



University of Southern Denmark

Intra- and interobserver agreement on polyp detection in colon capsule endoscopy evaluations

Buijs, Marleen; Krøijer, Rasmus ; Kobæk Larsen, Morten; Spada, Cristiano; Fernandez-Urien, Ignacio; Steele, Robert; Baatrup, Gunnar

Published in:
United European Gastroenterology Journal

DOI:
10.1177/2050640618798182

Publication date:
2018

Document version:
Final published version

Document license:
CC BY

Citation for pulished version (APA):
Buijs, M., Krøijer, R., Kobæk Larsen, M., Spada, C., Fernandez-Urien, I., Steele, R., & Baatrup, G. (2018). Intra- and interobserver agreement on polyp detection in colon capsule endoscopy evaluations. *United European Gastroenterology Journal*, 6(10), 1563-1568. <https://doi.org/10.1177/2050640618798182>

Go to publication entry in University of Southern Denmark's Research Portal


Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Intra and inter-observer agreement on polyp detection in colon capsule endoscopy evaluations

Maria Magdalena Buijs¹ , Rasmus Kroijer², Morten Kobaek-Larsen¹, Cristiano Spada^{3,4}, Ignacio Fernandez-Urien⁵, Robert JC Steele⁶ and Gunnar Baatrup¹

Abstract

Background: Colon capsule endoscopy is a promising technique for evaluation of the colon, but its reproducibility is still unknown.

Objective: This study assesses intra and inter-observer agreement in evaluations of colon capsule endoscopy videos.

Methods: Forty-two complete colon capsule endoscopy investigations were analysed by three experts and two beginners. Intra-observer agreement was assessed in paired readings of two experts and two beginners. Agreement was determined by the intraclass correlation coefficient: poor (<0.5), moderate (0.5–0.75), good (0.75–0.9) and excellent (>0.9).

Results: Agreement on ‘indication for a following colonoscopy’ based on the number and size of detected polyps and bowel cleansing quality was poor among all observers. Agreement among experts on the detection of large polyps and number of polyps was moderate, but agreement on bowel cleansing quality was poor. Beginners were in moderate agreement with the experts on polyp detection. Intra-observer agreement in experts was moderate to excellent for the detection of large polyps (≥ 10 mm), excellent for the number of polyps, and poor to moderate for bowel cleansing quality. Intra-observer agreement in beginners was poor to moderate for all variables.

Conclusions: This study shows a poor agreement on ‘indication for a following colonoscopy’, but a high intra and inter-observer agreement for polyp detection among experts, as well as a moderate agreement between beginners and experts.

Trial registration: NCT02303756.

Keywords

Colon capsule endoscopy, intra-observer agreement, inter-observer agreement, polyp detection, bowel cleansing

Received: 27 March 2018; accepted: 2 August 2018

Key summary

Summarise the established knowledge on this subject

- Colon capsule endoscopy is a promising diagnostic investigation.
- The reliability and repeatability of colon capsule endoscopy evaluation is currently unknown.

What are the significant and/or new findings of this study?

- Intra-observer agreement on polyp detection is moderate to excellent in experts.

¹Department of Surgery, Odense University Hospital, Svendborg, Denmark

²Department of Clinical Research, University of Southern Denmark, Odense, Denmark

³Digestive Endoscopy Unit, Catholic University, Rome, Italy

⁴Digestive Endoscopy Unit, Poliambulanza Foundation, Brescia, Italy

⁵Department of Gastroenterology, Complejo Hospitalario de Navarra, Pamplona, Spain

⁶Department of Surgery, Ninewells Hospital and Medical School, Dundee, Scotland

Corresponding author:

Maria Magdalena Buijs, Forskningsenheden, Baagøes Allé 15, 5700 Svendborg, Denmark.

Email: Maria.Magdalena.Buijs@rsyd.dk

- Inter-observer agreement on polyp detection is moderate among experts.
- Beginners are moderately in agreement with experts on polyp detection.
- Agreement on bowel cleansing quality is poor among all readers.

Introduction

In recent years colon capsule endoscopy (CCE) has emerged as a promising diagnostic technique to assess the colonic and rectal mucosa, for example as a possible filter test in colorectal cancer screening.¹⁻³ Most studies on CCE have focused on its accuracy in comparison to the gold standard colonoscopy and on bowel preparation regimens.¹⁻⁸ These studies indicate that the accuracy of polyp detection in CCE is comparable to that of colonoscopy. In a recent clinical trial comparing the accuracy of CCE to colonoscopy 53 participants (21%) underwent a second colonoscopy to investigate 'false positive' polyps in CCE. In 85% of these participants new polyps were detected.⁹ A total of 82 polyps was detected, of which 24 were 10 mm or greater, which suggests that the accuracy of polyp detection in CCE might be higher than in colonoscopy.

The guideline on CCE by the European Society of Gastrointestinal Endoscopy (ESGE) recommends a colonoscopy, if polyps sized 6 mm or greater or three or more polyps regardless of size are detected.¹⁰

The quality and reproducibility of CCE evaluations has not previously been assessed and the learning curve for CCE readers after formal training is as yet unknown. Leighton and Rex have proposed a grading scale for cleanliness in CCE assessing five segments, on either a four-point scale based on the amount of faecal matter or a two-point scale based on the amount of bubbles that interfere with a reliable assessment of the mucosa.¹¹ These scales have not been validated or compared to the polyp detection rate, as the Boston bowel preparation scale (BBPS) has been for colonoscopy.^{12,13}

The primary objective of this study was to determine intra and inter-observer agreement among experts and beginners on 'indication for colonoscopy', based on the number and size of polyps, and bowel cleansing quality. Secondary outcomes were the intra and inter-observer agreement between experts and beginners on the prevalence of polyps sized 10 mm or greater, the total number of polyps and quality of bowel cleansing.

Materials and methods

Study design

This study was conducted as an intra and inter-observer study in 42 complete CCE investigations with varying levels of bowel cleansing quality according to the primary assessment. Inter-observer agreement was assessed among three experts and two medical doctors

with formal training but no previous experience with CCE evaluations ('beginners') between November 2016 and March 2017. Intra-observer agreement was determined in two experts and two beginners who evaluated all videos twice. All 84 videos were presented in an anonymised and random order, in order to prevent the identification of identical videos.

Study population

CCE videos were selected from a total of 136 completed CCE videos recorded in screening participants with a positive faecal occult blood test who participated in a clinical trial.⁹ Thirty-two videos with good or fair cleansing and 10 videos with poor or unacceptable cleansing in the primary evaluation were selected by a research nurse who duplicated and renamed all 84 videos in a random order with identical videos at least 20 videos apart.

CCE readers

The two 'experts' are internationally recognised for their experience in CCE (CS and IFU). Both had reviewed over 1500 CCE videos and performed over 10,000 colonoscopies prior to their participation in this study. Both use the BBPS to assess cleansing in optical colonoscopy, but were unable to disclose their adenoma detection rate as yet.

Data from the primary evaluation in the clinical trial, performed by a commercial company (Corporate Health, Hamburg, Germany) were used as a third expert assessment. At this company a nurse reviewed the entire video and selected areas of interest. These were subsequently evaluated by a gastroenterologist who decided on the final findings. The nurses and gastroenterologists had at least 2 years of experience with CCE. They were unaware of this inter-observer study during their evaluations and were only informed afterwards.

The two beginners were medical doctors (RK and MMB) with no previous experience with CCE who completed the e-course by Given Imaging, before evaluating the CCE videos.¹⁴ One had previous colonoscopy experience with over 2000 colonoscopies performed and knowledge of the BBPS, and the other did not.

Data collection

Evaluation of the videos was performed with Rapid Reader Software (Medtronic, USA) in which images

of both video heads of the capsule are displayed with time stamps corresponding to the time since the ingestion of the CCE capsule.

In each video the reviewer recorded: the first caecal image, the last rectal image and the last passage through both the hepatic and splenic flexure; the quality of cleansing in right, transverse, left and total colon as good, fair, poor or unacceptable; the number of detected polyps and the time in minutes spent on the evaluation. Details of each polyp were recorded: the time stamp on the first image of the polyp; which of the two video heads detected the polyp; the size in mm measured with the software's polyp size estimation tool, and polyp morphology as flat, sessile, pedunculated, large mass, or unknown. The 'indication for a colonoscopy' was defined as either three or more polyps, polyps of 10 mm or greater and/or unacceptable bowel cleansing. Data from the database are available from the corresponding author upon reasonable request.

Statistics

Sample size calculation was based on an intraclass correlation coefficient (ICC) of 0.85 (95% confidence interval (CI) 0.75–0.95) and indicated a minimum number

of 31 videos if reviewed by two independent observers. A total of 42 videos was included, 32 with good or fair bowel cleansing and 10 with poor or unacceptable bowel cleansing in the primary assessment.

The number and maximum size of polyps were described as mean and standard deviations. All videos were classified based on the largest detected polyp, as no polyps, small polyps (≤ 9 mm) and large polyps (≥ 10 mm). Intra and inter-observer agreement were analysed with Cohen's K for binary values and ICC for both binary and numerical variables. ICC estimates and their 95% CIs were based on an absolute agreement, two-way mixed-effects model. ICC values of less than 0.5 are considered of poor reliability, between 0.5 and 0.75 as moderate, between 0.75 and 0.9 as good and over 0.9 as excellent reliability.¹⁵

All calculations were performed using Stata IC 15.0.

Results

The results of the primary round of evaluations of all 42 videos and the inter-observer agreement among experts and beginners are presented in Tables 1 and 2. Inter-observer agreement on 'indication for colonoscopy' was poor among the three experts, two beginners,

Table 1. Results of the evaluations per CCE reader.

	Expert 1	Expert 2	Expert 3	Beginner 1	Beginner 2
Colonoscopy indicated ^a	30 Videos	29 Videos	25 Videos	29 Videos	31 Videos
Number of polyps ^b	3.5 ± 2.4 (34/42)	4.6 ± 2.8 (35/42)	3.0 ± 1.9 (36/42)	3.9 ± 2.5 (35/42)	4.2 ± 3.0 (37/42)
Videos with no polyps	8	7	6	7	5
Videos with small polyps	14	16	13	10	14
Videos with large polyps	20	19	23	25	23
Average maximal size ^b	10.6 ± 4.6 (34/42)	10.8 ± 5.0 (35/42)	11.8 ± 6.2 (36/42)	13.3 ± 6.0 (35/42)	10.9 ± 6.1 (37/42)
Unacceptable cleansing	6	0	1	5	4

CCE: colon capsule endoscopy.

Small and large polyps are defined as less than or at least 10 mm, respectively. The mean number and maximal size of the polyps are based on all videos with detected polyps, and presented with corresponding standard deviations.

^aIndication for therapeutic colonoscopy based on algorithm: ≥ 3 polyps, at least one polyp ≥ 10 mm and/or unacceptable bowel cleansing.

^bBased on videos with detected polyps, number of videos is noted in parentheses.

Table 2. Inter-observer agreement.

	Experts	Beginners	All
Colonoscopy indicated	0.51 (0.33–0.67)	0.54 (0.29–0.72)	0.55 (0.42–0.69)
Large polyps (≥ 10 mm)	0.81 (0.71–0.89)	0.62 (0.39–0.77)	0.74 (0.63–0.83)
Number of polyps	0.70 (0.52–0.82)	0.62 (0.40–0.78)	0.67 (0.55–0.78)
Bowel cleansing quality	0.53 (0.33–0.70)	0.70 (0.51–0.83)	0.53 (0.33–0.70)

CCE: colon capsule endoscopy; ICC: intraclass correlation coefficient.

Inter-observer agreement among three experts, two beginners and all five CCE readers, assessed with ICC with 95% confidence interval.

The values printed in bold show moderate agreement, the other values present poor agreement.

Table 3. Intra-observer agreement.

	Expert 1	Expert 2	Beginner 1	Beginner 2
Colonoscopy indicated	0.79 (0.64–0.88)	0.89 (0.81–0.94)	0.73 (0.55–0.85)	0.56 (0.31–0.74)
Large polyps (≥ 10 mm)	0.95 (0.92–0.97)	0.81 (0.68–0.89)	0.71 (0.53–0.84)	0.62 (0.39–0.78)
Number of polyps	0.98 (0.96–0.98)	0.99 (0.98–0.99)	0.76 (0.60–0.87)	0.74 (0.56–0.85)
Bowel cleansing quality	0.75 (0.57–0.85)	0.69 (0.49–0.82)	0.69 (0.50–0.82)	0.69 (0.49–0.82)

ICC: intraclass correlation coefficient.

Intra-observer agreement in two experts and two beginners, assessed with ICC with 95% confidence interval.

Values printed in bold show good and excellent agreement, the other values show moderate and poor agreement.

and all five readers together. Inter-observer agreement on which videos contained large polyps and the number of detected polyps was moderate among both experts and all five readers.

Agreement on the classification of bowel cleansing quality was moderate between beginners, and poor among experts and all five readers. A subanalysis on agreement on ‘indication for colonoscopy’ based only on the size and number of polyps did not improve agreement among the readers.

Intra-observer agreement was determined by evaluating the double reading of the videos and is presented in Table 3. Experts had a higher intra-observer agreement for ‘indication for colonoscopy’, detection of large polyps and the number of detected polyps than beginners. For both experts agreement on the number of polyps was excellent, as was the detection of large polyps in expert 1. Agreement on quality of bowel cleansing in experts was moderate and poor, similar to the beginners.

Discussion

Agreement on the calculated variable ‘indication for colonoscopy’ in this study was poor in both experts and beginners. A subanalysis that excluded bowel cleansing quality did not improve agreement on ‘indication for colonoscopy’, therefore disagreement on bowel cleansing quality is not the only factor that influences inter-observer agreement. Nonetheless, intra-observer agreement in expert 1 increased from moderate to excellent on the exclusion of bowel cleansing quality in this subanalysis (data not displayed).

Beginners had a lower intra and inter-observer agreement for the detection of large polyps and the total number of polyps, which shows that having experience with evaluating CCE investigations is likely to influence the number of detected polyps and their size measurement.

The excellent agreement on the detection of large polyps and the number of polyps by experts suggests that experience contributes to the consistent detection and measuring of polyps within one reader. The lower

intra-observer agreement on bowel cleansing quality indicates that experience does not necessarily contribute to a reproducible evaluation of bowel cleansing quality.

Bowel cleansing quality assessments need to be improved in order to increase the reproducibility and thereby reliability of CCE evaluations. Bowel cleansing quality is a more subjective assessment than detection and measuring of polyps and might become more consistent with training. Prior studies show that clearly defined guidelines and a short training increased the inter-observer agreement on bowel cleansing quality in colonoscopy.¹² In the current CCE e-course¹⁴ evaluation of bowel cleansing quality plays a minor role. A specific training on how to assess bowel cleansing quality might therefore contribute to a more consistent evaluation of cleanliness. In order to develop this training, consensus is needed on how to classify unacceptable bowel cleansing in CCE. A possibility would be to use the Leighton–Rex scale, nonetheless there are limitations to the use of this scale.¹¹ The Leighton–Rex scale has not been validated yet, assesses cleansing in five bowel segments that are not always easy to recognise in CCE, and provides no instructions on how to classify cleanliness of the whole video.

There are several limitations to the methods of this study. First, the relatively low number of paired videos could lead to both over and underestimation of agreement. It might also lead to enhanced intra-observer agreement due to observer recollection of a previously evaluated identical video. Second, this study only focusses on CCE evaluations, thereby not assessing variability between two separate CCE examinations in the same subject, which might have a poorer inter-observer agreement. Third, we only provided the observers with the choice of four different bowel classes without concrete guidelines specifying how to assess bowel cleansing, thus possibly causing some inter-observer variation. Fourth, we used a different classification for referral for a colonoscopy to the ESGE guidelines, instead of a 6 mm limit we used a 10 mm limit in this study.

A difficulty of this study is the lack of gold standards for CCE evaluations. It would make sense to use colonoscopy as the gold standard. However, inconsistencies between the CCE evaluations and colonoscopies might also be due to differences other than intra and inter-observer variability. Colonoscopy has been known to miss up to 24% of all polyps, whereas CCE might detect the same polyp multiple times due to the peristaltic movement of the capsule in the colon.^{16,17} Polyp size measurement in colonoscopy has been known to vary significantly between optical estimation and pathology measurements,^{18–20} whereas the quality of polyp size estimation in CCE is still unknown. These differences might lead to an under or overestimation of the variability, therefore we chose to assess the consistency of CCE evaluations by and within observers without taking a gold standard into account.

The intra and inter-observer agreement on the number of detected polyps has not yet been studied in colonoscopy. Consequently, we cannot compare the repeatability of CCE evaluations with that of colonoscopies. However, the intra and inter-observer agreement for polyp detection among experts is very high in this study, which indicates a high repeatability for CCE evaluations.

One study on inter-observer variation in polyp size has been published, presenting an ICC of visual estimation of 0.84 (95% CI 0.78–0.90) in experts and 0.69 (95% CI 0.60–0.79) in beginners.²¹ The inter-observer agreement in both beginners and experts is quite similar compared to our study; however, the number of observers is 40 and therefore not directly relatable to our study.

Future studies on CCE should focus on analysing the reliability of polyp size estimation and generating a higher agreement on bowel cleansing quality. The current polyp size estimation tool is likely to underestimate polyps that are far away and overestimate polyp size for polyps that are close to the camera. The quality of the size estimation might accordingly be poor despite the very good inter-observer agreement. Moreover, it is currently unknown how the polyp sizes in CCE correspond to polyp sizes in colonoscopy and pathology. The pathological polyp size could be used as a standard for polyp sizes in CCE.

A possibility in assessing bowel cleansing quality is the development of an objective evaluation of bowel cleansing quality, for example computerised analysis of the images. Additional training for assessing cleanliness or clearer guidelines on how to assess bowel cleansing quality might also improve inter-observer agreement.

Conclusion

This study shows a poor agreement on ‘indication for a following colonoscopy’ among all observers. The intra

and inter-observer agreement on the detection of large polyps and the number of polyps in CCE is high among experts, but not significantly higher than among doctors with formal training only. Evaluation of bowel cleansing is a more subjective assessment, with lower intra and inter-observer agreement in experts and beginners. Bowel cleansing assessment and polyp size estimations influence the indication for a following colonoscopy greatly. Consequently, future research should focus on investigating the quality of polyp size estimations and improving assessments of bowel cleansing quality in CCE in order to increase the reproducibility of CCE investigations.

Declaration of conflicting interests

Medtronic® funded the expert evaluations and provided financial support in the purchase of the colon capsules; but had no definitive influence on the study design and manuscript. Independent financial support was also obtained from the Danish Cancer Society, OUH research fund and the Region of Southern Denmark. There are no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval

All videos in this study were anonymised and could not be traced back to the corresponding screening participants, hence approval by the ethical committee was not required for this study. Transfer from the original trial of the anonymised CCE videos to this study was conducted in accordance with and approved by the data protection agency.

Funding

Medtronic funded the expert evaluations and provided financial support in the purchase of the colon capsules; but had no definitive influence on the study design and manuscript. Independent financial support was also obtained from the Danish Cancer Society, OUH research fund and the region of Southern Denmark. There are no other relationships or activities that could appear to have influenced the submitted work.

Informed consent

Screening participants signed an informed consent form for participation in the clinical trial and consecutive scientific use of their CCE videos.

ORCID iD

Maria Magdalena Buijs  <http://orcid.org/0000-0001-7206-0713>

References

1. Spada C, Pasha SF, Gross SA, et al. Accuracy of first- and second-generation colon capsules in endoscopic detection of colorectal polyps: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol: the official*

- clinical practice journal of the American Gastroenterological Association* 2016; 14: 1533–1543.
2. Rex DK, Adler SN, Aisenberg J, et al. Accuracy of capsule colonoscopy in detecting colorectal polyps in a screening population. *Gastroenterology* 2015; 148: 948–957.
 3. Holleran G, Leen R, O'Morain C, et al. Colon capsule endoscopy as possible filter test for colonoscopy selection in a screening population with positive fecal immunology. *Endoscopy* 2014; 46: 473–478.
 4. Spada C, Riccioni ME, Hassan C, et al. PillCam colon capsule endoscopy: a prospective, randomized trial comparing two regimens of preparation. *J Clin Gastroenterol* 2011; 45: 119–124.
 5. Rondonotti E and Pennazio M. Colorectal polyp diagnosis: results with the second-generation colon capsule (CCE-2). *Colorectal Dis: the official journal of the Association of Coloproctology of Great Britain and Ireland* 2015; 17(Suppl. 1): 31–35.
 6. Hartmann D, Keuchel M, Philipper M, et al. A pilot study evaluating a new low-volume colon cleansing procedure for capsule colonoscopy. *Endoscopy* 2012; 44: 482–486.
 7. Kakugawa Y, Saito Y, Saito S, et al. New reduced volume preparation regimen in colon capsule endoscopy. *World J Gastroenterol* 2012; 18: 2092–2098.
 8. Singhal S, Nigar S, Paleti V, et al. Bowel preparation regimens for colon capsule endoscopy: a review. *Therapeut Adv Gastroenterol* 2014; 7: 115–122.
 9. Kobaek-Larsen M, Kroijer R, Dyrvig AK, et al. Back-to-back colon capsule endoscopy and optical colonoscopy in colorectal cancer screening individuals. *Colorectal Dis: the official journal of the Association of Coloproctology of Great Britain and Ireland* 2017; 20: 479–485.
 10. Spada C, Hassan C, Galliche JP, et al. Colon capsule endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2012; 44: 527–536.
 11. Leighton JA and Rex DK. A grading scale to evaluate colon cleansing for the PillCam COLON capsule: a reliability study. *Endoscopy* 2011; 43: 123–127.
 12. Calderwood AH and Jacobson BC. Comprehensive validation of the Boston Bowel Preparation Scale. *Gastrointest Endosc* 2010; 72: 686–692.
 13. Parmar R, Martel M, Rostom A, et al. Validated scales for colon cleansing: a systematic review. *Am J Gastroenterol* 2016; 111: 197–204.
 14. Given Imaging. Course: Pillcam Colon. 2015. CCE e-learning. <http://eu.cce-learning.com> (accessed 30 June 2017).
 15. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropract Med* 2016; 15: 155–163.
 16. Rex DK, Cutler CS, Lemmel GT, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* 1997; 112: 24–28.
 17. Pickhardt PJ, Nugent PA, Mysliwiec PA, et al. Location of adenomas missed by optical colonoscopy. *Ann Intern Med* 2004; 141: 352–359.
 18. Schoen RE, Gerber LD and Margulies C. The pathologic measurement of polyp size is preferable to the endoscopic estimate. *Gastrointest Endosc* 1997; 46: 492–496.
 19. Anderson BW, Smyrk TC, Anderson KS, et al. Endoscopic overestimation of colorectal polyp size. *Gastrointest Endosc* 2016; 83: 201–208.
 20. Jin HY and Leng Q. Use of disposable graduated biopsy forceps improves accuracy of polyp size measurements during endoscopy. *World J Gastroenterol* 2015; 21: 623–628.
 21. Kim JH, Park SJ, Lee JH, et al. Is forceps more useful than visualization for measurement of colon polyp size? *World J Gastroenterol* 2016; 22: 3220–3226.