

Real-time, computer-aided, detection-assisted colonoscopy eliminates differences in adenoma detection rate between trainee and experienced endoscopists



Authors

Giuseppe Biscaglia^{*1}, Francesco Cocomazzi^{*1}, Marco Gentile¹, Iliaria Loconte², Alessia Mileti², Rosa Paolillo², Antonella Marra¹, Stefano Castellana³, Tommaso Mazza³, Alfredo Di Leo², Francesco Perri¹

Institutions

- 1 Division of Gastroenterology and Endoscopy, “Casa Sollievo della Sofferenza” Hospital, IRCCS, San Giovanni Rotondo, Italy
- 2 Section of Gastroenterology, Department of Emergency and Organ Transplantation, University of Bari, Italy
- 3 Laboratory of Bioinformatics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy

submitted 22.7.2021

accepted after revision 3.1.2022

Bibliography

Endosc Int Open 2022; 10: E616–E621

DOI 10.1055/a-1783-9678

ISSN 2364-3722

© 2022. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,

70469 Stuttgart, Germany

Corresponding author

Francesco Cocomazzi, MD, Division of Gastroenterology and Endoscopy, “Casa Sollievo della Sofferenza” Hospital, IRCCS, San Giovanni Rotondo, Italy

Fax: +39 0882410274

francescococomazzi@gmail.com

ABSTRACT

Background and study aims Adenoma detection rate (ADR) is a well-accepted quality indicator of screening colonoscopy. In recent years, the added value of artificial intelligence (AI) has been demonstrated in terms of ADR and adenoma miss rate (AMR). To date, there are no studies evaluating the impact of AI on the performance of trainee endoscopists (TEs). This study aimed to assess whether AI might eliminate any difference in ADR or AMR between TEs and experienced endoscopists (EEs).

Patients and methods We performed a prospective observational study in 45 subjects referred for screening colonoscopy. A same-day tandem examination was carried out for each patient by a TE with the AI assistance and subsequently by an EE unaware of the lesions detected by the TE. Besides ADR and AMR, we also calculated for each subgroup of endoscopists the adenoma per colonoscopy (APC), polyp detection rate (PDR), polyp per colonoscopy (PPC) and polyp miss rate (PMR). Subgroup analyses according to size, morphology, and site were also performed.

Results ADR, APC, PDR, and PPC of AI-supported TEs were 38%, 0.93, 62%, 1.93, respectively. The corresponding parameters for EEs were 40%, 1.07, 58%, 2.22. No significant difference was found for each analysis between the two groups ($P > 0.05$). AMR and PMR for AI-assisted TEs were 12.5% and 13%, respectively. Sub-analyses did not show any significant difference ($P > 0.05$) between the two categories of operators.

Conclusions In this single-center prospective study, the possible impact of AI on endoscopist quality training was demonstrated. In the future, this could result in better efficacy of screening colonoscopy by reducing the incidence of interval or missed cancers.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death worldwide

[1–2]. In the early 1970s, Basil Morson was the first to speculate that polyps could be the precursors of CRC [3]. Later, the National Polyp Study demonstrated how CRC could be prevented by removing adenomatous polyps during colonoscopy [4]. Because CRC becomes symptomatic only in advanced stages, screening programs are being implemented to increase early

* These authors equally contributed to this work.

diagnosis and reduce morbidity and mortality [2]. Colonoscopy is the primary screening and surveillance diagnostic tool and its clinical impact in preventing advanced neoplasia depends on the endoscopist's ability to identify not only easily recognizable polypoid lesions but also non-polypoid flat or depressed superficial neoplastic lesions (SNL) [5]. In fact, evidence has shown that CRC also can develop from non-polypoid or serrated-type lesions, which are predominant in the right colon [6, 7]. Inadequate interpretation of these lesions would explain, at least in part, one of the causes of failure to prevent right or interval colon cancer [8].

During colonic examination, the first step is identifying all mucosal irregularities with white light endoscopy (WLE) [9]. The operator's determination to search for the slightest variation in appearance of the mucosa, use of high-definition (HD) endoscopes, adequate instrumental retraction time (at least 6–8 minutes), a good exploration technique, and complete intestinal cleansing are fundamental to detect a colonic lesion [6, 9]. All these elements, required for a "quality colonoscopy," are in any case insufficient if not supported by the knowledge of the existence of various types of lesions. It is, in fact, essential to know what and where to look [10].

Use of traditional chromoendoscopy has proven to improve detection of lesions, especially those that are not polypoid. On the other hand, comparison between HD-WLE and virtual chromoendoscopy did not show significant differences either in the polyp detection rate (PDR) or in the adenoma detection rate (ADR), even if a recent meta-analysis showed a slight but marginal superiority of the latter [10–12]. In the last few years, new approaches have been proposed to improve ADR and other quality colonoscopy indicators, including: devices to be mounted on the instrument (e. g., Endocuff, Endorings); delayed-release capsules based on methylene blue (MB-MMX); systematic changes of posture in the retraction phase; double inspection of the right colon; use of the water-exchange technique; and use of artificial intelligence (AI) [10]. Despite these proposed approaches, which unfortunately are not available or used everywhere, high adenoma miss rates (AMR) and unacceptable ADRs among colonoscopists continue to persist, increasing the risk of interval cancer [13]. Several studies have shown the added value in terms of ADR, adenoma per colonoscopy (APC), and AMR using AI, indicating that it is influenced neither by the size nor by the morphology of lesions, unlike the human eye [14, 15]. However, no advantage in detecting an advanced adenoma was found [14, 16]. Reaction time in recognizing a lesion also was different with AI [13].

Tandem colonoscopy is considered the standard for assessing ADR and AMR. In the past, literature based on this protocol were made to evaluate trainees' required experience to develop competence in performing a screening examination. One of these reports showed that about 450 colonoscopies are needed to achieve an AMR of less than 25% in patients older than 60 years, which is close to 22% to 26% of the literature [15–17]. Wang et al. [18], in a tandem colonoscopy study with the same endoscopist, demonstrated the advantage of AI in reducing the overall miss rate for adenomas. Lui et al. [15] have already shown that AI may benefit AMR for less experienced colo-

noscopists, but no one has demonstrated its role in a tandem colonoscopy model with trainees.

In the present study, a tandem model was adopted in which a trainee endoscopist (TE) carried out a colonoscopy with the help of AI, and his/her performance in terms of ADR, APC, AMR, PDR, PPC, and PMR was compared with that of an experienced endoscopist (EE) without AI.

Patients and methods

Study design

This prospective, single-center, observational study was performed at the Division of Gastroenterology and Digestive Endoscopy of "Casa Sollievo della Sofferenza" Hospital – IRCCS in San Giovanni Rotondo (Italy). Our center is part of the endoscopic training network of the Postgraduate School of Gastroenterology of the University of Bari (Italy). Patients aged 50 to 69 years undergoing colonoscopies for screening were recruited. History of inflammatory bowel disease, colorectal cancer, poor bowel cleansing, previous colorectal surgery, or contraindications for polypectomy were exclusion criteria. Our Ethics Committee approved the protocol, and informed consent was obtained from each participant before the investigation.

Study procedure

Seven investigators, four TEs (I.L., A.M., R.P., A.M.), and three EEs (G.B., M.G., F.P.) performed examinations. Their specialty or in progress specialty was gastroenterology. The mean age of the TEs was 33; that of the EEs was 51. TEs had initial experience with at least 200 colonoscopies (200–400 exams) while each EE had carried out at least 10,000 exams in his or her professional life. After patients received conscious sedation with meperidine and midazolam, we sequentially performed a tandem colonoscopy with HD colonoscopes (CF-Q 180, CF-H 185, CF-H 190, and CF-HQ 190, Olympus Medical Systems, Tokyo, Japan) in all eligible individuals. The TE first carried out a complete investigation with a computer-aided polyp detection (CADE) system assistance (GI-Genius; Medtronic, Minneapolis, Minnesota, United States). Cecal intubation was demonstrated by identification of the appendix orifice and the ileocecal valve. A withdrawal time of at least 6 minutes was required. Then, the EE performed a standard, white light, blinded colonoscopy, removing all detected lesions. Patient age, bowel preparation using the Boston Bowel Preparation Scale (BBPS), location, number, size, Paris [5], and NICE [19] classifications of all detected SNL were reported by each investigator and recorded by the sedation nurse. The site was divided into proximal and distal according to the lesion distance from splenic flexure. Diminutive polyps (≤ 5 mm) were not classified according to the Paris classification but considered as a separate morphologic category.

Lesions found by the TE in the first exam and "rediscovered" by the EE were considered as "found lesions," whereas lesions detected only during the second procedure were considered as "missed" ones. Each procedure was videotaped and promptly watched at the end of the tandem protocol to avoid any lesions being forgotten and not removed. For the same reason,

SNL features and sites reported by each endoscopist were compared. The colonic segment was reevaluated if missed lesions were found by an endoscopist but not the other. Histopathology was adopted as the reference standard.

When the EE had to intervene because the TE had difficulty completing the examination, the procedure was continued by the EE and the patient was consequently excluded from the tandem protocol and the analysis; the same procedure was followed when the BBPS was less than 6.

Outcomes

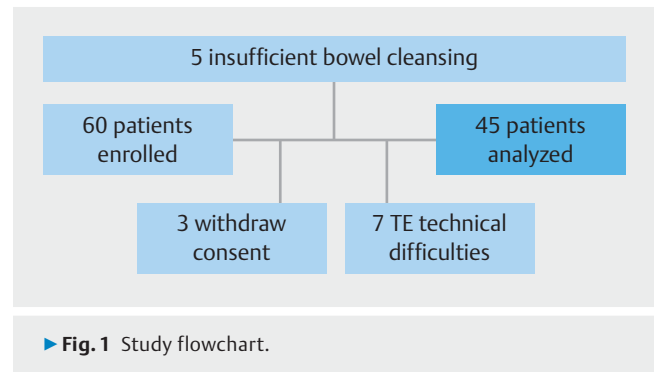
Considering the additional benefit of AI in detecting adenomas irrespective of size, site, and morphology [14], our primary outcome was to demonstrate that AI-supported TE ADR was not statistically lower than that for an EE without AI. For this purpose, we calculated ADR for each endoscopist's category and compared the results. ADR was defined as the proportion of subjects with at least one histologically proven adenoma [20]. APC, PDR, and polyp per colonoscopy (PPC) for each group of endoscopists were also calculated and compared. APC was defined as the total number of adenomas divided by the number of colonoscopies performed [18, 20]. PDR was defined as the proportion of colonoscopies with at least one polyp [21]. PPC was defined as the total number of polyps divided by the total number of patients [18]. AI-assisted TE AMR and PMR were evaluated. AMR was defined as the number of missed adenomas divided by the total number of adenomas found in both procedures [17–18]. PMR was defined as the number of missed polyps divided by the total number of polyps detected in both passes in which the unresected, diminutive, NICE 1 lesions of the rectum were also included [18]. Sub-analyses according to size, morphology, and site were also performed.

Statistical analysis

Observed proportions were compared using the Chi-Square Statistic and the `chisqtestClust` method of the `hstestClust R` package [22], when appropriate. The latter involves a set of reweighted marginal hypothesis tests for clustered data and, in particular, a chi-squared contingency table test and goodness-of-fit for clustered data based upon which counts referred to which patient. Statistical evaluations were performed using R version 3.6.0. Data are shown as counts of success events and $P < 0.005$ was considered statistically significant.

Results

The TE had between 8 and 14 months of endoscopic training. A total of 60 patients were enrolled in this study. In five of them, intestinal cleansing was considered insufficient; seven times the EE had to intervene due to difficulties on the part of the TE; three patients withdrew their consent during the procedure. Finally, 45 patients were considered eligible and analyzed (30 men; mean age: 59 years) (► Fig. 1). This sample size allowed this study to capture medium effect sizes ($d = 0.53$) with a power = 0.8, and α error probability of 0.05. Overall, the TEs found 87 lesions: 20 polyps (>5 mm), 20 non-polypoid lesions (NPL) (>5 mm), and 47 diminutive ones (► Fig. 2). The PPC was 1.93. Forty-two of 87 lesions were found in 17 subjects and his-



tologically characterized as adenomas [ADR: 38% (17/45); APC: 0.93]. Of 45 non-adenomatous lesions, at least one was detected in the other 11 patients who had no adenomas with a PDR of 62% (28/45).

The EE found 100 lesions: 22 polyps, 28 NPL, and 50 diminutive lesions. Of these, 48 adenomas were detected in 18 patients. The PPC was 2.22 and the ADR was 40% (18/45) with an APC of 1.07. Of 52 non-adenomatous lesions, at least one was detected in the other eight patients who had no adenomas with a PDR of 58% (26/45).

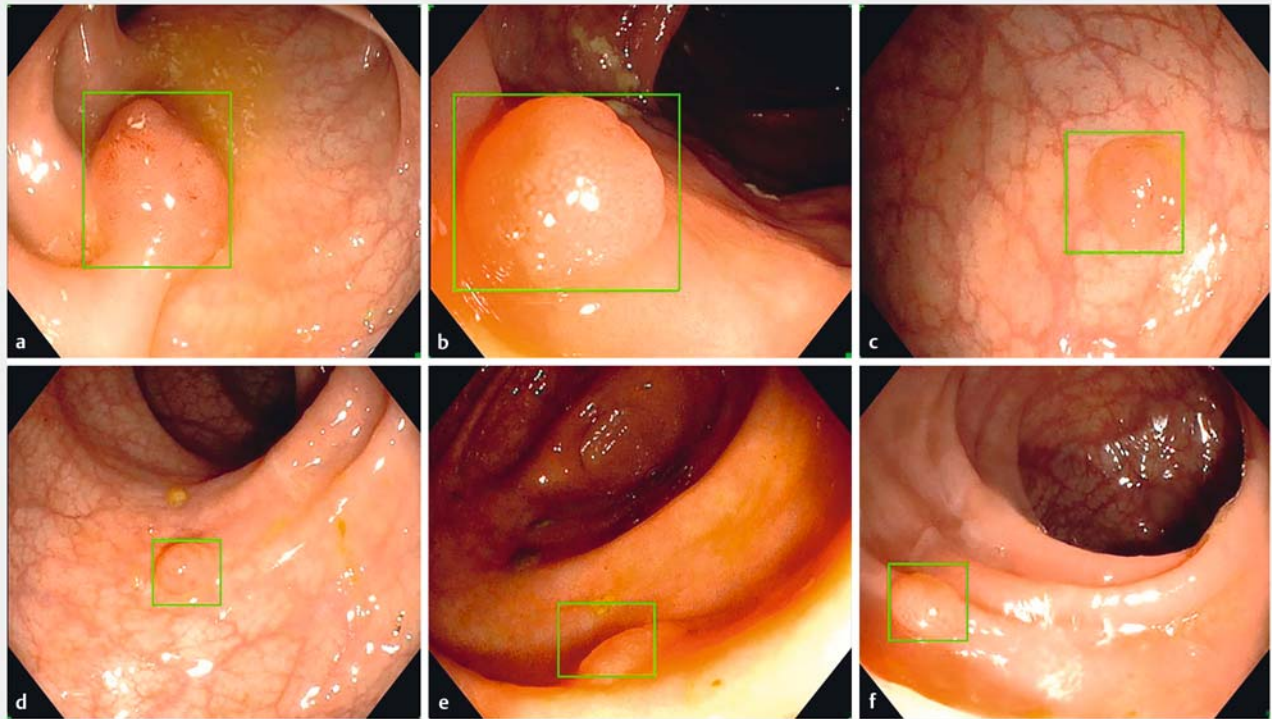
When the findings of the TEs were compared with those of the EE, the AMR and the PMR for the TEs were 12.5% (6/48) and 13% (13/100), respectively. No significant difference in the rates of all “quality” indicators between CAde-assisted TEs and EEs was found (► Fig. 3).

Regarding morphology, the rate of NPL (23% for TEs vs. 28% for EE) was slightly different but not statistically significant ($P = 0.86$). The rates of polyps (23% for TEs vs. 22% for EEs) and diminutive (54% for TEs vs. 50% for EEs) lesions were also comparable between the two groups ($P = 1$ and 0.91, respectively). Regarding the size, diminutive polyps were the predominant lesions (47/87 for TEs and 50/100 for EEs). However, no significant difference was found between lesions more than (46% for TEs vs. 50% for EEs) or ≤ 5 mm ($P = 0.90$ and 0.91, respectively). Finally, as regards location, detected lesions were mainly distal (54% for TEs vs. 55% for EEs), without any significant difference between the two categories of endoscopists. These results are summarized in ► Fig. 3.

Discussion

CAde systems are revolutionary devices that are increasingly being used in clinical setting for location and characterization of lesions in medical images. In gastrointestinal endoscopy, these systems, based on AI, mainly have been used during colonoscopy because of their ability to distinguish between abnormal and normal mucosa and detect and characterize colonic lesions [23].

One of the most challenging tasks for endoscopists is identifying and removing all adenomatous colonic lesions during colonoscopy. This activity can be evaluated by the ADR, defined as the rate at which a physician finds one or more precancerous polyps during a normal screening colonoscopy procedure for patients over 50 years old. Professional societies have deter-



► **Fig. 2** Examples of lesions detected with artificial intelligence(video stills).

► **Fig. 3** Summary of results.

		Trainees + AI group (45)	Expert endoscopists group (45)	P value	
	ADR	38% (17)	40% (18)	1	
	APC	0.93 (42)	1.07 (48)	1	
	PDR	62% (28)	58% (26)	0.72	
	PPC	1.93 (87)	2.22 (100)	0.69	
	AMR	12.5%			
	PMR	13%			
		Detections (87)	Detections (100)		
Morphology	}	NPL	23% (20)	28% (28)	0.86
		Polyps	23% (20)	22% (22)	1
Size	}	Diminutive	54% (47)	50% (50)	0.91
		Lesions > 5 mm	46% (40)	50% (50)	0.90
Location	}	Proximal	46% (40)	45% (45)	1
		Distal	54% (47)	55% (55)	1

AI, artificial intelligence; ADR, adenoma detection rate; APC, adenoma per colonoscopy; PDR, polyp detection rate; PPC, polyp per colonoscopy; AMR, adenoma miss rate; PMR, polyp miss rate; NPL, non-polypoid lesion.

mined that the benchmark rate should be at least 25% (30% in men, 20% in women) [24]. In our single-center study, we found 38% and 40% values for AI-assisted TEs and EEs, demonstrating their comparative performance. Although calculated among

TEs, ADR in our CADe group was in line with the pooled AI-ADR reported in the literature (36.6%) [14], but lower than the only GI-Genius reported ADR for EEs [20]. Studies calculating the TE-ADR without AI reported a value similar to ours and a

time-learning effect, with an increase in this parameter as training years increased [24].

Our calculated APC, PDR, and PPC also were comparable between the two groups. Our CADe APC was lower than the 1.07 calculated by Repici et al. [20] in their EE group, but higher than the pooled one reported in the literature [14]. Regarding PDR, we found a value higher than the pooled one (50.3%) reported in the literature with AI support [14]. This rate probably is related to further technological improvements in these devices, as shown by other authors who found a value comparable with ours [18]. In our study, some lesions were found by the TEs and not by the EEs (62% vs. 58%). However, these lesions were diminutive NICE 1 rectal polyps, as confirmed by reviewing the videotapes.

We found an AMR value of 12.5%, much lower than the 27% reported by Munroe et al [17] in a group of TEs without AI and partially in line with the literature tandem ones (10%-30%) [18]. However, our AMR and PMR were comparable to those for the same parameters calculated with CADe colonoscopy in expert hands (13.89% and 12.98%, respectively) [18].

As in previous studies, we did not find differences in detection of lesions with different morphology, size, and location [14]. Moreover, we did not calculate a meaningful prolongation of withdrawal time in the TE group to explain the false positive by AI, probably because our TEs were already trained in SNL characterization, as reported in two previous works of our group [25,26]. The EE withdrawal time was higher due to their diagnostic and interventional examination.

Limitations

First, this was a single-center study. Our sample size was relatively small because we excluded all cases in which the EE had to intervene because a TE had difficulty to completing the examination. Moreover, the device, which our service soon will be acquiring, was tested for only a short trial period. As soon as we have the equipment available, we will implement our data even if we think that a multicenter study with a larger sample size is needed to confirm these results. Second, because the different endoscopists were aware of this ongoing study of their own performance, the study results could have overestimated the actual parameters, especially for the experienced group. In fact, studies have shown limitations of the human eye related to fatigue, distraction, and level of alertness during examination and the added value of AI to overcome them [13]. Although AI has been shown to be superior to the human eye in searching for lesions in several studies, adequate intestinal cleansing and optimal withdrawal time are indispensable prerequisites for maximizing AI results [13, 15]. Third, we carried out our study with HD-WLE. We cannot exclude better performance with the synergistic effect of other techniques or tools (e.g., chromoendoscopy, Endocuff). As already proposed by other authors [20], future studies should compare neoplasia detection with CADe over chromoendoscopy with CADe alone.

Although limited by a small sample, we think our results are in line with those reported in the literature in expert hands precisely because our TEs were already trained to recognize and characterize colonic lesions. This greatly limited the number of

false positive and false negative results, as the TEs already knew what to look for.

Conclusions

In conclusion, we demonstrated a possible significant (practical relevance) impact that AI might have in colonoscopy training in terms of quality. We think that adding a CADe device during TE examination could help shorten and improve their learning curve for quality performance (e.g., understanding where a lesion may be found, when they need to observe with better attention, and features of polyps), and contribute to making up for any inattention of an EE during the exam supervision. In the future, this could also result in better efficacy of screening colonoscopy by reducing the incidence of interval or missed cancers. However, we must not forget that AI is only a helpful tool in addition to the human eye. A good endoscopist is one who can reach the cecum and detect polyps without the help of AI and definitively remove them.

Competing interests

The authors declare that they have no conflict of interest.

References

- [1] Puig I, Mårholm C, Bustamante M. Endoscopic imaging techniques for detecting early colorectal cancer. *Curr Opin Gastroenterol* 2019; 35: 432–439
- [2] Dekker E, Tanis PJ, Vleugels JLA et al. Colorectal cancer. *Lancet* 2019; 394: 1467–1480
- [3] Morson B. President's address. The polyp-cancer sequence in the large bowel. *Proc R Soc Med* 1974; 67: 451–457
- [4] Winawer SJ, Zauber AG, Ho MN. The National Polyp Study Workgroup. et al. Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med* 1993; 329: 1977–1981
- [5] The Paris endoscopic classification of superficial neoplastic lesions. esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003; 58: S3–S43
- [6] Kudo S, Lambert R, Allen JI et al. Nonpolypoid neoplastic lesions of the colorectal mucosa. *Gastrointest Endosc* 2008; 68: S3–S47
- [7] Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* 2010; 138: 2088–2100
- [8] Lambert R, Kudo SE, Vieth M et al. Pragmatic classification of superficial neoplastic colorectal lesions. *Gastrointest Endosc* 2009; 70: 1182–1199
- [9] Lambert R, Saito H, Saito Y. High-resolution endoscopy and early gastrointestinal cancer...dawn in the East. *Endoscopy* 2007; 39: 232–237
- [10] Bisschops R, East JE, Hassan C et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2019. *Endoscopy* 2019; 51: 1155–1179
- [11] Kamiński MF, Hassan C, Bisschops R et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2014; 46: 435–449

- [12] East JE, Vleugels JL, Roelandt P et al. Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review. *Endoscopy* 2016; 48: 1029–1045
- [13] Hassan C, Wallace MB, Sharma P et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. *Gut* 2020; 69: 799–800
- [14] Hassan C, Spadaccini M, Iannone A et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021; 93: 77–85.e6
- [15] Lui TKL, Hui CKY, Tsui VWM et al. New insights on missed colonic lesions during colonoscopy through artificial intelligence-assisted real-time detection (with video). *Gastrointest Endosc* 2021; 93: 193–200.e1
- [16] Barua I, Vinsard DG, Jodal HC et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* 2021; 53: 277–284
- [17] Munroe CA, Lee P, Copland A et al. A tandem colonoscopy study of adenoma miss rates during endoscopic training: a venture into uncharted territory. *Gastrointest Endosc* 2012; 75: 561–567
- [18] Wang P, Liu P, Glissen Brown JR et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. *Gastroenterology* 2020; 159: 1252–1261.e5
- [19] Hayashi N, Tanaka S, Hewett DG et al. Endoscopic prediction of deep submucosal invasive carcinoma: validation of the narrow-band imaging international colorectal endoscopic (NICE) classification. *Gastrointest Endosc* 2013; 78: 625–632
- [20] Repici A, Badalamenti M, Maselli R et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020; 159: 512–520.e7
- [21] Zorron Cheng Tao Pu L, Singh G, Rana K et al. Polyp detection rate as a surrogate for adenoma and sessile serrated adenoma/polyp detection rates. *Gastrointest Tumors* 2020; 7: 74–82
- [22] Williamson JM, Datta S, Satten GA. Marginal analyses of clustered data when cluster size is informative. *Biometrics* 2003; 59: 36–42
- [23] Sinagra E, Badalamenti M, Maida M et al. Use of artificial intelligence in improving adenoma detection rate during colonoscopy: Might both endoscopists and pathologists be further helped. *World J Gastroenterol* 2020; 26: 5911–5918
- [24] Krigel A, Patel A, Kaplan J et al. Anesthesia assistance in screening colonoscopy and adenoma detection rate among trainees. *Dig Dis Sci* 2020; 65: 961–968
- [25] Cocomazzi F, Gentile M, Perri F et al. Interobserver agreement of the Paris and simplified classifications of superficial colonic lesions: a Western study. *Endosc Int Open* 2021; 9: E388–E394
- [26] Cocomazzi F, Gentile M, Perri F et al. Accuracy and inter-observer agreement of the nice and kudo classifications of superficial colonic lesions: a comparative study. *Int J Colorectal Dis* 2021; 36: 1561–1568