Gastroenterology Report, 11, 2023, goac081

https://doi.org/10.1093/gastro/goac081
Original Article

ORIGINAL ARTICLE

Artificial intelligence empowers the second-observer strategy for colonoscopy: a randomized clinical trial

Pu Wang [1,†], Xiao-Gang Liu[1,†], Min Kang[2], Xue Peng[3], Mei-Ling Shu[4], Guan-Yu Zhou[1], Pei-Xi Liu[1], Fei Xiong[1], Ming-Ming Deng[2], Hong-Fen Xia[2], Jian-Jun Li[3], Xiao-Qi Long[4], Yan Song[1] and Liang-Ping Li[1,*]

[1]Department of Gastroenterology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, Sichuan, P. R. China; [2]Department of Gastroenterology, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, P. R. China; [3]Department of Gastroenterology, Xinqiao Hospital, Third Military Medical University, Chongqing, P. R. China; [4]Department of Gastroenterology, Suining Central Hospital, Suining, Sichuan, P. R. China

*Corresponding author. Department of Gastroenterology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, No.32 West Second Section, First Ring Road, Chengdu, Sichuan 610072, China. Tel: +86-28-8739 3927; Email: llpsamsph001@outlook.com
†These authors contributed equally to this study.

Abstract

Background: In colonoscopy screening for colorectal cancer, human vision limitations may lead to higher miss rate of lesions; artificial intelligence (AI) assistance has been demonstrated to improve polyp detection. However, there still lacks direct evidence to demonstrate whether AI is superior to trainees or experienced nurses as a second observer to increase adenoma detection during colonoscopy. In this study, we aimed to compare the effectiveness of assistance from AI and human observer during colonoscopy.
Methods: A prospective multicenter randomized study was conducted from 2 September 2019 to 29 May 2020 at four endoscopy centers in China. Eligible patients were randomized to either computer-aided detection (CADe)-assisted group or observer-assisted group. The primary outcome was adenoma per colonoscopy (APC). Secondary outcomes included polyp per colonoscopy (PPC), adenoma detection rate (ADR), and polyp detection rate (PDR). We compared continuous variables and categorical variables by using R studio (version 3.4.4).
Results: A total of 1,261 (636 in the CADe-assisted group and 625 in the observer-assisted group) eligible patients were analysed. APC (0.42 vs 0.35, $P = 0.034$), PPC (1.13 vs 0.81, $P < 0.001$), PDR (47.5% vs 37.4%, $P < 0.001$), ADR (25.8% vs 24.0%, $P = 0.464$), the number of detected sessile polyps (683 vs 464, $P < 0.001$), and sessile adenomas (244 vs 182, $P = 0.005$) were significantly higher in the CADe-assisted group than in the observer-assisted group. False detections of the CADe system were lower than those of the human observer (122 vs 191, $P < 0.001$).
Conclusions: Compared with the human observer, the CADe system may improve the clinical outcome of colonoscopy and reduce disturbance to routine practice (Chictr.org.cn No.: ChiCTR1900025235).

Key words: artificial intelligence; colon cancer screening; adenoma; early detection; computer-aided detection

Submitted: 13 May 2022; **Revised:** 15 November 2022; **Accepted:** 17 November 2022

Introduction

Colorectal cancer (CRC) is the leading cause of cancer-related death worldwide [1]. As the major precursors of CRC, adenomas and serrated lesions are frequently missed during colonoscopy [2–4]. Recent tandem colonoscopy studies have reported non-negligible miss rates from 26% to 41% for adenomas and 27% for serrated lesions [5–8].

Unexposed polyps are one of the leading causes of missed diagnoses. The problem could be addressed to some extent by expanding the visual field and improving mucosa exposure [2, 9, 10]. Misdiagnosis caused by non-recognition of visible lesions on the monitor is considered another equally important issue [11]. Accordingly, to add a second observer during colonoscopy is the most effective way to reduce the possibility of non-recognition. Evidence has shown that with either a trainee or an experienced nurse acting as a second observer during colonoscopy, the polyp detection rate (PDR) will increase significantly, but the effect on the adenoma detection rate (ADR) is still controversial [12–14]. In addition, the second-observer strategy mainly benefits to low-to-moderate adenoma detectors with an ADR of <35% [15].

In recent years, artificial intelligence (AI) has achieved promising success in gastrointestinal (GI) endoscopy [3, 16–19]. An expert-level AI computer-aided detection (CADe) [20] system can be an ideal standardized second observer during colonoscopy as it provides real-time visual alerts on visible lesions, including those that appear briefly in the visual field. Evidence has shown that high-performance CADe systems could effectively increase the detection of colon adenomas of all levels of adenoma detectors in real clinical settings [9, 18, 21–25]. However, there is still a lack of direct evidence to demonstrate whether a CADe system is superior to trainees or experienced nurses as a second observer to increase adenoma detection during colonoscopy.

In the present study, we aimed to explore whether a high-performance CADe system can serve as a better assistant than trainees or experienced nurses to improve adenoma detection during colonoscopy. The false detections of CADe systems and observers are also compared in the study.

Materials and methods

Study design and patients

A multicenter, open-labeled, randomized–controlled study was conducted at four endoscopy centers (Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, Sichuan, China; The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China; Xinqiao Hospital of Third Military Medical University, Chongqing, China; Suining Central Hospital, Suining, Sichuan, China) in China from 2 September 2019 to 29 May 2020.

We enrolled patients aged between 18 and 75 years who underwent symptomatic, screening, or surveillance colonoscopy. The exclusion criteria included a history of inflammatory bowel disease (IBD), CRC, polyposis syndromes, colorectal surgery, contraindication for biopsy, current lower gastrointestinal bleeding, poor general condition, or those who did not consent before randomization. During colonoscopy, patients who were highly suspected of suffering from polyposis syndromes, IBD, and CRC mass, or their cecum were not reached were excluded from the study. After colonoscopy, patients who withdrew their consent and failed the pathology lab due to insufficient tissue from cold-forceps biopsy were also excluded. Written informed consent was provided from all participants before the colonoscopy procedure.

The protocol was approved by the Hospital Institutional Review Board Ethics Committees of Sichuan Provincial People's Hospital (No. 2019168–1), Affiliated Hospital of Southwest Medical University (No. KY2019055), Xinqiao Hospital of Third Military Medical University (No.AF/SC-08/1.0), and Suining Central Hospital (No. LLSNCH20200011).

The trial was registered on 17 August 2019 at Chictr.org.cn (No. ChiCTR1900025235).

Randomization

All eligible patients were randomized to either the CADe-assisted group or the observer-assisted group in a 1:1 ratio. Block randomization with a block size of four was used to determine the assignment of each participant. Patients were blinded to the grouping. Operating endoscopists were informed of the group allocation before each procedure.

Interventions

The CADe system [26] (EndoScreener V1.0.1, Shanghai Wision AI Co., Ltd, China) used in the CADe-assisted group is a real-time automatic polyp detection system (Supplementary Figure 1) developed on a deep-learning algorithm. The system was validated to have a per-image sensitivity of 94.38%, per-image specificity of 95.92%, and an area under the receiver-operating characteristic curve of 0.984 to detect colon polyps in colonoscopy report images [16, 20, 27]. The original colonoscopy video stream was presented on the primary monitor with no latency (Supplementary Video 1) and the CADe system augmented the single monitor with detection results (the blue hollow box) at a latency of 23.28 ± 1.39 ms on subsequent frames (Supplementary Video 2).

In the observer-assisted group, second human observers and endoscopists shared the same monitor throughout the procedure. Observers reported any identifiable polyp presented on the endoscopy monitor with a laser pointer during withdrawal. They were qualified in colonoscopy diagnosis and had considerable experience of 100–500 colonoscopy procedures. Each center assigned one qualified observer for the trial.

Procedures

All eligible patients were randomized (1:1) to the white-light colonoscopy group either with assistance of the CADe system or with assistance of a human observer. Eight experienced endoscopists (>2,000 colonoscopy screenings) from four endoscopy centers participated in the trial.

In the CADe-assisted group, the CADe system processed each frame of the video stream synchronously and reported any detected polyp location by showing a hollow alert box in an augmented-reality manner on the endoscopy screen with a simultaneous alarm sound (Supplementary Video 3). The CADe system was operating during withdrawal only. Endoscopists were required to check and verify every area within the alert box based on their own clinical judgement, whereas in the observer-assisted group, a routine white-light colonoscopy was performed with one trainee involved as a second observer. Trainees were selected as second human observers because they represented the most common and qualified candidate to act as the second human observer in the endoscopy workflow. Trainees outperformed nurses in terms of endoscopic

knowledge and expertise, and were more available than experienced endoscopists. The trainee was required to report any plausible polyp by spotting the location using a laser pointer on the screen and any discussion of the diagnosis was prohibited. The operating endoscopist was required to verify any area pointed out by the trainee if the endoscopist did not notice the same area the first time round. The diagnosis was based on the endoscopist's own judgement. If the operating endoscopist detected one lesion no later than the observer, the interaction would not affect the colonoscopy procedure. In both groups, endoscopists were not allowed to withdraw until the colon lumen was fully inflated.

All polyps verified by the operating endoscopist were biopsied or removed by using cold-forceps biopsy. All biopsied tissues were sent for pathological examination. Diminutive ($\leq$2 mm) rectal polyps deemed by the endoscopist to be hyperplastic in nature [7] under blue laser imaging (BLI) or Fuji Intelligent Chromoendoscopy (FICE) mode according to type 1 of the NBI International Colorectal Endoscopic (NICE) Classification [28] were not biopsied.

Colonoscopies were performed using Fujifilm model LASEREO and VP4450HD (Fujifilm, Tokyo, Japan), high-definition colonoscopes (EC-L590, EC-580, EC-590), and high-definition monitors. Each colonoscopy was performed with white light only, except for the NICE classification for certain detected lesions when the BLI or FICE mode was activated for a short interval during the colonoscopy.

The clean withdrawal time referred to the withdrawal time excluding the biopsy time. The withdrawal time of the procedures with no polyp detected was referred to as the withdraw time of patients with no polyp detected.

The bowel-preparation method was oral administration of 2 L of polyethylene glycol with 6 mL of simethicone solution, given in split doses. Both patients undergoing anesthetized and non-anesthetized colonoscopy were included (Supplementary Tables 1 and 2). Anesthetics, including midazolam, fentanyl, or propofol, were delivered and supervised by an anesthesiologist during each colonoscopy examination.

Bowel cleanliness was measured by using the Boston Bowel Preparation Scale during colonoscopy. The insertion time to the cecum, withdrawal time, and biopsy time for each lesion were recorded by a staff assistant with a stopwatch during each colonoscopy procedure. Polyp size was estimated by the endoscopist using open-biopsy forceps.

In the CADe-assisted colonoscopy, both missed polyps and false detections by the CADe system were recorded. A missed polyp by the CADe system was defined as a polyp that was verified by the endoscopist but not reported by the CADe system in any frame. A false detection by the CADe system was defined as occurring when an alert box kept tracking an object but it was not deemed as a polyp after closer observation by endoscopists. In observer-assisted colonoscopy, the false detection of the observers, which was defined as areas pointed out by the observer but not deemed as a polyp by the endoscopist, was recorded. Since observers standing by would notice the polyp simultaneously once endoscopists identified the suspicious area, no missed polyps were caused by observers. Any complication during the procedure or recovery was also recorded.

Outcome measurements

The primary outcome was adenoma per colonoscopy (APC). It was defined as the total number of detected adenomas divided by the total number of colonoscopy procedures of each group. Secondary outcomes included polyp per colonoscopy (PPC), ADR, and PDR. PPC was defined as the total number of detected polyps divided by the total number of colonoscopy procedures of each group, including non-biopsied hyperplastic polyps in the rectum. ADR was defined as the proportion of individuals with at least one adenoma detected among all patients. PDR was defined as the proportion of individuals with at least one polyp detected among all patients. We defined advanced adenomas as any adenoma of $\geq$10 mm in size, or containing villous histology, or with high-grade dysplasia [29, 30].

Statistical analysis

A two-sample *t*-test with a two-sided α level of 0.05 and a statistical power of 80% was used to estimate the sample size. To detect a 0.1 difference (0.45 vs 0.35) in APC, assuming a standard deviation (SD) of 0.64, 1,430 patients were needed to be enrolled into the study with a 10% buffer for potential exclusions or dropouts. Each participant was randomized to undergo CADe-assisted colonoscopy or observer-assisted colonoscopy. Statistical analysis was performed per protocol using R software (version 3.4.4).

Descriptive statistics were calculated for all measured variables and derived parameters. For continuous variables such as time to reach the cecum and colonoscope withdrawal time, we calculated means, medians, IQRs, SDs, and ranges. For categorical variables, summary statistics were counts and percentages. We used *t*-tests to compare continuous variables. For categorical variables, we used Fisher's exact test or χ^2 test to compare detection rates between groups. For estimates of proportions, we calculated 95% exact binomial confidence intervals (CIs). All tests applied were two-tailed. A two-sided *P*-value of 0.05 was the threshold for statistical significance.

Poisson regression was used to evaluate the number of adenomas and polyps detected by CADe-assisted diagnosis in colonoscopy. Logistic regression analysis was performed to evaluate the effect of CADe-assisted diagnosis for colonoscopy on the adenoma/PDR. The response variable was the binary outcome of whether an adenoma/polyp was detected. The covariate was the group variable indicating whether the patient belonged to the CADe-assisted group.

Results

Baseline and demographic data

A total of 1,261 eligible patients were analysed, with 636 patients in the CADe-assisted group and 625 patients in the observer-assisted group (Figure 1). The clean withdrawal time was 7.06 vs 6.87 min ($P = 0.055$) and the withdrawal time of procedures with no polyp detected was 6.85 vs 6.82 min in the CADe-assisted group and the observer-assisted group, respectively ($P = 0.793$) (Table 1).

There were no statistical differences between the two groups in demographic data, insertion time, bowel-preparation level, indication for colonoscopy (Tables 1 and 2), or adenoma risk factors (Supplementary Table 3). No complications were reported.

Polyp-level analysis APC, PPC, and polyp characteristics

A total of 1,229 polyps were detected, including 487 (39.6%) adenomas and 20 (1.6%) sessile serrated lesions (SSLs). Of these, 720 (58.6%) and 509 polyps (41.4%) were found in the CADe-assisted group and the observer-assisted group, respectively (Table 3). The APC was 0.42 in the CADe-assisted group and 0.35 in the

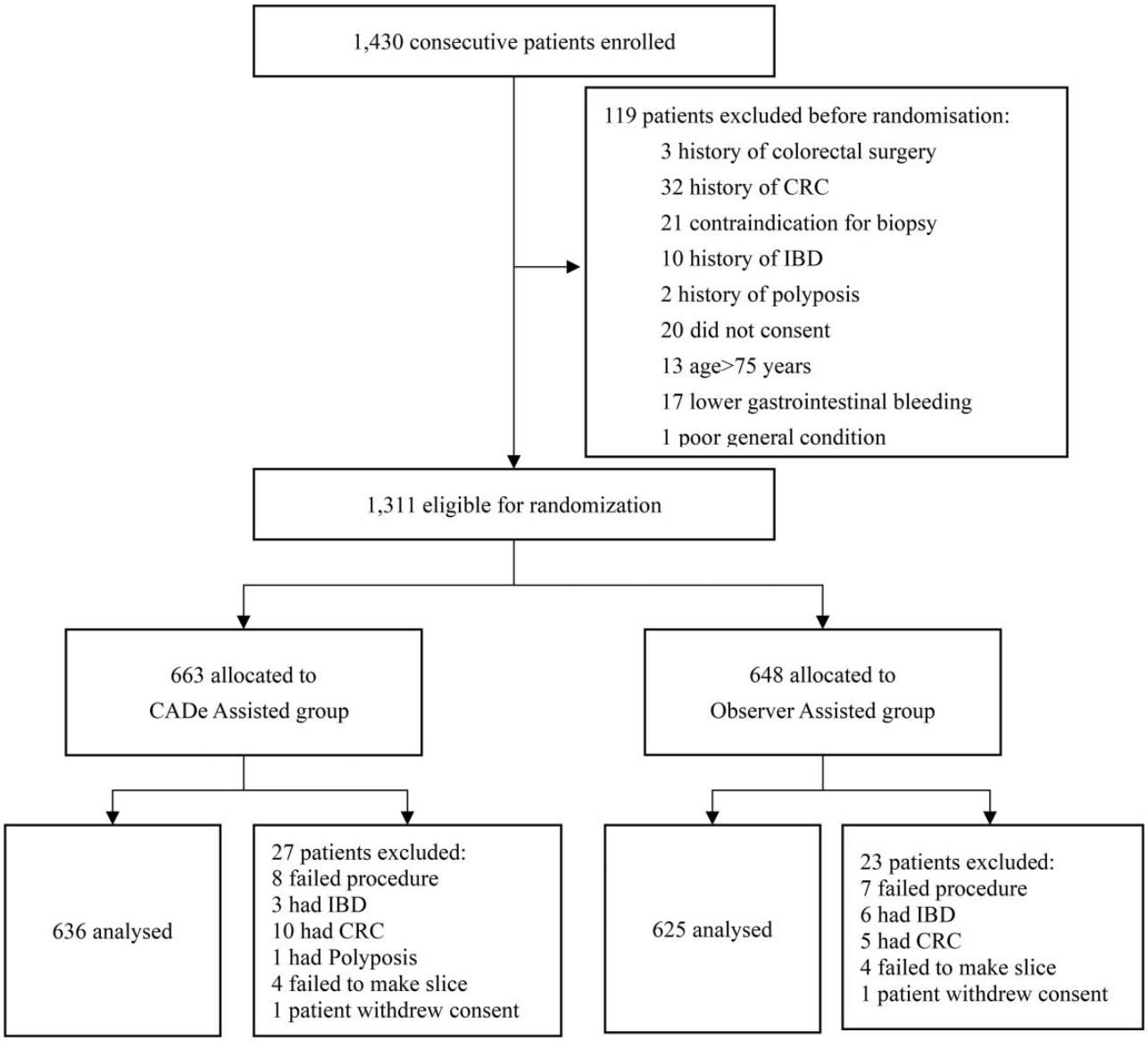

Figure 1. Flow diagram of enrollment

observer-assisted group (fold change [FC] 1.213, 95% CI 1.014 to 1.450, $P = 0.034$). The PPC was 1.13 in the CADe-assisted group and 0.81 in the observer-assisted group (FC 1.390, 95% CI 1.241 to 1.557, $P < 0.001$) (Table 4). There was no statistical difference in the detection of advanced adenomas ($P = 0.562$) or SSLs ($P = 0.199$) between the two groups.

In terms of morphology, the number of detected sessile polyps (683 vs 464, $P < 0.001$) and that of sessile adenomas (244 vs 182, $P = 0.005$) were significantly higher in the CADe-assisted group than in the observer-assisted group. In terms of the size, the number of detected diminutive polyps was significantly higher in the CADe-assisted group than in the observer-assisted group (0–5 mm, 616 vs 425, $P < 0.001$); the numbers of detected diminutive and small adenomas were also higher in the CADe-assisted group than in the observer-assisted group, but with no statistical significance (0–5 mm, 192 vs 159, $P = 0.11$; 6–10 mm, 72 vs 50, $P = 0.059$). No significant difference was observed for lesions (including polyps and adenoma) of >10 mm ($P = 0.594$ and 0.278). In terms of the location, more polyps were identified in the sigmoid colon and rectum while more adenomas were identified in the sigmoid colon in the CAD-assisted group than in the observer-assisted group (Table 3).

Patient-level analysis ADR and PDR

The ADR was 25.8% in the CADe-assisted group and 24.0% in the observer-assisted group (relative risk [RR] 1.100, 95% CI 0.852 to 1.421, $P = 0.464$). The PDR was 47.5% in the CADe-assisted group and 37.4% in the observer-assisted group (RR 1.511, 95% CI 1.207 to 1.892, $P < 0.001$) (Table 4).

False detections of the CADe system and second human observer

False detection (based on the judgement of operating endoscopists) of the CADe system and observers were 122 and 191, respectively ($P < 0.001$). Observers reported more false detections on wrinkled mucosa (119 vs 35, $P < 0.001$). The CADe system reported more false detections on local inflammation (23 vs 9,

Table 1. Characteristics of colonoscopy procedure and bowel cleansing

Characteristic	CADe-assisted group ($n = 636$)	Observer-assisted group ($n = 625$)	P-value[a]
Withdrawal time, min, mean (SD)	8.20 (± 2.77)	7.67 (± 2.53)	<0.001
Withdrawal time except biopsy, min, mean (SD)	7.06 (± 1.59)	6.87 (± 1.90)	0.055
Insertion time, min, mean (SD)	6.09 (± 5.60)	6.42 (± 4.92)	0.266
No polyp withdrawal time, min, mean (SD)	6.85 (± 1.27)	6.82 (± 1.70)	0.793
Procedure time of day			0.641
Morning, *n* (%)	223 (35.06)	227 (36.32)	
Afternoon, *n* (%)	413 (64.94)	398 (63.68)	
Endoscope type			0.988
EC-590ZW/M	23 (3.62)	24 (3.84)	
EC-L590WM	141 (22.17)	142 (22.72)	
EC-580RD/M	26 (4.09)	24 (3.84)	
EC-590WM	48 (7.55)	43 (6.88)	
EC-L590ZW	398 (62.58)	392 (62.72)	
Anesthesia[b]			0.973
No	144 (22.64)	142 (22.72)	
Yes	492 (77.36)	483 (77.28)	
Boston score, mean (SD)	6.83 (± 1.28)	6.87 (± 1.34)	0.678
Boston score rank, *n* (%)			0.463
Unqualified (sum <6.0 or anyone <2.0)	108 (16.98)	116 (18.56)	
Qualified (sum ≥6.0 and everyone ≥2.0)	528 (83.02)	509 (81.44)	

No polyp withdrawal time, withdrawal time during those colonoscopies where no polyp was detected or removed.
[a]P-value from χ^2 test (or Fisher's exact test, as appropriate).
[b]Anesthesia was administered with midazolam, fentanyl by an anesthesiologist there to monitor for complications.
CADe, computer-aided detection; SD, standard deviation.

Table 2. Sociodemographics of the participants

Characteristic	CADe-assisted group ($n = 636$)	Observer-assisted group ($n = 625$)	P-value
Age, years, median (IQR)	46.00 (36.75–54.00)	47.00 (37.00–55.00)	0.173
Indication, *n* (%)			0737
Screening	112 (17.61)	102 (16.32)	
Symptomatic	487 (76.57)	490 (78.40)	
Surveillance	37 (5.82)	33 (5.28)	
Sex, *n* (%)			0.07
Female	272 (42.77)	299 (47.84)	
Male	364 (57.23)	326 (52.16)	
BMI, median (IQR)	23.28 (21.22–25.37)	22.96 (20.89–24.97)	0.199
BMI category, *n* (%)			0.293
<25	462 (72.64)	473 (75.68)	
25 ≤ BMI < 30	162 (25.47)	137 (21.92)	
≥30	12 (1.89)	15 (2.40)	

CADe, computer-aided detection; BMI, body mass index; NSAID, non-steroidal anti-inflammatory drug.

Table 3. Characteristics of polyps and adenomas detected in each group

Characteristic	CADe-assisted group ($n = 720$)	Observer-assisted group ($n = 509$)	P-value
Pathology, *n* (%)			0.144
Carcinoma	0 (0.00)	0 (0.00)	1
SSL	13 (1.81)	7 (1.38)	0.199
Adenoma			
Advanced adenoma	1 (0.14)	2 (0.39)	0.562
Others	268 (37.22)	216 (42.44)	0.030
Benign lesions			
Hyperplastic and inflammatory	438 (60.83)	284 (55.80)	<0.001
Hamartoma	0 (0.00)	0 (0.00)	1
Normal colon mucosa	0 (0.00)	0 (0.00)	1
Polyp location, *n* (%)			0.030
Cecum	26 (3.61)	17 (3.34)	0.191
Ascending	106 (14.72)	89 (17.49)	0.274
Transverse	99 (13.75)	90 (17.68)	0.593
Descending	87 (12.08)	65 (12.77)	0.095
Sigmoid	176 (24.44)	99 (19.45)	<0.001
Rectum	226 (31.39)	149 (29.27)	<0.001
Polyp shape, *n* (%)			0.042
Pedunculated	33 (4.58)	42 (8.25)	0.266
Sessile	683 (94.86)	464 (91.16)	<0.001
Non-polypoid (LST)	4 (0.56)	3 (0.59)	0.723
Polyp size, mm, mean (SD)	3.47 (± 2.09)	3.90 (± 2.82)	0.233
Polyp size category, mm, *n* (%)			0.137
0–5[a]	616 (85.56)	425 (83.50)	<0.001
6–10	97 (13.47)	75 (14.73)	0.119
>10	7 (0.97)	9 (1.77)	0.594
Adenoma location, *n* (%)			0.427
Cecum	14 (5.20)	5 (2.29)	0.052
Ascending	60 (22.30)	53 (24.31)	0.572
Transverse	59 (21.93)	49 (22.48)	0.384
Descending	47 (17.47)	44 (20.18)	0.817
Sigmoid	71 (26.39)	48 (22.02)	0.045
Rectum	18 (6.69)	19 (8.72)	0.828
Adenoma shape, *n* (%)			0.051
Pedunculated	22 (8.18)	33 (15.14)	0.124
Sessile	244 (90.71)	182 (83.49)	0.005
Non-polypoid (LST)	3 (1.12)	3 (1.38)	0.983
Adenoma size, mm, mean (SD)	4.60 (± 2.34)	5.13 (± 3.52)	0.055
Adenoma size category, mm, *n* (%)			0.234
0–5	192 (71.38)	159 (72.94)	0.110
6–10	72 (26.77)	50 (22.94)	0.059
>10	5 (1.86)	9 (4.13)	0.278

[a]0–5 mm in diameter refers to diminutive in size.
CADe, computer-aided detection; SSL, sessile serrated lesion; LST, laterally spreading tumor; SD, standard deviation.

$P = 0.019$). There was no statistical difference between the two groups in terms of false detection on bubble, debris, circular blood vessels, or capsules (Table 5).

Of all detected polyps in the CADe-assisted group, none was missed by the CADe system.

Discussion

In the multicenter open-labeled randomized study, we found a significant improvement in APC in the CADe-assisted colonoscopy compared with the observer-assisted colonoscopy. APC was chosen over ADR as the primary outcome to directly

Table 4. Adenoma detection rate and adenomas detected per colonoscopy

Clinical outcome	CADe-assisted group ($n = 636$)	Observer-assisted group ($n = 625$)	**P-value**[a]	FC/RR	95% CI
PDR	47.5%	37.4%	<0.001	1.511	1.207–1.892
ADR	25.8%	24.0%	0.464	1.100	0.852–1.421
PPC	1.13	0.81	<0.001	1.390	1.241–1.557
APC	0.42	0.35	0.034	1.213	1.014–1.450

CADe, computer-aided detection; PDR, polyp detection rate; ADR, adenoma detection rate; PPC, polyps detected per colonoscopy; APC, adenomas detected per colonoscopy; RR, relative risk; FC, fold change.
[a]P-value from χ^2 test (or Fisher's exact test, as appropriate) or t-test.

Table 5. Characteristics of false alarms and missed polyps by CADe system and observer

False alarm	CADe-assisted group ($n = 636$)	Observer-assisted group ($n = 625$)	*P*-value
Total	122 (100.00)	191 (100.00)	<0.001
Bubble	3 (2.46)	9 (4.71)	0.094
Feces	13 (10.66)	11 (5.76)	0.715
Undigested debris	13 (10.66)	8 (4.19)	0.298
Wrinkled mucosa	35 (28.69)	119 (62.30)	<0.001
Circular blood vessel	34 (27.87)	35 (18.32)	0.847
Local inflammation	23 (18.85)	9 (4.71)	0.019
Local bleeding	1 (0.82)	0 (0.00)	0.997
Drug capsules	0 (0.00)	0 (0.00)	1
Diverticulum	0 (0.00)	0 (0.00)	1
Missed polyp	0	NA	NA

All values are presented as number of patients followed by percentage in parentheses.
NA, not applicable; CADe, computer-aided detection.

compare the performance of the second human observer and the CADe system in the detection of every single adenoma. Although ADR is a guideline-recommended quality indicator, it can lead to the "one and done" phenomenon that endoscopists may subsequently perform the procedure with less intensity after the identification of the first adenoma [31]. APC is considered a better-quality indicator as it reflects removal of all adenomas per colonoscopy and has been demonstrated to be significantly associated with ADR [32]. Increasing APC is as effective as increasing ADR in reducing the risk of interval CRC [33, 34]. Additionally, the correlation coefficient between the APC and the advanced adenoma detection rate was also higher than that of ADR [35].

With the development of medical imaging technology, the limitations of human-eye diagnosis are underlined. Higher resolution and a larger visual field provide more information as well as challenges to human vision. Polyps that are originally visible but missed are mainly because they are non-obvious, briefly visible, partially blocked, or flashed at the edge of the screen [9]. A second human observer as a primary attempt to facilitate polyp detection may only partly mitigate the misdiagnosis. It is still not worldwide practice in the endoscopy workflow because of the natural defects of the human eye such as "inattentional blindness" [36, 37] and "change blindness" [38]. Using high-performance CADe systems that provide real-time pixel-level analysis of each frame in the video stream consistently as an assist during performing colonoscopy is an ideal approach to address the misdiagnosis of visible lesions.

In the study, the overall APC and PPC of a standard colonoscopy was improved more with CADe assistance than with human-observer assistance (APC, 0.42 vs 0.35, $P = 0.034$; PPC 1.13 vs 0.81, $P < 0.001$). The CADe system maintained high sensitivity to subtle lesions and assisted in detecting more sessile adenomas with sizes of <1 cm. Detecting the true number of polyps per patient facilitates the timely removal of high-risk precancerous lesions that are small in size and provides a better understanding of the risk of metachronous cancer and the surveillance interval [9, 39].

We further analysed each operating endoscopist of the four centers. A similar trend of improved polyp detection with CADe-assisted colonoscopy was found across the baseline detection rates of different endoscopists and their observers (Supplementary Table 4). In the study, the CADe-assisted group achieved higher APC with fewer false detections, providing more accurate assistance to performing endoscopists. The overall false detections of the CADe system were lower than those of human observers (122 vs 191, $P < 0.001$), indicating that the increase in adenoma detection did not come at the cost of an increase in false-alarm rates.

The CADe system is considered to involve a large number of false alerts during the endoscopy because some camera-captured distant areas alerted by AI's tracking box are confirmed as not being polyps when observed closely [40]. However, some of the alert boxes for distantly suspected polyps may contribute to the increase in ADR, thus we propose a new concept of "meaningful suspicion." The diagnostic process of endoscopists is "suspect then confirm" or "suspect then exclude," but the latter is often ignored under the human version of "false positive," as no one would often reflect on their false suspicions, whereas the CADe system faithfully analyzes all the suspected polyp areas and provides an alert box for anything that looks like a polyp away from the camera, no matter whether it is a true polyp or something else [21].

As an independent human second observer, trainees could fully focus on diagnosis without performing colonoscopy themselves. Interestingly, they reported more false detections than the CADe system. The correct and wrong suspicions of the performing endoscopist, together with the true and false detections from the assistance of the human observers or the CADe system, contribute to the eventual adenoma detection. The deep-learning-based CADe system should not pursue 100% specificity, but provide meaningful suspicions. It is critical to increase adenoma detection by closer observation for any area that looks like a polyp but cannot be identified due to limited pixels at a distance [16, 17, 21]. Research and development can be devoted to evaluating and categorizing the quality of false positives.

The study has several limitations. First, there was no blank control group. Previous studies have shown that CADe systems and observer assistance could improve the baseline ADR or PDR of a single endoscopist [9, 12, 17, 18]. The aim of this study was to directly compare CADe systems and observer assistance. Second, the non-blinded design may introduce subjective bias. Operating endoscopists may be more attentive due to a "competitive spirit" [31] or unduly rely on the alert of either CADe systems or human observers. Large-scale randomization and the participation of multiple endoscopists may minimize the bias. Third, we did not investigate the case of trainees as colonoscopy operators, nor senior endoscopists, trainees, and nurses as second observers. Different combinations of colonoscopy operators and second observers with different experience and diagnostic modalities need to be further investigated to derive the best indications for the application of the CADe system. Fourth, as the study was based on a Chinese population with a younger age than the guideline-recommended screening population, the result may not be generalizable to Western screening populations. Intercontinental multicenter studies involving endoscopists of different diagnostic levels as colonoscopy operators are needed to provide external validity.

In conclusion, results of the study indicate the advantage of the CADe system as a second observer. The APC and PPC of diminutive polyps were significantly higher with the assistance of the CADe system than with the assistance of the second human observer. Also, false detections of the CADe system were lower than those of observers. The compensating effect of AI on human vision cannot be achieved by increasing the number of human observers alone. The development of AI in colonoscopy should not be limited to fields in which humans are already skilled, but should also enhance human detectability by providing meaningful alarms.

Supplementary Data

Supplementary data is available at *Gastroenterology Report* online.

Authors' Contributions

All authors contributed to study concept. P.W. and X.G.L. contributed to the study design. P.X.L., P.W., G.Y.Z., X.P., M.K., M.L.S., F.X., M.M.D., H.F.X., J.J.L., X.Q.L., and Y.S. contributed to the acquisition of data. P.W. contributed to the interpretation of data and drafting of the manuscript. All authors read and approved the final manuscript.

Funding

The study has no funding.

Acknowledgements

We thank Dr Wenfei Zhang for the advice on statistical analysis. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional IRB and with the Helsinki Declaration. Informed consent was obtained from all individual participants enrolled in the study.

Conflict of Interest

None declared.

References

1. American Cancer Society. *Cancer Facts and Figures: 2017*. Atlanta, GA: American Cancer Society, 2017.
2. 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.e5.
3. Mahmud N, Cohen J, Tsourides K *et al.* Computer vision and augmented reality in gastrointestinal endoscopy. *Gastroenterol Rep (Oxf)* 2015;**3**:179–84.
4. Berzin TM, Topol EJ. Adding artificial intelligence to gastrointestinal endoscopy. *Lancet* 2020;**395**:485.
5. 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.e11.
6. 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.
7. 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.
8. Glissen Brown JR, Mansour NM, Wang P, *et al.* Deep learning computer-aided polyp detection reduces adenoma miss rate: a United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial) [published online ahead of print, 2021 Sep 14]. Clin Gastroenterol Hepatol 2022;**20**(7):1499–1507.e4. .
9. Douglas R, Alessandro R, Seth G *et al.* High-definition colonoscopy versus Endocuff versus EndoRings versus full-spectrum endoscopy for adenoma detection at colonoscopy: a multicenter randomized trial. *Gastrointest Endosc* 2018;**88**: 335–44
10. Uraoka T, Tanaka S, Saito Y *et al.* Computer-assisted detection of diminutive and small colon polyps by colonoscopy using an extra-wide-area-view colonoscope. *Endoscopy* 2021; **53**:E102–3.
11. 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.
12. Aslanian H, Shieh FK, Chan FW *et al.* Nurse observation during colonoscopy increases polyp detection: a randomized prospective study. *Am J Gastroenterol* 2013;**108**:166–72.
13. Lee CK, Park DI, Lee SH *et al.* Participation by experienced endoscopy nurses increases the detection rate of colon polyps during a screening colonoscopy: a multicenter, prospective, randomized study. *Gastrointest Endosc* 2011;**74**:1094–102.
14. Buchner AM, Shahid MW, Heckman MG *et al.* Trainee participation is associated with increased small adenoma detection. *Gastrointest Endosc* 2011;**73**:1223–31.
15. Tziatzios G, Gkolfakis P, Triantafyllou K. Effect of fellow involvement on colonoscopy outcomes: a systematic review and meta-analysis. *Dig Liver Dis* 2019;**51**:1079–85.
16. Alagappan M, Brown JRG, Mori Y *et al.* Artificial intelligence in gastrointestinal endoscopy: the future is almost here. *World J Gastrointest Endosc* 2018;**10**:239–49.

17. Misawa M, Kudo SE, Mori Y *et al*. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018;**154**:2027–9.e3.
18. Weigt J, Repici A, Antonelli G *et al*. Performance of a new integrated computer-assisted system (CADe/CADx) for detection and characterization of colorectal neoplasia. Endoscopy 2022;**54**(2):180–184.
19. Mori Y, Kudo SE, Berzin TM *et al*. Computer-aided diagnosis for colonoscopy. *Endoscopy* 2017;**49**:813–9.
20. Wang P, Xiao X, Glissen Brown JR *et al*. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* 2018;**2**:741–8.
21. 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
22. 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.e7.
23. 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.
24. Liu P, Wang P, Glissen Brown JR *et al*. The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study. *Therap Adv Gastroenterol* 2020;**13**:175628482097916.
25. Lei S, Wang Z, Tu M *et al*. Adenoma detection rate is not influenced by the time of day in computer-aided detection colonoscopy. *Medicine (Baltimore)* 2020;**99**:e23685.
26. Glissen Brown JR, Bilal M, Wang P *et al*. Introducing computer-aided detection to the endoscopy suite. *VideoGIE* 2020;**5**:135–7.
27. Zhou G, Xiao X, Tu M *et al*. Computer aided detection for laterally spreading tumors and sessile serrated adenomas during colonoscopy. *PLoS One* 2020;**15**:e0231880.
28. 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**:625–32.
29. Rutter MD, East J, Rees CJ *et al*. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. *Gut* 2020;**69**:201–23.
30. 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.
31. Wang HS, Pisegna J, Modi R *et al*. Adenoma detection rate is necessary but insufficient for distinguishing high versus low endoscopist performance. *Gastrointest Endosc* 2013;**77**:71–8.
32. Park SK, Kim HY, Lee CK *et al*. Comparison of adenoma detection rate and adenoma per colonoscopy as a quality indicator of colonoscopy. *Scand J Gastroenterol* 2016;**51**:886–90.
33. 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.
34. Kaminski MF, Regula J, Kraszewska E *et al*. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010;**362**:1795–803.
35. Park SK, Kim HY, Lee CK *et al*. Comparison of adenoma detection rate and adenoma per colonoscopy as a quality indicator of colonoscopy. *Scand J Gastroenterol* 2016;**51**:886–90.
36. Memmert D, Unkelbach C, Ganns S. The impact of regulatory fit on performance in an inattentional blindness paradigm. *J Gen Psychol* 2010;**137**:129–39.
37. Simons DJ, Chabris CF. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 1999;**28**: 1059–74.
38. Simons DJ, Rensink RA. Change blindness: past, present, and future. *Trends Cogn Sci* 2005;**9**:16–20.
39. Robertson DJ, Kaminski MF, Bretthauer M. Effectiveness, training and quality assurance of colonoscopy screening for colorectal cancer. *Gut* 2015;**64**:982–90.
40. Holzwanger EA, Bilal M, Glissen Brown JR *et al*. Benchmarking definitions of false-positive alerts during computer-aided polyp detection in colonoscopy. *Endoscopy* 2021;**53**:937–40.