Clinical Gastroenterology and Hepatology 2021;■:■–■

Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial)

Jeremy R. Glissen Brown,[*] Nabil M. Mansour,[‡] Pu Wang,[§] Maria Aguilera Chuchuca,[*] Scott B. Minchenberg,[||] Madhuri Chandnani,[*] Lin Liu,[¶] Seth A. Gross,[#] Neil Sengupta,[**] and Tyler M. Berzin[*]

[*]Center for Advanced Endoscopy, Division of Gastroenterology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts; [‡]Section of Gastroenterology and Hepatology, Baylor College of Medicine, Houston, Texas; [§]Department of Gastroenterology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, China; [||]Department of Internal Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts; [¶]Institute of Natural Sciences, MOE-LSC, School of Mathematical Sciences and SJTU-Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai, China; [#]Division of Gastroenterology and Hepatology, New York University Langone Health System, New York, New York; [**]Section of Gastroenterology, University of Chicago Medicine, Chicago, Illinois

BACKGROUND & AIMS: **Artificial intelligence-based computer-aided polyp detection (CADe) systems are intended to address the issue of missed polyps during colonoscopy. The effect of CADe during screening and surveillance colonoscopy has not previously been studied in a United States (U.S.) population.**

METHODS: **We conducted a prospective, multi-center, single-blind randomized tandem colonoscopy study to evaluate a deep-learning based CADe system (EndoScreener, Shanghai Wision AI, China). Patients were enrolled across 4 U.S. academic medical centers from 2019 through 2020. Patients presenting for colorectal cancer screening or surveillance were randomized to CADe colonoscopy first or high-definition white light (HDWL) colonoscopy first, followed immediately by the other procedure in tandem fashion by the same endoscopist. The primary outcome was adenoma miss rate (AMR), and secondary outcomes included sessile serrated lesion (SSL) miss rate and adenomas per colonoscopy (APC).**

RESULTS: **A total of 232 patients entered the study, with 116 patients randomized to undergo CADe colonoscopy first and 116 patients randomized to undergo HDWL colonoscopy first. After the exclusion of 9 patients, the study cohort included 223 patients. AMR was lower in the CADe-first group compared with the HDWL-first group (20.12% [34/169] vs 31.25% [45/144]; odds ratio [OR], 1.8048; 95% confidence interval [CI], 1.0780-3.0217; $P = .0247$). SSL miss rate was lower in the CADe-first group (7.14% [1/14]) vs the HDWL-first group (42.11% [8/19]; $P = .0482$). First-pass APC was higher in the CADe-first group (1.19 [standard deviation (SD), 2.03] vs 0.90 [SD, 1.55]; $P = .0323$). First-pass ADR was 50.44% in the CADe-first group and 43.64 % in the HDWL-first group ($P = .3091$).**

CONCLUSION: **In this U.S. multicenter tandem colonoscopy randomized controlled trial, we demonstrate a decrease in AMR and SSL miss rate and an increase in first-pass APC with the use of a CADe-system when compared with HDWL colonoscopy alone.**

Keywords: Adenoma Detection Rate; Adenoma Miss Rate; Computer-aided Detection; Deep Learning; Randomized Tandem Colonoscopy Study.

Abbreviations used in this paper: **ADR, adenoma detection rate; AI, artificial intelligence; AMR, adenoma miss rate; APC, adenomas per colonoscopy; BPPS, Boston Bowel Preparation Scale; CADe, computer-aided detection; CONSORT, Consolidated Standards of Reporting Trials; CRC, colorectal cancer; HDWL, high-definition white light; PDR, polyp detection rate; PMR, polyp miss rate; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; SSL, sessile serrated lesions; SSLPC, sessile serrated lesions per colonoscopy; U.S., United States.**

1542-3565
https://doi.org/10.1016/j.cgh.2021.09.009

Studies have shown that increased adenoma detection during colonoscopy is associated with a decreased risk of interval colon cancer.[1] However, adenoma detection rates (ADRs) vary significantly among physicians,[1] and tandem colonoscopy studies have demonstrated that adenoma miss rates (AMRs) may also vary greatly, between 6% and 41%.[2,3]

Missed adenomas can be broadly categorized into adenomas that remain obscured from the visual field (eg, behind mucosal folds or debris) and adenomas that appear partially or fully in the visual field but are missed by the endoscopist. For the former, devices that reduce blind spots, such as distal scope attachments and panoramic colonoscopy, may increase an individual endoscopist's ADR.[3,4] For adenomas that appear in the visual field but are missed by the endoscopist, several solutions have been proposed. A second observer during colonoscopy, such as a nurse observer or a gastroenterology trainee, has shown variable benefit in increasing adenoma detection as well.[5] Computer-aided detection (CADe) – the use of machine learning or deep learning for lesion detection – has recently been applied successfully during colonoscopy both retrospectively in still-image and video data[6] and prospectively in randomized clinical trials in China, Italy, and Japan.[5,7,8] There are no prospective data on the efficacy of a CADe system in a diverse population of patients in the United States (U.S.).

The aim of the Computer Aided Detection Tandem Colonoscopy Study (CADeT-CS) was to assess the comparative AMR for CADe-assisted colonoscopy when compared with high-definition white light (HDWL) colonoscopy alone.

What You Need to Know

Background

Prior studies have shown that deep learning computer-aided detection systems may increase adenoma detection rate and reduce adenoma miss rate, but little is known about how these technologies might function in a diverse patient population presenting for screening or surveillance colonoscopy in the United States.

Findings

This study showed a decrease in adenoma miss rate, polyp miss rate, and sessile serrated lesion miss rate and an increase in adenomas per colonoscopy in patients who underwent colonoscopy with the assistance of a deep learning computer-aided detection system compared with high definition white light colonoscopy.

Implications for patient care

In this United States multicenter randomized controlled trial, we showed that computer-aided detection has the potential to decrease inter-provider variability in colonoscopy quality by reducing adenoma miss rate, even in experienced providers.

Methods

This was a prospective, multi-center, single-blind randomized tandem colonoscopy study. Patients were enrolled across 4 university endoscopy centers in the U.S. from May 7, 2019, through November 24, 2020. This was an investigator-initiated study, with research software and study funding provided by Wision LLC. The protocol was registered on ClinicalTrials.gov (NCT03925337) and was approved by the Beth Israel Deaconess Medical Center Institutional Review Board and the institutional review boards of each participating site. All authors had access to the study data and reviewed and approved the final manuscript.

Study Population

We included adult patients ($\geq$22 years) presenting for colonoscopy for colorectal cancer (CRC) screening or surveillance. Patients were excluded if they were undergoing diagnostic colonoscopy (for indications such as gastrointestinal hemorrhage). We also excluded patients with inflammatory bowel disease, patients who were found to have colorectal masses $>$2 cm in size, patients referred for endoscopic mucosal resection, and patients with standard contraindications to colonoscopy such as acute diverticulitis and known or suspected perforation. Incomplete colonoscopies (those where endoscopists did not successfully intubate the cecum due to technical difficulties or poor bowel preparation) and patients found to have a Boston Bowel Preparation Scale (BPPS) score of 0 to 1 in any of 3 segments were excluded from the primary analysis.

Randomization

All eligible patients were randomized via computer-generated randomization to receive either CADe colonoscopy first or HDWL colonoscopy first, followed immediately by the other procedure via block randomization with a block size of 10. Randomization was achieved using a digital random number generator, and patient assignments were contained in sealed, opaque envelopes. Patients were blinded to the result of their randomization. Provider participants were informed of group allocation directly prior to the start of the colonoscopy procedure. The same provider-participant performed both colonoscopies.

Study Intervention

Patients underwent both HDWL colonoscopy and CADe-assisted colonoscopy during the trial period. The CADe system (EndoScreener, Shanghai Wision AI Co, Ltd,

Shanghai, China) is an automatic polyp detection system that was developed using a deep neural network based on SegNet architecture (Supplementary Methods).[9] The system has also been studied prospectively in 4 randomized clinical trials in China.[5,7,10,11] The CADe system was installed on a separate computer system, and the output of the system was projected on a second monitor that was connected to the primary monitor via a serial digital interface cable. The CADe system was programmed to output a blue bounded box on polyp instances during colonoscopy (Supplementary Figure 1).

Colonoscopy Procedure

Each colonoscopy was performed using Olympus CLV 190-series colonoscopes. Patients underwent sedation at the discretion of the gastroenterologist using a combination of benzodiazepine and narcotic medications or with propofol under the supervision of a trained anesthesiologist.

During insertion and standard withdrawal, the second (CADe) monitor was positioned facing away from the endoscopist and nursing staff. During artificial intelligence (AI)-assisted withdrawal, the screen was turned to face the endoscopist (Supplementary Figure 1).

Bowel preparation was evaluated and graded by the endoscopist using the BPPS.[12] Subjects with a BPPS of 0 to 1 in any of 3 segments (descending colon, transverse colon and ascending colon) were excluded from the primary analysis. After cecal intubation, withdrawal time was measured by the research assistant in real-time using a stop watch. Withdrawal time included time spent on polypectomy or therapeutic procedures. Polyp morphology was evaluated using the Paris Classification of superficial neoplastic lesions for colon polyps.[13] Polyp size, location, and polypectomy method were recorded. All polyps were removed at the moment of first identification by the endoscopist, using standard polypectomy techniques. Nonresectable lesions were biopsied. Diminutive, hyperplastic polyps in the rectum were left in situ at the discretion of the endoscopist if deemed clinically insignificant. Specimens were processed using standard methods and evaluated by clinical pathologists who were blinded to group allocation.

During CADe-assisted withdrawal, true positives, false positives, and false negatives were recorded. True positives were defined as lesions detected for $\geq$2 seconds by the research software and deemed to be consistent in appearance with a polyp by the endoscopist. False positives were defined as lesions detected for $\geq$2 seconds by the research software but ultimately deemed by the endoscopist to have a gross appearance not consistent with polyp. False negatives were defined as lesions that were not detected or detected for $<$2 seconds by the research software and were deemed by the endoscopist to be consistent with polyp. Any suspicious lesions encountered during CADe-assisted colonoscopy were inspected by the colonoscopist on the primary endoscopy monitor, and a final determination was made by the clinician on whether or not to remove the lesion.

Outcomes

The primary outcome was AMR, calculated as the number of histologically confirmed adenomas detected during the second colonoscopy in either arm divided by the total number of adenomas detected during both procedures. Polyp miss rate (PMR) – calculated as the number of polyps detected during the second colonoscopy in either arm divided by the total number of polyps detected during both procedures – was included as a secondary outcome. Other secondary outcomes included hyperplastic polyp and sessile serrated lesion (SSL) miss rates, which were calculated in identical fashion to AMR. ADR was calculated for the first colonoscopy, the second colonoscopy, and in total for each group, and was defined as the proportion of colonoscopies where at least 1 adenoma was detected during the relevant procedure. Polyp detection rate (PDR) was calculated as the proportion of colonoscopies where at least 1 polyp was detected. The mean number of adenomas and polyps detected in each group was also included in the analysis. Adenomas per colonoscopy (APC), polyps per colonoscopy, and sessile serrated lesions per colonoscopy were also calculated for the first-pass colonoscopy, the second-pass colonoscopy, and total in each group.

Statistical Analysis

A sample size of 196 patients was calculated (with a 1:1 randomization, hence 98 patients per each group) to assume a 10% adenoma miss rate (the primary endpoint) for AI-first, and a 35% adenoma miss rate for standard-first, allowing 80% power with a 2-sided alpha level of 0.05. A total of 234 patients were enrolled to account for additional withdrawal and participant dropout. The expected dropout rate was 10%.

For baseline sociodemographic and clinical characteristics and colonoscopy quality parameters, comparison between the 2 groups was performed using a *t*-test for continuous variables and a χ^2 test for categorical variables. PDRs and ADRs were compared using logistic regression, and polyps per colonoscopy and APC were compared using Poisson regression. We used multivariate logistic regression to analyze the factors associated with missed adenomas. Randomization group, patient age, gender, adenoma size and location, and BBPS were treated as fixed effects. Odds ratios (ORs) and 95% confidence intervals (CIs) were presented as well. A *P*-value of less than .05 was considered to signify statistical significance. Statistical analysis was performed using R (version 4.0.3).

Results

Study Population

A total of 234 patients were deemed eligible for assessment and were consented for the study. Two patients were excluded from randomization (Supplementary Methods; Figure 1). A total of 232 patients entered the study. Of these, 116 patients were randomized to the AI-assisted CADe-colonoscopy first group and 116 patients were randomized to the HDWL colonoscopy-first group. After the exclusion of 9 patients, the study cohort included 223 individuals. Of the patients, 45.3% (101/223) were female, 67.7% (151/223) were Caucasian, and 21% were African American (47/223) (Table 1). The procedure indication was primary CRC screening in 59.6% (133/223) and post-polypectomy surveillance in 40.4% (90/223). No difference was detected in procedure indication in each group.

Intraprocedural Characteristics

There was no significant difference in bowel cleansing in the first-pass colonoscopy in either group (Table 1). Median withdrawal time (inclusive of polypectomy and other intervention) was 9 minutes and 31 seconds (interquartile range [IQR], 07:42–13:45) on the first withdrawal in the CADe-first arm and 8 minutes and 30 seconds (IQR, 07:00–10:59) in the HDWL colonoscopy-first arm, and this difference was statistically significant ($P = .0098$). Second withdrawal time was longer in the HDWL colonoscopy-first arm (07:28; IQR, 06:20–09:16 vs 06:30; IQR, 05:45–07:14; $P = .02$). Excluding colonoscopies where polypectomy or other interventions were performed, there was no statistically significant difference in median withdrawal time. Median withdrawal time was 8 minutes and 28 seconds (IQR, 06:48–12:11) on the first withdrawal in the CADe colonoscopy-first arm and 7 minutes and 18 seconds (IQR, 06:17-08:26) on the first withdrawal in the HDWL colonoscopy-first arm ($P = .0614$) (Table 1).

Missed Lesion Analysis

In the CADe-first group, 34 adenomas were missed out of 169 total adenomas detected, for an AMR of 20.12% (34/169) compared with an AMR of 31.25% (45/144) in the HDWL-first group ($P = .0247$) (Table 2; Figure 2) with an OR of 1.8048 (95% CI, 1.0780–3.0217). The PMR was also significantly lower in the CADe-first group compared with the HDWL-first group (20.70% vs 33.71%; $P = .0007$). The miss rate of SSLs was also significantly lower in the CADe-first group compared with the HDWL-first group (7.140% vs 42.11%; $P = .0482$). There was no significant difference in the hyperplastic PMR or advanced AMR.

Characteristics of Missed Adenomas

There was no statistically significant difference in missed adenomas stratified by size (<5 mm, 5–9 mm and ≥10 mm) or location between groups. There was no statistically significant difference in lesion shape of missed adenomas between groups as well (Table 3). There were 2 missed SSLs that were ≥5 mm (Supplementary Table 2). These were 5 mm and 6 mm in size.

Factors Associated with Missed Adenomas

In a multivariate logistic regression analysis, 3 factors reached statistical significance as factors associated with missed adenomas: randomization to HDWL-first vs CADe (OR, 1.8830; $P = .0214$), age ≤65 years old (OR, 1.7390; $P = .0451$), and right colon vs other location (OR, 1.7865; $P = .0436$) (Supplementary Table 1; Supplementary Figure 2).

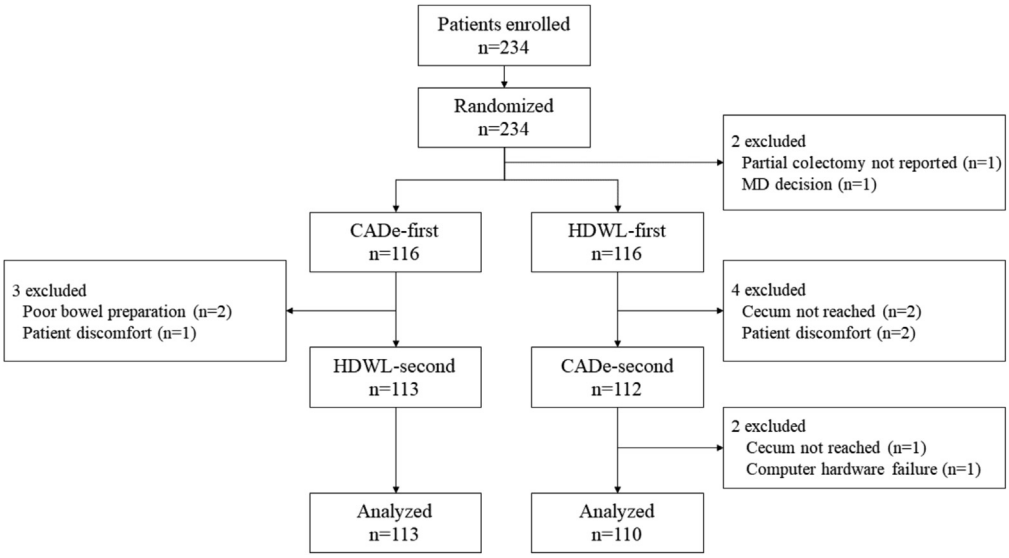


Figure 1. Study outline.

Table 1. Sociodemographic and Clinical Characteristics of the Patients

Characteristics	CADe-first (n = 113)	HDWL-first (n = 110)	*P*-value
Age, y	61.18 (9.83)	60.51 (8.45)	.5883
Sex, female	59 (52.21)	42 (38.18)	.0353
Race or ethnic group[a]			
Asian	7 (6.19)	8 (7.27)	.748
African American	25 (22.12)	22 (20.00)	.6974
Caucasian (non-Hispanic)	80 (70.80)	71 (64.55)	.3182
Other	1 (0.88)	9 (8.18)	.0085
Hispanic ethnic group	4 (3.54)	13 (11.82)	.0199
Procedure indication			.8687
Screening	68 (60.18)	65 (59.09)	
Surveillance	45 (39.82)	45 (40.91)	
BBPS[b]	9.00 (8.00–9.00)	9.00 (8.00–9.00)	.1176
Boston score rank			.427
Inadequate (sum <6.0 or any segment <2.0)	4 (3.54)	2 (1.82)	
Adequate (sum ≥6.0 and every segment ≥2.0)	109 (96.46)	108 (98.18)	
Withdrawal time, mm:ss[c]			
1st withdrawal time	09:31 (07:42–13:45)	08:30 (07:00–10:59)	.0098
2nd withdrawal time	06:30 (05:45–07:14)	07:28 (06:20–09:16)	< .0001
Withdrawal times without intervention,[d] mm:ss[b]			
1st withdrawal time	08:28 (06:48–12:11)	07:18 (06:17–08:26)	.0614
2nd withdrawal time	06:31 (05:55–07:13)	06:24 (06:08–07:20)	.9848

Note: Data presented as number (%), mean (SD), or median (IQR).
BBPS, Boston Bowel Prep Score; CADe, computer-aided detection; HDWL, high-definition white light; IQR, interquartile range; SD, standard deviation.
[a]Race and ethnic group were reported by the participant.
[b]BBPS range from 0 to 9, with higher scores indicating better bowel cleanliness.
[c]Minutes:seconds.
[d]Withdrawal times for procedures done without any intervention.

Detection Rates and Lesions per Colonoscopy

There was no statistically significant difference in PDR or ADR during the first procedure between both groups (ie, first-pass PDR and ADR) (Table 4). There was a significant difference in APC: APC was 1.19 (standard deviation [SD], 2.03) in the CADe-first group and 0.90 in the HDWL-first group (SD, 1.55) ($P = .0323$). Baseline ADR in the HDWL-first group was 43.64 on the first pass.

False Positives and False Negatives

There were 107 false positives during CADe colonoscopy in the CADe-first group and 96 false positives during CADe colonoscopy in the HDWL-first group ($P = .2128$). The majority of false positives in both groups were mucosal folds (Supplementary Table 4). There were 3 false negatives in the CADe-first group, defined as polyps detected by the endoscopist that

Table 2. Analysis of Per-lesion Miss Rate

Characteristic	CADe-first (n = 113)	HDWL-first (n = 110)	*P*-value	OR	95% CI
Polyp, total	285	264	.5612[a]	0.9516	0.8049–1.1250
Miss rate, %	20.70 (59/285)	33.71 (89/264)	.0007	1.9481	1.3273–2.8592
Adenoma, total	169	144	.2403[5]	0.8753	0.7009–1.0932
Miss rate, %	20.12 (34/169)	31.25 (45/144)	.0247	1.8048	1.0780–3.0217
Hyperplastic polyp, total	55	41	.1959[5]	0.7658	0.5111–1.1475
Miss rate, %	23.64 (13/55)	39.02 (16/41)	.1071	2.0677	0.8546–5.0029
Sessile serrated lesions	14	19	.3455[5]	1.3942	0.6990–2.7805
Miss rate, %	7.14 (1/14)	42.11 (8/19)	.0482	9.4545	1.0181–87.7969
Advanced adenoma,[b] total	9	5	.3146[5]	0.5707	0.1913–1.7029
Miss rate, %	11.11 (1/9)	0.00 (0/5)	.9971	<0.0001	<0.0001–inf

CADe, Computer-aided detection; CI, confidence interval; HDWL, high-definition white light; OR, odds ratio.
[a]Calculated using Poisson regression.
[b]Advanced adenoma defined as adenoma size ≥10 mm.

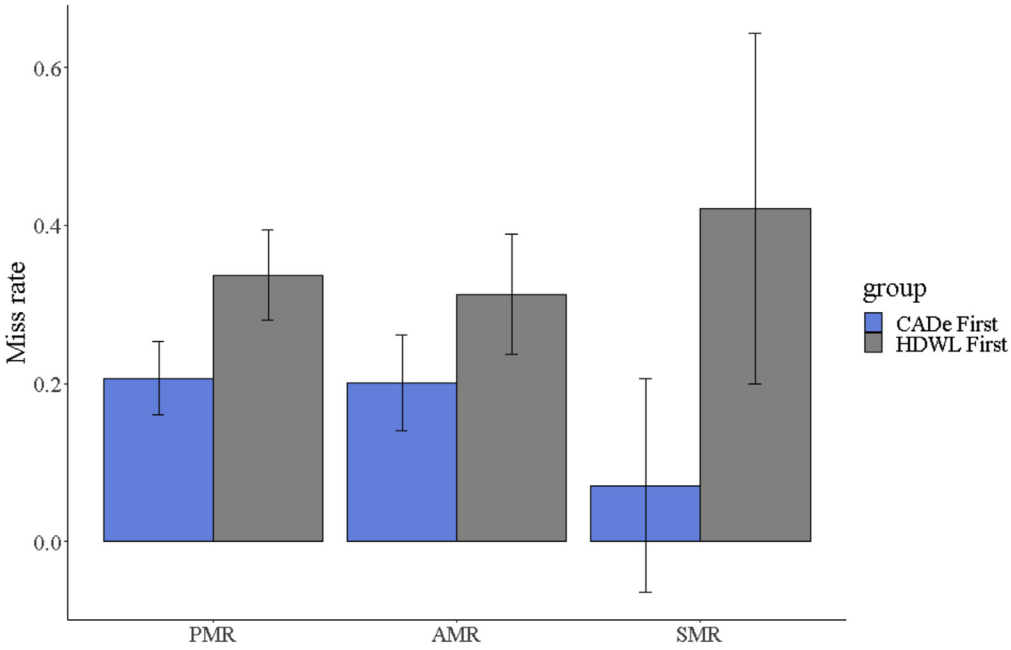


Figure 2. Per-lesion miss rates.

were not recognized by the CADe system (Supplementary Table 3).

Adverse Events

There were no immediate adverse events in the CADe-first group or the HDWL-first group.

Discussion

In this study, we compared AMRs using a CADe system during colonoscopy with standard-of-care HDWL colonoscopy in a randomized, single-blind, multi-center tandem colonoscopy study in a diverse patient population presenting for screening or surveillance colonoscopy in the U.S. We observed a relative reduction of 35.61% in AMR in the CADe-first group compared with the HDWL-first group and an absolute difference of 11.09%. In addition, PMR and SSL miss rates were significantly lower in the CADe-first group, and first-pass APC was significantly higher in the CADe-first group.

This is the first U.S. study to demonstrate the potential benefit of using a deep learning CADe system during screening and surveillance colonoscopy. Several prospective, randomized single-center trials in China have shown increases in ADR with the use of CADe systems.[14] These studies have limited generalizability to U.S. screening populations for several reasons, including low

Table 3. Characteristics of Missed Adenomas

Characteristic	CADe-first (n = 34), n (%)	HDWL-first (n = 45), n (%)	*P*-value
Adenoma size, mm			.4487
<5	24 (70.59)	35 (77.78)	
5–9	9 (26.47)	10 (22.22)	
≥10	1 (2.94)	0 (0.00)	
Adenoma location			.1024
Cecum	0 (0.00)	4 (8.89)	
Ascending colon	11 (32.35)	9 (20.00)	
Hepatic flexure	4 (11.76)	1 (2.22)	
Transverse colon	10 (29.41)	11 (24.44)	
Splenic flexure	0 (0.00)	1 (2.22)	
Descending colon	4 (11.76)	13 (28.89)	
Rectosigmoid	5 (14.71)	6 (13.33)	
Lesion shape			.5872
Ip	1 (2.94)	1 (2.22)	
Is	27 (79.41)	32 (71.11)	
IIa	6 (17.65)	10 (22.22)	
IIb	0 (0.00)	2 (4.44)	

CADe, Computer-aided detection; HDWL, high-definition white light.

Table 4. Polyp and Adenoma Detection Rates

Characteristic	CADe-first (n = 113)	HDWL-first (n = 110)	*P*-value[a]	OR	95% CI
Whole process					
PDR, %	75.22	76.36	.8422	1.0643	0.5765–1.9647
ADR, %	55.75	52.73	.6504	0.8852	0.5225–1.4997
APC, mean (SD)	1.50 (2.30)	1.31 (2.04)	.2403	0.8753	0.7009–1.0932
PPC, mean (SD)	2.52 (2.86)	2.40 (2.63)	.5612	0.9516	0.8049–1.1250
SSLPC, mean (SD)	0.12 (0.46)	0.17 (0.55)	.3455	1.3942	0.6990–2.7805
First pass					
PDR, %	70.80	65.45	.3923	0.7816	0.4444–1.3747
ADR, %	50.44	43.64	.3091	0.7606	0.4489–1.2887
APC, mean (SD)	1.19 (2.03)	0.90 (1.55)	.0323	0.7533	0.5812–0.9764
PPC, mean (SD)	2.00 (2.45)	1.59 (1.87)	.023	0.7955	0.6530–0.9690
SSLPC, mean (SD)	0.12 (0.46)	0.10 (0.33)	.7323	0.8692	0.3894–1.9402
Second pass					
PDR, %	38.05	45.45	.2629	1.3566	0.7954–2.3138
ADR, %	22.12	25.45	.5593	1.202	0.6482–2.2288
APC, mean (SD)	0.30 (0.62)	0.41 (0.93)	.1764	1.3596	0.8710–2.1225
PPC, mean (SD)	0.52 (0.78)	0.81 (1.31)	.0091	1.5496	1.1151–2.1534
SSLPC, mean (SD)	0.01 (0.09)	0.07 (0.35)	.047	8.2182	1.0279–65.7069

ADR, Adenoma detection rate; APC, adenomas per colonoscopy; CADe, computer-aided detection; CI, confidence interval; HDWL, high-definition white light; OR, odds ratio; PDR, polyp detection rate; PPC, polyps per colonoscopy; SD, standard deviation; SSLPC, sessile serrated lesions per colonoscopy.
[a]Calculated using logistic/Poisson regression.

baseline ADR for participating endoscopists and colonoscopy indications that were not limited to CRC screening/surveillance. A multi-center study in Italy showed a similar benefit in ADR and APC in a provider-participant pool with a high baseline ADR, along with higher proportions of both detected diminutive polyps and detected polyps 6 to 9 mm in size.[8] However, this study also included colonoscopy indications other than CRC screening/surveillance, and detailed demographic information regarding race and ethnicity were not reported.[8] Recently, Kamba et al performed a randomized tandem colonoscopy study across 4 endoscopy centers in Japan comparing standard white light colonoscopy with CADe-assisted colonoscopy and similarly found a lower AMR in the CADe-first group (13.8%) compared with standard colonoscopy (36.7%).[15]

In our multivariate logistic regression model (Supplementary Table 1), the predictors of missed polyps included use of HDWL-first rather than CADe-first colonoscopy, age ≤65 years, and location (right colon vs other location) (Supplementary Table 1). In addition, although there was not a statistically significant difference in missed diminutive or small polyps, the majority of missed adenomas in both groups were <5 mm or 5 to 9 mm in size. There is considerable debate regarding the clinical relevance of small vs diminutive adenomas and the effect that increased detection of diminutive adenomas might have on future surveillance intervals.[16]

In the present study, AMR in the experimental group (CADe-first) was 20.12%. Although this rate was significantly lower than in the HDWL-first group, it still represents missed adenomas. In a post-hoc analysis of our prior tandem CADe colonoscopy study, 3 senior endoscopists reviewed video recordings of the CADe system to understand the phenomenon of missed polyps when utilizing CADe, and found that the majority of polyps 'missed' during first-pass CADe colonoscopy were not present in the visual field during the first-pass colonsocopy.[17] In addition, in the present study, in which CADe detected 285 polyps, there were only 3 false negatives (defined as polyps that were visualized by the endoscopist but not by the CADe system). Overall, this suggests that the 'missed polyps' in the CADe arm may have been obscured behind folds rather than in the visual field. Further research is needed on combining CADe technologies with mucosal exposure devices, as the benefits of these tools for polyp detection may be additive.

In this study, there were 107 AI false positives in 113 colonoscopies in the CADe-first-first group and 96 AI false positives in 110 colonoscopies in the HDWL-first group. The rate of false positives was consistent across both groups, and the number of false positives was consistent with false positive rates in prior studies.[5,14] However, it is worth noting that the reported rate of false positives can change depending on the clinical definition used, and no gold standard has yet been established for a CADe false positive.[18]

Another strength of the present study is the finding of a decrease in SSL miss rate. Miss rates in general for flat adenomas and serrated polyps are higher than for polypoid adenomas, and SSLs contribute disproportionately to the histology of post-colonoscopy CRCs.[19] Most prospective trials have only shown an increase in ADR through an increase in diminutive (size ≤5 mm) and

small (5–9 mm) adenomas, with no increase in SSL detection rate. Our study, along with the study by Kamba et al,[15] is among the first to show the potential benefit for CADe specifically in reducing SSL miss rate. Larger studies are needed to see if this relationship bears out.

In this study, median withdrawal time for the first withdrawal was longer by 61 seconds in the CADe-first group compared with the HDWL-first group. The withdrawal time for the second withdrawal (the CADe withdrawal) was also significantly longer by 58 seconds in the HDWL-first group. There are several potential reasons for this finding. First, our protocol did not "pause the stopwatch" during polypectomy; thus, withdrawal time in this study reflects both mucosal inspection time as well as time required for interventions. Because more polyps were found during CADe-first colonoscopy, more polypectomies were also undertaken. In a subgroup analysis of procedures in which no polyps were found, withdrawal time was not significantly longer in the CADe group (Table 1). Because CADe was a new technology for all endoscopists involved in the study, this also may have affected withdrawal time, and the use of 2 monitors during CADe-assisted withdrawal could also be a factor.

This study has several limitations. First, our study was not powered to detect a difference in ADR. Second, the tandem colonoscopy design used in this study reveals important insights regarding CADe performance, but is somewhat limited in terms of generalizability to the real-world clinical setting. Endoscopists could not be blinded to a patient's group assignment while conducting each withdrawal. It is possible that endoscopist performance was influenced by being observed or that endoscopists who participated for the length of the study became over-reliant on CADe during withdrawal, leading to an overestimation or underestimate of CADe performance. However, these effects should generally have been balanced across both randomization groups.

Third, this study only included experienced endoscopists with a high baseline ADR at U.S. academic medical centers. It is less clear how CADe-assisted colonoscopy will affect endoscopy performance for trainees, for junior endoscopists, and in community settings. Some studies suggest the most benefit for CADe for endoscopists with limited experience or high procedure volume and for patients who present with a high polyp burden.[20] In this study, first-pass ADR in the HDWL-first group was 0.44. This value is well above the current American Society for Gastrointestinal Endoscopy benchmark of 0.25 for quality colonoscopy.[21] This may be seen as a particular strength of the current study as it shows efficacy of a CADe system in a provider-participant pool with a high baseline ADR.

Fourth, this study utilized a second monitor adjacent to the primary endoscopy monitor, similar to other early trials.[5,7] However, recent studies, including our previous tandem colonoscopy study, have utilized a single-monitor setup.[8,17] Although a dual-monitor setup may be less burdensome if latency of the CADe system is above a detectable visible threshold, it may also have negative effects on endoscopist gaze pattern.[22] We suspect a single-monitor setup may be preferred in the long run as it allows for easier integration of the technology.

Conclusion

In this U.S. multicenter randomized control trial, we showed a decrease in AMR with the use of a deep-learning CADe system when compared with HDWL colonoscopy alone. In addition, we showed a decrease in PMR and SSL miss rate and an increase in first-pass APC. CADe has the potential to decrease interprovider variability in colonoscopy quality by reducing AMR, even in experienced providers.

Clinicaltrials.gov, Number: NCT03925337.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of Clinical Gastroenterology and Hepatology at www.cghjournal.org, and at http://doi.org/10.1016/j.cgh.2021.09.009.

References

1. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014; 370:1298–1306.
2. Ahn SB, Han DS, Bae JH, et al. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. Gut Liver 2012;6:64–70.
3. 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;88:854–864.
4. Ngu WS, Bevan R, Tsiamoulos ZP, et al. Improved adenoma detection with Endocuff Vision: the ADENOMA randomised controlled trial. Gut 2019;68:280–288.
5. 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;68:1813–1819.
6. 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;2:741–748.
7. 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;5:343–351.
8. 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.
9. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–2495.
10. 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;13:1756284820979165.

11. Chen P-J, Lin M-C, Lai M-J, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018;154:568–575.
12. Lai EJ, Calderwood AH, Doros G, et al. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009;69(3 Pt 2):620–625.
13. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc 2003;58(6 Suppl):S3–S43.
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. Kamba S, Tamai N, Saitoh I, et al. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. J Gastroenterol 2021; 56:746–757.
16. Anderson JC, Rex DK, Robinson C, et al. Association of small versus diminutive adenomas and the risk for metachronous advanced adenomas: data from the New Hampshire Colonoscopy Registry. Gastrointest Endosc 2019;90:495–501.
17. 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.
18. Holzwanger EA, Bilal M, Glissen Brown JR, et al. Benchmarking false positive definitions for computer aided polyp detection in colonoscopy. Endoscopy 2021;53:937–940.
19. Pohl H, Srivastava A, Bensen SP, et al. Incomplete polyp resection during colonoscopy-results of the Complete Adenoma REsection (CARE) study. Gastroenterology 2013;144:74–80.e1.
20. 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.
21. Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. Gastrointest Endosc 2015;81:31–53.
22. Almansa C, Shahid MW, Heckman MG, et al. Association between visual gaze patterns and adenoma detection rate during colonoscopy: a preliminary investigation. Am J Gastroenterol 2011;106:1070–1074.

Reprint requests
Address requests for reprints to: Jeremy R. Glissen Brown, MD, Center for Advanced Endoscopy, Division of Gastroenterology and Hepatology, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Ave, Boston, Massachusetts 02130. tel: (617) 754-8888. e-mail: jglissen@bidmc.harvard.edu.

Acknowledgment
The authors thank the following research coordinators for their efforts: Roya Dastdjerdi, MPH, at Beth Israel Deaconess Medical Center; Olga Solonowicz, BS, at the University of Chicago; Ghoncheh Ghiasian, MPH, at New York University Langone Medical Center; and Chandra Kovvali, MHA, Michael Mercado, and Vanessa Balonwu at Baylor College of Medicine. In addition, the authors thank Joseph Feuerstein, MD, and Mandeep Sawhney, MD, at Beth Israel Deaconess Medical Center as well as Sonia Kupfer, MD, at the University of Chicago for their participation as provider-participants in the study. We also thank J. Thomas Lamont for his contribution to manuscript review.

CRediT Authorship Contributions

Jeremy R. Glissen Brown, MD (Conceptualization: Equal; Data curation: Supporting; Investigation: Supporting; Project administration: Supporting; Supervision: Supporting; Writing – original draft: Lead; Writing – review & editing: Equal)

Nabil M. Mansour, MD (Data curation: Supporting; Investigation: Equal; Supervision: Supporting; Writing – review & editing: Equal)

Pu Wang, MD (Conceptualization: Equal; Methodology: Equal; Writing – review & editing: Equal)

Maria Aguilera Chuchuca, MD (Data curation: Equal; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Writing – review & editing: Supporting)

Scott Minchenberg, MD, PhD (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Writing – review & editing: Equal)

Madhuri Chandnani, MD (Data curation: Supporting; Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting)

Lin Liu, PhD (Formal analysis: Lead; Methodology: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Seth A. Gross, MD, FACG, FASGE, AGAF (Data curation: Supporting; Investigation: Equal; Project administration: Equal; Writing – review & editing: Equal)

Neil Sengupta, MD, FACG (Data curation: Supporting; Investigation: Equal; Methodology: Supporting; Writing – review & editing: Equal)

Tyler M. Berzin, MD, MS, FASGE (Conceptualization: Lead; Data curation: Supporting; Formal analysis: Supporting; Investigation: Lead; Methodology: Supporting; Project administration: Supporting; Writing – review & editing: Lead)

Conflicts of interest
These authors disclose the following: Seth Gross has served as a consultant for Olympus. Tyler Berzin has served as a consultant for Wision AI, Fujifilm, and Medtronic. The remaining authors disclose no conflicts.

Funding
This was an investigator-initiated study, with research software and study funding provided by Wision LLC.
Lin Liu gratefully acknowledges funding support by Shanghai Municipal Science and Technology Major Project No.2021SHZDZX0102, Major Program of National Natural Science Foundation of China No.12090024 and Shanghai Natural Science Foundation Grant No.21ZR1431000.

Supplementary Methods

The study was designed taking into account the original Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) reporting guidelines[1] and artificial intelligence (AI)-specific extensions to SPIRIT designed for the reporting of clinical trials involving AI-specific interventions.[2] The manuscript was prepared according to CONSORT (Consolidated Standards of Reporting Trials) reporting guidelines for the dissemination of trial reports[3] as well as recently published AI-extensions to CONSORT.[4]

Study Intervention

The system was developed on 5545 colonoscopy images from 1290 patients in a single endoscopy center in the Endoscopy Center of Sichuan Provincial People's Hospital between January 2007 and December 2015 and validated on a dataset of 27,113 images from 1138 different patients; a public database of 612 polyp-containing images from Hospital Clinic of Barcelona, Barcelona, Spain; a dataset consisting of video of 138 polyp instances; and a collection of 54 unaltered full-length colonoscopy videos.[5] Validation of the algorithm showed high per-image sensitivity (94.38%), high per-image specificity (95.92%), and a low false positive rate.[5] The iteration of the research software used in the current study processed >30 frames per second with a latency of 46.56 ± 2.79 ms (Version 1.0.0).

During insertion and standard withdrawal, the second (CADe) monitor was positioned facing away from the endoscopist and nursing staff. During AI-assisted withdrawal, the screen was turned to face the endoscopist (as seen in Supplementary Figure 1). A research assistant was in the room and had visual access to the second screen only when the second screen was visible to the rest of the providers.

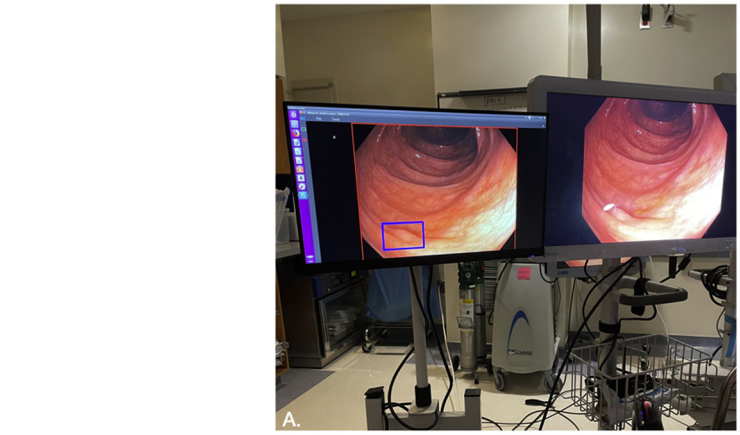
Supplementary Results

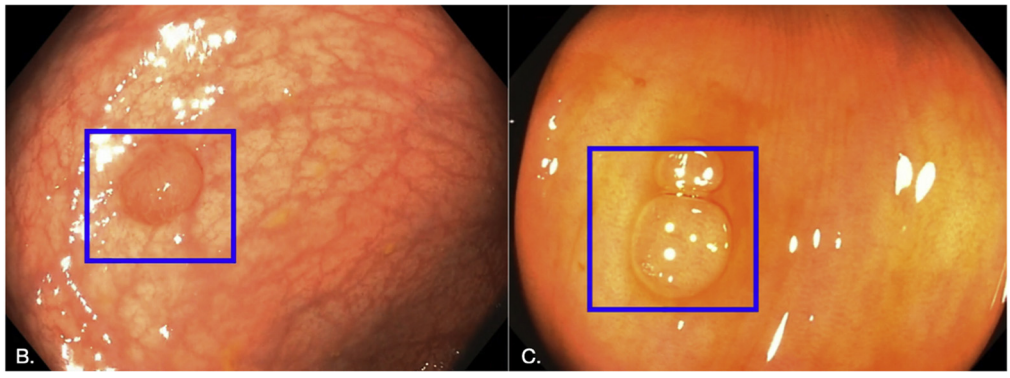
Study Population

Two patients were excluded from randomization (Figure 1): 1 due to altered anatomy encountered during the procedure (partial colectomy that was not reported in the history) and the second due to late recognition of poor bowel preparation.

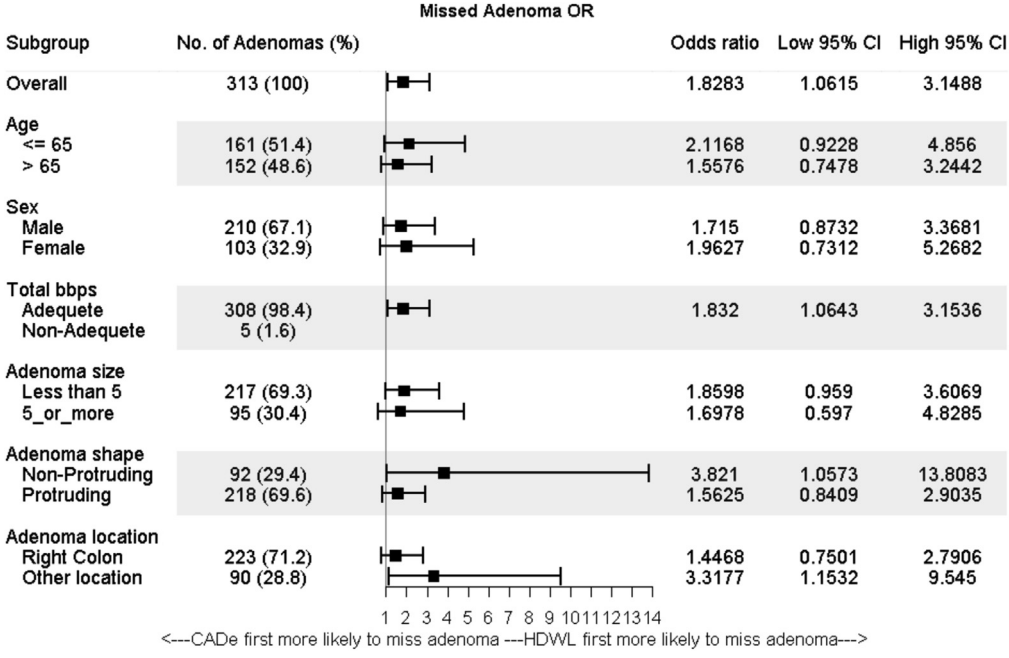
Supplementary References

1. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158:200–207.
2. Cruz Rivera S, Liu X, Chan AW, et al. SPIRIT-AI and CONSORT-AI Working Group; SPIRIT-AI and CONSORT-AI Steering Group; SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 2020;26:1351–1363.
3. Schulz KF, Altman DG, Moher D, et al. , CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
4. Liu X, Rivera SC, Moher D, et al. , SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020;370:m3164.
5. 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;2:741–748.





Supplementary Figure 1. An example of the dual-monitor set up employed in our study **(A)**; bounded box output of the CADe system on a polyp instance **(B)**; and a false positive (*bubble*) **(C)**.



Supplementary Figure 2. Subgroup analysis. Subgroup analyses were performed using multivariate logistic regression.

Supplementary Table 1. Multivariate Analysis of Factors Associated With Missed Adenoma

Predictors	OR	95% CI	*P*-value
Randomization: (CADe-first vs HDWL-first)	1.8261	1.0671–3.1247	.0280
Age, y: (≤65 vs >65)	1.7163	1.0025–2.9385	.0490
Location: (right colon vs other)	1.8467	1.0486–3.2522	.0336
Size, mm: (<5 vs ≥5)	0.7242	0.3991–1.3143	.2886
BBPS: (sum ≥6.0 vs other)	0.7084	0.0744–6.7472	.7643
After adjusting for gender:			
Randomization: (CADe-first vs HDWL-first)	1.8830	1.0984–3.2278	.0214
Age, y: (≤65 vs >65)	1.7390	1.0122–2.9876	.0451
Gender: (female vs male)	1.0268	0.5816–1.8127	.9273
Location: (right colon vs other)	1.7865	1.0167–3.1392	.0436
Size, mm: (<5 vs ≥5)	0.8868	0.4816–1.6330	.6998
BBPS: (sum ≥6.0 vs other)	0.7780	0.0819–7.3940	.8270

BBPS, Boston Bowel Preparation Scale; CADe, computer-aided detection; CI, confidence interval; HDWL, high-definition white light; OR, odds ratio.

Supplementary Table 2. Characteristics of Missed Sessile Serrated Lesions

Characteristic	CADe-first (n = 1), n (%)	HDWL-first (n = 8), n (%)	*P*-value
Adenoma size, mm			
<5	1 (100)	6 (75.00)	
≥5	0 (0.00)	2 (25.00)	.5707
Adenoma location			
Cecum	1 (100)	1 (12.50)	
Ascending colon	0 (0.00)	6 (75.00)	
Hepatic flexure	0 (0.00)	1 (12.50)	
Transverse colon	0 (0.00)	0 (0.00)	
Splenic flexure	0 (0.00)	0 (0.00)	
Descending colon	0 (0.00)	0 (0.00)	
Rectosigmoid	0 (0.00)	0 (0.00)	.1396
Lesion shape			
Ip	0 (0.00)	0 (0.00)	
Is	1 (100)	5 (62.50)	
IIa	0 (0.00)	3 (37.50)	
IIb	0 (0.00)	0 (0.00)	.4533

CADe, Computer-aided detection; HDWL, high-definition white light.

Supplementary Table 3. False Negative Rate

Characteristic	All polyps detected in CADe procedures (n = 315)
False negative rate,[a] n (%)	3 (0.95)

CADe, Computer-aided detection.
[a]Detected by an endoscopist but not by CADe..

Supplementary Table 4. False Positive Rate

Characteristic	CADe-first (n = 113)	HDWL-first (n = 110)	*P*-value
False detection, n (%)	107 (100)	96 (100)	.2128
Mucosa fold	66 (61.68)	66 (68.75)	
Feces debris	7 (6.54)	1 (1.04)	
Bubble	5 (4.67)	7 (7.29)	
Suction polyp	11 (10.28)	8 (8.33)	
Prior polypectomy site	2 (1.87)	6 (6.25)	
Pill/medication	1 (0.93)	0 (0.00)	
Other (eg, mucosal trauma, inverted diverticulum, blood vessel, etc)	15 (14.02)	8 (8.33)	

CADe, Computer-aided detection; HDWL, high-definition white light.