoidoscopy for a period of eight months prior to the study. Self-assessment in this group was carried out using the same RAF-C forms. The learning curve of this cohort was corrected for baseline colonoscopy experience.

Statistical analysis

The primary outcome parameter for learning curve evaluation was the CIR. Learning curves were plotted as a moving average per twenty procedures, in accordance with our historical cohort. The learning curves of the study group and historical cohort were compared and analyzed with linear regression. Pearson's rho correlation coefficients and paired t-tests were used to analyze the correlations and differences in assessment scores by trainee, expert and video-expert. A two-sided p-value of <0.05 was considered significant. Analyses were carried out in IBM SPSS 21. Graphs and trendlines were created with standard software.

RESULTS

From January 2011 to April 2014, sixteen trainees were included in this study. Ten trainees started their endoscopy training in a tertiary hospital, whereas six started in a regional hospital. All of them were included during their third year of residency, after two initial years of internal medicine training.

The total number of colonoscopies performed for this study was 3857; the mean number of procedures per trainee was 241 (range 104 to 509). The group learning curve was compared with the learning curve of the historical cohort, as moving averages per 20 procedures (Figure 3). In the study group, a CIR of ≥90% was achieved after an average number of 240 colonoscopies. The progression of the learning curve was significantly steeper in the study assessment group ($p<0.001$). Baseline CIR was significantly lower in the study group compared to the historical cohort (43.6% vs. 64.1%; $p<0.001$).

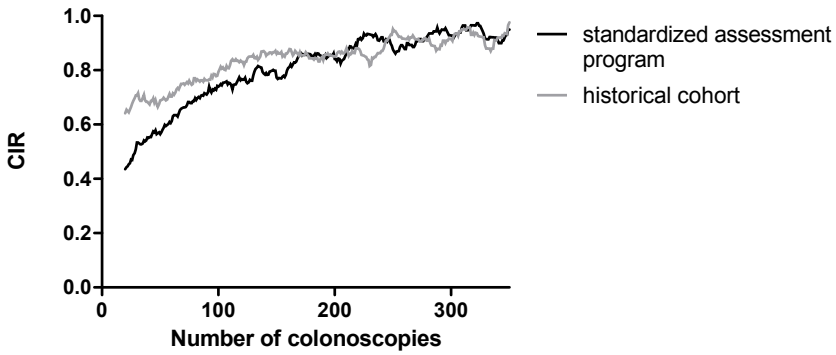


Figure 3. Learning curve of the study group compared with the historical cohort. Cecal intubation rate is plotted as a moving average per 20 procedures.

Patient-based assessments (objective scores)

Ninety patient-based colonoscopies (range 3-10 per trainee) were assessed during the standardized assessment program. Video recordings were obtained for 82 procedures. Eight colonoscopies were not recorded due to technical problems. Figure 4 shows the cecal intubation rate during the assessments, per group of 50 previously performed colonoscopies. CIR is significantly different between the groups (p=0.047). The mean time to cecal intubation dur-

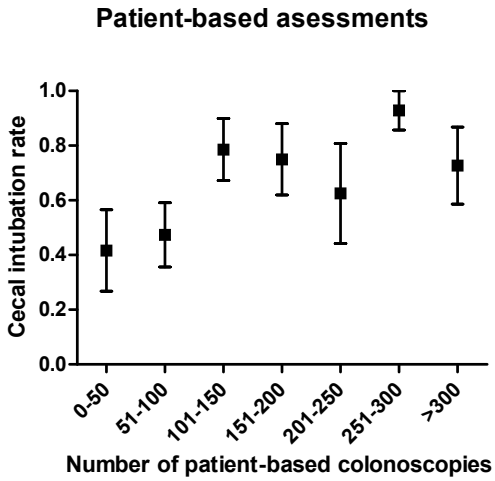


Figure 4. Cecal intubation rate during patient-based assessments, plotted per group of 50 previously performed patient-based colonoscopies (mean + standard error).

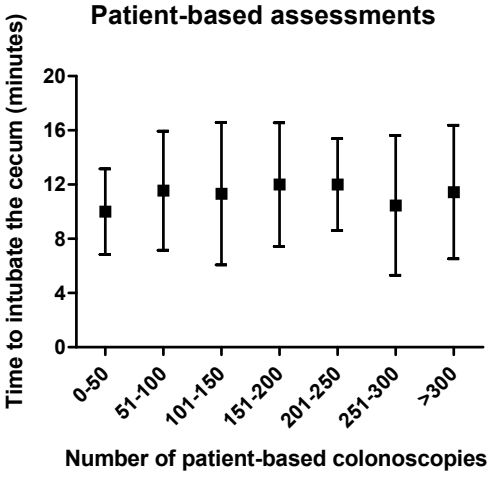


Figure 5. Mean time to cecal intubation (mean + standard deviation) during patient-based assessments.

ing patient-based assessments is graphically depicted in Figure 5. There was no significant improvement in time needed to intubate the cecum.

Patient-based assessments (subjective scores)

The scores of trainee, expert and video-expert on RAF parameters of the procedure were compared (Table 1). The supervisor gave significant higher scores than both the trainee and the video-expert. Furthermore, the correlations of scores given by trainee, expert and video-expert were evaluated (Table 2). We analyzed the increment in the overall performance scores in relation to the number of colonoscopies. Linear regression showed that the number of colonoscopies was significantly associated with the scores on overall performance ($p<0.001$). This applied to scores that were self-assessed as well as expert- and video-assessed.

The correlation between the nurse-assessed patient's comfort and the total number of colonoscopies performed was not significant (Pearson's rho 0.10; $p=0.3$). There was a significant negative correlation between the time needed to intubate the cecum and the comfort of the patient (Pearson's rho -0.38; $p=0.003$).

Trainees experienced the standardized assessment program as valuable and rated the provided feedback as useful (mean Likert score of 4 for both items). Overall, they felt secure during the assessments (Likert score of 4 as well). All participants would recommend this program to a fellow trainee in the same stage of endoscopy training.

Table 1. Mean scores of the patient-based assessments, scored on a visual analogue scale (VAS) from 0-100 by trainee, expert and video-expert (first 3 columns). The last three columns provide insight in the differences between the given scores (paired t-tests).

	Self-assessment	Expert assessment	Video-assessment	Self vs. expert	Expert vs. video	Self vs. video
Time to reach cecum	55	74	49	p<0.001	p<0.001	p=0.20
Handling of the scope	57	79	56	p<0.001	p<0.001	p=0.95
Solving loopings	52	71	55	p<0.001	p<0.001	p=0.23
Time with clear view	66	82	57	p=0.2	p<0.001	p=0.26
Overall performance	56	78	58	p<0.001	p<0.001	p=0.32

Table 2. Correlations between the scores given by trainee, expert and video-expert, based on the patient-based assessments, were analyzed. This table shows the correlation coefficients of the assessment scores given by the three groups (Pearson's rho).

	Self vs. expert	Expert vs. video	Self vs. video
Time to reach cecum	0.66*	0.53*	0.50*
Handling of the scope	0.47*	0.47*	0.49*
Solving loopings	0.56*	0.56*	0.54*
Time with clear view	0.23*	0.29*	0.13
Overall performance	0.56*	0.49*	0.56*

* significant ($p<0.05$)

DISCUSSION

This study describes the impact of a standardized assessment and feedback program, including expert- and video-assessment, on the learning curve of colonoscopy trainees. Those participating in such a program, show a steeper learning curve than trainees who only follow their regular endoscopy training. In the study group, a steady 90% CIR was reached at an average of 240 procedures, in contrast to 280 procedures in the historical cohort. This happened despite the fact that the study group had received less endoscopy training at baseline. Based on these results, we advocate the implementation of such regular standardized assessment with integrated feedback.

The baseline cecal intubation rate between the current study group and the historical cohort differed significantly, suggesting dissimilarities between the two groups in baseline endoscopy experience. This difference may be explained by the fact that our historical cohort consisted of trainees from a single center, who all followed an identical upper GI- and sigmoidoscopy training program before starting with performing colonoscopy at inclusion. Our study group consisted of trainees from different training centers, both tertiary and district hospitals and mostly without any endoscopic experience.

Recently published data give insight in the results of colonoscopy training courses in the USA and UK.[65,124] In the UK study, describing a cohort of 297 trainees, a CIR of 90% was found after 233 procedures using moving average analysis. The study from the Mayo clinic with 41 trainees, reported a CIR of 85% after an average of 275 procedures. These results were both achieved with intensive training programs: considerable investments and improvements of the training program in the UK and introduction of a skill assessment tool in the Mayo study. Although historical cohorts are lacking in these studies, these results also support an intensified training program.[124,127,128]

In this study, we analyzed the utility of video-assessment. Our results show that there were significant differences in the mean RAF scores given by the video-expert and supervisor. The supervisor tended to give higher scores to the trainee than the video-expert. This may be explained by a difference in the personal standard of scoring, but also by the lack of audio-feedback and the difference in observation position for the video-expert. More importantly, the scores given by the two experts correlated highly. These correlations support the use of video-assessment in a training setting. Assessing competency levels by more than one assessor strengthens the measurement and the generalizability of the conclusions.

Carrying out a study like this is challenging and inevitably leads to loss of data. With participants from different centers throughout the country, the planning of the various assessments required effort. Unexpected findings during colonoscopy leading to abortion of the procedure, and last-minute cancellations of colonoscopies by patients gave rise to further practical hurdles. As a result, some patient-based assessments were not carried out at the predetermined interval, or were limited to observation of a single colonoscopy. This is by definition related to the challenges of field-research, the effects were minimized by strict protocol adherence and effort from study team and trainees under study.

In accordance with previous studies, the results of this study underline the fact that the use of learning curves is a valuable tool to measure skill development in individuals as well as a group of trainees. Cecal intubation rate is a well-established marker of performance and the use of this marker as primary outcome is rational.[129,130] On top of the continuous assessment of the learning curve, a regular assessment program during colonoscopy training provides

trainees with concrete feedback and focus points for improvement of their colonoscopy skills. The positive experience of the trainees with this program underlines this even more.

Currently, we are working on an extensive output module for the electronic portfolio for trainees. With this module, real-time feedback on learning progress and performance can be given, while at the same time allowing a trainee to compare individual performance to the peer group. This reflection may enhance progression of the learning curve in itself.

Conclusions

This study showed that the learning curve of trainees for colonoscopy can be positively influenced by a standardized assessment program, where experts provide a colonoscopy trainee with concrete support for improvement of performance on top of standardized self-assessment.

Chapter 6

Patient comfort and quality in colonoscopy

Vivian E. Ekkelenkamp, Kevin Dowler, Roland M. Valori, Paul Dunckley

World Journal of Gastroenterology 2013;19:2355-61.

ABSTRACT

Introduction: Cecal intubation rate (CIR), use of sedation and adenoma detection rate are key performance indicators for colonoscopy. CIR is the most widely recognised measure of performance. Patient comfort is not routinely assessed; it is unknown whether higher intubation rates are achieved at the expense of greater patient discomfort, deeper sedation and possibly higher risk. The aim was to explore the relationship of patient comfort and experience to commonly used performance indicators for colonoscopy.

Methods: All colonoscopies performed in our four endoscopy centers are recorded in two reporting systems that log key performance indicators. From 2008 to 2011 all procedures performed by qualified endoscopists were evaluated; procedures performed by trainees were excluded. The following variables were measured: CIR, nurse-reported comfort levels (NRCL) on a scale from 1 to 5, polyp detection rate (PDR), patient experience (PE) of the procedure (worse than expected, as expected, better than expected), and use of sedation and analgesia. Pearson's correlation coefficient was used to identify relationships between performance indicators.

Results: A total of 17,027 colonoscopies were performed by 23 independent endoscopists between 2008 and 2011. Cecal intubation rate varied from 79.0% to 97.8%, with 18 out of 23 endoscopists achieving a CIR of >90%. The percentage of patients experiencing significant discomfort during their procedure (defined as NRCL of 4 or 5) ranged from 3.9 to 19.2% with an average of 7.7%. CIR was negatively correlated with NRCL-45 ($r=-0.61$; $p<0.005$), and with poor patient experience ($r=-0.54$; $p<0.01$). The average dose of midazolam (mean 1.9 mg, with a range of 1.1 to 3.5 mg) given by the endoscopist was negatively correlated with CIR ($r=-0.59$; $p<0.01$). CIR was positively correlated with PDR ($r=0.44$; $p<0.05$), and with the numbers of procedures performed by the endoscopists ($r=0.64$; $p<0.01$).

Conclusions: The best colonoscopists have a higher CIR, use less sedation, cause less discomfort and find more polyps. Measuring patient comfort is valuable in monitoring performance.

BACKGROUND

Colonoscopy is a very common procedure performed to investigate colonic symptoms and screen for cancer and polyps.[131] It has always been known that colonoscopy can cause harm and even death, but poor quality colonoscopy has only been linked to other important outcomes in the last decade. Back-to-back colonoscopies identified important missed lesions[132], fast withdrawal times were associated with lower adenoma detection rates[133,134], and low adenoma detection rates are associated with higher rates of missed cancer.[100] Several studies have shown that colonoscopy misses, and fails to 'protect' individuals from, cancer.[135-139] Thus there has been increasing attention on the quality of colonoscopy[140,141], especially in the context of colorectal cancer screening where there is potential for causing harm to otherwise healthy people.

In order to assess quality, the British Society of Gastroenterology (BSG) has defined a set of indicators and auditable outcomes for colonoscopy.[130] Important key performance indicators are an unadjusted cecal intubation rate (CIR) of >90% and an adenoma detection rate of > 10%. CIR is globally recognised as the main measure of competence in colonoscopy in a non-screening setting and is one of the key measures used in a colorectal cancer screening. It is an absolute requirement for total colonoscopy, and poor completion rates may be one reason why colonoscopy does not prevent cancer in the right colon.[142-144] However, there are several factors that can influence the CIR and thus the performance of an endoscopist.[145]

A possible consequence of having CIR as a prime indicator of quality is that individuals with poor technique may push harder and persist for longer to achieve the standard. This could lead to more pain and the administration of more sedation. Clearly this could cause unnecessary harm to patients, including more perforations and sedation related complications.[146]

To prevent this eventuality the BSG proposed that other key performance indicators should be sedation and comfort.[130] Standards were set for sedation, particularly for older patients, but there is no standard for comfort so it was designated an essential 'auditable outcome': a standard that should be measured, reviewed and acted upon, but not one for which an absolute performance level could be defined.

Various studies have addressed patient pain or discomfort during colonoscopy, and identified predictive factors of pain.[147-150] However, none have explored the use of sedation and patient comfort as measures of performance.

This study aims to analyse the different factors affecting an individual's performance in diagnostic colonoscopy and to explore the use of patient comfort scores as performance indicators for colonoscopy.

METHODS

All colonoscopies performed in the four endoscopy units in one healthcare organisation are recorded on two electronic endoscopy reporting systems (SQL scope and Unisoft), which log the key performance indicators defined by the BSG: CIR; polyp detection rate (PDR) (adenomatous and hyperplastic); and sedation (invariably opiates and midazolam). Colonoscopies performed by all independently practicing endoscopists during the four year period of 2008 to 2011 were included in the analysis. Throughout the UK (and in this study) an unadjusted CIR is used: the rate is not adjusted at all, even for obstructions and poor bowel preparation.

Table 1. Five-point scale of nurse-reported comfort levels.

Nurse reported comfort levels		Descriptors
1	No discomfort	Talking/comfortable throughout
2	Minimal discomfort	1 or 2 episodes of mild discomfort with no distress
3	Mild discomfort	More than 2 episodes of discomfort without distress
4	Moderate discomfort	Significant discomfort experienced several times with some distress
5	Severe discomfort	Frequent discomfort with significant distress

Table 2. Data completeness on colonoscopies performed from 2008-2011.

Variable	Total number of colonoscopies with missing data	% of colonoscopies with complete data
CIR	0	100
NRCL	520	95
PE	1647	84
Midazolam	62	99
Opiates	65	99
Polyp detection	3863	71

Comfort is assessed using nurse-reported comfort levels (NRCL) on a 5-point scale, which is shown in Table 1. The attending endoscopy nurses assess the comfort of the patient during the procedure without discussing it with the endoscopist, and record it immediately. For this study, significant discomfort was defined as a NRCL of either level 4 or 5 (NRCL-45).

The patient experience (PE) is captured by the recovery nurse before the patient leaves the unit. Patients are asked whether their experience was: better than expected, as expected, or worse than expected. Both the comfort scores and the PE are recorded on the hospital administration system. The colonoscopists are identified in the reporting system so that all data can be linked to individuals.

The influence of midazolam and opiate analgesia on NCRL and worse patient experience (PE-W) was also explored. A further variable used in this analysis was PDR. The dataset for PDR was less complete as our endoscopic reporting systems did not mandate the input of PDR until September 2010.

A complete dataset was not available for all variables. Table 2 lists the numbers of colonoscopies where data was not documented.

Statistical analysis

Relationships of CIR to comfort (NRCL-45), sedation and PE-W were explored using Pearson's correlation coefficient. The relationship between the number of procedures performed per year and CIR was also studied using Pearson's correlation coefficient. Only endoscopists performing colonoscopies for the full four year period were included in this analysis. A Mann-Whitney U test was used to assess whether there was a difference in the number of colonoscopies performed by those with a higher CIR.

RESULTS

During the four year period from 1 January 2008 to 28 December 2011, 17,027 colonoscopies were performed by 23 colonoscopists; 88.8% of procedures were performed on service lists; 11.2% of procedures were performed on bowel cancer screening lists. Data is reported as performance data for these colonoscopists.

Colonoscopy completion

CIR varied from 79.0% to 97.8%, with 18 out of 23 endoscopists achieving >90%. Four endoscopists completed colonoscopy in 85-89% of the procedures and 1 locum endoscopist in 79%. The effect of the number of colonoscopies performed on CIR was studied. Only endoscopists performing colonoscopy during the whole period were included in this analysis alone (n=16). CIR was positively correlated with the average number of procedures performed per annum ($r=0.64$; $p<0.01$) (Figure 1a). The average CIR for these 16 endoscopists was 94.3%. Endoscopists with a CIR of less than 94.3% performed an average of 139.9 colonoscopies per year whereas those with a CIR of greater than 94.3% performed an average of 245.9 procedures ($p<0.05$).

Patient comfort

The percentage of patients experiencing significant discomfort during their procedure (defined as NRCL of 4 or 5) ranged from 3.9 to 19.2% with an average of 7.7%. There was significant negative correlation between NRCL-45 and CIR ($r=-0.61$; $p<0.005$) (Figure 1b).

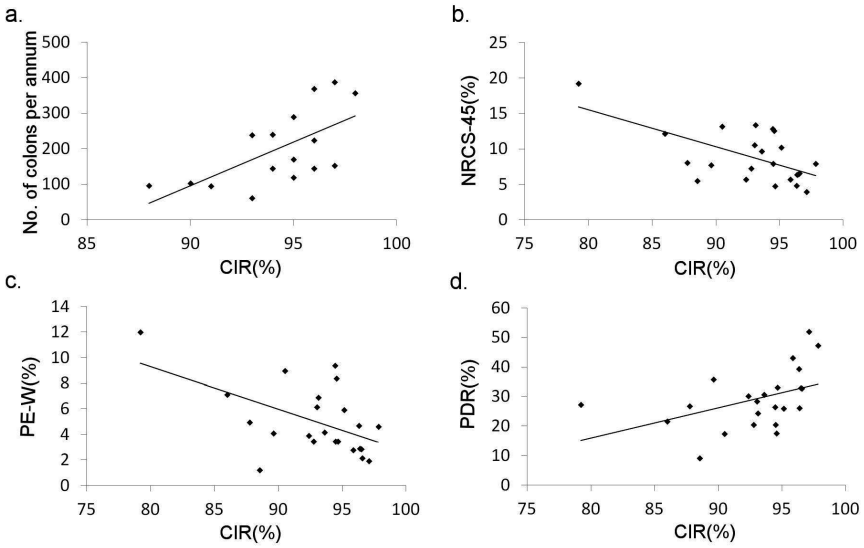


Figure 1. The figure shows correlations of cecal intubation rate (CIR) with number of annual colonoscopies (a); nurse-reported comfort level of 4-5 (b); patient experience worse than expected (c) and polyp detection rate (d).

Patient experience

A worse than expected patient experience (PE-W) was recorded in 4.3% of procedures (1.2-12.0%). PE-W correlated negatively with CIR (r=-0.54; p<0.01) (figure 1c). There was strong correlation between NRCL-45 and PE-W (r=0.92, p<0.0001). Only 2% of patients with a NRCL of 1, 2 or 3 rated the procedure as worse than expected compared to 28% of patients with a NRCL of 4 or 5.

Sedation

The sedation used in our endoscopy units for colonoscopy is usually a combination of an opiate (either pethidine or fentanyl) and midazolam. An increasing proportion of procedures is done without sedation.

The average amount of midazolam used per procedure was 1.9 mg, varying from 1.1 mg to 3.5 mg. Average dose of midazolam was negatively correlated with CIR (r=-0.59; p<0.01). To assess whether this was due to higher doses of midazolam being used by colonoscopists with worse CIRs or to a higher rate of no sedation being used by those with better CIRs, the analysis was repeated for the sedated colonoscopies only. In this sedated group (n=14870) there was a significant correlation between average midazolam usage and CIR (r=-0.60, p<0.005). The percentage of colonoscopies performed without sedation was not significantly correlated with CIR (r=0.30, p=0.13). There was also a correlation between midazolam dose and NRCL-45 (r=0.54, p<0.01) but not for midazolam and PE-W (r=0.37, p=0.08). In unsedated patients, there was no correlation between CIR with either NRCL-45 (r=-0.09, p>0.05) or PE-W (r=-0.01, p>0.05). However, the numbers were smaller in this group, especially for colonoscopists who rarely performed colonoscopy without sedation. Furthermore, the more uncomfortable procedures would have led to patients being given sedation thereby introducing bias.

There were 4 endoscopists who used fentanyl and 19 who used pethidine as their opiate of preference. To ensure uniformity, the endoscopists using fentanyl were excluded from the analysis on analgesia. There was no significant correlation between average pethidine dose, and CIR (r=-0.39, p>0.05), NRCL-45 (r=0.17 p>0.05) or PE-W (r=0.06, p>0.05).

Polyp detection

In this study, the average PDR (including both hyperplastic and adenomatous polyps) was 31.8% (range 9.2-51.9%). There was a positive correlation between PDR and CIR (r=0.44; p<0.05) (Figure 1d).

Performance indicators over time

Table 3 shows data on the CIR, NRCL-45, PE-W, midazolam usage and PDR for each year. A consistent improvement is seen in all variables between 2008 and 2011.

Table 3. Improvements in key performance indicators between 2008-2011.

	CIR (%)	NRCL-45 (%)	PE-W (%)	Midazolam (mean dose - mg)	PDR (%)
2008	93.3	10.0	5.6	2.3	29.6
2009	93.4	7.8	4.2	2.0	27.4
2010	94.6	7.6	4.1	1.8	31.9
2011	95.9	5.8	3.7	1.7	37.7

DISCUSSION

In this study we explored factors that predict high performance in colonoscopy. Ideally a colonoscopy should be safe, complete and comfortable. It should also detect and remove safely and completely all important lesions. The CIR has become the most universally recognised performance indicator. While striving to achieve and exceed target CIRs there is a potential danger that a colonoscopist will cause more discomfort, or put the patient at risk of perforation and excessive sedation. The results of this study indicate the reverse: those colonoscopists with the highest CIR use less sedation, cause less discomfort and achieve a better patient experience. Furthermore, it appears they are more vigilant, identifying more polyps than those with lower intubation rates. The results also show that better colonoscopists perform more colonoscopies. In this study, colonoscopists with a CIR of greater than 94.3% performed an average of 245.9 procedures per annum compared with 139.9 for the endoscopists with a CIR lower than 94.3%. This is consistent with previously published data.[151] This study adds further weight to the argument that there should be a minimum number of procedures performed by an endoscopist per annum to maintain their skills.

There are very large variations in the use of sedation across the world ranging from virtually none in Scandinavian countries to increasing use of deep sedation with propofol in Australia, France, Germany and the USA. The use of sedation is still not as safe as we would like.[152] In the USA, it is now common to perform a colonoscopy with propofol and it has been shown that patient satisfaction is higher than with other types of sedation.[153,154] Conversely, a Scandinavian study showed that high sedation rates were not associated with less painful colonoscopies.[149] Another Scandinavian group showed that sedation is not necessary for screening individuals, and an American group clearly believes unsedated colonoscopy has a place and has coined the phrase 'sedation-risk-free colonoscopy'.[155]

In our study, the average midazolam dose used was negatively correlated with CIR: the more often the cecum was reached, the less midazolam was used and, furthermore, patients did not experience more discomfort. These findings demonstrate that colonoscopy can be performed without deep sedation and without significant discomfort in the majority of patients.

Sedation alters the perception and recollection of discomfort experienced during colonoscopy. Thus the patient cannot necessarily provide an accurate guide of pain during the procedure. An alternative to the patient assessing discomfort is for the endoscopist or endoscopy nurse to make the assessment. We ask the nurse to make this assessment because they are more likely to be objective and have the benefit of observing all colonoscopists perform colonoscopy. Our comfort scale has not been formally validated but it assesses three components of discomfort: severity, frequency and the extent to which it is distressing the patient. Interestingly there was strong correlation of this nurse-assessed scale with patient reports ($r=0.92$, $p<0.0001$). Only 2% of patients with a NRCL of 1, 2 or 3 rated the experience as worse than expected. It is likely that different nurses rate discomfort differently but that discrepancy would be applied to all colonoscopists. There are always two nurses in the procedure room during a colonoscopy and the nurses are encouraged to discuss the comfort score with each other before making a final decision.

The assessment of patient experience is different from that of discomfort by a health professional. Because of the effect of sedation on experience and recall, we chose not to ask patients to rate comfort but to rate their experience of the procedure compared to what they expected. This measure was chosen on the assumption that a worse experience than

expected was unacceptable and a better or as expected experience was acceptable. Clearly a patient's rating will be affected by the way they are prepared for the procedure and hearsay. It is possible that the patients of a colonoscopist who routinely tells them that they will experience terrible pain will rarely report the experience worse than expected. We cannot control or assess this possibility. It seems very unlikely that the colonoscopists with high CIR tell their patients that they will have a bad experience when the nurses rate them as causing less pain than their colleagues.

Sedation practice varies but the majority use a combination of opiates and sedatives, and an increasing number use no sedation. It is therefore difficult to make meaningful comparisons. However, whichever way the data was examined the same conclusion was drawn: colonoscopists with high CIR use less sedation (midazolam). One argument against using CIR (especially an unadjusted rate) as a performance indicator is that endoscopists may use excessive force to ensure that the cecum is intubated. However, data from this study shows that comfort scores were better in colonoscopists with a higher CIR and there was no evidence that they were using more opiate analgesia.

A possible bias in this study is case mix. It is possible that the colonoscopists with the highest CIR were colonoscoping the easiest patients. Previous studies have identified factors that predict lower CIR: female sex, older patient and the presence of diverticular disease.[129,147] Until recently our reporting system was not capturing diagnoses according to a recognised coding system so it is not possible to determine the proportion of patients with diverticular disease in each of the colonoscopist cohorts. About 30% of patients listed for colonoscopy are pooled and listed with the endoscopist that is first available. This sharing of patients reduces the likelihood that an individual will be scoping a particularly difficult group of patients. Furthermore, colonoscopists with a higher CIR are often asked to scope 'difficult' patients meaning case mix is more likely to affect them adversely. Another possible source of case mix bias is bowel cancer screening (FOBT positive) patients because only accredited colonoscopists are allowed to colonoscope them. These patients are usually asymptomatic and may therefore be easier to colonoscope; there is however no data available on this topic. They certainly have more polyps than other patients, which may bias polyp detection data. Whilst only 10% of all colonoscopies are performed on screen positive patients, up to 50% of the procedures performed by the bowel cancer screening colonoscopists are on screened patients. However, only 2 of the 23 colonoscopists for the majority of the study period were screening accredited and several of the high performing (high CIR, low sedation, low discomfort) colonoscopists were not screening colonoscopists. Another possible confounder is the use of unadjusted CIR instead of the CIR being adjusted for poor bowel preparation or obstruction. CIR would invariably have been higher if adjusted. We chose to use unadjusted CIR as this is standard practice in the UK for quality assessment. The number of cases with poor bowel preparation or obstruction was probably low and there is no reason to believe that one endoscopist was exposed to all those cases especially as the bowel preparation was standardised across all four units. Therefore, we feel that it is unlikely that the use of adjusted CIR would influence the main findings in this study.

Adenoma detection rate is a key performance indicator and has been shown to be related to the chance of post colonoscopy colorectal cancer.[100] Ideally, adenoma detection rate should be recorded but linking endoscopic with pathology databases is difficult, and late entry of pathology data into an endoscopic database is fraught with problems. In view of this difficulty, we have used polyp rather than adenoma detection in this study whilst recognising the limita-

tions of this approach. However, recent studies have shown that PDR can be used as a marker for ADR because they are highly correlated.[156, 157] A recent study of colonoscopies performed on the UK Bowel Cancer Screening programme also found a positive correlation between adenoma detection rate and cecal intubation rate.[158]

In each of the endoscopy units included in this study there is a robust quality assurance process for colonoscopy. All colonoscopists are fed back their performance indicators on a quarterly basis. If any colonoscopist underperforms, the endoscopy lead will discuss this with them and, if appropriate, offer further support and training. Furthermore, most of the colonoscopists in this study have completed a training the trainer course during which there is detailed discussion of colonoscopy technique and ways to improve it. These approaches are likely to have contributed to the consistent improvements in CIR, patient comfort/experience and PDR. One aspect of quality assurance we did not address in this study is occurrence of complications in colonoscopy. Our study explores the intubation performance, not performance of therapy. There were no diagnostic perforations during the period of this study and no procedure related deaths. Literature tells us that less than 1:1000 patients will suffer from a complication of colonoscopy without biopsies or polypectomy.[159] A much larger sample size would be required to test the relationship of key performance indicators and complication rates.

In conclusion, this study demonstrates that the best colonoscopists are doing more colonoscopies per year, get to the cecum more often, use less sedation, cause less discomfort, achieve a better patient experience and find more polyps. We believe that measurement of patient comfort and experience, use of sedation, together with CIR, could provide a richer picture of a colonoscopist's performance, at least of intubation skills.

Take home message

This study shows that the best colonoscopists, i.e. the ones that have the highest CIR and PDR, also have the best comfort scores, despite using less sedation. Measurement of patient comfort during sedated or non-sedated colonoscopy may provide useful information on endoscopist performance.

Section III

Learning curves and quality assessment of ERCP

Chapter 7

Competence development in ERCP – the learning curve of novice trainees

Vivian E. Ekkelenkamp, Arjun D. Koch, Erik. A.J. Rauws, Gerard J.J.M. Borsboom, Robert A. de Man, Ernst J. Kuipers

Endoscopy 2014; in press.

ABSTRACT

Introduction: Measures for competence in ERCP during training are poorly defined. Currently, various training and accreditation programs base competence on a minimum number of procedures. There is a general awareness that procedural competence certification should be based on objective performance criteria. Continuous self-assessment using a Rotterdam Assessment Form for ERCP (RAF-E) can provide insight in performance of trainees. The aim of this study was to express competence development in ERCP as a learning curve.

Methods: Trainees in ERCP in a tertiary referral center were invited to participate. All procedures were appraised using RAF-E. The indication for each ERCP was classified, as well as presence of a virgin papilla. Complexity was graded on a 3-point scale. The primary outcome parameter was common bile duct (CBD) cannulation success rate. Success of the intended therapeutic interventions was additionally expressed as a learning curve.

Results: Fifteen trainees from one tertiary medical center in the Netherlands were included. 1541 ERCPs (624 procedures in native papillary anatomy) were assessed through RAF-E. Unassisted CBD cannulation success rate improved from 36% at baseline to 85% after 200 procedures ($p<0.001$). CBD cannulation success in 624 patients with a virgin papilla improved from 22% at baseline to 68% after 180 procedures ($p<0.001$). Learning curves for therapeutic interventions showed significant improvements for successful sphincterotomy and stent placement ($p=0.01$ and $p<0.001$, respectively).

Conclusions: Learning curves are a valuable means to assess competence in ERCP. Differences in learning curves can be shown with RAF-E. Competence should be based on actual performance, instead of minimum numbers.

INTRODUCTION

Endoscopic retrograde cholangiopancreatography (ERCP) is considered to be one of the most challenging procedures in gastrointestinal endoscopy. In order to reach competence in performing this procedure, extensive training is necessary. Trainees differ considerably in the rate with which they acquire their endoscopy skills.[60] Moreover, methods to measure ERCP competence are lacking and poorly defined. At present, competence is merely assumed on a minimum number of ERCP procedures, mostly ranging in various training programs from 100 to 200. The ASGE recognized that a number of 180-200 ERCPs were required for competence, based on the first study regarding this subject.[60] This important study used a composite endpoint for procedure success, and evaluated a relatively small number of procedures per trainees (average 85), with only 3 trainees reaching the level of 180 procedures. The evaluation of trainee performance was carried out by trainers. Additional data are scarce and threshold numbers do not necessarily reflect competence. Various guidelines state that competence should be established by objective performance criteria.[160-162] Outcome measures such as common bile duct (CBD) cannulation rates and therapeutic success give insight in performance. The ASGE guidelines state that a trainee should be able to reach an 80-85% CBD cannulation success rate at the end of ERCP training.[161] There is however no standardized method to assess such criteria.

Continuous self-assessment may be one way to gain insight in performance of trainees. This method has already proven its value for competence measurement in colonoscopy. We previously developed and assessed the Rotterdam Assessment Form for Colonoscopy (RAF-C), with which learning curves can be plotted and individual as well as group performance analyzed and followed.[61] A similar method can be used for competence assessment in ERCP.

The aim of this study was to express competence development as a learning curve for different types of procedures and to assess the performance of residents starting with ERCP training.

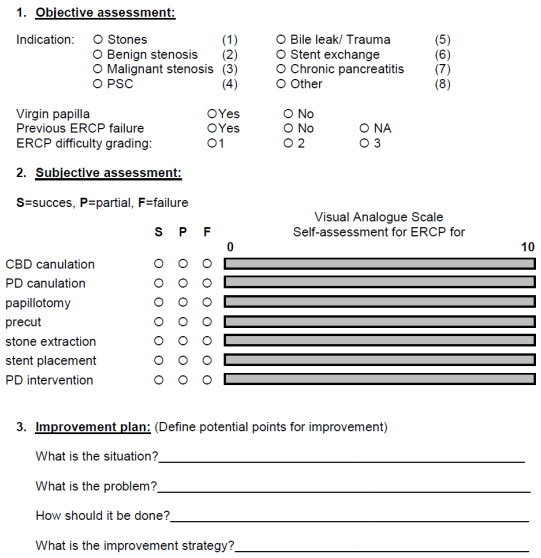
METHODS

Program

From January 2008 to March 2013 we performed a prospective evaluation of individual and group ERCP performance in our academic medical center. The gastroenterology training program in the Netherlands starts with a two-year residency in internal medicine followed by a four-year program in gastroenterology. Trainees participate in endoscopy throughout these four years. After reaching competence in upper GI endoscopy and colonoscopy, residents start with ERCP training. All successive trainees performing ERCP in our department were included in this study. They were in year four to six of their training and had reached competence in basic endoscopic procedures. Trainees were supposed to complete a newly developed self-assessment form after every ERCP.

Self-assessment form

For this program, the Rotterdam Assessment Form for ERCP (RAF-E) was used. This form is in itself an unvalidated assessment instrument, but it is partially based on previously validated assessment tools.[52,61,125,163] Also, it has been used to assess the performance of experienced


1. **Objective assessment:**

Indication: ○ Stones (1) ○ Bile leak/ Trauma (5)
○ Benign stenosis (2) ○ Stent exchange (6)
○ Malignant stenosis (3) ○ Chronic pancreatitis (7)
○ PSC (4) ○ Other (8)

Virgin papilla ○Yes ○ No
Previous ERCP failure ○Yes ○ No ○ NA
ERCP difficulty grading: ○1 ○ 2 ○ 3

2. **Subjective assessment:**

S=succes, **P**=partial, **F**=failure

	S	P	F	Visual Analogue Scale Self-assessment for ERCP for 0 – 10
CBD canulation	○	○	○	
PD canulation	○	○	○	
papillotomy	○	○	○	
precut	○	○	○	
stone extraction	○	○	○	
stent placement	○	○	○	
PD intervention	○	○	○	

3. **Improvement plan:** (Define potential points for improvement)

What is the situation?_____

What is the problem?_____

How should it be done?_____

What is the improvement strategy?_____

Figure 1. The Rotterdam Assessment Form for ERCP (RAF-E).

endoscopists.[164] The assessed items in the RAF-E comprise the proposed quality indicators for ERCP.[161] Figure 1 shows the form, which consists of three parts. The first part covers objective parameters such as procedural indication, technical difficulty degree based on Schutz's classification and previous ERCP failure.[165] Table 1 shows this modified classification. The second part of the form contains success or failure options for different parts of the procedure such as cannulation of the CBD or pancreatic duct (PD), stent placement, sphincterotomy or stone extraction. These parameters can be scored as successfully completed, partial success, or failure. Regarding cannulation, successful completion was defined as deep cannulation of the desired duct, partial success was defined as opacification of the duct or

Table 1. Degrees of difficulty based on Schutz's classification.

Difficulty degree	Biliary procedures	Pancreatic procedures
Grade 1	• Diagnostic cholangiography • Biliary cytology • Stone extraction ≤ 10 mm • Dilatation of stenosis/ stent placement/ nasobiliary drain in extrahepatic strictures	• Diagnostic pancreatography • Pancreatic cytology
Grade 2	• Stone extraction > 10 mm • Dilatation of stenosis/ stent placement/ nasobiliary drain in hilar tumours or benign intrahepatic strictures	• Cannulation of papilla minor
Grade 3	• BII anatomy • Intrahepatic stone extraction • Stone extraction with lithotrypsy	• Therapeutic pancreatic procedures including pseudocyst drainage

passing of a guidewire, but without deep cannulation with a catheter. Partial successful stenting was defined as achievement of partial or incomplete drainage of the desired segments. However, in the analysis all partially successful aspects of the procedure were considered to be failures, in order to avoid any discussion on definitions of partial success. These various aspects of the procedure are followed by the request to complete an improvement plan after every 10 procedures with a four-step approach. This is based on the Osborn-Parnes Creative Problem Solving Process developed in the 1950s.[166] An example of such a detailed plan would include detailed address to questions on problem identification, solutions, and improvement strategy. The subjective scores were not taken into account in the statistical analysis, nor the outcome of the improvement plan. The value of the subjective assessment was in creating self-awareness to enable reflection on performance rather than in providing evidence for quality measurements.

Trainees only registered the parts of the procedure where they were actively involved.

Patients

All ERCPs performed in this study were part of the regular training program, supervised by a staff endoscopist. Patients were referred for ERCP for a broad range of indications such as gall stones, benign or malignant strictures and stent placement. Patients were routinely sedated using fentanyl and midazolam.

Main outcome measures

The primary outcome measure was the rate of successful common bile duct cannulation. This parameter was used to create individual and group learning curves. Success of therapeutic interventions such as stone extraction or stent placement was calculated as well. Since all procedures performed during a training period have an effect on the learning process, we firstly evaluated all procedures combined per endoscopist, irrespective of their degree of technical difficulty. If the supervisor had to intervene in the procedure for any reason, the part of intervention counted as failure, and the further part of the procedure was considered not applicable for the trainee. When CBD cannulation was successfully performed by the trainee, and stone extraction failed, only this last part counted as a failure. Procedures performed in patients with a virgin papilla were analyzed separately regarding CBD cannulation, since this is the step that an ERCP performing endoscopist in an average practice should be able to achieve.

Statistical analysis

For all analyses, we used binary outcome measures (success vs. failure). Partial success was regarded as failure. We plotted the probability of a successful common bile duct cannulation against number of ERCPs with 95% CI, obtained from a generalized linear mixed model (logistic regression for longitudinal data) and we analyzed the significance of the increment in the learning curve this way. The probability was modeled with a cubic spline having two interior knots equally spaced over the available range of ERCP numbers. Correlations between the repeated measurements of the individual endoscopists were accounted for by including a random intercept term for the endoscopists in these models. Similar analyses were performed for therapeutic interventions such as stone extraction and stent placement. Differences in proportions were analyzed using X^2 tests. A two-sided p-value of <0.05 was considered significant. Analyses were carried out in SAS version 9.2.

RESULTS

Fifteen trainees were included in this study. From January 2008 to March 2013 they filled out 1541 RAF-E's. Three trainees already started with ERCP training before January 2008, but the number of ERCPs performed before participating in this study was taken into account in the analysis. The median number of RAF-E's filled out per trainee was 90 (range 20-218). Adherence to completion of the forms was 82.9%. Table 2 gives an overview of the different indications for which the ERCPs were performed. Table 3 shows the distribution of the procedures per difficulty degree (this information was available for 98.1% of the procedures).

Table 2. Overview of indications.

Indication	Number of procedures (%) (n = 1541)
Stones	361 (23.5)
Stenosis (benign or malignant)	730 (47.6)
Primary sclerosing cholangitis	63 (4.1)
Biliary leakage or trauma	71 (4.6)
Endoprosthesis change	43 (2.8)
Chronic pancreatitis	172 (11.2)
Other	94 (6.1)

Table 3. Distribution of ERCPs per difficulty degree.

Difficulty degree	Number of procedures (%) (n = 1511)
1	1029 (68.1)
2	222 (14.7)
3	260 (17.2)

Learning curves

During the assessment period of each individual endoscopist, the overall CBD cannulation success rate gradually increased with the amount of ERCPs performed (Figure 2a). This increment in the learning curve was significant ($p<0.001$). We analyzed differences in individual performance as well. One trainee who performed under average achieved a CBD cannulation success rate of 60% after 100 procedures, while another who performed above the group curve, reached a success percentage of 84% after 100 ERCPs (X^2 test; $p<0.005$). Figure 2b provides insight in the learning curve of two other randomly selected trainees plotted against the group average.

We analyzed CBD cannulation success in patients with a virgin papilla separately. The number of ERCPs performed in patients with native papillary anatomy was 624 (40.5%). Table 4 gives an overview of the CBD cannulation success rate per block of 20 ERCPs in patients with native papillary anatomy. The obtained learning curve in these patients is shown in Figure 3. The improvement over time was significant ($p<0.001$). However, the success rates for CBD cannulation in this subgroup of patients were remarkably lower than in all patients together.

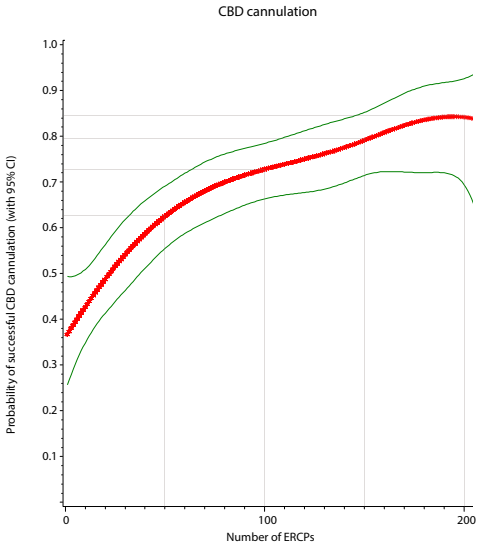


Figure 2a. Probability of achieving a successful cannulation of the CBD plotted against the number of ERCPs performed. The figure shows the group curve with a 95% CI (red and green lines).

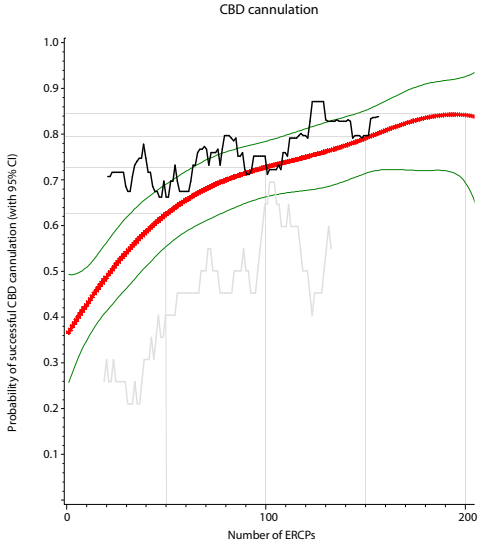


Figure 2b. Learning curves of two randomly selected trainees plotted against the group curve. This provides insight in individual versus group performance.

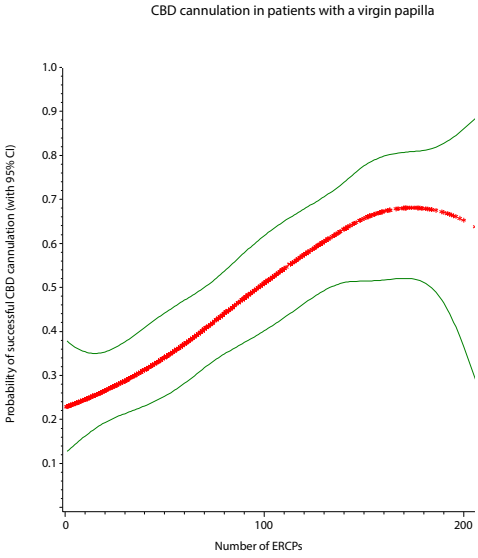


Figure 3. Probability of a successful CBD cannulation in patients with a virgin papilla.

Table 4. CBD cannulation success in patients with a virgin papilla per 20 procedures (n = 624 ERCPs).

Number of ERCPs performed	N (trainees)	Successful CBD cannulation (%)	Range (%)
1 to 20	12	26.7	0 – 40.0
21 to 40	11	23.4	0 – 50.0
41 to 60	11	30.0	0 – 60.0
61 to 80	10	36.0	0 – 60.0
81 to 100	8	52.6	0 – 70.0
101 to 120	8	57.0	20.0 – 80.0
121 to 140	4	56.5	25.0 – 80.0
141 to 160	4	58.6	33.0 – 86.0
161 to 180	2	60.7	50.0 – 71.0
181 to 200	1	87.5	n/a

The learning curves for therapeutic interventions were addressed as well. These are shown in Figure 4a, b and c. The increment in the learning curve for sphincterotomy and stent placement was also significant (p=0.01 and p<0.001, respectively). The success of stone extraction did not significantly increase over time (p=0.44).

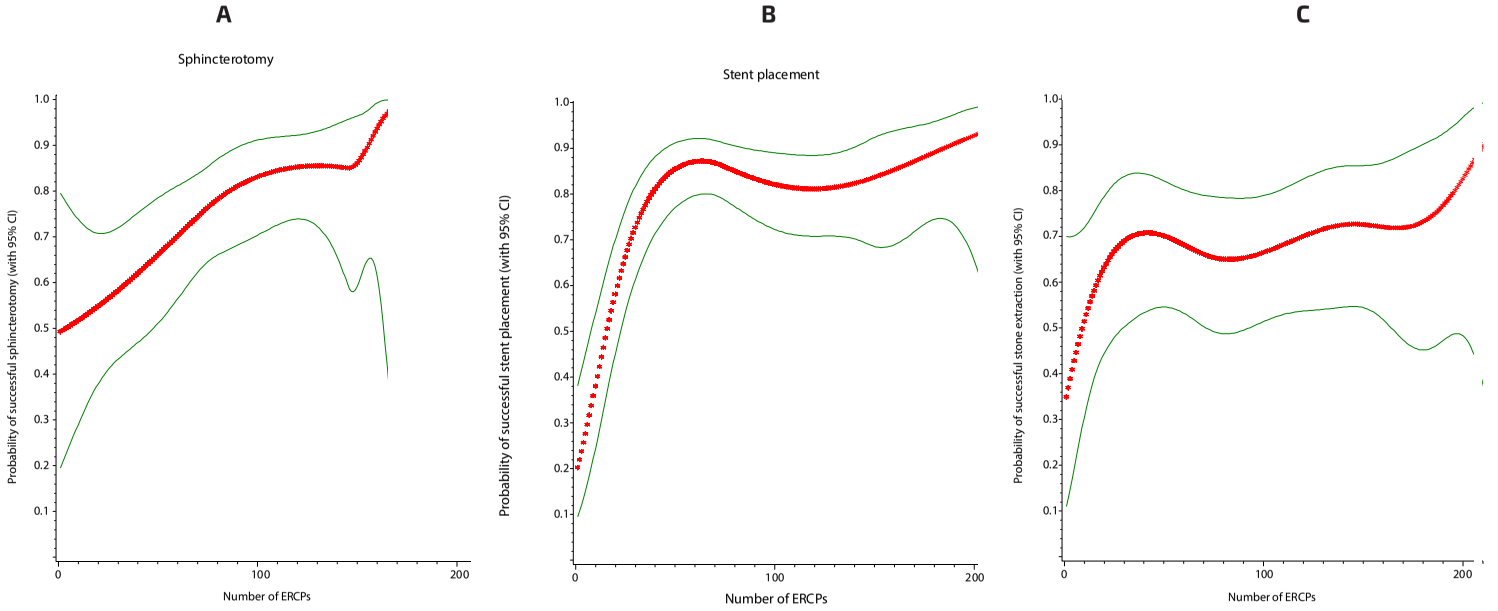


Figure 4a,b,c. Learning curves for therapeutic success. 4a shows the curve for sphincterotomy, 4b for stone extraction and 4c for stenting. The x-axis represents the number of ERCPs performed, the y-axis is the probability of achieving a successful intervention.

DISCUSSION

In this prospective study, we evaluated the performance of gastroenterology residents starting with ERCP training using a self-assessment form. With this relatively simple method we were able to provide insight in the learning curve of trainees for common bile duct cannulation as well as further outcome parameters for ERCP such as success of therapy. The self-assessment form allowed to determine the progress of every individual trainee on these different aspects of the procedure, but simultaneously provided insight in average group progression.

In this study, all trainees followed the same program regarding endoscopy training. All trainees started with a basic flexible endoscopy course, which is a theoretical and hands-on training for EGD (esophagogastroduodenoscopy) and colonoscopy. Hereafter, the following endoscopy training consisted of performing procedures on lists of EGD and sigmoidoscopy. After this period, trainees started performing colonoscopies. When they were considered competent in these 'basic' procedures after extensive exposure, they entered the ERCP training program.

Up until now, as in most countries worldwide, training and certification for ERCP in the Netherlands is based on a minimum number of procedures performed. There is however growing awareness that unvalidated threshold numbers are an inadequate method to determine competence and reach certification. The assessment of procedural competence by using individual learning curves is likely a more thorough method. Regarding procedural competence, the ASGE recommends a CBD cannulation success percentage of 80-85% after completing ERCP training.[161] This number has no real scientific basis but seems intuitively right.[167] In our study, trainees did not reach this success percentage after the 100 ERCPs required for certification in the Netherlands. However, since performance obviously varies per trainee, individual assessment of performance for defining competence is more robust than minimum numbers. The Rotterdam Assessment Form for ERCP is useful for monitoring competence development and can show at what point objective performance standards are being met.

Only a few studies have addressed development of technical competence in ERCP.[68,71,168-170] Our results showed a CBD cannulation success rate of >80% after approximately 160 procedures; other numbers range from a 85% CBD cannulation success after 148 ERCPs to 80% success after 300-400 procedures.[68,169]

The study of Jowell et al. was the first to provide insight in the learning curve of trainees regarding ERCP.[170] Based on this study, several guidelines were adapted regarding training and threshold numbers. There are nonetheless significant differences between this study and ours, with as main dissimilarity the involvement of trainers in performance evaluation as opposed to self-assessment and development of learning curves in our study. In current training, we feel that individual assessment of performance by learning curves, as shown in our study, is much more robust and relevant for the individual trainee than simple threshold numbers. This can easily be monitored through self-assessment and we were able to follow quite a number of trainees from the beginning of their training. Since adequate exposure is essential for becoming competent as well, we would advocate to base competence on both numbers and success.

CBD cannulation can be regarded as a surrogate marker for procedural competence. The curve of CBD cannulation was comparable with those of therapeutic interventions such as stent placement and sphincterotomy. This means that the learning curve for overall CBD

cannulation success can be used to provide a quick insight on whether an individual trainee is progressing according to the expected group learning curve or not. It will, to some extent, reflect the same learning curve on therapeutic interventions. Stent placement showed an 84% success rate after 160 ERCPs. This curve revealed a slight decline around this number of procedures. The learning curve for stone extraction did not significantly increase over time. After approximately 160 ERCPs performed, stone extraction was successful in 72% of the patients. We can only speculate about reasons for this flattening of the learning curve, especially around 100 ERCPs. The curve was created for all difficulty degrees together. The learning curve might have been better when only level 1 stone extractions were analyzed. However, numbers were too small to perform the analysis for this subgroup. Most learning curves show such a decrease at some point; this might be explained by the fact that trainees are getting overconfident and want to act more independently, with this slight decline in performance as a result. There is nonetheless no real scientific evidence for this theory.

An important remark is that the success rate of CBD cannulation in patients with a virgin papilla was remarkably lower than the overall success rate. This is not an unexpected finding: a previous successful sphincterotomy often makes cannulation much easier. The number of patients with a virgin papilla was smaller as well; that has to be taken into account when drawing conclusions. We do recognize that a successful CBD cannulation in a patient with a virgin papilla is one of the basic steps in ERCP that every endoscopist should be able to execute.

With continuous assessment using RAF-E, in the near future real-time feedback can be provided on a trainee's performance in comparison with peers. When a trainee then deviates from the group curve, an intervention, such as additional training, is possible. This system provides trainers solid grounds for assessment of performance.

When assessing procedural success in relation to numbers of procedures performed, the adherence to filling out the forms has to be taken into account. In this study, adherence to filling out the self-assessment forms was 82.9%. The median number of missing procedures per trainee was 17 (range 0-60). Based on previous studies and practical experience[61,164], these missing procedures are most likely random, due to lack of time and busy programs. The influence on the learning curve will therefore be minimal. This is also described in other studies on endoscopic learning curves, in which the missing data were thought to have no influence on the outcome.[61,65,170]

However, when we do take the missing procedures into account, it seems plausible that the endpoints of the learning curve, i.e. CBD cannulation success rate of 85%, will be reached after slightly more (n + 17) procedures. Considering the missing procedures as failures might be another way to deal with this limitation. Nonetheless, for the previous mentioned reasons, we regarded them as randomly missing. A different limitation of this study is that not all trainees get the opportunity to perform more than 100 ERCPs during formal training. The median number of RAF-E's filled out per trainee was 90. However, three trainees already started their ERCP training before participating in the study and did reach the threshold of >100 ERCPs. As a consequence, the learning curve after 100 ERCPs is based on smaller numbers. This limitation is a direct result of the present design of the training program for ERCPs in the Netherlands. Up until recently, most GI fellows were enrolled in the ERCP training program, which created a capacity problem. However, GI teaching programs are changing and shifting from a broad, relatively superficial training to specific focus areas such as interventional endoscopy or gastrointestinal oncology. Consequently, less residents will enter the ERCP

training program and, hopefully, those that are going to do ERCPs, will get the opportunity to be exposed to a sufficient number of procedures to gain competence.

Furthermore, complications were not taken into account in this study. The focus of this study was to gain insight in procedural competence. Nevertheless, in the end, overall outcome and patient safety is very important. In the Netherlands, there is already a compulsory complication registration. The relation between good performance and occurrence of complications is likely to be inverse. There are however no data yet to support this.

This self-assessment program was used as an addition next to the traditional master-apprentice model. This method is subject to the knowledge of the trainee and therefore might be biased. On the other hand, the trainer might have influence on the evaluation. Moreover, self-reflection is more likely to raise awareness of one's own performance and provides therefore an add-on effect in optimizing a learning process. There was no predetermined protocol on the involvement of the supervising attending. As pointed out, the self-assessment program was additional to the regular training program for ERCP. It is very difficult to implement a standardized protocol for the whole ERCP training program, which still for a great deal consists of the master-apprentice model. A strict protocol is therefore practically not feasible. Trainer assessment by the RAF-E was not part of this project, but forms an interesting follow up study. We would propose an extended form for trainers, with room for feedback. Furthermore, the comparison of subjective scores given by trainees and trainers may provide additional insight in the value of self-reflection. We evaluated the RAF-E in a post-training setting as well, where it has proven its value for assessment of performance of experienced endoscopists.[164]

In concordance with the American Society for Gastrointestinal Endoscopy, we believe that in the current era of training, competence should be based on learning curves instead of threshold numbers alone.[171] The Rotterdam Assessment Form for ERCP is an easy-to-use device to monitor competence development in ERCP in individual as well as group performance. This form is now available in an electronic portfolio for all gastroenterology residents in the Netherlands. We believe that with further development of the e-portfolio, where real-time feedback can be provided regarding own performance compared to the peer group, assessment of performance through RAF-E enhances quality in ERCP training.

Chapter 8

Quality evaluation through self-assessment: a novel method to gain insight into ERCP performance

Vivian E. Ekkelenkamp, Arjun D. Koch, Jelle Haringsma, Jan-Werner Poley, Henk R. van Buuren, Ernst J. Kuipers, Robert A. de Man

Frontline Gastroenterology, 2014;5:10-16.

ABSTRACT

Introduction: The American Society for Gastrointestinal Endoscopy (ASGE) Committee on Outcomes Research has recommended monitoring nine ERCP-specific quality indicators for quality assurance in ERCP. With the development of a self-assessment tool for ERCP (Rotterdam Assessment Form for ERCP - RAF-E), key indicators can easily be assessed. The aim of this study was to test in daily practice an easy-to-use form for assessment of procedural quality in ERCP and to determine ERCP quality outcomes in a tertiary referral hospital.

Methods: This was a prospective study carried out in a tertiary referral hospital. In January 2008 a quality self-assessment program was started. Five qualified endoscopists participated in this study. All ERCPs were appraised using RAF-E. Primary parameters were common bile duct (CBD) cannulation rate and procedural success. The indication was classified and procedural difficulty was graded; success rates of therapeutic interventions were measured for all different difficulty degrees.

Results: A total number of 1691 ERCPs were performed. 1515 (89.6%) of these were appraised using RAF-E. Median CBD cannulation success rate was 94.1%. Successful sphincterotomy was accomplished in almost all patients (median 100%; range 98.2-100%). Stent placement was successful in 97.8% and complete stone extraction, if indicated, was achieved in 86.8%.

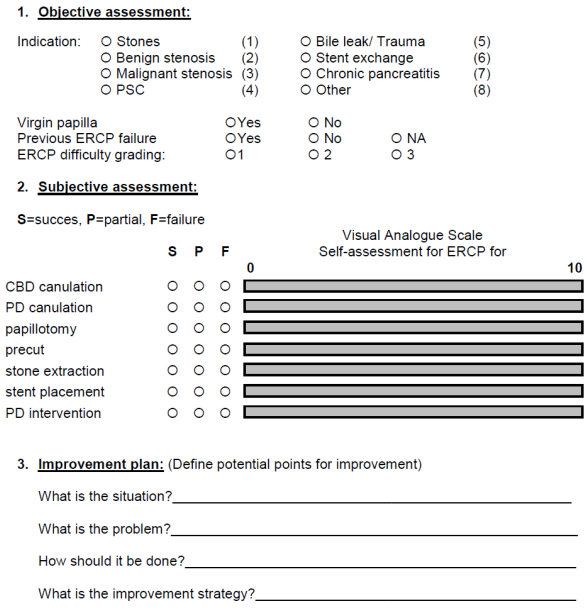
Conclusions: Quality indicators for ERCP can be measured using the Rotterdam self-assessment program for ERCP. Outcome data in ERCPs obtained with this RAF-E provide insight in the quality of individual as well as group performance and can be used to assess and set standards for quality control in ERCP.

BACKGROUND

In endoscopic procedures, the awareness on quality has increased over the last decade. Important landmarks are documenting whether the intended outcome of the procedure was achieved under the restriction of acceptable side effects, patient comfort and costs. Endoscopic retrograde cholangiopancreatography (ERCP) is a complex and challenging procedure for diagnosis and treatment of biliary and pancreatic disorders. It carries a relatively high risk of complications, and thus, assurance of quality is pivotal.[172-174]

However, monitoring and enhancing the quality of endoscopic procedures is impossible without methods to register specific parameters. Therefore, the American Society for Gastrointestinal Endoscopy (ASGE) Committee on Outcomes Research has defined a set of quality indicators for ERCP.[161] These quality indicators roughly fall in three separate domains: preprocedural quality indicators (appropriateness of indication), procedural quality indicators (i.e. common bile duct cannulation rates, therapeutic success), and post-procedural outcomes (i.e. documentation, complications). There have been several publications on complications as a marker of quality[175-179], but procedural success or patient-related outcomes are less often described.[126]

In addition, self-assessment of procedural performance in ERCP might enhance quality by stimulating active reflection on one's actions. Questions on topics such as individual versus group performance, personal performance level compared to that of colleagues, and how to learn from mistakes made, can be answered through proper self-assessment. Furthermore, specific objective outcomes such as CBD cannulation rate can easily be calculated. In order to gain insight in quality of ERCPs, especially focusing on endoscopists' performance, we developed a self-assessment program for ERCP. A similar method has proven its value for


1. **Objective assessment:**

Indication: ○ Stones (1) ○ Bile leak/ Trauma (5)
○ Benign stenosis (2) ○ Stent exchange (6)
○ Malignant stenosis (3) ○ Chronic pancreatitis (7)
○ PSC (4) ○ Other (8)

Virgin papilla ○Yes ○ No
Previous ERCP failure ○Yes ○ No ○ NA
ERCP difficulty grading: ○1 ○ 2 ○ 3

2. **Subjective assessment:**

S=succes, **P**=partial, **F**=failure

	S	P	F	Visual Analogue Scale Self-assessment for ERCP for 0 — 10
CBD canulation	○	○	○	
PD canulation	○	○	○	
papillotomy	○	○	○	
precut	○	○	○	
stone extraction	○	○	○	
stent placement	○	○	○	
PD intervention	○	○	○	

3. **Improvement plan:** (Define potential points for improvement)

What is the situation?_____

What is the problem?_____

How should it be done?_____

What is the improvement strategy?_____

Figure 1. The Rotterdam Assessment Form for ERCP.

colonoscopy.[61] Based on this study, as well as on previously validated assessment tools, the Rotterdam Assessment Form for ERCP (RAF-E) was developed (Figure 1).[52, 125, 163]

The aim of this study was to test in daily practice an easy-to-use form for assessment of procedural quality in ERCP. A major drawback of an easy-to-use form to be completed after every procedure without taking up too much time, is the number of parameters that can be measured. Secondly, we aimed to determine ERCP quality outcomes in a tertiary referral center.

METHODS

From January 2008 to December 2011, a prospective, single-center evaluation of ERCP performance was carried out in the Erasmus MC – University Medical Center in Rotterdam, the Netherlands. A quality assessment program was carried out through the use of self-assessment. All staff-members of the department performing ERCP participated. All ERCP procedures performed by these five qualified endoscopists were included in the analysis. Both scheduled and emergency procedures were included.

Participants completed the RAF-E after each ERCP. The form, as shown in Figure 1, contains an objective and a subjective part. Objective parameters include indication, difficulty degree adapted from Schutz's classification (Table 1)[165], previous failed attempts at cannulation in a different center, and presence of a "virgin" papilla or previous sphincterotomy, as well as common bile duct (CBD) cannulation success and success of various therapeutic procedures. In the subjective section endoscopists are asked to grade their performance on a visual analogue scale. After each completion of a set of ten assessment forms, an improvement plan can be formulated, based on the Osborn-Parnes Creative Problem Solving Process as used in other fields.[166, 180] The subjective scores are not taken into account in the statistical analysis, nor the outcome of the improvement plan. The value of the subjective assessment lies in creating self-awareness to enable self-reflection on performance rather than in providing evidence for quality measurements. To exclude registration bias, all non-self-assessed

Table 1. Degrees of difficulty based on Schutz's classification.

Difficulty degree	Biliary procedures	Pancreatic procedures
Grade 1	• Diagnostic cholangiography • Biliary cytology • Stone extraction ≤ 10 mm • Dilatation of stenosis/ stent placement/ nasobiliary drain in extrahepatic strictures	• Diagnostic pancreatography • Pancreatic cytology
Grade 2	• Stone extraction > 10 mm • Dilatation of stenosis/ stent placement/ nasobiliary drain in hilar tumours or benign intrahepatic strictures	• Cannulation of papilla minor
Grade 3	• BII anatomy • Intrahepatic stone extraction • Stone extraction with lithotrypsy	• Therapeutic pancreatic procedures including pseudocyst drainage

cases were evaluated as well. The number and outcome of these were studied by checking reports from the endoscopy database and assessed according to the same RAF-E criteria.

A questionnaire with questions on the endoscopists' experience with the form, why some forms were not completed and suggestions for improvement was filled out by all participating endoscopists to gain insight in their opinion on the added value of this quality measurement program.

Statistical analysis

Data were analysed using the statistical software package PASW 17. Descriptive statistics and graphs were used to characterize the data. Chi-square tests were used in order to test for differences between two groups. Regarding all results, a two-sided p-value < 0.05 was considered statistically significant. Performance data on cannulation rate are expressed as group medians plus ranges unless stated otherwise.

RESULTS

From January 2008 to December 2011, the five qualified endoscopists performing ERCP in our department participated in this study. A total number of 1691 ERCPs was performed by those endoscopists in this period. Fifteen hundred fifteen procedures were appraised using RAF-E. Adherence to completion of the self-assessment forms was 89.6%. All 176 non-assessed ERCPs were checked manually by deriving reports from our electronic endoscopy reporting system which captures all procedures. The CBD cannulation success rate of non-assessed procedures was not significantly different from self-assessed ERCPs (95.3% vs. 94.1%; χ^2=0.774, p=0.379).

The appraised procedures were further analysed. The average number of ERCPs per endoscopist per year was 116 (range 56-184). Table 2 presents an overview of indications and complexity for ERCP in this hospital. The distribution of difficulty degrees is shown in Table 3. Figure 2 shows the ERCP difficulty degree distribution per endoscopist. Table 4 shows the results of performance data of all endoscopists who participated in this study (total procedural outcomes as well as separate per difficulty degree).

Table 2. Overview of indications for ERCP.

Indication	Number of procedures (total n = 1515) (%)
Stones	346 (22.8)
Malignant stenosis	359 (23.7)
Benign stenosis	272 (18.0)
Chronic pancreatitis	190 (12.5)
Primary sclerosing cholangitis (PSC)	90 (5.9)
Biliary leakage or trauma	79 (5.2)
Endoprosthesis change	41 (2.7)
Other	138 (9.1)

Table 3. Distribution of difficulty degrees.

Difficulty degree	n (%)
1	910 (60.1)
2	230 (15.2)
3	375 (24.8)

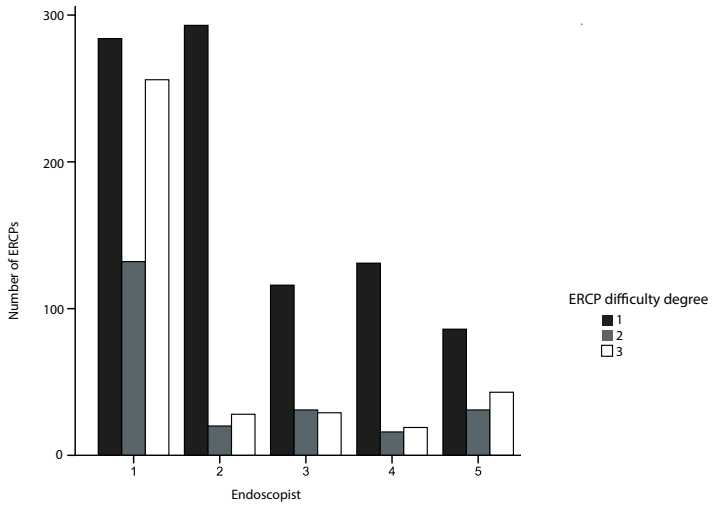


Figure 2. ERCP difficulty degree distribution per endoscopist.

Indications per difficulty level

The most frequent indications in level 1 ERCPs were common bile duct stones (n=251; 27.6%), benign strictures (n=216; 23.7%), or malignant strictures (n=240; 26.4%). Malignant stricture was in the level 2 group the most frequent indication for ERCP; this accounted for 41.7% of the total number (n=96). The major part of the level 3 ERCPs was performed for pancreatic indications, such as endoscopic therapy in chronic pancreatitis (47.2% of all level 3 ERCPs).

Success of procedures

Procedural success was calculated for the most common indications: stone extraction and benign or malignant strictures. In 148 patients with stones, both sphincterotomy and stone extraction was attempted after cannulation of the CBD. In these cases, sphincterotomy was successfully performed in all cases as well and complete stone extraction was successful in 92.3%. Procedural success in cases in which the ERCP was indicated for suspicion of CBD stones was therefore 92.3% (range 84.2-94.4%). In patients with strictures (either benign or malignant; n=482), CBD cannulation success rate was 98.4% and stent placement was

Table 4. Performance data.

ERCP difficulty	Intent	Papilla (n)	Success (%)	Partial success* (%)	Failure(%)
All	CBD cannulation	All (1317)	94.1(84.2-97.2)	0.6 (0-1.8)	5.6 (2.3-15.5)
		Virgin (602)	87.0 (79.2-95.5)	1.4 (0-4.0)	10.7 (3.2-20.2)
		Sphincterotomy (713)	98.4 (90.5-100)	0	1.6 (0-9.5)
		Previous failure (266)	86.7 (65.4-96.2)	0 (0-3.4)	13.3 (3.2-34.6)
		No previous failure (733)	96.2 (89.8-99.2)	0 (0-1.0)	3.6 (0.8-10.2)
	Sphincterotomy	n= 377	100 (98.2-100)	0	0 (0-1.8)
	Stenting	n= 808	97.8 (90.2-98.5)	1.0 (0-4.6)	2.1 (0-5.2)
	Stone extraction	n= 296	86.8 (74.1-89.6)	6.6 (0-14.8)	8.3 (6.3-11.1)
1	CBD cannulation	n= 892	91.9 (85.0-96.0)	0.7 (0-1.8)	7.0 (3.2-14.7)
	Sphincterotomy	n= 254	100	0	0
	Stenting	n= 498	97.4 (91.2-100)	0 (0-5.4)	2.6 (0-4.9)
	Stone extraction	n= 201	91.3 (83.9-100)	0 (0-8.9)	5.7 (0-13.3)
2	CBD cannulation	n= 229	100 (90.0-100)	0 (0-3.2)	0 (0-10.0)
	Sphincterotomy	n= 46	100	0	0
	Stenting	n= 163	97.0 (90.0-100)	0 (0-4.8)	0 (0-10.0)
	Stone extraction	n= 53	75.0 (50.0-100)	16.7 (0-37.5)	0 (0-13.3)
3	CBD cannulation	n= 196	85.7 (62.5-96.7)	0 (0-0.8)	14.3 (2.5-37.5)
	PD cannulation	n= 266	85.7 (65.7-91.7)	0 (0-2.7)	14.3 (8.1-34.3)
	Spincterotomy	n= 76	100 (83.3-100)	0	0 (0-16.7)
	Stenting	n= 147	95.7 (81.3-100)	0 (0-4.5)	1.3 (0-18.8)
	Stone extraction	n= 42	75.0 (50.0-100)	12.5 (0-25.0)	4.5 (0-25.0)

*Partial success regarding cannulation was defined as passage of contrast fluid or successful cannulation with a guidewire, but cannulation with the catheter was not achieved.

successful in 96.8% of the cases, resulting in an overall 95.3% (range 82.4-100%) procedural success rate.

The CBD cannulation success rate in patients with a virgin papilla compared with patients who had undergone a previous sphincterotomy was significantly different (87.0 (range 79.2-95.5) vs. 98.4% (range 90.5-100); χ^2=36.66, p<0.01). Analysing CBD cannulation success rate in previous ERCP failure versus no failed procedure before (86.7 vs. 96.2%) showed a significant difference as well (χ^2=13.88, p<0.01).

Questionnaire

The participating endoscopists were asked to give their opinion about the self-assessment program through a short questionnaire. The common opinion was that this program was valuable and useful to gain insight in performance. The form was found to be easy to fill out with clear questions. It took the participating endoscopists approximately one minute to fill out the form. The endoscopists were also asked for reasons why some procedures were not assessed. Main reason for this was that they forgot to fill out the RAF-E due to time pressure or busy programs. Receiving feedback on performance was stated as important in order to achieve and maintain good adherence.

DISCUSSION

In this study, we prospectively evaluated the performance of endoscopists with respect to ERCPs in a tertiary referral hospital using a self-assessment method. With a simple form we were able to present a descriptive analysis of indications, difficulty degrees, cannulation success rates and therapeutic success. This quality assessment of 1515 procedures gave insight in performance of individual endoscopists as well as group performance. The self-assessment program seems a reliable method to monitor quality and performance. With the rising attention for quality assurance and the expectancy that health care inspection will take measures within the near future in order to assess and assure endoscopic quality, this method of self-assessment seems a useful instrument to register performance. In addition to complication registration, which is already compulsory for all endoscopy departments in our country, this procedural registration will address a different and valuable aspect of procedural quality.

The demand for quality assessment in endoscopic procedures is increasing, but up until now, there is still no gold standard to assess the quality of ERCPs. Dutch guidelines state that an endoscopist is certified for ERCP when he has performed 100 procedures (Dutch Association of Gastroenterologists). Obviously, there is no scientific basis for this threshold number and the quality of those procedures remains unknown. ASGE guidelines state that a CBD cannulation success rate of ≥ 90% is an overall appropriate target including experts, and that most endoscopists should be able to achieve a success rate of 85%.[161] In general, gastroenterologists in smaller regional hospitals usually perform ERCPs with a level 1 difficulty degree. Technically more challenging procedures with difficulty level 2 or 3 are more often performed in tertiary referral centers. Questions regarding quality are being raised, such as the number of successful CBD cannulations and whether there was a difference in success between virgin papillas and patients with previous sphincterotomy. Numbers on the most frequently performed types of procedures are interesting as well. Another matter might be

about a case with successful CBD cannulation, but failed stone extraction. On top of these questions, we wanted to encourage critically reflecting on one's performance. Our aim was to develop a self-assessment form that was easy to fill out and addressed the problems raised. A major drawback of too extensive evaluation forms is the inversely correlated drop in adherence to filling them out as we experienced in a pilot study. As reported by the participants, the most common reason not to fill out the form was lack of time during a busy program. It was thus necessary to find a balance between the number of questions asked in the form and ease of completing it. Unfortunately, this implies an inevitable trade-off for a number of parameters that would have been interesting to monitor as well. In order to achieve compliance as high as possible and to develop a practical assessment tool for the future, the most important outcome parameters were determined through expert opinion and this process resulted in the development of the RAF-E form used in this study. The final version of the form combines registration of procedural intention and outcome; we have used parts of previously validated assessment tools and metrics as described by Peter Cotton.[181,182] Parameters that were scored as 'partially successful', were considered to be failures in the analysis in order to avoid any discussions on definitions of partial success. The results have shown that RAF-E is an easy to use device that provides insight in performance of individual endoscopists as well as larger groups. Unfortunately, the improvement plan was sparsely filled out by the experts, so it is difficult to make a statement on whether performance has improved afterwards. However, it might be worth considering linking the RAF-E to the electronic endoscopy reporting system. This is a desirable step to take within the near future, which ensures the reliability of this method and provides easy accessible data for analysis. Since too much registration seems to be a burden for doctors in general, as well as for endoscopists, we believe that this linking of systems will result in even more reliable data.

Previous studies on ERCP quality focused mainly on complications, an accepted outcome parameter to assess quality.[183-185] However, next to complication registration, we believe that the procedural quality is of equal importance as stated above. Some study groups have published their performance data on i.e. cannulation success with numbers similar to our center (92-94%).[176,177,186] Since performing ERCP only for diagnostic purposes, for example in suspected common bile duct stones[187,188], is considered obsolete and risky, less invasive methods such as MRCP or EUS are recommended for diagnostics. ERCP should thus be limited to therapeutic interventions. Taking this into account, therapeutic or procedural success is an important outcome parameter as well. The ASGE states that clearance of common bile duct stones should be achieved in >85% of the cases. In our center the overall success rate was 86.8%, which is adequate. Stent placement was successful in 97.8% of patients, which is in line with the ASGE recommendations as well (>90% success rate). These data show that our RAF-E provides insight in performance criteria such as those formulated by the ASGE. We used a modified version of the Schutz classification in this study. An ASGE working party has proposed a revised complexity score with four difficulty categories. This score was not yet available at the start of this study.[189] The results of our study are in line with the findings of the ERCP Quality Network[190], which is an electronic system where endoscopists can enter their data anonymously. As a result of this anonymity, no data verification can be performed and data is thus subject to bias. Moreover, one cannot recollect whether the ERCPs entered in the database are the only procedures performed by the participating endoscopists (registration bias). On the other hand, since the ERCP Quality Network is anonymous, there is less incen-

tive to leave out failed cases. Our study was single-center; the ERCP Quality Network enables endoscopists from different centers to enter their procedures in the database.

This is the first study to show the rationale for using a self-assessment program in order to assess quality of ERCPs and the prospective design is one of the strengths. In the Netherlands, no methods are available to gain insight in quality and performance of ERCP, apart from retrospective database research. This results in the unique character of our data.

There are some limitations to this study. One possible source of bias is registration bias. The endoscopists themselves are responsible for filling out the assessment forms correctly. One might state that the forms are sensitive for falsification since they are not linked to an endoscopy report database. An endoscopist could just leave out the procedures that failed. Therefore, reports of all ERCPs performed in the timeframe were extracted to quantify and evaluate non-assessed procedures. These non-assessed procedures were checked manually on procedural outcomes such as CBD cannulation success. There were no significant differences in outcome between assessed and non-assessed procedures, which makes it unlikely that procedures were left out on purpose. Another limitation of this study is that it was performed in a single tertiary referral center. The performance numbers cannot be extrapolated to the general endoscopist in smaller hospitals with a different workload and case mix. Moreover, there was quite some variation in numbers and casemix between the endoscopists in this study alone. This might have had some influence on the results, but this is a reflection of the work- and caseload in our endoscopy unit and probably many other endoscopy units across the world. However, a clear relation between numbers of ERCPs performed and outcome has not been established yet.[191] The numbers in this study are too small for a clear point of view on this topic, but when looking closer at our analysis, there was no correlation between volume and performance of the participating endoscopists. The one with the lowest volume was not the worst performer and vice versa. Finally, the impact of trainees on procedural outcome or success was not established in this study. Since this study was performed in a teaching hospital, trainees were involved in most ERCPs. It would be interesting to analyse their contribution in a new prospective study.

This study shows that a simple self-assessment form is a successful device to provide insight in quality of ERCPs, on an individual base as well as group performance. However, we experienced that in order to achieve and maintain a good adherence, reporting data to the participants on an individual base was eminent, but this required time and dedication as well as one of the endoscopists stimulating the others to fill out the RAF-E forms. The next step is to roll out this self-assessment program nationwide in the Netherlands, which will provide insight in quality and performance regarding ERCPs across the country. Next to investigating quality of ERCPs performed by experienced endoscopists, including trainees in this self-assessment program will provide additional information on learning curves on top of quality assessment.

Chapter 9

Prospective evaluation of ERCP performance – results of a nationwide quality registry

Vivian E. Ekkelenkamp, Robert A. de Man, Frank ter Borg, Pieter C.J. ter Borg, Marco J. Bruno, Marcel J.M. Groenen, Antonie J.P. van Tilburg, Erik A.J. Rauws, Arjun D. Koch

Submitted

ABSTRACT

Introduction: Despite significant interest by health authorities, patient organizations and insurance companies, data on procedural outcome and quality of ERCP in general and academic practice are sparse. The aim of this study was to assess procedural outcome of ERCP within a large prospective registry in the Netherlands and to evaluate associations between endoscopist-related factors and procedural outcome.

Methods: All endoscopists performing ERCP in the Netherlands were invited to register their ERCPs over a one-year period using the Rotterdam Assessment Form for ERCP (RAF-E). Primary outcome measure was procedural success. A priori difficulty level of the procedure was classified according to Schutz. Baseline characteristics of the endoscopist, e.g. previous experience, were recorded at study entry. Regression analysis was performed to identify predictors of procedural outcome.

Results: A total of 8575 ERCPs was registered by 171 endoscopists from 61 centers. This entails about 50% of all ERCPs performed in that period nationwide. Overall procedural success was 85.8%. An intact papillary anatomy was present in 5106 patients (59.5%): procedural success in this subset of patients was 83.4% versus 89.4% after sphincterotomy ($p<0.001$). Multivariable logistic regression identified “degree of difficulty”, “intact papillary anatomy”, and “previous ERCP failure” to be independently associated with procedural failure. “Yearly volume of ERCPs” and “trainee involvement” were independently associated with success.

Conclusions: Our nationwide prospective RAF-E registry proved to be a valuable tool to gain insight in procedural outcome of ERCPs. The overall procedural success rate for ERCP was 85.8%. Factors predictive of outcome include Schutz classification (degree of difficulty), papillary anatomy, previous ERCP failure, ERCP volume, and trainee involvement.

INTRODUCTION

The awareness on quality assurance in healthcare has increased over the last decade. Quality audits intend to increase transparency, maintain or achieve individual as well as collective excellence and provide patients with the best care possible.

This awareness also applies to gastrointestinal endoscopic procedures. In 2006, the American Society for Gastrointestinal Endoscopy (ASGE) published quality metrics for endoscopic retrograde cholangiopancreatography (ERCP).[161] ERCP is one of the technically more challenging procedures in gastrointestinal endoscopy with a relatively high complication risk, ranging between 3 to 15%.[176,178,183,185] Measurement of procedural outcome is intended to help the development of quality improvement initiatives.

Procedural quality can roughly be divided into three categories: pre-procedural, procedural and post-procedural. Whereas pre- and post-procedural outcomes focus on appropriateness of the indication, informed consent, documentation and complications, procedural quality relies mainly on the technical performance of the endoscopist. The ASGE stated that an endoscopist should be able to reach a minimal success rate of 85% for the most commonly performed procedures such as stone extraction or relief of biliary obstruction. Most evidence for these numbers is derived from single-center studies, which might not represent common ERCP practice.[167,168]

The Rotterdam Assessment Form for ERCP (RAF-E), a self-assessment registry tool, provided insight in ERCP performance in a tertiary referral center.[164] However, the earlier mentioned pitfall applies to this study as well: the results cannot be extrapolated to general practice.

The aims of the present study therefore were to evaluate procedural quality and outcome of ERCP in the Netherlands by means of a voluntary, one-year online registry and to identify predictors of procedural outcome.

METHODS

From November 2012 to November 2013, a prospective, nationwide ERCP quality registry, for which all endoscopists performing ERCP were invited, was carried out. Endoscopists consenting to participate in the study gained secured access to a web-based registry containing the RAF-E. All endoscopists that logged on to the web-based registry were invited to fill out a form with baseline characteristics, such as yearly and lifetime number of ERCPs, dexterity and center volume.

Although the RAF-E was originally designed to monitor progression of training and construct learning curves, it also includes parameters which can be used to evaluate procedural outcome including the a priori difficulty degree of a procedure (Table 1), previous failed attempts, trainee involvement, presence of a native papillary anatomy, inadvertent pancreatic duct cannulation, success of common bile duct (CBD) cannulation and procedural success.[165,181]

The quality of the data was verified. A random sample of procedures from a total of eight centers was drawn and the results were manually cross-checked with the original endoscopy reports. Twenty percent of the registered ERCPs per endoscopist were cross-checked. Moreover, the outcomes of all ERCP procedures that had not been entered into the registry were examined as well and these were analyzed to assess registration bias.

Table 1. Degrees of difficulty based on Schutz's classification.

Difficulty degree	Biliary procedures	Pancreatic procedures
Grade 1	• Diagnostic cholangiography • Biliary cytology • Stone extraction ≤ 10 mm • Dilatation of stenosis/ stent placement/ nasobiliary drain in extrahepatic strictures	• Diagnostic pancreatography • Pancreatic cytology
Grade 2	• Stone extraction > 10 mm • Dilatation of stenosis/ stent placement/ nasobiliary drain in hilar tumours or benign intrahepatic strictures	• Cannulation of papilla minor
Grade 3	• BII anatomy • Intrahepatic stone extraction • Stone extraction with lithotrypsy	• Therapeutic pancreatic procedures including pseudocyst drainage

Data on total number of ERCPs performed in the Netherlands were obtained through Dutch Hospital Data (DHD), a national foundation responsible for the collection and management of all sorts of data from all Dutch hospitals. The total number of ERCPs performed in the Netherlands could be calculated, based on unique procedural codes that need to be supplied for the financial reimbursement of the interventions.

Statistical analysis

Descriptive statistics were used to characterize the data. An initial univariate logistic regression analysis was performed to identify possible explanatory variables associated with the primary endpoint 'procedural failure'. Variables of potential significance ($p \leq 0.10$) were entered in a multivariate model. Stepwise backwards selection was used to fit models. The multivariate analysis was carried out by a generalized estimating equations model (GEE), in order to correct for the repeated measurements for endoscopists and center. Fisher's exact test and Chi-square tests were used for analysis of proportions. Analyses were carried out using IBM SPSS Statistics 21 and SAS®.

RESULTS

During the one-year study period, a total of 8575 ERCPs was entered in the registry by 171 different endoscopists working in 61 hospitals. According to figures retrieved from the Dutch Hospital Data, nationwide 16910 ERCPs were performed in 2012 (86% of these were coded as therapeutic). Accordingly, the present study included approximately 50% of all ERCPs.

Figure 1 shows a flowchart of the participating centers and endoscopists. The mean number of ERCPs registered per endoscopist was 50 (range 1-366). The mean estimated number of procedures performed per year, as filled out at baseline, was 70 with a range from 20 to 400. The sum of the estimated yearly number of ERCPs of all participants was 11000. This is approximately 65% of the truly performed procedures. Approximately 65% of all invited endoscopists participated in this study, meaning that the individual estimate of yearly volume corresponds very well with the definite number. The mean number of ERCPs per center was 285 (range 50-1200).

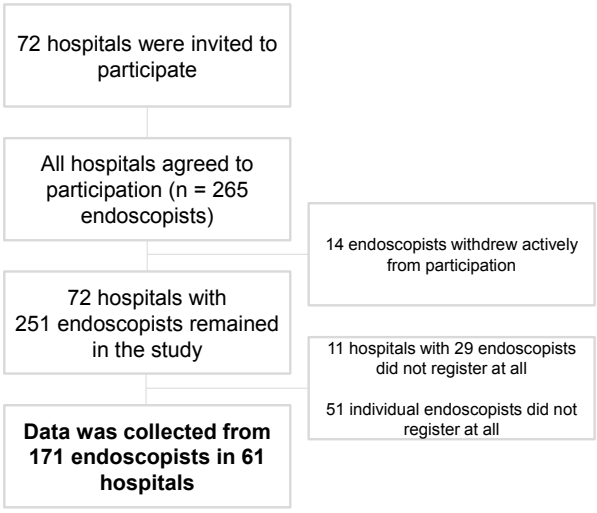


Figure 1. Flowchart of participating endoscopists.

Table 2. Characteristics of the registered ERCPs.

Intention	N	%	Procedural success (N)	%
Complete stone extraction CBD	4388	51.2	3740	85.2
Endoprosthesis – stenosis CBD	1829	21.3	1576	86.2
Metal stent – stenosis CBD	545	6.4	476	85.7
Endoprosthesis bile leakage	292	3.4	256	87.7
Therapy chronic pancreatitis	186	2.2	146	78.5
Other	1335	15.6	1175	88.0
Total	8575	100	7360	85.8
Difficulty degree	N	%	Procedural success (N)	%
1	5676	66.3	4999	88.1
2	1989	23.2	1676	84.3
3	890	10.4	669	75.2
Native papillary anatomy	N	%	Procedural success (N)	%
Yes	5106	59.5	4259	83.4
No	3469	40.5	3101	89.4

Overall procedural success was 85.8%. Table 2 provides insight in the ERCP characteristics, such as intention of the procedure, degree of difficulty and presence of native papillary anatomy, and shows success rates in those subgroups. Procedural success in patients with native papillary anatomy was 83.4%, which is significantly lower than the success rate in patients who had undergone previous sphincterotomy (89.4%; $p<0.001$). Figure 2 displays the overall and individual success rates per number of RAF-E's.

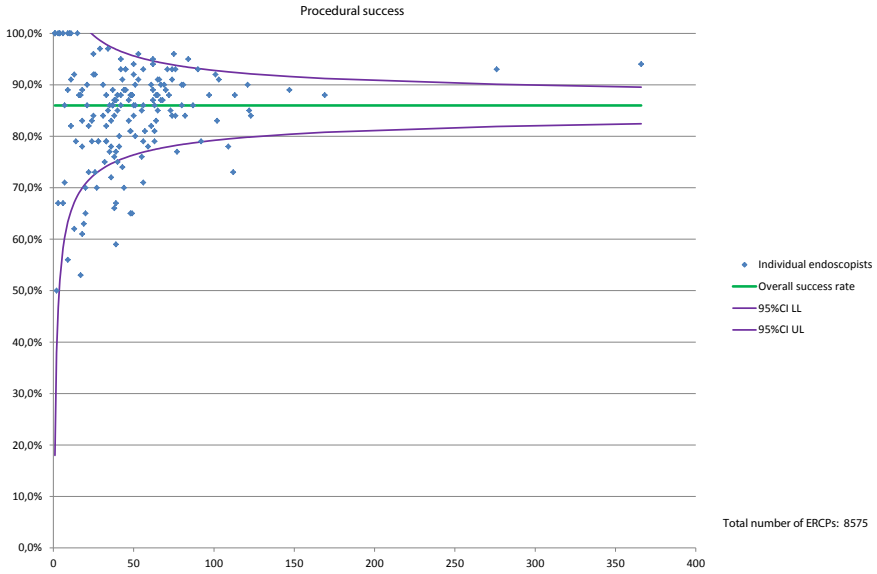


Figure 2. Overall procedural success and individual performance.

Multivariate analysis (predictors of procedural failure)

A multivariate logistic regression analysis was performed in order to identify endoscopist- as well as procedure-related predictors for procedural outcome. After initial univariate regression analysis, the following covariates were included in the multivariate model: trainee involvement; yearly ERCP volume per endoscopist; lifetime number of ERCPs performed; difficulty degree; native papillary anatomy; ASA classification; and previous ERCP failure. All these covariates were included in the model. However, due to a high correlation between the lifetime number of ERCPs performed and yearly volume of ERCPs (ρ=0.836; $p<0.001$), lifetime number was excluded. Figure 3 is a graphic depiction of this correlation showing that endoscopists who have performed a large number of ERCPs lifetime, also have a high yearly volume. The results of the multivariate regression analysis are shown in Table 3. We corrected for endoscopist within center in this model, in order to adjust for repeated measurements. The risk of ERCP failure plotted by the level of difficulty, and the endoscopists' volume, is shown in Figure 4.

Subsequently, we tested for interaction between the variable "difficulty degree" and the other variables. There was no significant interaction; therefore the model can be applied to only level 1 & 2 procedures as well. Since every endoscopist performing ERCPs should be able to complete those types of procedures, we focused on this specific subset.

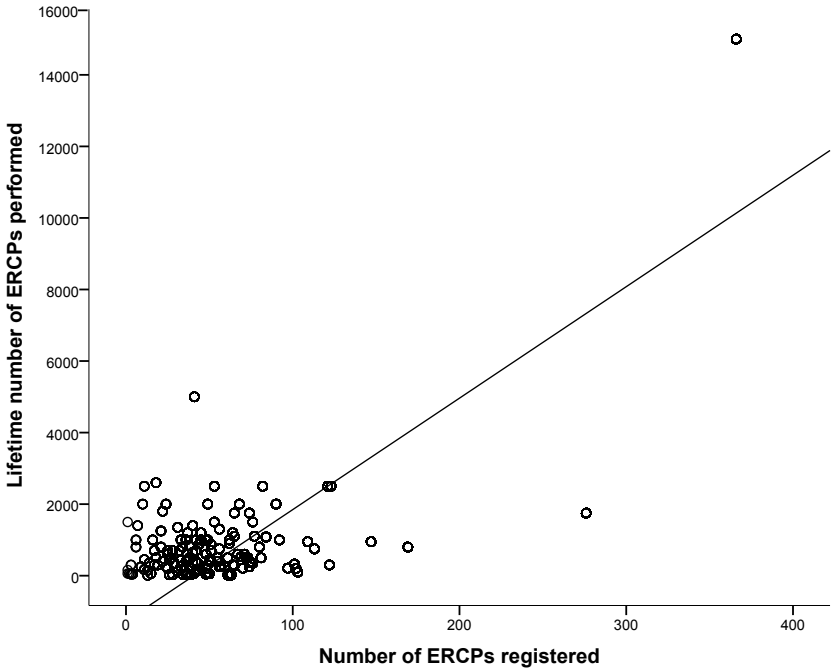


Figure 3. Distribution of endoscopists regarding ERCPs performed lifetime and volume of ERCPs performed yearly.

Table 3. Generalized Estimating Equation model with predictors for procedural failure.*

Variables	Estimate for failed procedure (95% CI)	p-value
Endoscopist-related		
Yearly volume of ERCPs		
<50 vs. 51-100	-0.348 (-0.573- -0.123)	0.002
<50 vs. 101-200	-0.653 (-0.945- -0.361)	<0.001
<50 vs. >200	-1.687 (-1.999- -1.375)	<0.001
Procedure-related		
Trainee involvement	-0.189 (-0.365- -0.014)	0.04
Difficulty degree		
2 vs. 1	0.449 (0.281-0.617)	<0.001
3 vs. 1	1.335 (1.065-1.604)	<0.001
Native papillary anatomy	0612 (0.422-0.803)	<0.001
Previous ERCP failure	1.099 (0.882-1.312)	<0.001

*In this model, we adjusted for endoscopist within center in order to correct for the repeated measurements.

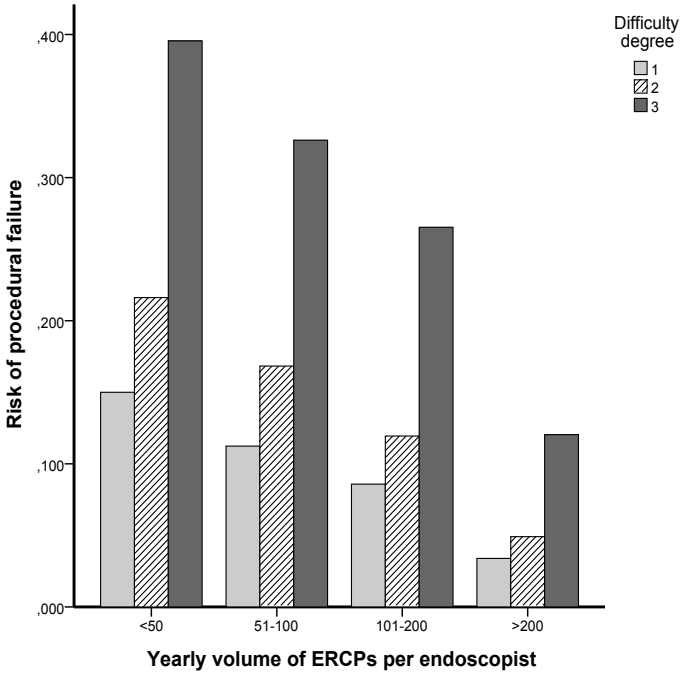


Figure 4. Risk of procedural failure per degree of difficulty, plotted by yearly volume of ERCPs.

Data verification

Eight randomly selected hospitals were visited during the study period. First, 20% of the RAF-E's per endoscopists were randomly selected and verified. This resulted in a total of 281 data entries in the database that were manually cross-checked with the original endoscopy report for accuracy. All but one ERCP were registered correctly.

Second, we checked for ERCPs that had not been entered into the database. A total of 441 ERCPs had not been registered. These procedures were examined using the endoscopy reports and procedural outcome was captured. The overall success rate in the non-registered ERCPs was 76.0%. This was significantly lower than the overall success rate derived from the RAF-E's (85.8%; $p<0.001$). This difference originated mainly from the missing procedures performed by endoscopists that actively participated in the registry, rather than non-participating endoscopists.

DISCUSSION

In this study, we explored the quality and procedural outcome of ERCP in the Netherlands by means of a one-year, nationwide, web-based registration. Overall procedural success was almost 86% based on a total number of 8575 registered procedures. We were able to identify endoscopist- as well as procedure-related predictors of procedural failure in ERCP. Moreover,

this study showed that ERCP volume indeed was associated with procedural outcome. With the rising attention for quality assurance and the expectancy that healthcare authorities will take measures within the near future to assure endoscopic quality, a quality registry like this seems useful for assessment of performance and procedural outcome.

Outcome measures such as procedural success, CBD cannulation success or success of intended therapy have earned their place in quality assurance.[176,177,190,192] A recent meta-analysis was published on procedural ERCP quality, which was in good standing.[193] However, most included studies were performed in tertiary centers and this clearly affects the generalizability of the results. Due to the limited number of studies performed in a broader perspective, the absence of data still hampers the development of clear recommendations on ERCP quality and outcome. We believe that our nationwide registry is a valuable and unique addition to existing studies providing insight in overall ERCP quality in different hospital settings, including the smaller low-volume centers.

Our data revealed endoscopist- and procedure-related predictors of procedural outcome in ERCP. In contrast to the nationwide study performed in the United Kingdom[191], we did find a significant association between ERCP volume per endoscopist and procedural outcome. We showed that a yearly volume of ≥50 ERCPs – as an arbitrary cut-off level - was significantly associated with a lower risk of procedural failure. This is also in accordance with other studies.[175,194,195] However, some nuances are in place here. There was a high correlation between lifetime number of ERCPs and yearly volume of ERCPs performed per endoscopist. It seems therefore not desirable to ensure quality by establishing a threshold of 50 procedures per year alone. Experience is equally important, but the lack of solid data hampers the ability to draw firm conclusions.

An unexpected finding of our study is that the involvement of trainees is significantly associated with a lower risk of failure, in contrast to the conclusions of other papers.[191,193,194] One possible explanation for this is that the involvement of trainees represents the quality of the center in our study; in order to meet the requirements set for training centers, quality of care may be higher.

The occurrence of registration bias is a concern. During data verification, it emerged that the quality of the entered data was good and the content correlated highly with the endoscopy reports. Data verification evidently revealed that the non-registered procedures included proportionately more failures than those that were entered in the web-based registry (success rate of 76% versus 86%). Of course a random sample has its drawbacks, but when we consider this as the reference for all missing ERCPs (approximately 49% nationwide), the overall procedural success rate will in all likelihood be lower than the reported 86%. The only way to solve this issue and to be able to draw valid conclusions, full coverage with one hundred percent adherence is necessary. This adherence rate will never be reached in a voluntary registry. Nonetheless, this limits somewhat the findings of our study. A threshold of 50 procedures per year per endoscopist seems an easy to apply quality criterion. Nevertheless, from our analysis it becomes also clear that in this registry there were endoscopists doing less procedures but with good procedural outcome and vice versa. Case mix, previous experience, hospital volume and cumulative experience are important in this regard. The question therefore remains whether quality assurance should be guided by imposing a minimal number of annual procedures or by continuous monitoring of procedural outcome per individual endoscopist. Only the latter method truly monitors and safeguards procedural outcome. Addition of a registry on

pre-procedural quality as well as complications, to a registry on procedural outcome like this, provides an even more complete picture of ERCP quality.[179]

Conclusions

This nationwide prospective registry of ERCP shows an overall procedural success rate for ERCP in the Netherlands of 85.8%. Factors predictive of outcome include Schutz classification (degree of difficulty), native papillary anatomy, previous ERCP failure, ERCP volume, and trainee involvement. Our nationwide online RAF-E registry proves to be a valuable tool to gain insight in procedural ERCP outcome.

Section IV

General discussion

Chapter 10

General conclusions and future perspectives

Training procedural skills in gastrointestinal (GI) endoscopy has gradually shifted from threshold numbers to a more competency-based approach. Although the GI endoscopy community worldwide is largely convinced that threshold numbers poorly reflect individual competence, the answer on how to assess individual competence and how to implement individual-based approaches in a training curriculum remains enigmatic.

The aims of this thesis were 1) to investigate learning curves for colonoscopy, identify factors and methods to influence this curve and to assess the performance of specialists; and 2) to examine the value of an assessment tool for ERCP used by trainees as well as experienced endoscopists in order to gain insight in learning curves and procedural quality. In this final chapter the main conclusions from the studies presented are summarized and put into perspective with clear recommendations and directions for future research.

Quality assessment in colonoscopy

Patients, doctors and health care authorities agree that any medical intervention needs to be of high quality. This also accounts for colonoscopy, even more so when applied in a screening program when offered to presumably asymptomatic subjects with low-risk of the disease. Quality indicators such as cecal intubation rate (CIR) and adenoma detection rate (ADR) are established markers that are being used worldwide. However, other important domains contribute to high-quality colonoscopy. In **chapter 3** we explored the opinion of experts regarding the factors that are considered of importance, if not obligatory, for high-quality colonoscopy. We were able to compose an ultimate list of items for colonoscopy by carrying out a Delphi survey among experts with an excellent reputation in the field. Not only technical skills, such as the previously mentioned CIR and ADR, were considered essential, but the experts valued factors with respect to patient safety and professional behaviour very highly as well. The attention for core knowledge complements the assessment of purely technical skills. The high inter-expert agreements on specific aspects that define quality in colonoscopy justify the implementation of the factors that resulted from this study, in an assessment device for trainees in colonoscopy. This could provide trainers more support in concrete competence assessment.

Ideally, the assessment and teaching of procedural skills involves a structured training curriculum. Trainees go through a transition from novice to competence to excellence. The ultimate goal is to reach (procedural) competence or even excellence quickly, predictably, and with the least harm being done to patients. A pre-patient curriculum using virtual reality (VR) simulators in a skills lab environment seems a soundj139

start for colonoscopy training. In other medical specialties, such as anaesthesiology, surgery and urology, simulator training has earned a widely integrated role in training curricula. By supervised hands-on training on a simulator, a trainee is able to master the basic skills of colonoscopy, without any ‘unnecessary’ patient contact or discomfort. A number of simulators have been validated for colonoscopy and studies have shown that there is a progression in simulator-based performance after practice. While this holds a promise, the transfer of those mastered skills to patient-based colonoscopy is of course the ultimate goal. This skill transfer was studied in **chapter 4**. We were able to prove that novices who followed a simulator training curriculum for colonoscopy showed improved performance in patient-based colonoscopy. An important finding: without an increment in the learning curve in patient-based performance, especially in the early training phase, simulator-based training

would be useless. This study provides a rationale for intensive simulator training in the early learning curve of novices training in colonoscopy.

A recent paper from the UK described the learning curve of 297 trainees based on 36,730 colonoscopies.[124] The British trainees reached a CIR of ≥90% after having performed an average of 230 procedures. This threshold is reached at an earlier stage than previous studies from the Netherlands and USA have shown.[61,65] This difference may be related to the set-up of the respective training programs. In the UK, the endoscopy training curriculum dramatically changed over the last years, with intensification of training. For example, a supervisor has to be in the room during all colonoscopies. The intensified training program resulted in a steeper learning curve and earlier competence. In **chapter 5**, we aimed to influence the learning curve in a comparable manner. We studied the influence of a standardized assessment program on the learning curve of novices in colonoscopy. Comparing this curve with a historical cohort that did not participate in an assessment program, showed that the learning curve was steeper in the study group as well, in accordance with the British study. This strengthens the notion that supervision with structured feedback, also during the later stages of training, is beneficial. Finally, we could conclude that video-assessment by an expert is comparable to the assessment of an expert in the room, and that self-assessment is valuable as well. Based on these results, we would advocate that an assessment program, as described in this chapter, is implemented in the colonoscopy training curriculum in the Netherlands.

In order to be able to define the competence criteria or goals for trainees, it is important to benchmark performance of certified endoscopists first. Without a benchmark, it is rather difficult to determine procedural competence for training. There have been several publications regarding cecal intubation rate as a surrogate marker of quality, but evidence for other identifiers for quality of skills are lacking. What if an endoscopist pursuits a CIR of >90% at all costs, and jeopardizes a patient's comfort, or even worse, safety? In **chapter 6** we examined the relationship between the patient's comfort levels and the performance of the endoscopist in terms of CIR and polyp detection rate (PDR). Patient comfort was measured by the assisting nurses, and patient experience of the procedure was surveyed after the procedure in the recovery unit. Our most important finding was that the best colonoscopists achieved the highest success rates of cecal intubation and polyp detection, used less sedatives and still received the best comfort scores (both nurse-assessed and experience of the patient himself). We therefore believe that measurement of patient comfort during colonoscopy (sedated or non-sedated) complements the assessment of determined quality indicators of technical skills.

Quality assessment in ERCP

With ERCP as one of the most technically challenging and risky procedures being performed in the average clinical gastroenterology practice, insight in performance is needed. In the Netherlands, up until now, a minimum number of 100 ERCPs is required for certification. However, evidence for this threshold number is lacking. On top of that, international guidelines propose a steady ≥80% or ≥85% success rate as endpoint of ERCP training instead of threshold numbers.[161] In **chapter 7** we evaluated ERCP performance of novice trainees. Using a self-assessment form, the Rotterdam Assessment Form for ERCP (RAF-E), CBD cannulation success was plotted as a learning curve. A CBD cannulation rate of ≥80% was only reached after an average of 160 procedures, meaning that some trainees not even were competent after having performed that number of procedures. The previously mentioned conventional

threshold number of 100 ERCPs used in the Netherlands seems therefore inadequate. We advocate the use of learning curves for assessment of performance and monitor progression over time. The learning curves for different therapeutic aspects of the procedures were examined as well. These learning curves showed similar development of skills, in a parallel progression over time, for those various aspects; hence we would propose that continuous assessment of CBD cannulation can even act as a valuable surrogate marker for evaluating overall competence in ERCP.

Similar to colonoscopy benchmarking, performance of certified endoscopists in ERCP was assessed. We used the RAF-E in **chapter 8** in order to gain insight in performance of expert endoscopists carrying out ERCP. This study showed that the RAF-E is a simple and easy-to-use device to gain insight in quality and procedural outcome for experienced doctors. Implementing a quality registry using the RAF-E in a multicenter, even nationwide, setting may be the first step in improving ERCP quality and is described in **chapter 9**. Quality and outcome in ERCP is a hot topic amongst different parties, such as government, Health Care Inspectorate as well as the Dutch association of gastroenterologists (NVMDL). However, the debate lacks scientific grounds, since there are no data on quality and outcome. For this reason, a nationwide quality registry was carried out. The RAF-E was implemented in an electronic portfolio, used by both trainees as well as gastroenterologists. All ERCP-performing gastroenterologists in the Netherlands were invited to participate in this one-year quality registration. This study revealed that overall procedural quality of ERCP in the Netherlands, based on 8575 procedures, was sufficient with a procedural success rate of nearly 86%. Predictors of procedural failure could be identified. The ongoing discussion regarding volume and quality still remains largely unanswered. In this registry, there was a significant correlation between the yearly volume of ERCPs performed and the risk of procedural failure. This is in contradiction with a recent paper from the UK and a meta-analysis. In these studies, volume was not identified as an independent predictor of procedural failure.[191,193] It is nonetheless plausible that when exposure and experience in a procedure increase, there is a positive influence on performance and therefore, quality improvement. Finally, the increasing awareness that rises through a registration like this, may already influence procedural quality in a positive way.

General recommendations

Based on the results presented in this thesis, we make several recommendations. First of all, training of novices in colonoscopy should start with a pre-patient training curriculum. Skills acquired on a virtual reality simulator are transferable to patient-based procedures. When the learning curve on the simulator reaches a plateau phase, the time is right to start with patient-based colonoscopies. Secondly, competence during patient-based training in colonoscopy needs to be measured and continuously (self-) assessed. The derived learning curves provide insight in progression of skills at any point in time. Furthermore, we propose the implementation of a standardized assessment program for colonoscopy in order to repeatedly assess trainees during their competence development. An important aspect to bear in mind is that both technical as well as behavioural skills merit attention from trainer and trainee. The same repetitive self-assessment in ERCP is worth pursuing as well. A learning curve of CBD cannulation reflects the overall skill development of a trainee. Real-time feedback by monitoring these learning curves should be provided and allows for early identification of trainees deviating from the group average. To conclude, we propose a permanent quality registration for certified endoscopists as well. Such a registry creates awareness and direct feedback,

which in itself already may have a positive impact on procedural quality. The time-consuming process of entering procedures in a database and the limited time a doctor has, hamper the effectuation of such a registry. Therefore it is desirable that the information needed for quality registration can be directly and automatically extracted in a standardised manner from the endoscopy reports.

Future perspectives

In the Netherlands, the Rotterdam Assessment Form for colonoscopy and ERCP (RAF-C and RAF-E) have been implemented in a web-based portfolio, which is available for all gastroenterology trainees. Efforts are made in the development of software that enables to demonstrate personal learning curves, plotted against group benchmarks. The Rotterdam Assessment Form for ERCP (RAF-E) is made available for both trainees and expert endoscopists. Trainees can gain insight in their learning curve in a similar way as described for colonoscopy. The proposed procedural quality registration for ERCP will be implemented in the Netherlands as a continuation of our study. In the meantime, the proposed linking of endoscopy reporting systems with the e-portfolio where procedural quality is registered, is being carried out and will be available in the near future for all gastroenterologists, both certified and in training. This will allow for automated registration of all procedures and provide data for future studies.

Appendices

REFERENCES

1. Luursema JM, Buzink SN, Verwey WB, et al. Visuo-spatial ability in colonoscopy simulator training. Adv Health Sci Educ 2010;15:685-694.
2. Centre for EBM Levels of Evidence. Oxford Centre for Evidence-based Medicine - levels of evidence. Volume 2013: http://www.cebm.net/index.aspx?o=1025, 2011.
3. Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005;19:1523-1532.
4. Datta V, Mandalia M, Mackay S, et al. The PreOp flexible sigmoidoscopy trainer: Validation and early evaluation of a virtual reality based system. Surg Endosc 2002;16:1459-1463.
5. Bhutani MS, Aveyard M, Stills Jr HF. Improved model for teaching interventional EUS. Gastrointest Endosc 2000;52:400-403.
6. Bhutani MS, Hoffman BJ, Hawes RH. A swine model for teaching endoscopic ultrasound (EUS) imaging and intervention under EUS guidance. Endoscopy 1998;30:605-609.
7. Bittner Iv JG, Mellinger JD, Imam T, et al. Face and construct validity of a computer-based virtual reality simulator for ERCP. Gastrointest Endosc 2010;71:357-364.
8. Fayez R, Feldman LS, Kaneva P, et al. Testing the construct validity of the Simbionix GI Mentor II virtual reality colonoscopy simulator metrics: module matters. Surgical endoscopy 2010;24:1060-1065.
9. Felsher JJ, Olesevich M, Farres H, et al. Validation of a flexible endoscopy simulator. Am J Surg 2005;189:497-500.
10. Ferlitsch A, Glauninger P, Gupper A, et al. Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. Endoscopy 2002;34:698-702.
11. Frimberger E, von Delius S, Rosch T, et al. A novel and practicable ERCP training system with simulated fluoroscopy. Endoscopy 2008;40:517-520.
12. Grantcharov TP, Carstensen L, Schulze S. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. JSLS 2005;9:130-133.
13. Haycock AV, Bassett P, Bladen J, et al. Validation of the second-generation Olympus colonoscopy simulator for skills assessment. Endoscopy 2009;41:952-958.
14. Kim S, Spencer G, Makar GA, et al. Lack of a discriminatory function for endoscopy skills on a computer-based simulator. Surg Endosc Interv Tech 2010;24:3008-3015.
15. Koch AD, Buzink SN, Heemskerk J, et al. Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. Surg Endosc 2008;22:158-162.
16. Koch AD, Haringsma J, Schoon EJ, et al. A second-generation virtual reality simulator for colonoscopy: Validation and initial experience. Endoscopy 2008;40:735-738.
17. Leung JW, Lee JG, Rojany M, et al. Development of a novel ERCP mechanical simulator. Gastrointest Endosc 2007;65:1056-1062.
18. Leung JW, Wang D, Hu B, et al. A head-to-head hands-on comparison of ERCP mechanical simulator (EMS) and ex-vivo porcine stomach model (PSM). J Intervent Gastroenterol 2011;1:108-113.
19. MacDonald J, Ketchum J, Williams RG, et al. A lay person versus a trained endoscopist: can the preop endoscopy simulator detect a difference? Surg Endosc 2003;17:896-898.
20. Mahmood T, Darzi A. A study to validate the colonoscopy simulator: It is usefully discriminatory for more than one measurable outcome. Surg Endosc 2003;17:1583-1589.
21. McConnell RA, Kim S, Ahmad NA, et al. Poor discriminatory function for endoscopic skills on a computer-based simulator. Gastrointest Endosc 2012.
22. Moorthy K, Munz Y, Jiwanji M, et al. Validity and reliability of a virtual reality upper gastrointestinal simulator and cross validation using structured assessment of individual performance with video playback. Surg Endosc 2004;18:328-333.

23. Neumann M, Mayer G, Ell C, et al. The Erlangen endo-trainer: Life-like simulation for diagnostic and interventional endoscopic retrograde cholangiography. Endoscopy 2000;32:906-910.
24. Plooy AM, Hill A, Horswill MS, et al. Construct validation of a physical model colonoscopy simulator. Gastrointest Endosc 2012;76:144-150.
25. Ritter EM, McClusky Iii DA, Lederman AB, et al. Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 2003;7:871-878.
26. Sedlack RE, Baron TH, Downing SM, et al. Validation of a colonoscopy simulation model for skills assessment. Am J Gastroenterol 2007;102:64-74.
27. Sedlack RE, Kolars JC. Validation of a computer-based colonoscopy simulator. Gastrointest Endosc 2003;57:214-218.
28. Sugden C, Aggarwal R, Banerjee A, et al. The development of a virtual reality training curriculum for colonoscopy. Ann Surg 2012;256:188-192.
29. von Delius S, Thies P, Meining A, et al. Validation of the X-Vision ERCP Training System and Technical Challenges During Early Training of Sphincterotomy. Clin Gastroenterol Hepatol 2009;7:389-396.
30. Sedlack R, Petersen B, Binmoeller K, et al. A direct comparison of ERCP teaching models. Gastrointest Endosc 2003;57:886-890.
31. Tuggy ML. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. J Am Board Fam Pract 1998;11:426-433.
32. Van Sickle KR, Buck L, Willis R, et al. A multicenter, simulation-based skills training collaborative using shared GI mentor II systems: Results from the Texas association of surgical skills laboratories (TASSL) flexible endoscopy curriculum. Surg Endosc Interv Tech 2011;25:2980-2986.
33. Ahad S, Boehler M, Schwind CJ, et al. The effect of model fidelity on colonoscopic skills acquisition. A randomized controlled study. J Surg Educ 2013;70:522-527.
34. Ahlberg G, Hultcrantz R, Jaramillo E, et al. Virtual reality colonoscopy simulation: A compulsory practice for the future colonoscopist? Endoscopy 2005;37:1198-1204.
35. Barthet M, Gasmil M, Boustiere C, et al. EUS training in a live pig model: Does it improve echo endoscope hands-on and trainee competence? Endoscopy 2007;39:535-539.
36. Buzink SN, Koch AD, Heemskerk J, et al. Acquiring basic endoscopy skills by training on the GI Mentor II. Surg Endosc Interv Tech 2007;21:1996-2003.
37. 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.
38. Eversbusch A, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: A randomized trial using a virtual reality endoscopy trainer. Surg Endosc 2004;18:1514-1518.
39. Ferlitsch A, Schoefl R, Puespoek A, et al. Effect of virtual endoscopy simulator training on performance of upper gastrointestinal endoscopy in patients: A randomized controlled trial. Endoscopy 2010;42:1049-1056.
40. Gerson LB, Van Dam J. A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. Endoscopy 2003;35:569-575.
41. Giulio ED, Fregonese D, Casetti T, et al. Training with a computer-based simulator achieves basic manual skills required for upper endoscopy: A randomized controlled trial. Gastrointest Endosc 2004;60:196-200.
42. Haycock A, Koch AD, Familiari P, et al. Training and transfer of colonoscopy skills: a multinational, randomized, blinded, controlled trial of simulator versus bedside training. Gastrointest Endosc 2010;71:298-307.

43. Kaltenbach T, Leung C, Wu K, et al. Use of the colonoscope training model with the colonoscope 3d imaging probe improved trainee colonoscopy performance: A pilot study. Dig Dis Sci 2011;56:1496-1502.
44. Kruglikova I, Grantcharov TP, Drewes AM, et al. The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity virtual-reality simulation: A randomised controlled trial. Gut 2010;59:181-185.
45. Kruglikova I, Grantcharov TP, Drewes AM, et al. Assessment of early learning curves among nurses and physicians using a high-fidelity virtual-reality colonoscopy simulator. Surg Endosc Interv Tech 2010;24:366-370.
46. Lim BS, Leung JW, Lee J, et al. Effect of ERCP mechanical simulator (EMS) practice on trainees ERCP performance in the early learning period: US multicenter randomized controlled trial. Am J Gastroenterol 2011;106:300-306.
47. Mahmood T, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: No feedback, no learning. Surg Endosc 2004;18:1224-1230.
48. Maiss J, Millermann L, Heinemann K, et al. The compactEASIE(registered trademark) is a feasible training model for endoscopic novices: A prospective randomised trial. Dig Liver Dis 2007;39:70-78.
49. Sedlack RE, Kolars JC. Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: Results of a pilot study. Am J Gastroenterol 2004;99:33-37.
50. Sedlack RE, Kolars JC, Alexander JA. Computer simulation training enhances patient comfort during endoscopy. Clin Gastroenterol Hepatol 2004;2:348-352.
51. Snyder CW, Vandromme MJ, Tyra SL, et al. Retention of colonoscopy skills after virtual reality simulator training by independent and proctored methods. Am Surg 2010;76:743-746.
52. Park J, MacRae H, Musselman LJ, et al. Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg 2007;194:205-211.
53. Ende A, Zopf Y, Konturek P, et al. Strategies for training in diagnostic upper endoscopy: A prospective, randomized trial. Gastrointest Endosc 2012;75:254-260.
54. Elvevi A, Cantu P, Maconi G, et al. Evaluation of hands-on training in colonoscopy: Is a computer-based simulator useful? Dig Liver Dis 2012;44:580-584.
55. Moorthy K, Munz Y, Orchard TR, et al. An innovative method for the assessment of skills in lower gastrointestinal endoscopy. Surg Endosc 2004;18:1613-1619.
56. Phitayakorn R, Marks JM, Reynolds HL, et al. Expert benchmark for the GI Mentor IITM. Surg Endosc Interv Tech 2009;23:611-614.
57. Sarker SK, Albrani T, Zaman A, et al. Procedural performance in gastrointestinal endoscopy: live and simulated. World J Surg 2010;34:1764-1770.
58. Biau DJ, Williams SM, Schlup MM, et al. Quantitative and individualized assessment of the learning curve using LC-CUSUM Review. Br J Surg 2008;95:925-929.
59. Chung JI, Kim N, Um MS, et al. Learning curves for colonoscopy: A prospective evaluation of gastroenterology fellows at a single center. Gut Liver 2010;4:31-35.
60. Jowell PS, Baillie J, Branch MS, et al. Quantitative assessment of procedural competence: A prospective study of training in endoscopic retrograde cholangiopancreatography. Ann Intern Med 1996;125:983-989.
61. Koch AD, Haringsma J, Schoon EJ, et al. Competence measurement during colonoscopy training: The use of self-assessment of performance measures. Am J Gastroenterol 2012;107:971-975.
62. Lee SH, Chung IK, Kim SJ, et al. An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. Gastrointest Endosc 2008;67:683-689.

63. Marshall JB. Technical proficiency of trainees performing colonoscopy: A learning curve. Gastrointest Endosc 1995;42:287-291.
64. Parry BR, Williams SM. Competency and the colonoscopist: a learning curve. AUST NEW ZEALAND J SURG 1991;61:419-422.
65. Sedlack RE. Training to competency in colonoscopy: Assessing and defining competency standards. Gastrointest Endosc 2011;74:355-366+366.e351-366.e352.
66. Selvasekar CR, Holubar SD, Pendlimari R, et al. Assessment of screening colonoscopy competency in colon and rectal surgery fellows: A single institution experience. J Surg Res 2012;174:e17-e23.
67. Tassios PS, Ladas SD, Grammenos I, et al. Acquisition of competence in colonoscopy: The learning curve of trainees. Endoscopy 1999;31:702-706.
68. Verma D, Gostout CJ, Petersen BT, et al. Establishing a true assessment of endoscopic competence in ERCP during training and beyond: a single-operator learning curve for deep biliary cannulation in patients with native papillary anatomy. Gastrointest Endosc 2007;65:394-400.
69. Waller HM, Connor SJ. Cumulative sum (Cusum) analysis provides an objective measure of competency during training in endoscopic retrograde cholangio-pancreatography (ERCP). Hpb 2009;11:565-569.
70. Wani S, Cote GA, Keswani R, et al. Learning curves for EUS by using cumulative sum analysis: Implications for American Society for Gastrointestinal Endoscopy recommendations for training. Gastrointest Endosc 2013;77:558-565.
71. Watkins JL, Etzkorn KP, Wiley TE, et al. Assessment of technical competence during ERCP training. Gastrointest Endosc 1996;44:411-415.
72. Meenan J, Anderson S, Tsang S, et al. Training in radial EUS: What is the best approach and is there a role for the nurse endoscopist? Endoscopy 2003;35:1020-1023.
73. Cass OW, Freeman ML, Peine CJ, et al. Objective evaluation of endoscopy skills during training. Ann Intern Med 1993;118:40-44.
74. Chak A, Cooper GS, Blades EW, et al. Prospective assessment of colonoscopic intubation skills in trainees. Gastrointest Endosc 1996;44:54-57.
75. Church J, Oakley J, Milsom J, et al. Colonoscopy training: The need for patience (patients). ANZ J Surg 2002;72:89-91.
76. Hawes R, Lehman GA, Hast J. Training resident physicians in fiberoptic sigmoidoscopy. How many supervised examinations are required to achieve competence? Am J Med 1986;80:465-470.
77. Leyden JE, Doherty GA, Hanley A, et al. Quality of colonoscopy performance among gastroenterology and surgical trainees: A need for common training standards for all trainees? Endoscopy 2011;43:935-940.
78. Spier BJ, Benson M, Pfau PR, et al. Colonoscopy training in gastroenterology fellowships: determining competence. Gastrointest Endosc 2010;71:319-324.
79. Spier BJ, Durkin ET, Walker AJ, et al. Surgical resident's training in colonoscopy: Numbers, competency, and perceptions. Surg Endosc Interv Tech 2010;24:2556-2561.
80. Vassiliou MC, Kaneva PA, Poulose BK, et al. How should we establish the clinical case numbers required to achieve proficiency in flexible endoscopy? Am J Surg 2010;199:121-125.
81. Vitale GC, Zavaleta CM, Vitale DS, et al. Training surgeons in endoscopic retrograde cholangiopancreatography. Surg Endosc Interv Tech 2006;20:149-152.
82. 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.
83. Boyle E, Al-Akash M, Patchett S, et al. Towards continuous improvement of endoscopy standards: Validation of a colonoscopy assessment form. Colorectal Dis 2012;14:1126-1131.

84. Hope WW, Hooks Iii WB, Kilbourne SN, et al. Assessing resident performance and training of colonoscopy in a general surgery training program. Surg Endosc Interv Tech 2013;27:1706-1710.
85. Mohamed R, Shaheen AA, Raman M. Evaluation of colonoscopy skills - How well are we doing? Can J Gastroenterol 2011;25:198-200.
86. Sarker SK, Albrani T, Zaman A, et al. Procedural performance in gastrointestinal endoscopy: An assessment and self-appraisal tool. Am J Surg 2008;196:450-455.
87. 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.e1123.
88. Shah SG, Thomas-Gibson S, Brooker JC, et al. Use of video and magnetic endoscope imaging for rating competence at colonoscopy: validation of a measurement tool. Gastrointest Endosc 2002;56:568-573.
89. Thomas-Gibson S, Rogers PA, Suzuki N, et al. Development of a video assessment scoring method to determine the accuracy of endoscopist performance at screening flexible sigmoidoscopy. Endoscopy 2006;38:218-225.
90. Vassiliou MC, Kaneva PA, Poulose BK, et al. Global assessment of gastrointestinal endoscopic skills (GAGES): A valid measurement tool for technical skills in flexible endoscopy. Surg Endosc Interv Tech 2010;24:1834-1841.
91. Gotzberger M, Rosch T, Geisenhof S, et al. Effectiveness of a novel endoscopy training concept. Endoscopy 2011;43:802-807.
92. Harewood GC, Murray F, Winder S, et al. Evaluation of formal feedback on endoscopic competence among trainees: The EFFECT trial. Ir J Med Sci 2008;177:253-256.
93. Neumann M, Hahn C, Horbach T, et al. Score card endoscopy: A multicenter study to evalaute learning curves in 1-week courses using the Elangen Endo-Trainer. Endoscopy 2003;35:515-520.
94. Neumann M, Meining A, Buerschaper C, et al. Training in GI endoscopy: Can we objectively predict the performance of trainees? A prospective correlation study. Z Gastroenterol 2005;43:445-450.
95. Proctor DD, Price J, Dunn KA, et al. Prospective evaluation of a teaching model to determine competency in performing flexible sigmoidoscopies. Am J Gastroenterol 1998;93:1217-1221.
96. Suzuki N, Thomas-Gibson S, Vance M, et al. Efficacy of an accelerated colonoscopy training week: Audit from one national colonoscopy training center in the UK. DIG ENDOSC 2006;18:288-293.
97. Thomas-Gibson S, Bassett P, Suzuki N, et al. Intensive training over 5 days improves colonoscopy skills long-term. Endoscopy 2007;39:818-824.
98. Grimm H, Binmoeller KF, Soehendra N. Endosonography-guided drainage of a pancreatic pseudocyst. Gastrointest Endosc 1992;38:170-1.
99. Lee TJ, Rutter MD, Blanks RG, et al. Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. Gut 2012;61:1050-7.
100. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010;362:1795-1803.
101. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. American Journal of Gastroenterology 2006;101:873-885.
102. Cass OW. Training to competence in gastrointestinal endoscopy: A plea for continuous measuring of objective end points. Endoscopy 1999;31:751-754.
103. Freeman ML. Training and competence in gastrointestinal endoscopy. Rev Gastroenterol Disord 2001;1:73-86.
104. Valori R, Sint Nicolaas J, de Jonge V. Quality assurance of endoscopy in colorectal cancer screening. Best Pract Res Clin Gastroenterol 2010;24:451-64.
105. van Hove PD, Tuijthof GJ, Verdaasdonk EG, et al. Objective assessment of technical surgical skills. Br J Surg 2010;97:972-87.

106. Michels ME, Evans DE, Blok GA. What is a clinical skill? Searching for order in chaos through a modified Delphi process. Med Teach 2012;34:e573-81.
107. Pearce J, Jones C, Morrison S, et al. Using a Delphi process to develop an effective train-the-trainers program to train health and social care professionals throughout Europe. J Trauma Stress 2012;25:337-43.
108. Searle RD, Howell SJ, Bennett MI. Diagnosing postoperative neuropathic pain: a Delphi survey. Br J Anaesth 2012;109:240-4.
109. Wallengren J. Identification of core competencies for primary care of allergy patients using a modified Delphi technique. BMC Med Educ 2011;11:12.
110. Milholland AV, Wheeler SG, Heieck JJ. Medical assessment by a Delphi group opinion technic. N Engl J Med 1973;288:1272-5.
111. Dunckley P, Elta G. Quality assurance of training. Best Pract Res Clin Gastroenterol 2011;25:397-407.
112. J.A.G. Joint Advisory Group on GI Endoscopy. Colonoscopy. DOPS assessment form. http://www.thejag.org.uk/TrainingforendoscopistsDOPSforms.aspx.
113. Armstrong D, Barkun A, Bridges R, et al. Canadian Association of Gastroenterology consensus guidelines on safety and quality indicators in endoscopy. Can J Gastroenterol 2012;26:17-31.
114. Baillie J, Evangelou H, Jowell P, et al. The future of endoscopy simulation: a Duke perspective. Endoscopy 1992;24 Suppl 2:542-3.
115. Gerson LB, Van Dam J. Technology review: The use of simulators for training in GI endoscopy. Gastrointest Endosc 2004;60:992-1001.
116. Frank JR, Langer B. Collaboration, communication, management, and advocacy: teaching surgeons new skills through the CanMEDS Project. World J Surg 2003;27:972-8; discussion 978.
117. Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2002;97:1296-308.
118. Grantcharov TP, Reznick RK. Teaching procedural skills. BMJ 2008;336:1129-31.
119. Walsh CM, Sherlock ME, Ling SC, et al. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev 2012;6:CD008237.
120. Gillies D, Haritsis A, Williams C. Computer simulation for teaching endoscopic procedures. Endoscopy 1992;24 Suppl 2:544-8.
121. Mahmood T, Darzi A. A study to validate the colonoscopy simulator. Surg Endosc 2003;17:1583-9.
122. Verdaasdonk EG, Stassen LP, van Wijk RP, et al. The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. Surg Endosc 2007;21:214-9.
123. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. Acad Med 2005;80:S46-54.
124. 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.
125. Reznick R, Regehr G, MacRae H, et al. Testing technical skill via an innovative “bench station” examination. Am J Surg 1997;173:226-30.
126. Ekkelenkamp VE, Dowler K, Valori RM, et al. Patient comfort and quality in colonoscopy. World J Gastroenterol 2013;19:2355-2361.
127. Bowles CJ, Leicester R, Romaya C, et al. A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening tomorrow? Gut 2004;53:277-83.
128. Gavin DR, Valori RM, Anderson JT, et al. The national colonoscopy audit: A nationwide assessment of the quality and safety of colonoscopy in the UK. Gut 2013;62:242-249.

129. Aslinia F, Uradomo L, Steele A, et al. Quality assessment of colonoscopic cecal intubation: An analysis of 6 years of continuous practice at a University hospital. Am J Gastroenterol 2006;101:721-731.
130. R Valori BR. BSG Quality and safety indicators for endoscopy. JAG 2007/8.
131. Lieberman DA, Weiss DG, Bond JH, et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. N Engl J Med 2000;343:162-8.
132. Rex DK, Cutler CS, Lemmel GT, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology 1997;112:24-8.
133. Barclay RL, Vicari JJ, Greenlaw RL. Effect of a time-dependent colonoscopic withdrawal protocol on adenoma detection during screening colonoscopy. Clin Gastroenterol Hepatol 2008;6:1091-8.
134. Barclay RL, Vicari JJ, Doughty AS, et al. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. N Engl J Med 2006;355:2533-41.
135. Bressler B, Paszat LF, Chen Z, et al. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. Gastroenterology 2007;132:96-102.
136. Ferrandez A, Navarro M, Diez M, et al. Risk factors for advanced lesions undetected at prior colonoscopy: not always poor preparation. Endoscopy 2010;42:1071-6.
137. Singh H, Nugent Z, Demers AA, et al. Rate and predictors of early/missed colorectal cancers after colonoscopy in Manitoba: A population-based study. Am J Gastroenterol 2010;105:2588-2596.
138. Pohl H, Robertson DJ. Colorectal cancers detected after colonoscopy frequently result from missed lesions. Clin Gastroenterol Hepatol 2010;8:858-64.
139. Leung K, Pinsky P, Laiyemo AO, et al. Ongoing colorectal cancer risk despite surveillance colonoscopy: the Polyp Prevention Trial Continued Follow-up Study. Gastrointest Endosc 2010;71:111-7.
140. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. Am J Gastroenterol 2006;101:873-85.
141. 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.
142. Leaper M, Johnston MJ, Barclay M, et al. Reasons for failure to diagnose colorectal carcinoma at colonoscopy. Endoscopy 2004;36:499-503.
143. Lieberman DA, Weiss DG, Veterans Affairs Cooperative Study G. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. N Engl J Med 2001;345:555-60.
144. Neerincx M, Terhaar sive Droste JS, Mulder CJ, et al. Colonic work-up after incomplete colonoscopy: significant new findings during follow-up. Endoscopy 2010;42:730-5.
145. Radaelli F, Meucci G, Sgroi G, et al. Technical performance of colonoscopy: the key role of sedation/analgesia and other quality indicators. Am J Gastroenterol 2008;103:1122-30.
146. Arrowsmith JB, Gerstman BB, Fleischer DE, et al. Results from the American Society for Gastrointestinal Endoscopy/U.S. Food and Drug Administration collaborative study on complication rates and drug use during gastrointestinal endoscopy. Gastrointest Endosc 1991;37:421-7.
147. Chung YW, Han DS, Yoo KS, et al. Patient factors predictive of pain and difficulty during sedation-free colonoscopy: A prospective study in Korea. Dig Liver Dis 2007;39:872-876.
148. Kim WH, Cho YJ, Park JY, et al. Factors affecting insertion time and patient discomfort during colonoscopy. Gastrointest Endosc 2000;52:600-5.
149. Seip B, Bretthauer M, Dahler S, et al. Patient satisfaction with on-demand sedation for outpatient colonoscopy. Endoscopy 2010;42:639-646.
150. Eckardt AJ, Swales C, Bhattacharya K, et al. Open access colonoscopy in the training setting: Which factors affect patient satisfaction and pain? Endoscopy 2008;40:98-105.
151. Harewood GC. Relationship of colonoscopy completion rates and endoscopist features. Dig Dis Sci 2005;50:47-51.

152. Bell GD. Preparation, premedication, and surveillance. Endoscopy 2004;36:23-31.
153. Standards of Practice Committee, Lichtenstein DR, Jagannath S, et al. Sedation and anesthesia in GI endoscopy. Gastrointest Endosc 2008;68:205-16.
154. Cohen LB, Delegge MH, Aisenberg J, et al. AGA Institute review of endoscopic sedation. Gastroenterology 2007;133:675-701.
155. Thiis-Evensen E, Hoff GS, Sauar J, et al. Patient tolerance of colonoscopy without sedation during screening examination for colorectal polyps. Gastrointest Endosc 2000;52:606-10.
156. Francis DL, Rodriguez-Correa DT, Buchner A, et al. Application of a conversion factor to estimate the adenoma detection rate from the polyp detection rate. Gastrointest Endosc 2011;73:493-7.
157. Williams JE, Le TD, Faigel DO. Polypectomy rate as a quality measure for colonoscopy. Gastrointest Endosc 2011;73:498-506.
158. Lee TJ, Rutter MD, Blanks RG, et al. Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. Gut 2011.
159. Levin TR, Zhao W, Conell C, et al. Complications of colonoscopy in an integrated health care delivery system. Ann Intern Med 2006;145:880-6.
160. Chutkan RK, Ahmad AS, Cohen J, et al. ERCP core curriculum. Gastrointest Endosc 2006;63:361-76.
161. Baron TH, Petersen BT, Mergener K, et al. Quality indicators for endoscopic retrograde cholangiopancreatography. Am J Gastroenterol 2006;101:892-7.
162. NVMDL. Herstructurering opleiding Maag-Darm-Leverziekten - Eindtermen voor de opleiding tot Maag-Darm-Leverarts: http://www.mdl.nl/uploads/240/486/HOM_definitieve_versie_t.b.v._opleidersbijeenkomst.pdf [Dutch]. 2007 [Accessed 12 May 2011].
163. The Joint Advisory Group on GI Endoscopy. Colonoscopy - DOPS assessment form: www.thejag.org.uk, 2008 [Accessed 30 May 2013].
164. Ekkelenkamp VE, Koch AD, Haringsma J, et al. Quality evaluation through self-assessment: a novel method to gain insight into ERCP performance. Frontline Gastroenterol 2014;5:10-16.
165. Schutz SM, Abbott RM. Grading ERCPs by degree of difficulty: A new concept to produce more meaningful outcome data. Gastrointest Endosc 2000;51:535-539.
166. Osborn A, Parnes SJ. Osborn-Parnes approach to Creative Problem Solving.: www.cpsb.com, 1960 [Accessed 11 January 2012].
167. Jowell PS. Endoscopic Retrograde Cholangiopancreatography: Toward a Better Understanding of Competence. Endoscopy 1999;31:755-757.
168. Schlup MM, Williams SM, Barbezat GO. ERCP: a review of technical competency and workload in a small unit. Gastrointest Endosc 1997;46:48-52.
169. Vitale GC, Zavaleta CM, Vitale DS, et al. Training surgeons in endoscopic retrograde cholangiopancreatography. Surg Endosc 2006;20:149-52.
170. Jowell PS, Baillie J, Branch MS, et al. Quantitative assessment of procedural competence. A prospective study of training in endoscopic retrograde cholangiopancreatography. Ann Intern Med 1996;125:983-9.
171. ASGE/ACG. ASGE/ACG Taskforce on Quality in Endoscopy. Ensuring Competence in Endoscopy: http://www.asge.org/WorkArea/showcontent.aspx?id3384, 2011 [Accessed 13 September 2013].
172. Bjorkman DJ, Popp JW, Jr. Measuring the quality of endoscopy. Gastrointest Endosc 2006;63:S1-2.
173. Naylor G, Gatta L, Butler A, et al. Setting up a quality assurance program in endoscopy. Endoscopy 2003;35:701-7.
174. Williams EJ, Taylor S, Fairclough P, et al. Are we meeting the standards set for endoscopy? Results of a large-scale prospective survey of endoscopic retrograde cholangio-pancreatograph practice. Gut 2007;56:821-9.

175. Kapral C, Duller C, Wewalka F, et al. Case volume and outcome of endoscopic retrograde cholangiopancreatography: results of a nationwide Austrian benchmarking project. Endoscopy 2008;40:625-630.
176. Colton JB, Curran CC. Quality indicators, including complications, of ERCP in a community setting: a prospective study. Gastrointest Endosc 2009;70:457-467.
177. Enochsson L, Swahn F, Arnelo U, et al. Nationwide, population-based data from 11,074 ERCP procedures from the Swedish Registry for Gallstone Surgery and ERCP. Gastrointest Endosc 2010;72:1175-84, 1184 e1-3.
178. Cotton PB, Garrow DA, Gallagher J, et al. Risk factors for complications after ERCP: a multivariate analysis of 11,497 procedures over 12 years. Gastrointest Endosc 2009;70:80-8.
179. Jeurnink SM, Siersema PD, Steyerberg EW, et al. Predictors of complications after endoscopic retrograde cholangiopancreatography: a prognostic model for early discharge. Surg Endosc 2011;25:2892-900.
180. Francis NK, Hanna GB, Cresswell AB, et al. The performance of master surgeons on standard aptitude testing. Am J Surg 2001;182:30-3.
181. Cotton PB. Income and outcome metrics for the objective evaluation of ERCP and alternative methods. Gastrointest Endosc 2002;56:S283-90.
182. Cotton PB, Hawes RH, Barkun A, et al. Excellence in endoscopy: Toward practical metrics. Gastrointest Endosc 2006;63:286-291.
183. Christensen M, Matzen P, Schulze S, et al. Complications of ERCP: a prospective study. Gastrointest Endosc 2004;60:721-31.
184. Committee ASoP, Anderson MA, Fisher L, et al. Complications of ERCP. Gastrointest Endosc 2012;75:467-73.
185. Salminen P, Laine S, Gullichsen R. Severe and fatal complications after ERCP: analysis of 2555 procedures in a single experienced center. Surg Endosc 2008;22:1965-70.
186. Suissa A, Yassin K, Lavy A, et al. Outcome and early complications of ERCP: a prospective single center study. Hepatogastroenterology 2005;52:352-5.
187. Lee YT, Chan FK, Leung WK, et al. Comparison of EUS and ERCP in the investigation with suspected biliary obstruction caused by choledocholithiasis: a randomized study. Gastrointest Endosc 2008;67:660-8.
188. Petrov MS, Savides TJ. Systematic review of endoscopic ultrasonography versus endoscopic retrograde cholangiopancreatography for suspected choledocholithiasis. Br J Surg 2009;96:967-74.
189. Cotton PB, Eisen G, Romagnuolo J, et al. Grading the complexity of endoscopic procedures: results of an ASGE working party. Gastrointest Endosc 2011;73:868-74.
190. Cotton PB, Romagnuolo J, Faigel DO, et al. The ERCP quality network: a pilot study of benchmarking practice and performance. Am J Med Qual 2013;28:256-60.
191. Williams EJ, Ogollah R, Thomas P, et al. What predicts failed cannulation and therapy at ERCP? Results of a large-scale multicenter analysis. Endoscopy 2012;44:674-683.
192. Kapral C, Muhlberger A, Wewalka F, et al. Quality assessment of endoscopic retrograde cholangiopancreatography: results of a running nationwide Austrian benchmarking project after 5 years of implementation. Eur J Gastroenterol Hepatol 2012;24:1447-54.
193. DeBenedet AT, Elmunzer BJ, McCarthy ST, et al. Intraprocedural quality in endoscopic retrograde cholangiopancreatography: a meta-analysis. Am J Gastroenterol 2013;108:1696-704; quiz 1705.
194. Peng C, Nietert PJ, Cotton PB, et al. Predicting native papilla biliary cannulation success using a multinational Endoscopic Retrograde Cholangiopancreatography (ERCP) Quality Network. BMC Gastroenterol 2013;13:147.
195. Cote GA, Imler TD, Xu H, et al. Lower provider volume is associated with higher failure rates for endoscopic retrograde cholangiopancreatography. Med Care 2013;51:1040-7.

SUMMARY IN DUTCH

Introductie

Het opleiden van medisch specialisten heeft de afgelopen jaren veel aandacht gekregen. Enerzijds heeft de samenleving te maken met bezuinigingen en dient de specialistenopleiding gemiddeld genomen met een aantal maanden verkort te worden, anderzijds moeten er wel goede en bekwame artsen afgeleverd worden. Naast deze politieke discussie is ook de manier van opleiden onderwerp van debat.

Voor de opleiding tot Maag-Darm-Leverarts (MDL-arts) is dit niet anders. Gastro-intestinale endoscopie (kijkonderzoek van het maagdarmstelsel) vormt een belangrijk deel van de uiteindelijke beroepsuitoefening van de MDL-arts en heeft dus ook een groot aandeel in de opleiding. Tot op heden vindt certificering voor endoscopie plaats op basis van het aantal gedane procedures. De kwaliteit daarvan, of de kunde van de arts, wordt daarin niet meegenomen. De vraag is of de huidige criteria voldoende garantie bieden voor kwaliteit van endoscopie, of dat een meer competentiegerichte opleiding nodig is, waarbij de individuele bekwaamheid doorslaggevend is.

In dit proefschrift hebben we onderzocht hoe de leercurve van MDL-artsen in opleiding verloopt met betrekking tot colonoscopie (dikke darm onderzoek) en endoscopische retrograde cholangiopancreatografie, ook wel ERCP genoemd (onderzoek van de galwegen en/of alvleesklier). Daarnaast hebben we getracht factoren vast te stellen die van invloed zijn op die leercurve. Door middel van een anonieme enquête hebben we experts gevraagd wat zij nou belangrijke factoren vinden die een rol spelen bij een kwalitatief hoogwaardige colonoscopie. Naast het focus op het endoscopie-opleidingstraject hebben we ook de uitvoering van specialisten onderzocht met betrekking tot colonoscopie en ERCP, waarbij we onder andere hebben getracht voorspellers van falen te identificeren.

Belangrijkste resultaten

Allereerst wordt in **hoofdstuk 1** beschreven wat er zoal bekend is over opleiden en beoordelen binnen de gastro-intestinale endoscopie. We hebben een literatuuronderzoek verricht, waarbij we op een systematische manier de resultaten hebben beoordeeld en beschreven. Het resultaat is een samenvatting van de bestaande bewijzen voor verschillende meet- en toetsingsmethoden voor de endoscopie-opleiding, waarbij we onder andere concluderen dat gevalideerde ‘virtual reality’ simulatoren het leren van vaardigheden kunnen versnellen. Daarnaast worden leercurven als waardevol geacht bij het beoordelen van competentie binnen de endoscopie.

Colonoscopie

Hoofdstuk 3 beschrijft de uitkomsten van een enquête-onderzoek gehouden onder acht experts op het gebied van colonoscopie. Door middel van anonieme enquêtes en tussentijdse terugkoppeling van de uitkomsten aan de experts, is er consensus bereikt over wat de belangrijkste factoren zijn voor een kwalitatief hoogwaardige colonoscopie. Factoren op het gebied van zowel technische vaardigheden, als kennis en gedrag komen hierin naar voren. Ons advies is om deze factoren te implementeren in een beoordelingsformulier.

In **hoofdstuk 4** hebben we aangetoond dat het trainen van colonoscopie op een ‘virtual reality’ simulator een verbetering van de prestaties laat zien bij dokters die nog niet eerder

een scoop in handen hebben gehad. Dit is ook te vertalen naar verbeterde prestaties bij colonoscopie bij patiënten. In de praktijk betekent dit dat een arts in opleiding het eerste gedeelte van de leercurve zou kunnen doorlopen op een simulator, in plaats van meteen te starten met colonoscopie bij patiënten.

De invloed van een gestandaardiseerd beoordelings- en feedbackprogramma op de colonoscopie-leercurve wordt beschreven in **hoofdstuk 5**. Zestien artsen in opleiding (AIOS) tot MDL-arts uit verschillende ziekenhuizen hebben hun colonoscopieën geregistreerd. Ze ondergingen een beoordeling door een expert na blokken van ongeveer 50 gedane procedures. Tijdens de beoordeling deden de AIOS twee colonoscopieën bij patiënten. Op basis van deze procedures werd gericht en gestructureerd feedback gegeven. Vergeleken met een andere groep AIOS die dit beoordelingsprogramma niet hadden, hebben we aangetoond dat de leercurve met zo'n beoordeling steiler verloopt. Wij willen er voor pleiten om een dergelijke terugkerende, gestandaardiseerde beoordeling in de endoscopie-opleiding te implementeren.

In **hoofdstuk 6** hebben we gekeken naar de kwaliteit van colonoscopieën in een ziekenhuis in Gloucester, UK. Verschillende kwaliteitsindicatoren (volledigheid van de procedure, aantal poliepen gezien en verwijderd, gebruik van een roesje of pijnstilling) worden standaard geregistreerd. Daarnaast hebben de verpleegkundigen het comfort van de patiënt tijdens de procedure bijgehouden. Deze studie heeft aangetoond dat de 'beste dokters' de hoogste succespercentages halen met betrekking tot volledigheid van de colonoscopie en het aantal geziene poliepen. Ze gebruiken daarnaast minder medicatie voor een roesje en pijnstilling van de patiënt, en de patiënten van deze dokters waren juist het meest comfortabel tijdens het onderzoek. Ofwel: de beste dokters voeren kwalitatief de beste procedures uit en hebben de meest tevreden patiënten.

ERCP

Hoofdstuk 7 omvat het verloop van de leercurve van MDL-artsen in opleiding voor het endoscopische onderzoek van de galwegen en/of alvleesklier, de ERCP. We hebben hier aangetoond dat het gebruik van een zelfbeoordelingsformulier, het 'Rotterdam Assessment Form for ERCP (RAF-E)', inzicht geeft in de individuele en groepsprestaties van de dokters in opleiding. Met de gegevens uit het formulier kan een leercurve gemaakt worden voor verschillende onderdelen van de procedure. Dit biedt perspectief voor het meer individueel op maat opleiden van MDL-artsen. In **hoofdstuk 8** beschrijven we het gebruik van hetzelfde formulier, RAF-E, alleen dan door medisch specialisten. Aangezien de ERCP één van de meest lastige en risicovolle procedures is die een MDL-arts uitvoert, is het van belang om de uitkomsten van deze ERCP's in kaart te brengen. In deze studie hebben de MDL-artsen van één afdeling hun procedures geregistreerd. Met gegevens uit deze formulieren is inzicht verkregen in de procedurele kwaliteit van ERCP's en heeft de RAF-E zich nuttig bewezen voor het meten van prestaties van specialisten. Vervolgens beschrijft **hoofdstuk 9** een studie in het verlengde van hoofdstuk 8, namelijk het gebruik van RAF-E om inzicht te verkrijgen in de landelijke ERCP-praktijk. Alle MDL-artsen in Nederland die deze procedure verrichten, zijn uitgenodigd om een jaar lang hun ERCP's te registreren. We hebben de deelnamegraad van deze vrijwillige registratie in kaart gebracht en met ruim 8500 geregistreerde procedures hebben we zowel succespercentages als voorspellers van procedureel falen kunnen vaststellen. Het aantal gedane procedures per jaar door een MDL-arts heeft onder andere invloed op de uitkomst,

evenals de moeilijkheidsgraad. De basis voor een discussie over de toekomst van de ERCP's in Nederland is hiermee gelegd.

Conclusies

Op basis van de resultaten die gepresenteerd worden in dit proefschrift, kunnen we een aantal aanbevelingen doen.

Allereerst zou de endoscopie-opleiding moeten beginnen met een curriculum op de 'virtual reality' simulator. Vaardigheden die zijn aangeleerd op een simulator, vertalen zich naar colonoscopie bij patiënten. Wanneer op een simulator geen vooruitgang meer wordt geboekt, kan gestart worden bij patiënten.

Vervolgens moet competentie-ontwikkeling gemeten worden, bijvoorbeeld door middel van continue zelfbeoordeling. Op deze manier kan een leercurve gecreëerd worden. Daarnaast stellen we voor om een gestandaardiseerd beoordelings- en feedbackprogramma te implementeren om AIOS gerichter adviezen te kunnen geven en de leercurve te versnellen. Zowel technische als gedragsfactoren verdienen aandacht tijdens zo'n beoordeling.

Voor ERCP zou een vergelijkbare opzet waardevol kunnen zijn. Bij deze procedure kan eveneens een leercurve gecreëerd worden. Deze leercurve dient gemonitord te worden; real-time terugkoppeling verschaft inzicht in de vaardighedenontwikkeling van een AIOS. Tot slot zijn we van mening dat een kwaliteitsregistratie voor ERCP ingevoerd dient te worden voor MDL-artsen. Hiermee wordt bewustwording gecreëerd, en inzicht in eigen prestaties kan op zich al een kwaliteitsverbetering teweeg brengen. De extra administratieve last vormt een drempel voor een dekkende registratie, maar een koppeling met het endoscopieverslagsysteem en het registratiesysteem neemt deze last weg en is dus zeer wenselijk.

CONTRIBUTING AUTHORS

Henk R. van Buuren
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Frank ter Borg
Department of Gastroenterology and Hepatology, Deventer Hospital, Deventer, the Netherlands

Pieter C.J. ter Borg
Department of Gastroenterology and Hepatology, Ikazia Hospital, Rotterdam, the Netherlands

Gerard J.J.M. Borsboom
Department of Public Health, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Marco J. Bruno
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Kevin Dowler
Department of Gastroenterology, Gloucestershire Hospitals NHS Trust, Gloucester, UK

Paul Dunckley
Department of Gastroenterology, Gloucestershire Hospitals NHS Trust, Gloucester, UK

Marcel J.M. Groenen
Department of Gastroenterology and Hepatology, Rijnstate Hospital, Arnhem, the Netherlands

Jelle Haringsma
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands (*Current affiliation: Maasstad Hospital, Rotterdam, the Netherlands)*

Arjun D. Koch
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Ernst J. Kuipers
Department of Gastroenterology and Hepatology and Chair of the Board of Directors, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Wilco Lesterhuis
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Robert A. de Man
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Jan-Werner Poley
Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Erik A.J. Rauws
Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, the Netherlands

Erik J. Schoon
Department of Gastroenterology and Hepatology, Catharina Hospital, Eindhoven, the Netherlands

Antonie J.P. van Tilburg
Department of Gastroenterology and Hepatology, Sint Franciscus Gasthuis, Rotterdam, the Netherlands

Roland M. Valori
Department of Gastroenterology, Gloucestershire Hospitals NHS Trust, Gloucester, UK

LIST OF PUBLICATIONS

Ekkelenkamp VE, Koch AD, de Man RA, Kuipers EJ. *Training and competence assessment in gastrointestinal endoscopy: a systematic review.*
Submitted.

Ekkelenkamp VE, Koch AD, Lesterhuis W, de Man RA, Kuipers EJ. *Evaluating the learning curve in colonoscopy – the value of a repetitive assessment and feedback program.*
Submitted.

Ekkelenkamp VE, de Man RA, ter Borg F, ter Borg PCJ, Bruno MJ, Groenen MJM, van Tilburg AJP, Rauws AJ, Koch AD. *Prospective evaluation of ERCP performance – results of a nation-wide quality registry.*
Submitted.

Koch AD, **Ekkelenkamp VE**, Haringsma J, Schoon EJ, de Man RA, Kuipers EJ. *Prolonged colonoscopy simulator training leads to improved performance during patient-based assessment.*
Submitted.

Ekkelenkamp VE, Koch AD, Rauws EAJ, Borsboom GJJM, de Man RA, Kuipers EJ. *Competence development in ERCP – the learning curve of novice trainees.*
Endoscopy 2014; in press.

Ekkelenkamp VE, Koch AD, Haringsma J, Poley JW, van Buuren HR, Kuipers EJ, de Man RA. *Quality evaluation through self-assessment: a novel method to gain insight in ERCP performance.*
Frontline Gastroenterol 2014;5:10-16.

Ekkelenkamp VE, Koch AD, Haringsma J, Kuipers EJ, de Man RA. *Endoscopist-related factors contributing to high-quality colonoscopy; results of a Delphi survey.*
Perspectives on medical education 2014;3(1):31-40.

Ekkelenkamp VE, Dowler K, Valori RM, Dunckley P. *Patient comfort and quality in colonoscopy.*
World Journal of Gastroenterology 2013;19(15):2355-61.

PHD PORTFOLIO

Promotor: prof.dr. E.J. Kuipers
Co-promotor: dr. R.A. de Man

Erasmus MC University Medical Center Rotterdam, department of Gastroenterology and Hepatology

General academic and research skills

Biomedical English Writing and Communication
Biostatistical Methods I: Basic Principles, NIHES 2011
MolMed course Adobe Photoshop & Illustrator 2013

Presentations

A nationwide quality registry – endoscopist-related predictors of procedural outcome in ERCP.
Oral presentation, Digestive Disease Week, Chicago, USA 2014

ERCP quality in the Netherlands – the PERK study.
Oral presentation, NVGE Voorjaarsvergadering, Veldhoven 2014

Training in colonoscopy: the value of a standardized assessment program.
Poster presentation, United European Gastroenterology Week, Berlin 2013

Self-assessment of colonoscopy performance by trainees is feasible and compares well to expert trainer-assessment.
Poster presentation, Digestive Disease Week, Orlando, USA 2013

Assessment of ERCP performance in novice trainees.
Poster presentation, Digestive Disease Week, Orlando, USA 2013
Oral presentation, NVGE Voorjaarsvergadering, Veldhoven 2013

Ontwikkeling van competentie in ERCP: evaluatie van de leercurve.
Oral presentation, Congres Nederlandse Vereniging voor Medisch Onderwijs, Maastricht 2012

ERCP quality assessment and outcomes in a tertiary referral center.
Poster presentation, Digestive Disease Week, San Diego, USA 2012

ERCP quality assessment and outcomes in a regional setting.
Poster presentation, United European Gastroenterology Week, Amsterdam 2012

Endoscopist-related factors contributing to a high-quality colonoscopy, results of a Delphi survey.
Poster presentation, United European Gastroenterology Week, Amsterdam 2012

Patient comfort in colonoscopy performance.
Oral presentation, British Society of Gastroenterology, Birmingham, UK 2011
Oral presentation, NVGE Najaarsvergadering, Veldhoven, 2011
Poster presentation, Digestive Disease Week, San Diego, USA 2012

Prolonged colonoscopy simulator training leads to improved performance during patient-based assessment.
Poster presentation, Digestive Disease Week, Chicago, USA 2011

Developing competence in ERCP.
Oral presentation, NVGE Najaarsvergadering, Veldhoven 2010

International conferences

United European Gastroenterology Week (UEGW), Berlin 2013
Digestive Disease Week (DDW), Orlando 2013
Digestive Disease Week (DDW), San Diego 2012
United European Gastroenterology Week (UEGW), Amsterdam 2012
Digestive Disease Week (DDW), Chicago 2011

Seminars and workshops

Training ‘Omgaan met groepen’ 2013
Workshop ‘Grant application’ 2011
Workshop ‘How to write scientific English’ 2011
Workshop ‘Literature search, basic and advanced’ 2011
Seminar ‘Methodology of patient-based research’ 2011

Educational activities and lecturing

Tutoring bachelor students (Faculty of Medicine, Erasmus University) 2013
Lecture ‘Meten is weten – kwaliteit in endoscopie’. Seminar Ikazia Ziekenhuis 2013
Lecturing bachelor students ‘basics in endoscopy’ (Faculty of Medicine, Erasmus University) 2012
Lecturing bachelor (minor) students ‘simulation training in endoscopy’ (Faculty of Medicine, Erasmus University) 2012

ACKNOWLEDGEMENTS

Allereerst wil ik graag iedereen die heeft bijgedragen aan één of meerdere van mijn studies bedanken voor zijn waardevolle input. Alle AIOS, MDL-artsen, de PERK stuurgroep, de NVMDL, de MDL-staf van het Erasmus MC: zonder jullie inzet was dit onderzoek niet mogelijk geweest. Daarnaast wil ik ook de endoscopie-verpleegkundigen, baliemedewerkers en secretaresses bedanken voor de prettige samenwerking.

Promoveren doe je niet alleen: het afronden van een traject als dit is onmogelijk zonder de steun van vele anderen. Graag wil ik een aantal mensen in het bijzonder bedanken.

In de eerste plaats natuurlijk mijn promotor, prof.dr. E.J. Kuipers. Beste Ernst, jouw enthousiasme en immer positieve houding werken aanstekelijk, en hebben gezorgd voor niet alleen een succesvol, maar ook een ontzettend leuk promotietraject. Dank daarvoor. Het is bewonderenswaardig dat je dit enthousiasme bij vrijwel iedereen met wie jij te maken krijgt, weet over te brengen. Ik voel het als een eer om onder jouw supervisie te mogen promoveren.

Mijn co-promotor dr. R.A. de Man, beste Rob, jij hebt als grote kwaliteit om structuur in het grote geheel aan te brengen. Ook de abstracte onderwijskundige onderdelen wist jij te concretiseren, resulterend in een verheldering voor alle betrokkenen. Daarnaast wil ik je uiteraard danken voor het in mij gestelde vertrouwen om me op te leiden tot Maag-Darm-Leverarts.

Dr. A.D. Koch, Arjun, tijdens mijn wetenschappelijke stage werd ik aan je gekoppeld en dat bleek de start van een mooi samenwerkings- en promotietraject. Ik waardeer het dat je zo nauw betrokken bent geweest bij alle studies en altijd laagdrempelig te bereiken was. Naast een productieve samenwerking (zie hier het resultaat) heb ik vooral ook veel plezier beleefd aan de vrijdagmiddagen op de endoscopie. Fijn dat je je promotie nu ook hebt weten af te ronden en ik hoop dat we in de toekomst nog veel zullen samenwerken.

I would like to thank all members of my PhD committee for their willingness to participate in my PhD defense. Dr. Valori, dear Roland, I am honored that you are willing to visit Rotterdam for my thesis defense. My enthusiasm for learning curves and quality in endoscopy grew even larger during my Gastroenterology internship in Gloucester, thanks to you and Paul.

I would like to thank all contributing authors for their valuable input and pleasant collaboration.

Lieve collega's, wat is het toch een feest geweest om op het dak te promoveren. Het is heel bijzonder om met zo'n leuke groep lief en leed te delen, en om daarnaast af en toe ook de wetenschap tot een hoger niveau te tillen. Bruggenloopjes, skitrips, borrels, congressen, de ERNST. redactie: een greep uit de mooie activiteiten die ik met jullie heb meegemaakt. Dank voor alle hilarische, mooie en waardevolle momenten, zowel tijdens als na werktijd.

Lieve Edmée, na jaren gezelligheid op het dak en daarna ben ik vereerd dat je naast me wilt staan 5 september. Ik kijk uit naar alle toekomstige gezamenlijke MDL-congressen! Lieve Anne, vanaf het eerste moment in groepje 1 was het top en ik ben blij dat je zo'n goed

vriendinnetje bent geworden. We hebben al veel meegemaakt samen; heel fijn dat je mijn paranimf wilt zijn.

Clubgenootjes, teamies, groepje 1, en alle andere lieve vrienden: dank voor jullie interesse en alle ontspannende etentjes, feestjes, weekendjes en zondagen op het hockeyveld. Dat er nog maar vele mooie momenten mogen volgen!

Bob, Els, Ginette, René en kleine Gaël: leuk om een tweede familie in de buurt te hebben! Dank voor jullie interesse de afgelopen jaren.

Sander, dit is nu het eindresultaat van datgene waar ik me de afgelopen drie jaar mee bezig heb gehouden; wie had gedacht dat ik een boek zou schrijven? Ook al zijn we zo verschillend, ik ben blij dat je mijn broer bent. Gelukkig hebben we samen mooie herinneringen aan papa. Veel geluk voor jou en Rebecca.

Lieve papa, ik vind het vreselijk jammer dat je alleen maar mijn eerste stappen in de onderzoekswereld hebt meegemaakt. Ik had dit proefschrift zo graag aan je willen overhandigen. Ik hoop dat je er ergens toch wat van meekrijgt en dat je trots bent.

Lieve mam, zonder de onvoorwaardelijke steun en liefde van jou en pap was ik niet zo ver gekomen. Jullie hebben me altijd gestimuleerd om kansen aan te grijpen en om het beste uit mezelf te halen, en dit ook mogelijk gemaakt. Dank dat je er altijd voor mij en Sander bent. Ik ben trots op hoe jij je de afgelopen jaren staande hebt gehouden, tijdens papa's ziektegeschiedenis die langer is dan dit dankwoord, maar ook na zijn overlijden. Ik hoop dat je nog een mooie toekomst tegemoet gaat samen met Jan.

Lieve Hugo, wat is het toch fijn dat we een team zijn. Dat we nog maar veel mooie avonturen mogen beleven samen!

Vivian

ABOUT THE AUTHOR

Vivian Elske Ekkelenkamp was born on February 2nd, 1986 in Norg, the Netherlands. She finished her pre-university education in 2003 and started medical school at the University of Groningen in the same year. After obtaining her bachelor's degree, she did clinical rotations in several hospitals in the Netherlands and abroad in Korogwe (Tanzania) and Gloucester (UK). During these rotations, she developed a specific interest in Gastroenterology and Hepatology. Consequently, she performed research for her master's thesis in this field at the Erasmus MC University Medical Center Rotterdam. In 2011, she obtained her medical degree. Following her master's research, she started her PhD program at the department of Gastroenterology and Hepatology as described in this thesis. As of June 2014, she started with her two-year Internal Medicine residency at Reinier de Graaf Gasthuis in Delft (program director dr. E.F.M. Posthuma) as part of the formal postgraduate training. Hereafter, she will continue her training in Gastroenterology and Hepatology, both at Reinier de Graaf Gasthuis (program director dr. J.T. Brouwer) and Erasmus MC University Medical Center Rotterdam (program director dr. R.A. de Man).

The ultimate goal of a training and quality program in medicine is delivering competent doctors, who will improve to excellence after certification. This goal applies to training in gastrointestinal (GI) endoscopy as well. Nonetheless, evidence-based methods to accomplish competence and excellence in GI endoscopy, are lacking.
Therefore, Vivian Ekkelenkamp explored the use of learning curves for competence assessment of trainees regarding GI endoscopic procedures, such as endoscopic retrograde cholangiopancreatography (ERCP) and colonoscopy. Additionally, she investigated the value of a repetitive assessment and feedback program for colonoscopy trainees. She evaluated the quality of endoscopic procedures performed by certified gastroenterologists as well. The results of this research provide a rationale for the development of an improved training and quality program for GI endoscopy.
The research presented in this thesis was conducted from 2011 to 2014 at the department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, the Netherlands.

Competence & quality assessment
The future of training in GI endoscopy