




Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial

Shunsuke Kamba¹  · Naoto Tamai¹ · Iduru Saitoh¹ · Hiroaki Matsui¹ · Hideka Horiuchi¹ · Masakuni Kobayashi³ · Taku Sakamoto⁴ · Mai Ego⁴ · Akihiro Fukuda⁵ · Aya Tonouchi⁵ · Yuki Shimahara⁵ · Masako Nishikawa² · Haruo Nishino⁶ · Yutaka Saito⁴ · Kazuki Sumiyama¹

Received: 15 May 2021 / Accepted: 27 June 2021 / Published online: 3 July 2021
© Japanese Society of Gastroenterology 2021

Abstract

Background We have developed the computer-aided detection (CADe) system using an original deep learning algorithm based on a convolutional neural network for assisting endoscopists in detecting colorectal lesions during colonoscopy. The aim of this study was to clarify whether adenoma miss rate (AMR) could be reduced with CADe assistance during screening and surveillance colonoscopy. **Methods** This study was a multicenter randomized controlled trial. Patients aged 40 to 80 years who were referred for colorectal screening or surveillance at four sites in Japan were randomly assigned at a 1:1 ratio to either the “standard colonoscopy (SC)-first group” or the “CADe-first group” to undergo a back-to-back tandem procedure. Tandem colonoscopies were performed on the same day for each participant by the same endoscopist in a

preassigned order. All polyps detected in each pass were histopathologically diagnosed after biopsy or resection.

Results A total of 358 patients were enrolled and 179 patients were assigned to the SC-first group or CADe-first group. The AMR of the CADe-first group was significantly lower than that of the SC-first group (13.8% vs. 36.7%, $P < 0.0001$). Similar results were observed for the polyp miss rate (14.2% vs. 40.6%, $P < 0.0001$) and sessile serrated lesion miss rate (13.0% vs. 38.5%, $P = 0.03$). The adenoma detection rate of CADe-assisted colonoscopy was 64.5%, which was significantly higher than that of standard colonoscopy (53.6%; $P = 0.036$).

Conclusion Our study results first showed a reduction in the AMR when assisting with CADe based on deep learning in a multicenter randomized controlled trial.

Keywords Computer-aided detection · Deep learning · Adenoma miss rate · Adenoma detection rate · Colonoscopy

✉ Shunsuke Kamba
kanba@jikei.ac.jp

- ¹ Department of Endoscopy, The Jikei University School of Medicine, 3-25-8 Nishi-Shimbashi, Minato-ku, Tokyo 105-8461, Japan
- ² Clinical Research Support Center, The Jikei University School of Medicine, 3-25-8 Nishi-Shimbashi, Minato-ku, Tokyo 105-8461, Japan
- ³ Department of Endoscopy, The Jikei University School of Medicine Third Hospital, 4-11-1 Izumihoncho, Komae-shi, Tokyo 201-8601, Japan
- ⁴ Endoscopy Division, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
- ⁵ LPIXEL Inc., 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan
- ⁶ Coloproctology Center, Matsushima Hospital, 3-138 Isecho, Nishi-ku, Yokohama-shi, Kanagawa 220-0045, Japan

Abbreviations

ADR	Adenoma detection rates
AMR	Adenoma miss rate
CADe	Computer-aided detection
CADx	Computer-aided diagnosis
FUSE	Full-spectrum endoscopy
MAP	Mean number of adenomas per procedure
PPS	Per protocol set
PPSL	Per protocol set of lesions
PDR	Polyp detection rate
PMR	Polyp miss rate
PCCRC	Post-colonoscopy colorectal cancer
RCT	Randomized controlled trial
SSL	Sessile serrated lesion
SC	Standard colonoscopy

Introduction

Colonoscopy is a rational screening method for colorectal cancer (CRC) that can be performed consistently from detection to treatment of both precancerous and cancerous lesions. A large cohort study has previously demonstrated that colonoscopy could reduce CRC-related mortality by 68% [1]. However, the performance and quality of colonoscopy are heavily dependent on the skill, knowledge, and experience of the endoscopist; the majority of adenomatous lesions overlooked at the index colonoscopy become the cause of post-colonoscopy CRC. The adenoma detection rate (ADR) is the percentage of examinations in which one and more adenomatous colorectal lesions are found, which is now widely accepted as the most reliable performance indicator for each endoscopist and each examination. A series of clinical trials provided strong evidence that the ADR was inversely correlated to the incidence and mortality of CRC after index colonoscopy [2, 3]. To improve the detectability of colorectal lesions regardless of diagnostic yield among endoscopists, various novel technologies have been developed such as high definition endoscopy [4], image enhancement endoscopy [5], and cap attachment with flaps, which facilitate the flattening of colonic folds [6]. Computer-aided diagnosis systems based on deep learning, one of the machine learning methods of artificial intelligence, has become one of the top research interests in the field of endoscopy globally as they can be easily and inexpensively adapted to conventional endoscopy systems [7, 8]. Computer-aided diagnosis is subclassified into computer-aided detection (CADe), which supports lesion detection, and computer-aided diagnosis (CADx), which supports the differential diagnosis of lesions [9]. The deep learning-based CADe systems have been most profoundly explored in the colorectal field to reduce the number of missed lesions during colonoscopy. Some provide auditory and visual alerts indicating suspected areas of lesions that appear in the endoscopic image, such as the system studied in this study, have cleared regulations for clinical use as a medical device [10]. Several randomized controlled trials (RCTs) have been conducted for deep learning-based CADe for colonoscopy; all except for one analyzed the ADR as the primary endpoint and consequently showed improved ADR by CADe assistance [11]. Although the ADR value has been applied as a target or threshold to guarantee the performance of examination, ADR values for both groups with and without CADe assistance widely varied among the precedingly published RCTs. When the ADR is analyzed in a prospective fashion, a so-called “one and done” phenomenon, where insufficient attention is paid to detecting additional polyps after one likely adenoma is detected and removed, may inevitably occur [12]. It is also technically difficult to quantitatively

evaluate the rate or number of missed lesions from the ADR value. The adenoma miss rate (AMR) is another widely accepted performance indicator for colonoscopy that is calculated in sets of repeated two colonoscopies on the same patient and by counting the number of lesions missed at the first examination but found at the second examination [13]. The AMR is considered more suitable to compare the diagnostic support technologies by analyzing the number of lesions found during colonoscopy [14]. The AMR can be more sensitive than the ADR to investigate differences in lesion detectability even between endoscopists with a high ADR of 40% or more [15].

When evaluating novel technologies such as CADe that directly process medical images in real time, it cannot be denied that the technology may negatively affect the clinical judgment of the endoscopist. Wang et al. reported that the AMR of colonoscopy with assistance of CADe was significantly lower than that of standard colonoscopy (13.89% vs. 40.0%, $P < 0.0001$) in a single-center study [16]. However, the ADR simultaneously evaluated in that report did not show a significant difference compared with that of the control group. Additionally, the design of their study, which was performed by only three expert endoscopists in a single facility, cannot objectively deny the possibility of subjective bias. It may be inappropriate to apply their results to general clinical practice, including non-experts and endoscopists with various clinical skills.

We have developed the original CADe system that applied the deep learning algorithm. The system facilitates the detectability of colorectal polyps by presenting a bounding box with a beep sound in areas suspected of polyps during colonoscopy. The aim of this study was to clarify whether the AMR of screening and surveillance colonoscopy can be significantly reduced by assistance with our CADe system compared to standard colonoscopy. This study is the first prospective multicenter RCT to investigate AMR reduction by CADe applying artificial intelligence technology. By including both expert and non-expert endoscopists as operators, this study was designed with consideration for minimizing bias and generalizing the results.

Materials and methods

Study design

This multicenter RCT, designed with a parallel group for comparison using a back-to-back tandem approach, took place during August 2019–January 2020 at four sites in Japan: two advanced endoscopy centers of tertiary medical centers (The Jikei University Hospital, National Cancer Center Hospital), one secondary medical center (The Jikei University Third Hospital), and one private clinic

specialized in endoscopy (Matsushima Clinic). The study protocol was approved by The Jikei University Certified Review Board and registered on the Japan Registry of Clinical Trial (jRCTs032190061; <https://jrct.niph.go.jp/en-latest-detail/jRCTs032190061>). The report was according to the guidelines of “CONSORT 2010” and “CONSORT-AI extension” for AI-related RCTs.

Patients

Subjects were patients aged 40 to 80 years who were referred for colonoscopy for colorectal screening or surveillance after endoscopic treatment (only patients who had all lesions removed in the previous colonoscopy). Patients with the following criteria were excluded: known inflammatory bowel disease or stenosis of the large intestine, known familial polyposis, known colon polyps or advanced cancer, history of post-colorectal surgery (excluding appendectomy or rectal surgery), blood coagulation disorders, serious organ failure, pregnancy, and ineligible for registration as judged by the operator. Participants provided informed consent with sufficient understanding after receiving a full explanation regarding their participation in this study. All colorectal image data of participants inputted from endoscopy systems were analyzed by the CADe system. There were no exclusion criteria at the level of the input data.

Randomization scheme

Central randomized allocation was performed using a minimization method to equalize the patient backgrounds using a web-based electric data capture system (DATATRAK[®]) by the study assistant. Patients were randomly assigned in a 1:1 ratio to either a “standard colonoscopy (SC)-first group” or an “CADe-first group” to undergo a back-to-back tandem procedure. Adjustments were made for following five minimization factors: institution, sex (male/female), age (≥ 60 / < 60 years), proficiency of endoscopists (number of colonoscopies performed: ≥ 5000 or < 5000 cases), and reason for examination (minor symptomatic/asymptomatic screening, surveillance after endoscopic therapy, positive fecal occult blood test). Results of randomized allocation were notified to endoscopists just before colonoscope insertion, but not to the patients.

CADe system

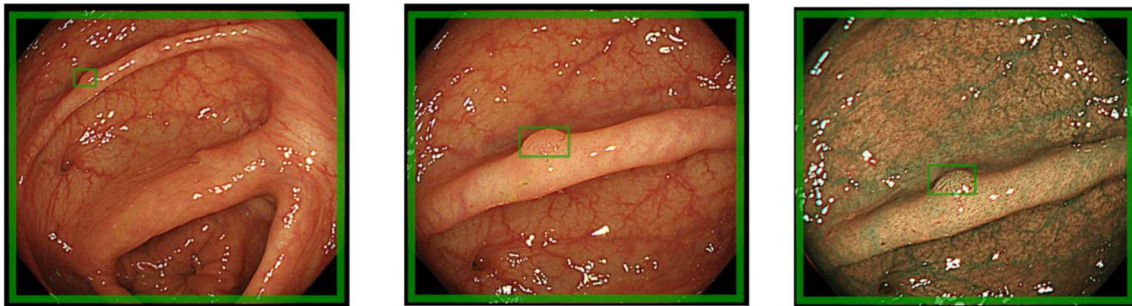
The CADe system is a software-based diagnostic support system for colonoscopy, which was developed by LPIXEL Inc. and The Jikei University School of Medicine under the funding support of Japan Agency for Medical Research and

Development. The CADe model was the latest version based on the original artificial intelligence algorithm used in a previous study that was created by modifying YOLOv3 (You-Only-Look-Once), which contains 53 layers of convolutional neural network architectures [17]. It was trained with 65,421 colonoscopic images (62,726 images with lesions and 2695 without lesions), which were collected from 4147 colonoscopy cases including 26,729 lesions. Images with lesions were annotated by enclosing the lesion area in a bounding box on each image by expert endoscopists and used as training data for objective recognition as “some colorectal lesions”. Images without lesions were used as training data of “normal structures” for reduction of false-positive detection. Evaluation of the per-image accuracy of the CADe system in the in-silico pilot test using 4158 still images (including 981 lesion-positive images and 3177 lesion-negative images), which were not used for the CADe training, revealed a sensitivity of 95.5%, and false-positive rate of 9.0%.

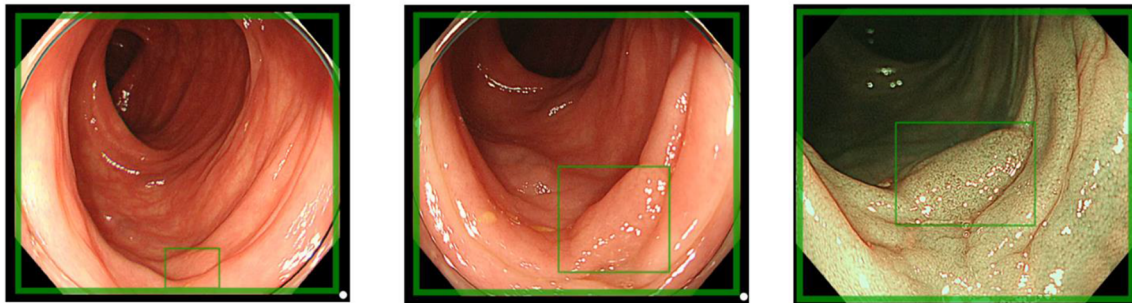
If the CADe system is used during a colonoscopy, a rectangular box is superimposed on the area suspected as a colorectal lesion on the screen of the primary monitor with a simultaneous warning sound to notify the endoscopist of the presence of a lesion (Fig. 1). The spec of the workstation with the CADe software installed was as follows: CPU Intel Core i5-8600 K [3.60 GHz/6Core/UHD630/TDP95W], MEMORY 32 GB, GPU NVIDIA TITAN V 12 GB, and Microsoft Windows 10 Home (64 bit) operating system. The real-time images were inputted via a High Definition-Serial Digital Interface (HD-SDI) cable in uncompressed digital format. All inputted images, including images with an excessive movement artifact or blur, were analyzed by the CADe system. All image frames of the video signals in HD-SDI format were resized to 544×544 pixels by cropping only the area of the endoscopic image. Following analysis by the CADe model, multiple overlapping bounding boxes were integrated using the Non-Maximum Suppression algorithm if intersection over union (IoU) (the area of overlap between the two bounding boxes/the area encompassed by the two bounding boxes) between the boxes was 0.1 or more. Furthermore, if the size of the detected bounding box occupied 30% or more of the area of the endoscopic image or if the detected bounding boxes in six consecutive video frames did not overlap in two or more frames ($\text{IoU} \geq 0.05$), the bounding boxes were not displayed as they were considered likely to be a false positive.

The same workstation and algorithm were used at all sites. Commonly used endoscopy systems and scopes could be applied to the CADe system. The analysis with the system was processed in real time with 20 ms latency and 30 frames per second average throughput. The CADe system was a prototype permitted to be used clinically only

CASE 1



CASE 2



A	B	C
D	E	F

Fig. 1 Detecting lesions missed during the first pass in the second pass of the SC-first group. Images of CASE 1 show a diminutive adenoma located at a hepatic flexure, which was detected by the CADe system. **A** An image when the lesion was detected by the CADe system. **B** The lesion was confirmed by the endoscopist from a short distance. **C** Narrow-band imaging. Images of

CASE 2 show a flat-elevated adenoma located at a descending colon, which was detected by the CADe system. **D** An area with a slightly different surface and color from the surrounding mucosa was detected by the CADe system. **E** The lesion was confirmed by the endoscopist from a short distance. **F** Narrow-band imaging.*CADe computer-aided detection

for this study by the Jikei University Certified Review Board.

Endoscopy system

The endoscopy system used in the study was a processor and light source of EVIS LUCERA ELITE (Olympus Medical Systems, Tokyo, Japan) and a scope limited to models with a 170° viewing angle such as CF-H290I/HQ290Z/HQ290I and PCF-H290I/H290Z (Olympus Medical Systems).

Endoscopists

Endoscopists who can perform screening colonoscopy and basic therapeutic procedures such as endoscopic mucosal resection and polypectomy without complications (endoscopists who can perform this task alone), including both experts with an experience of > 5000 colonoscopies and non-experts with an experience of < 5000 colonoscopies,

could participate as an operator for the endoscopic procedure. The operators had experienced at least one colonoscopy using the CADe system or watched the instructional video of the CADe system before the study began. The first and second procedures for a back-to-back tandem procedure were performed by the same endoscopist. If the operator could not insert the endoscope into the cecum due to difficulty of insertion, a senior endoscopist could help perform the insertion only.

Endoscopic procedure

A back-to-back tandem procedure was adopted to assess the AMR as the primary endpoint. In the SC-first group, standard colonoscopy (first pass) was performed first, followed by CADe-assisted colonoscopy (second pass); in the CADe-first group, CADe-assisted colonoscopy (first pass) was performed first, followed by standard colonoscopy (second pass).

Bowel preparation was performed by the standard method with 1.5–2 L of highly concentrated polyethylene glycol with ascorbate solution or 2–4 L of standard polyethylene glycol solution or 1.8–2.4 L of magnesium citrate solution according to the preparation protocol specified by the institution, which was administered in the morning of the day of colonoscopy. Assessment of the bowel cleansing score was rated by the operator during colonoscopy. After confirming a sufficient sedation level with diazepam, midazolam, flunitrazepam, or propofol with pethidine, a scope was inserted into the cecum for the first pass. After a close-up view of the appendiceal orifice or ileocecal valve was obtained, a colonoscopic observation was performed by withdrawing the scope to the anus using the first pass method of the allocated group. Following the observation of the first pass, the same scope was re-inserted into the cecum, and the observation of the second pass was performed. All polyps detected by endoscopists were biopsied or resected immediately, except for 1–5 mm white and flat-elevated lesions found between the rectum and sigmoid colon. The use of a distal transparent hood was allowed at the discretion of the endoscopist but the hood had to be attached in such a way that it did not impair the scope's field of view. Retroflexion of the scope was allowed only in the rectum. Colonoscopic observation was performed under white light imaging for at least 6 min on each pass. Dye use such as indigo carmine and crystal violet was not allowed. In each pass of the allocated group, the assistants recorded the size of the lesion, macroscopic morphology, the colorectal segment where it was located, the tissue-sampling method when the endoscopist detected a colorectal lesion, and the withdrawal inspection time, excluding the time used for polyp resection and tissue-sampling, using a stopwatch. The observation pass using the CADe system started after system activation, and the endoscopist searched for colorectal lesions using the visual and auditory assistance of the CADe system from the time of reaching the cecum. When the CADe surrounded the suspected lesion area with a bounding box, an operator judged whether there was an actual colorectal lesion in the boxed area.

Outcome measures

The AMR of standard colonoscopy and CADe-assisted colonoscopy was analyzed as the primary endpoint. Secondary endpoints were polyp miss rate (PMR), AMR per patient, PMR per patient, ADR at first pass, polyp detection rate (PDR) at first pass, and mean number of adenomas per procedure (MAP) at first pass. The miss rate was calculated by dividing the total number of lesions detected in the second pass by the total number of lesions detected in the first or second pass. The miss rate per patient was the

average value of the AMR calculated for each patient in each group. The detection rate was defined as the proportion of colonoscopies that detected one and more adenomatous colorectal lesions.

Other endpoints were the quality of bowel preparation based on the Aronchick scale [18], withdrawal time, and the percentage of patients with changed colonoscopy surveillance intervals according to the United States guidelines by the second pass [19]. Sub-group analyses were additionally performed using the following factors: histopathological type, macroscopic type classified with the Paris classification, lesion diameter (1–5, 6–9, ≥ 10 mm), lesion location, and endoscopist proficiency.

Histopathology definition

A colorectal adenoma was defined as a lesion with a final pathological diagnosis of tubular adenoma, tubulovillous adenoma, villous adenoma, adenocarcinoma, traditional serrated adenoma, or sessile serrated lesion (SSL). Based on this definition, the AMR, AMR per patient, ADR, and MAP were analyzed. All colorectal polyps were included in the PMR, PDR, and PMR per patient analyses. Advanced cancers and subepithelial tumors were excluded from the analysis. An advanced adenoma was defined as an adenoma that satisfied any of the following criteria: lesion diameter ≥ 10 mm, villous components, and high-grade atypia [20].

Sample size calculation

The comparison study in Japan reported that the AMR was 30.7% for standard colonoscopes and 20.5% for ultrawide-viewing colonoscopes [21]. We adopted a similar study design; thus, the AMR for standard colonoscopy was also estimated to be approximately 30%. If the AMR of CADe-assisted colonoscopy was assumed to be 15%, the required number of adenomas was 241 when calculated with a two-sided significance of 0.05 and a power of 0.8. Assuming the number of adenomas per patient was 0.75 and the dropout/exclusion rate from the analysis set of the primary endpoint was 10%, the required number of subjects was determined as 358. With 1:1 allocation, the number of subjects required per group was decided as 179.

Statistical analysis

For all data obtained as continuous values, basic statistics (maximum, median, minimum, quartiles, mean, and standard deviation) were calculated. For categorical data, frequencies were totalized for each category. All outcome measurements were calculated (point estimation) for the both groups, and the 95% confidence interval was

calculated using normal approximation. The AMR and PMR were analyzed using a permutation test, and a chi-square statistic was used as the difference measure in the permutation test. For the AMR per patient, PMR per patient, and MAP, analysis of covariance adjusted for allocation factors was performed. A chi-squared test was used for analyses of the ADR and PDR, sub-group analyses of the AMR. The mean withdrawal inspection time (the total observation time/number of cases) for withdrawing a scope was assessed with a one-way analysis of variance. A chi-squared test was used to analyze differences of bowel preparation quality evaluated by the Aronchick scale and the rate at which the surveillance intervals changed based on each guideline caused by the second pass between the two groups. The significance level was set as 5% on both sides. All statistical analyses were performed by SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

The analysis set for efficacy was defined separately as per protocol set (PPS) 1 for the group of subjects who completed the second pass and PPS2 for the group of subjects who completed the first pass. Sets of lesions detected by the CADe or an endoscopist and histologically diagnosed in subjects of PPS1 and PPS2 were defined as PPSL1 and PPSL2, respectively. Sets of lesions detected by the CADe or an endoscopist with or without histopathological diagnosis in subjects of PPS1 and PPS2 were defined as PPSL3 and PPSL4, respectively. In PPS1, the AMR and AMR per patient were analyzed using PPSL1 and the PMR and PMR per patient were analyzed using PPSL3. In PPS2, the ADR and MAP and the PDR were analyzed using PPSL2 and PPSL4, respectively. The full analysis set was used only for the analysis of an adverse event. All authors had access to the study data and reviewed and approved the final manuscript.

Results

Patient flow and background

A total of 358 eligible patients were enrolled at four study sites between August 2019 and January 2020; 179 each were assigned to the SC-first and CADe-first groups. Prior to colonoscopy of the first pass, three patients met the exclusion criteria. After starting colonoscopy, 14 patients were excluded for the following reasons: irrecoverable malfunction of the CADe system ($n = 3$), difficulty inserting into the cecum ($n = 3$), protocol violation ($n = 1$), severe abdominal pain due to colonoscopy ($n = 4$), inflammatory bowel disease ($n = 1$), and familial polyposis for the first time ($n = 1$). The final analysis sets included 172 patients in the CADe-first group and 174 in the SC-first group who could complete the first pass (PPS2) and 171 in

the CADe-first group and 173 in the SC-first group who could complete the second pass (PPS1). No serious adverse events occurred in either group. The patient flow diagram (CONSORT diagram) is shown in Fig. 2.

After randomized allocation, both groups had similar background data regarding age, sex, indication of colonoscopy, proficiency level of the operator in charge, and study site, which were related to CRC risk (Table 1).

Colonoscopy background

In PPS1, the Aronchick scale of 166 patients in the CADe-first group and 169 patients in the SC-first group was 1 to 3, which indicates a good level of bowel preparation ($P = 0.7222$). Most of the scopes used were small-diameter scopes such as PCF, reflecting the usual clinical situation. There was no significant difference in the withdrawal inspection time of the first pass (434 ± 81 vs. 449 ± 89 , $P = 0.1047$) and usage rate of an intestinal antispasmodic agent and a tip hood between both groups (Table 2). Colonoscopies were performed by 22 expert endoscopists and 10 non-expert endoscopists.

Miss rate of colorectal polyps

In the SC-first group, 219 adenomas were detected in the first pass and 127 missed adenomas were detected in the second pass assisted by CADe. In the CADe-first group, 244 adenomas were detected in the first pass and 39 missed adenomas were detected in the second pass with standard colonoscopy. The AMR of CADe-assisted colonoscopy was significantly lower than that of standard colonoscopy (13.8% vs. 36.7%, $P < 0.0001$). In addition, the SSL miss rate, which was 38.5% on standard colonoscopy, was reduced significantly to 13.0% using CADe assistance ($P = 0.032$). No advanced adenomas were missed by CADe-assisted colonoscopy, while two adenomas missed by standard colonoscopy were detected with CADe-assisted colonoscopy during the second pass ($P = 0.1887$). The PMR, including non-neoplastic polyps, was also significantly lower in CADe-assisted colonoscopy than in standard colonoscopy (14.2% vs. 40.6%, $P < 0.0001$) (Table 3).

The results of the sub-group analysis by lesion size revealed a significant reduction in the AMR in the size class of 1–5 mm. In the analysis by macroscopic morphology, the AMR of CADe-assisted colonoscopy was significantly lower in type-Is and IIa than that of standard colonoscopy, but there was no significant difference in type-Ip. The AMR in all segments, excluding the cecum and rectum, was significantly lower in CADe-assisted colonoscopy than in standard colonoscopy, and significant differences were more obvious in the right-sided colon

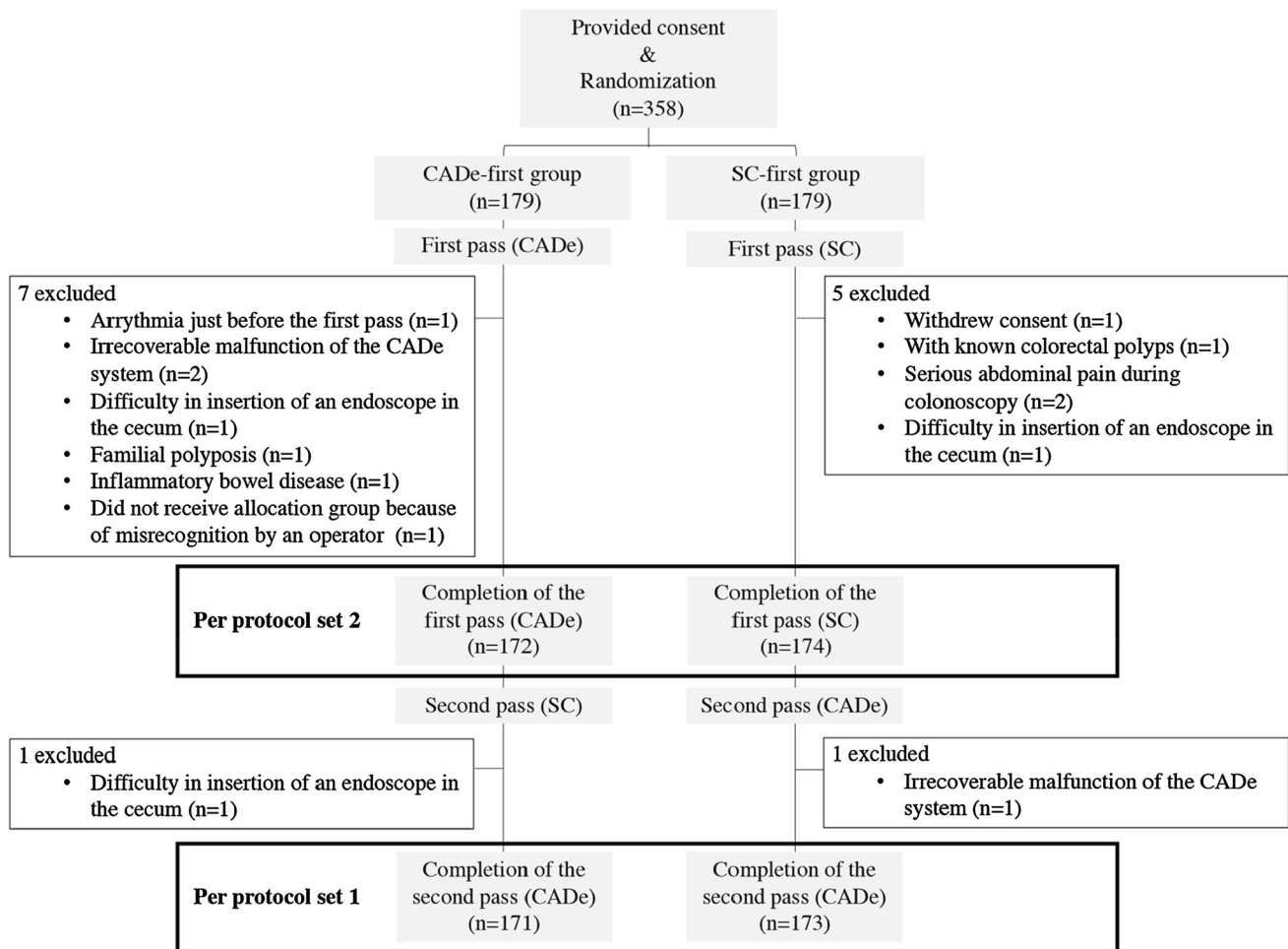


Fig. 2 Patient flow diagram (CONSORT diagram). *CADe computer-aided detection, SC standard colonoscopy

(ascending and transverse colon). When the AMR by proficiency level was analyzed, the AMR of both expert and non-expert endoscopists was significantly lower in CADe-assisted colonoscopy than in standard colonoscopy ($P < 0.01$) (Table 4).

The changing rate of the follow-up period based on the US colonoscopy screening guidelines defined according to the number of adenomas additionally detected by the second pass was 23.1% in the SC-first group, which was significantly higher than that in the CADe-first group (9.9%, $P = 0.001$).

Detectability of colorectal polyps

The ADR of CADe-assisted colonoscopy in the first pass of the CADe-first group was 64.5%, which was significantly higher than that of standard colonoscopy (53.6%) in the first pass of the SC-first group ($P = 0.036$). No significant difference was found in the detection rates of advanced

adenoma and SSL. In addition, the PDR did not show a significant difference between the CADe-first and SC-first groups (69.8% vs. 60.9, $P = 0.084$) (Table 5). The MAPs in the first pass did not differ significantly between the CADe-first and SC-first groups (1.42 ± 2.01 vs. 1.25 ± 1.80 , $P = 0.412$).

Discussion

This RCT demonstrated that the AMR of colonoscopy assisted with the CADe system developed by us was significantly lower than that of standard colonoscopy. This is the first report on a multicenter study showing the reducibility of AMR with CADe, wherein not only expert endoscopists but also non-expert endoscopists participated as operators. Our study design had the following distinguishing features compared to the previously reported single-center tandem study [16]. First, this study was

Table 1 Patient background information

	CADe-first group (n = 178)	SC-first group (n = 177)
Age, years (mean ± SD)	61.63 ± 9.89	61.44 ± 10.01
< 60 years no. (%)	75 (42.1)	77 (43.5)
≥ 60 years no. (%)	103 (57.9)	100 (56.5)
Sex no. (%)		
Male	136 (76.4)	136 (76.8)
Female	42 (23.6)	41 (23.2)
Indication of colonoscopy no. (%)		
1. Fecal occult blood test positive	31 (17.4)	31 (17.5)
2. Non- or mild-symptomatic screening	88 (49.4)	78 (44.1)
3. Surveillance after endoscopic treatment	59 (33.2)	68 (38.4)
Proficiency of endoscopists no. (%)		
Expert (≥ 5000 colonoscopies)	125 (70.2)	122 (68.9)
Non-expert endoscopy (< 5000 colonoscopies)	53 (29.8)	55 (31.1)
Study site no. (%)		
Site 1	72 (40.5)	74 (41.8)
Site 2	43 (24.2)	44 (24.9)
Site 3	39 (21.9)	39 (22.0)
Site 4	24 (13.5)	20 (11.3)

CADe computer-aided detection, SC standard colonoscopy

Table 2 Colonoscopy background data

	CADe-first group (n = 171)	SC-first group (n = 173)	P value
Bowel preparation scale no. (%)			
1–3	166 (97.1)	169 (97.7)	0.7222 ^a
4 or 5	5 (2.9)	4 (2.3)	
Scope-type no. (%)			
PCF-H290I	19 (11.1)	22 (12.7)	0.3381 ^a
PCF-H290Z	128 (74.9)	132 (76.3)	
CF-H290I	21 (12.3)	19 (11.0)	
CF-HQ290I	3 (1.8)	0 (0)	
Withdrawal inspection time of the first pass sec mean ± SD			
First pass	434 ± 81	449 ± 89	0.1047 ^b
Second pass	404 ± 63	431 ± 92	0.0021 ^b
Use of antispasmodic agent no. (%)	117 (68.4)	122 (70.5)	0.6725 ^a
Use of tip hood no. (%)	137 (80.1)	148 (85.6)	0.1814 ^a

CADe computer-aided detection, SC standard colonoscopy

^aChi-squared test

^bOne-way analysis of variance

conducted as a multicenter RCT, and its participating sites consisted of hospitals of various sizes and functions, such as a tertiary hospital (> 1000 hospital beds), a local core hospital (> 500 hospital beds), and a small clinic (without beds). Colonoscopy was performed under the same conditions and following the same procedures (e.g., patient position, a kind of scope and observation manner) as those

in a routine clinical setting at each site, except the protocol with respect to the withdrawal inspection time, which was limited to 6 min or longer. Second, 32 endoscopists participated in this study as colonoscopy operators; this is the largest number of endoscopists among all studies on endoscopic CADe reported to date [10, 22, 23]. Furthermore, this study was unique in that some of the

Table 3 Analysis of the miss rate

	The number of detected lesions		Miss rate of the first pass		Miss rate per patient of the first pass	
	First pass, no	Second pass, no	% (95% CI)	<i>P</i> value	% ± SD	<i>P</i> value
Adenoma						
CADe-first group	244	39	13.8 (9.8–17.8)	< 0.0001*	8.9 ± 21.9	< 0.0001***
SC-first group	219	127	36.7 (31.6–41.8)		25.4 ± 35.9	
Advanced adenoma						
CADe-first group	12	0	0 (0–0)	0.1887**	NA	NA
SC-first group	13	2	13.3 (0–30.5)		NA	
SSL						
CADe-first group	20	3	13.0 (0–26.8)	0.0332**	NA	NA
SC-first group	24	15	38.5 (23.2–53.7)		NA	
Polyp						
CADe-first group	315	52	14.2 (10.6–17.7)	< 0.0001*	9.62 ± 21.7	< 0.0001***
SC-first group	275	188	40.6 (36.1–45.1)		31.9 ± 36.4	

CADe computer-aided detection, CI confidence interval, SC standard colonoscopy, SSL sessile serrated lesion

*Permutation test

**Chi-squared test

***Analysis of covariance

Table 4 Sub-group analysis of the miss rate

	AMR of the first pass, %			AMR per patient of the first pass, %		
	CADe-first group	SC-first group	<i>P</i> value	CADe-first group	SC-first group	<i>P</i> value
Size (mm)						
1–5	13.89	41.33	< 0.0001*	7.8	25.5	< 0.0001**
6–9	17.65	20.37	0.7225*	4.0	4.1	0.8424**
10	0	19.05	0.0645*	0	1.7	0.0768**
Morphologic type						
Is	14.07	26.14	0.0113*	5.3	13.2	0.0083**
Ip	0	0	1*	0	0	1**
Ila	13.38	45.26	< 0.0001*	5.8	21.0	< 0.0001**
Location						
Ce	9.09	27.03	0.2140*	0.6	5.2	0.0108**
A	9.23	44.05	< 0.0001*	2.8	13.6	0.0002**
T	11.96	35.45	0.0001*	4.7	14.9	0.0007**
D	16.67	38.89	0.0353*	2.6	6.6	0.1187**
S	17.91	33.33	0.0426*	6.1	6.2	0.9001**
R	25.00	37.50	0.4834*	1.5	1.9	0.6811**
Operator						
Expert	14.14	39.15	< 0.0001*	9.26	23.7	0.0004**
Non-expert	13.04	32.84	0.0007*	8.03	29.2	0.0015**

CADe computer-aided detection, SC standard colonoscopy, AMR adenoma miss rate, Ce cecum, A ascending colon, T transverse colon, D descending colon, S sigmoid colon, R rectum

*Chi-squared test

**Analysis of covariance

Table 5 Analysis of detectability

	The number of patients with at least one lesion in the first pass, no	Detection rate, % (95% CI)	<i>P</i> value
Adenoma			
CADe-first group (<i>n</i> = 172)	111	64.5 (57.4–71.7)	0.036*
SC-first group (<i>n</i> = 174)	93	53.6 (46.0–60.9)	
Advanced adenoma			
CADe-first group (<i>n</i> = 172)	11	6.4 (2.7–10.1)	0.978*
SC-first group (<i>n</i> = 174)	11	6.3 (2.7–9.9)	
SSL			
CADe-first group (<i>n</i> = 172)	13	7.6 (3.6–11.5)	0.866*
SC-first group (<i>n</i> = 174)	14	8.1 (4.0–12.1)	
Polyp			
CADe-first group (<i>n</i> = 172)	120	69.8 (62.9–76.6)	0.084*
SC-first group (<i>n</i> = 174)	106	60.9 (53.7–68.2)	
	The number of adenomas detected in the first pass, no	MAP, mean ± SD	<i>P</i> value
CADe-first group (<i>n</i> = 172)	245	1.42 ± 2.01	0.412**
SC-first group (<i>n</i> = 174)	219	1.25 ± 1.80	

CADe computer-aided detection, CI confidence interval, SC standard colonoscopy, SSL sessile serrated lesion, MAP mean number of adenomas per procedure

*Chi-squared test

**Analysis of covariance

endoscopists were considered experts and some non-experts (experience < 5000 colonoscopies). Among 32 endoscopists, 22 have served in institutions unrelated to the development of the CADe system. Therefore, the study had been designed to minimize subjective bias, which facilitates the application of the results to common clinical practices.

In addition to the significant reduction in the AMR using our CADe system, the results of this RCT provided some interesting evidence. First, both the overall AMR in the analysis including SSL as the analysis set of lesions and the SSL miss rate in the sub-group analysis were significantly improved with the assistance of CADe. Although SSL, which is considered to be associated with a high risk of future colorectal cancer [24], tended to be excluded from the analysis set of the primary endpoint in previous trials of CADe because of its difficulty in detection [10, 16, 23], in this study, it was defined as an adenomatous lesion and included in the analysis set of the primary endpoint for the first time. In addition, the sub-group analysis according to the level of proficiency of the endoscopists also showed that the AMR of CADe-assisted colonoscopy was significantly lower than that of standard colonoscopy, and this was independent of operator expertise. Therefore, all endoscopists can benefit from the CADe system, regardless of their level of experience. Furthermore, according to the

results of the sub-group analysis based on the morphologic type and location of the lesions, significantly lower AMR in the CADe-first group was more obvious with flat lesions and those in the right-sided colon segment. The flat lesion in the right-sided colon is considered an important precursor of post-colonoscopy colorectal cancer (PCCRC) [25]; therefore, these results suggest that assistance by the CADe system may be effective in preventing PCCRC without the necessity of the skills and experience of an expert endoscopist [24, 25]. As the actual effect of the CADe system on preventing PCCRC remains unclear, further research is required to evaluate its long-term effectiveness. The AMR in the diminutive size class of 1–5 mm was significantly lower in the CADe-first group than in the SC-first group. This suggests that the CADe system in this study can detect more diminutive adenomas as with the meta-analysis of CADe previously reported [26].

Despite an increase in the number of diminutive polyps resected, no adverse events were observed.

The ADR of CADe-assisted colonoscopy (64.5%) was significantly higher than the baseline ADR (53.6%) of standard colonoscopy. Although the baseline ADR in this study was higher than that in previous RCTs [10, 23], it was equivalent to the results of an international multicenter cohort study from eight sites across the Asia-Pacific

region, which showed a comparable mean age of participants to our study [27].

The AMR per patient was set as the secondary endpoint; it was also analyzed as the primary endpoint in the J-FUSE study, which was conducted with a study design similar to our study design [21]. The AMR per patient in the J-FUSE study was 22.9% for standard colonoscopy and 11.7% for full-spectrum endoscopy (FUSE) with a 330° ultra-wide-viewing field ($P < 0.001$), which was equivalent to the results of this study (CADE 25.4% vs. standard colonoscopy 8.9%). Although the results of the sub-group analysis by location in the colon showed that FUSE significantly improved the AMR per patient only in the ascending colon, improvement was achieved in all segments of the right-sided colon (cecum, ascending colon, transverse colon) by CADE in this study. These results suggest that our CADE system might have the same or better lesion detectability as FUSE. The better outcomes could be partially explained by improved maneuverability since this CADE system can be implemented on a conventional endoscopic system; the CADE system may exert a significant preventive effect against detection failure of adenomas in multiple segments of a colon. While FUSE can only facilitate the detectability of lesions hidden behind folds in the colon, our CADE system can assist in detecting obscure lesions such as lesions having the same color as the background mucosa, diminutive lesions, and lesions exposed partially behind colonic folds.

There are several limitations to this study. First, this was an open-label study; therefore, subjective bias could have influenced the results. Additionally, there is growing discussion suggesting the AMR is an indicator that tends to favorably overrate novel technologies perhaps due to unavoidable biases when compared with the ADR of parallel studies [28]. The baseline AMR in previous tandem studies was reported as 40.0–48.0% [16, 29, 30]. Meanwhile, the AMR by standard colonoscopy in this study was 36.7%, despite the inclusion of non-expert endoscopists as operators. Considering that this study was also conducted in a multicenter clinical setting and included 22 endoscopists from three sites not involved in the development of the CADE system, including endoscopists with no experience of CADE-assisted colonoscopy in a clinical setting, the advantage of the use of CADE was shown under rigorous conditions. Second, since non-experts were defined as endoscopists who have performed < 5000 colonoscopies and can perform routine colonoscopy and endoscopic treatment alone, absolute novices were not included. Since a tandem study required two consecutive colonoscopies by the same endoscopist, we determined that physical invasions of patients would be excessive if novice endoscopists participated as operators in the study. This criterion (5000 colonoscopies) has been widely accepted because it was

also used in a past report on the detectability of colorectal lesions [31]. Third, since only the endoscopy system manufactured by Olympus medical systems was used in this study, it is uncertain whether similar results can be obtained with endoscopy systems manufactured by other vendors. Fourth, since the CADE system has been in the preparatory status for regulatory approval, its clinical use was limited to this study only. Fifth, the false-positive rate of the CADE system was not assessed in this study. It was impractical to assess the false-positive rate, which requires a retrospective video review for the analysis, because all data used in this study were collected prospectively.

In conclusion, this is the first report demonstrating that the miss rate of adenomatous lesions, including SSL, in colonoscopy assisted by the CADE system based on deep learning is significantly lower than that of standard colonoscopy in a multicenter RCT setting. Furthermore, sub-group analysis revealed a significantly reduced miss rate of the CADE system for flat lesions in the right-sided colon, considered an important precursor of PCCRC. These findings can be applied to common clinical practice as they were obtained from a multicenter RCT in which both experts and non-experts participated as operators across four sites with diverse clinical settings. Once the practical application of this CADE system is achieved, we hope that its widespread use will reduce the incidence of PCCRC in the near future.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00535-021-01808-w>.

Acknowledgements This research was supported by Japan Agency for Medical Research and Development under Grant Number 18ck0106272h0002.

Declarations

Conflict of interest The authors declare no conflicts of interest associated with this manuscript.

References

1. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med*. 2013;369:1095–105.
2. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med*. 2014;370:1298–306.
3. Kaminski MF, Wieszczy P, Rupinski M, et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. *Gastroenterology*. 2017;153:98–105.
4. Tziatzios G, Gkolfakis P, Lazaridis LD, et al. High-definition colonoscopy for improving adenoma detection: a systematic review and meta-analysis of randomized controlled studies. *Gastrointest Endosc*. 2020;91:1027–36.

5. Atkinson NSS, Ket S, Bassett P, et al. Narrow-band imaging for detection of neoplasia at colonoscopy: a meta-analysis of data from individual patients in randomized controlled trials. *Gastroenterology*. 2019;157:462–71.
6. Williet N, Tournier Q, Vernet C, et al. Effect of Endocuff-assisted colonoscopy on adenoma detection rate: meta-analysis of randomized controlled trials. *Endoscopy*. 2018;50:846–60.
7. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*. 2019;68:94–100.
8. Misawa M, Kudo SE, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology*. 2018;154:2027–9.
9. Vinsard DG, Mori Y, Misawa M, et al. Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointest Endosc*. 2019;90:55–63.
10. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68:1813–9.
11. Hassan C, Spadaccini M, Iannone A, et al. Performance of artificial intelligence for colonoscopy regarding adenoma and polyp detection: a meta-analysis. *Gastrointest Endosc*. 2021;93:77–85.
12. Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. *Gastrointest Endosc*. 2015;81:31–53.
13. Rex DK, Cutler CS, Lemmel GT, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology*. 1997;112:24–8.
14. Zhao S, Wang S, Pan P, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. *Gastroenterology*. 2019;156:1661–74.
15. Aniwani S, Orkoonsawat P, Viriyautsahakul V, et al. The secondary quality indicator to improve prediction of adenoma miss rate apart from adenoma detection rate. *Am J Gastroenterol*. 2016;111:723–9.
16. 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;159:1252–61.
17. Redmon J, Farhadi A. YOLOv3: an incremental improvement. <https://pjreddie.com/media/files/papers/YOLOv3.pdf>. Accessed 2020 April 29.
18. Saltzman JR, Cash BD, Pasha SF, et al. Bowel preparation before colonoscopy. *Gastrointest Endosc*. 2015;81:781–94.
19. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the us multi-society task force on colorectal cancer. *Gastroenterology*. 2012;143:844–57.
20. Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. *Gastrointest Endosc Clin N Am*. 2002;12:1–9.
21. Kudo T, Saito Y, Ikematsu H, et al. New-generation full-spectrum endoscopy versus standard forward-viewing colonoscopy: a multicenter, randomized, tandem colonoscopy trial (J-FUSE Study). *Gastrointest Endosc*. 2018;88:854–64.
22. Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADE-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol*. 2020;5:343–51.
23. Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology*. 2020;159:512–20.
24. East JE, Atkin WS, Bateman AC, et al. British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. *Gut*. 2017;66:1181–96.
25. Matsuda T, Fujii T, Sano Y, et al. Randomised comparison of postpolypectomy surveillance intervals following a two-round baseline colonoscopy: The Japan Polyp Study Workgroup. *Gut*. 2020. <https://doi.org/10.1136/gutjnl-2020-321996>.
26. Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy*. 2021;53:277–84.
27. Wong JCT, Chiu HM, Kim HS, et al. Adenoma detection rates in colonoscopies for positive fecal immunochemical tests versus direct screening colonoscopies. *Gastrointest Endosc*. 2019;89:607–13.
28. Zimmermann-Fraedrich K, Pohl H, Rösch T, et al. Designs of colonoscopic adenoma detection trials: more positive results with tandem than with parallel studies—an analysis of studies on imaging techniques and mechanical devices. *Gut*. 2021;70:268–75.
29. Gralnek IM, Siersema PD, Halpern Z, et al. Standard forward-viewing colonoscopy versus full-spectrum endoscopy: an international, multicentre, randomised, tandem colonoscopy trial. *Lancet Oncol*. 2014;15:353–60.
30. Dik VK, Gralnek IM, Segol O, et al. Multicenter, randomized, tandem evaluation of EndoRings colonoscopy—results of the CLEVER study. *Endoscopy*. 2015;47:1151–8.
31. Takeuchi Y, Sawaya M, Oka S, et al. Efficacy of autofluorescence imaging for flat neoplasm detection: a multicenter randomized controlled trial (A-FLAT trial). *Gastrointest Endosc*. 2019;89:460–9.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.