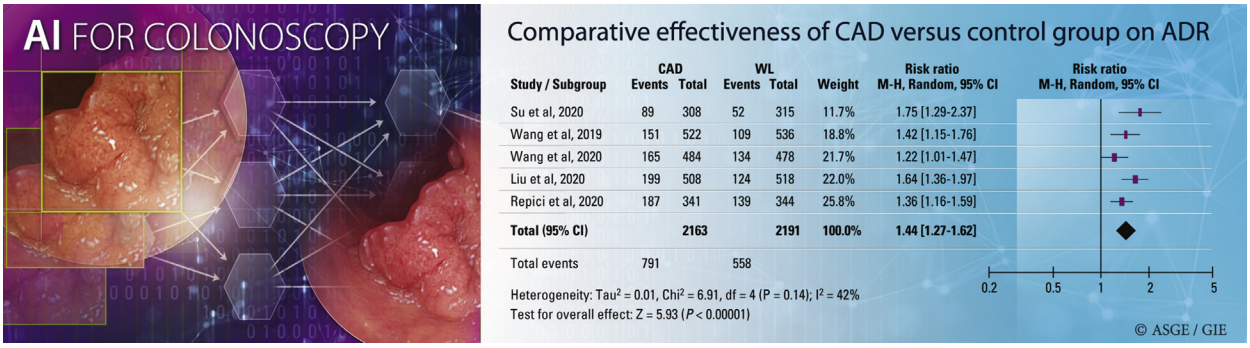
Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis

Cesare Hassan, MD, PhD, Marco,[1,*] Marco Spadaccini, MD,[2,3,*] Andrea Iannone, MD, PhD,[4] Roberta Maselli, MD, PhD,[2] Manol Jovani, MD,[5,6] Viveksandeep Thoguluva Chandrasekar, MD,[7] Giulio Antonelli, MD,[1] Honggang Yu, MD,[8] Miguel Areia, MD, PhD,[9] Mario Dinis-Ribeiro, MD,[10] Pradeep Bhandari, MD,[11] Prateek Sharma, MD, PhD,[7] Douglas K. Rex, MD,[12] Thomas Rösch, MD, PhD,[13] Michael Wallace, MD, PhD,[14] Alessandro Repici, MD[2,3]

Rome, Rozzano, Bari, Italy; Baltimore, Maryland; Boston, Massachusetts; Kansas City, Missouri; Indianapolis, Indiana; Jacksonville, Florida, USA; Wuhan, China; Coimbra, Porto, Portugal; Portsmouth, UK; Hamburg, Germany

GRAPHICAL ABSTRACT



Background and Aims: One-fourth of colorectal neoplasia are missed at screening colonoscopy, representing the main cause of interval colorectal cancer. Deep learning systems with real-time computer-aided polyp detection (CADe) showed high accuracy in artificial settings, and preliminary randomized controlled trials (RCTs) reported favorable outcomes in the clinical setting. The aim of this meta-analysis was to summarize available RCTs on the performance of CADe systems in colorectal neoplasia detection.

Methods: We searched MEDLINE, EMBASE, and Cochrane Central databases until March 2020 for RCTs reporting diagnostic accuracy of CADe systems in the detection of colorectal neoplasia. The primary outcome was pooled adenoma detection rate (ADR), and secondary outcomes were adenoma per colonoscopy (APC) according to size, morphology, and location; advanced APC; polyp detection rate; polyps per colonoscopy; and sessile serrated lesions per colonoscopy. We calculated risk ratios (RRs), performed subgroup and sensitivity analyses, and assessed heterogeneity and publication bias.

Results: Overall, 5 randomized controlled trials (4354 patients) were included in the final analysis. Pooled ADR was significantly higher in the CADe group than in the control group (791/2163 [36.6%] vs 558/2191 [25.2%]; RR, 1.44; 95% confidence interval [CI], 1.27-1.62; $P < .01$; $I^2 = 42\%$). APC was also higher in the CADe group compared with control (1249/2163 [.58] vs 779/2191 [.36]; RR, 1.70; 95% CI, 1.53-1.89; $P < .01$; $I^2 = 33\%$). APC was higher for ≤5-mm (RR, 1.69; 95% CI, 1.48-1.84), 6- to 9-mm (RR, 1.44; 95% CI, 1.19-1.75), and ≥10-mm adenomas (RR, 1.46; 95% CI, 1.04-2.06) and for proximal (RR, 1.59; 95% CI, 1.34-1.88), distal (RR, 1.68; 95% CI, 1.50-1.88), flat (RR, 1.78; 95% CI, 1.47-2.15), and polypoid morphology (RR, 1.54; 95% CI, 1.40-1.68). Regarding histology, CADe resulted in a higher sessile serrated lesion per colonoscopy (RR, 1.52; 95% CI, 1.14-2.02), whereas a nonsignificant trend for advanced ADR was found (RR, 1.35; 95% CI, .74-2.47; $P = .33$; $I^2 = 69\%$). Level of evidence for RCTs was graded as moderate.

Conclusions: According to available evidence, the incorporation of artificial intelligence as aid for detection of colorectal neoplasia results in a significant increase in the detection of colorectal neoplasia, and such effect is independent from main adenoma characteristics. (Gastrointest Endosc 2021;93:77-85.)

(footnotes appear on last page of article)

Interval colorectal cancer represents one of the most dismal consequences of screening colonoscopy, with an incidence of .5 to 1 per 1000 patient-years.[1,2] The main cause is represented by overlooked lesions that may be referred to recognition failure (when the endoscopist misses a lesion present on the screen) or incomplete mucosal exposure that depends on the complexity of the colorectal anatomy and/or suboptimal technique in the withdrawal phase of colonoscopy.[3-6]

By addressing these pitfalls, artificial intelligence is expected to reduce the risk of miss rate and consequently of interval colorectal cancer.[7,8] In detail, the adoption of convoluted neural networks (CNNs) or deep learning led to the technical feasibility of real-time computer-aided detection (CADe) capable of flagging the suspected lesion to the endoscopist with a visual and acoustic alarm. Different CNN systems are also able to alert the endoscopist any time the technical standard of withdrawal technique is suboptimal (ie, speed of withdrawal, inadequate level of cleansing, or slipping of the endoscope).

After successful artificial validation,[9] CADe systems have been applied in the clinical setting to assess their benefit of improved detection in terms of adenoma detection rate (ADR) or adenoma per colonoscopy (APC) as well as possible harms such as deskilling of the endoscopist or time wasting because of false-positive results.[10-12] Initial studies demonstrated favorable results on detection[10-12] but were generally underpowered to assess the relationship between increased detection and lesion characteristics, such as polyp size, morphology, location, or histology. In this regard, the miss rate of colorectal neoplasia at screening colonoscopy has been variably associated with small size, flat morphology, proximal location, and serrated histology in back-to-back studies.[3] In addition, there is uncertainty on whether the additional detection of neoplasia, namely an increase in ADR, is also associated with an increase in the detection of advanced adenomas, defined as either ≥10 mm or unfavorable histology.[13,14] The aim of our systematic review and meta-analysis was to assess the relationship between the increased detection led by CADe and the main features of the detected lesions.

METHODS

This systematic review was reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[15]

Data sources and search strategy

We performed a comprehensive literature search in MEDLINE, Cochrane Central Register of Controlled Trials, and EMBASE (up to March 31, 2020) electronic databases to identify randomized controlled trials (RCTs) evaluating the role of CADe systems in lesion detection or mucosal exposure. A specialist with expertise in systematic reviews of randomized trials designed the search strategy (Appendix 1, available online at www.giejournal.org). Electronic searches were supplemented by manual searches of references of included studies and review articles.

Selection process

Two review authors (M.S., A.I.) independently screened the titles and abstracts yielded by the search against the inclusion criteria. Full reports were obtained for all titles that appeared to meet the inclusion criteria or where there was any uncertainty. Review author pairs then screened the full text and abstract reports and decided whether these met the inclusion criteria. The reasons for excluding trials were recorded. Neither review author was blinded to the journal titles or to the study authors or institutions. For multiple articles for a single study, we used the latest publication and supplemented it, if necessary, with data from the more complete version.

Data extraction

Using standardized forms, 2 reviewers (M.S., A.I.) extracted data independently and in duplicate from each eligible study. Reviewers resolved disagreements by discussion. Unresolved disagreements were resolved by 2 arbitrators (C.H., A.R.). Data extracted for each study included publication status, study design and location, number of centers involved, number of patients, patient characteristics (mean/median age, gender), colonoscopy indication, ADR, polyp detection rate, number and characteristics (size,

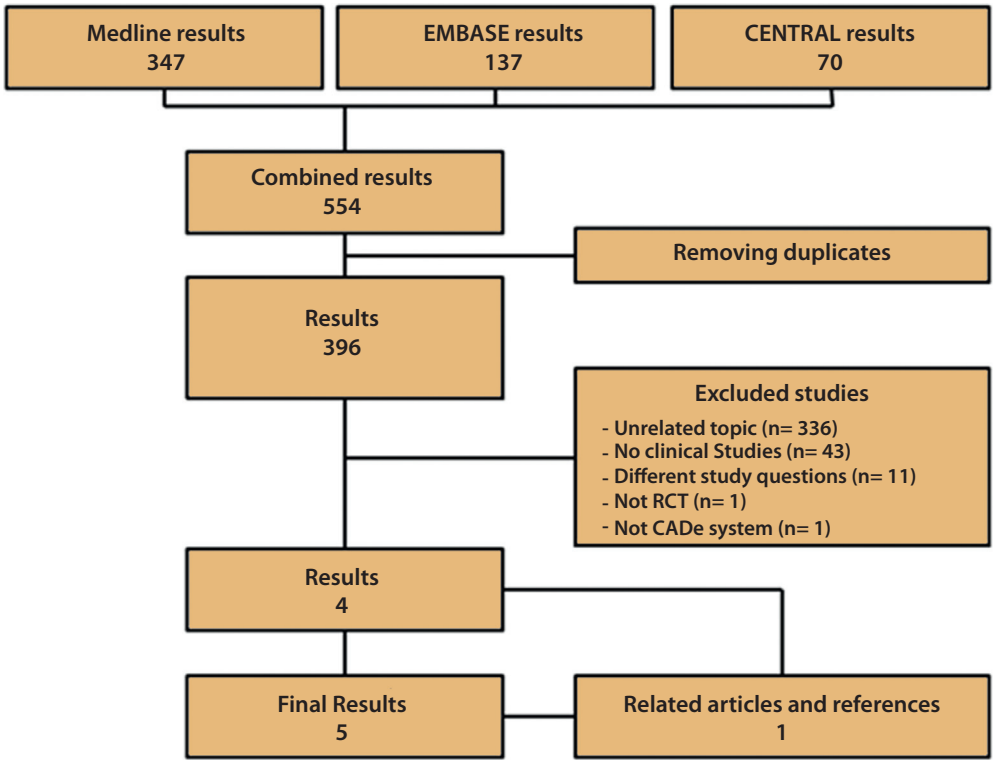


Figure 1. Study selection flow chart. *CADe*, Computer-aided polyp detection; *RCT*, randomized controlled trial.

location, and histology) of detected polyps, and withdrawal time. The corresponding authors of the included studies were asked for missing data. In the case of discrepancy or if data were missing, an attempt to contact the corresponding authors was done.

Inclusion and exclusion criteria

For the purpose of our meta-analysis, we screened all clinical studies for the following inclusion criteria: all adults (>18 years old) undergoing colonoscopy in a nonemergency setting, colonoscopy with high-definition endoscopes implemented with real-time CADe systems, colonoscopy with high-definition endoscopes, ADR and polyp detection rate, and study design (only RCTs were considered). Exclusion criteria were unavailable essential information, studies not published as full text articles, and studies including less than 50 patients in each group.

Outcomes

The primary outcome was the ADR, defined as the proportion of individuals undergoing a complete colonoscopy who had at least 1 adenoma detected (and removed). Secondary outcomes were as follows:

- APCs: number of APCs, calculated by dividing the total number of adenomas detected by the total number of colonoscopies. APC was investigated according to lesion characteristics, namely size, morphology, location, and histology.
- Polyp detection rate: proportion of individuals undergoing a complete colonoscopy who had at least 1 polyp detected.
- Polyps per colonoscopy: number of polyps per colonoscopy, calculated by dividing the total number of polyps detected by the total number of colonoscopies. Polyps per colonoscopy were investigated according to lesion characteristics, namely size, morphology, location, and histology.
- Sessile serrated lesions per colonoscopy: number of sessile serrated lesions per colonoscopy, calculated by dividing the total number of sessile serrated lesions detected by the total number of colonoscopies. The definition of sessile serrated lesion adopted by the authors was used for the purpose of our analysis.
- Advanced APCs: number of advanced APCs, calculated by dividing the total number of advanced adenomas detected by the total number of colonoscopies.
- Withdrawal time: the time spent in inspecting the colonic mucosa as the endoscope is withdrawn during a colonoscopy. Biopsy sampling or treatment time was excluded in all studies.

Quality assessment

Quality was assessed by the Cochrane Risk Bias Tool for randomized studies. Two reviewers (M.S., A.I.) assessed quality measures for included studies, and discrepancies were adjudicated by discussion.

TABLE 1. Study characteristics

Reference	Publication year	Country	Type of system	Colonoscopy indications	Endoscopists
Wang et al[11]	2019	China	CADe	Symptomatic: 974 Screening or surveillance: 84	2 senior endoscopists (>20,000 colonoscopies) 2 midlevel endoscopists (3000-10,000) 4 junior endoscopists (100-500)
Wang et al[21]	2020	China	CADe	Symptomatic: 804 Screening or surveillance: 158	4 senior endoscopists (>5 years' experience and >1000 colonoscopies per year)
Repici et al[10]	2020	Italy	CADe	Symptomatic: 161 Screening or surveillance: 524	6 experienced endoscopists (>2000 screening colonoscopies)
Liu et al[23]	2020	China	CADe	Symptomatic: 960 Screening or surveillance: 66	
Su et al[22]	2020	China	CADe + Quality	Symptomatic: 407 Screening or surveillance: 216	6 endoscopists (5000-8000 colonoscopies)

CAD, Computer-aided diagnosis; *CADe*, Computer-aided polyp detection; *ns*, not statistically significant.

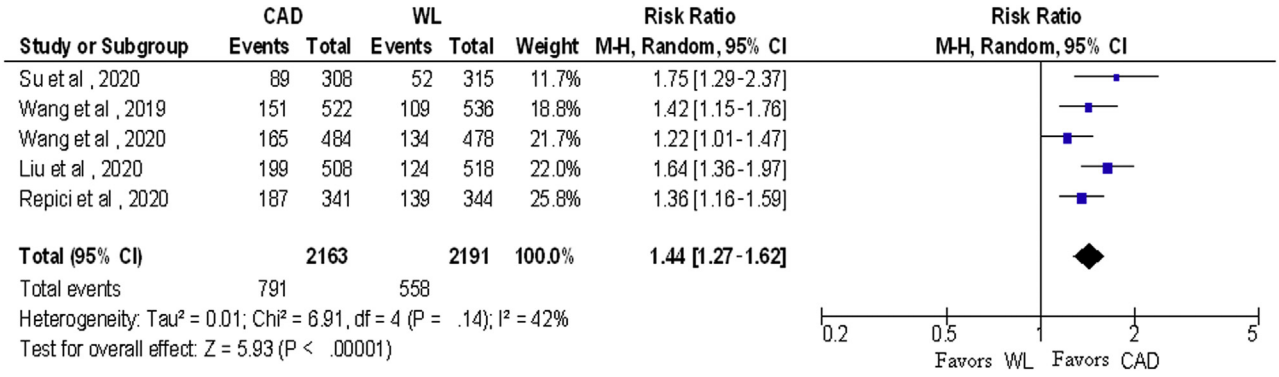


Figure 2. Comparative effectiveness of CAD versus control group on adenoma detection rate. *CAD*, Computer-aided diagnosis; *WL*, white light; *CI*, confidence interval.

Data synthesis and analysis

In individual trials, we estimated risk ratios (RRs) for dichotomous outcomes and mean differences for continuous outcomes, together with their 95% confidence intervals (CIs). When mean and standard deviation were not reported for continuous outcomes, we calculated these statistics from median and interquartile range, according to the methods described by Luo et al[16] and Wan et al.[17] We calculated pooled estimates using the DerSimonian and Laird random-effects model.[18]

We assessed heterogeneity of intervention effects among primary studies using the χ^2 test (Cochran Q) and I^2 statistic. We considered I^2 cut-off points of 25%, 50%, and 75% as indicative of low, moderate, and high heterogeneity, respectively.[19]

We conducted prespecified subgroup analyses for APCs and polyps per colonoscopy by colonic segment (right-sided colon segment, transverse colon, left-sided colon segment, or rectum), colonic site (proximal to splenic flexure or distal colon), size (≤5 mm, 6-9 mm, ≥10 mm, <10 mm, ≥10 mm), and morphology (polypoid or nonpolypoid). We estimated differences among subgroups by the Mantel-Haenszel test and heterogeneity using the χ^2 test and I^2 statistic.

We planned sensitivity analyses by the leave-1-out approach for the primary outcomes (ie, ADR and polyp detection rate) to investigate the influence of each individual trial on the overall effect estimate. We also performed sensitivity analysis for withdrawal time by excluding studies investigating systems that provide control on the endoscope withdrawal time.

We explored publication bias using funnel plots. We rated the quality of evidence according to the Grades of Recommendation, Assessment, Development and Evaluation approach.[20]

We performed leave-1-out sensitivity analyses using Stata (StataCorp LP, StataCorp, College Station, Tex, USA). All other analyses were carried out using RevMan 5.3 (The Cochrane Collaboration, Copenhagen, Denmark).

TABLE 1. Continued

No. of patients		Gender (male) n (%)		Age (y) Mean ± SD or median (interquartile range)		Withdrawal time Mean ± SD		
Control	CAD	Control	CAD	Control	CAD	Control	CAD	*P* value
536	522	249 (46)	263 (50)	49.9 ± 13.8	51.1 ± 13.2	6.1 ± 1.1	6.2 ± 1.4	ns
478	484	254 (53)	241 (50)	49.0 (40-56)	49.0 (39-60)	6.4 ± 1.1	6.5 ± 1.3	ns
344	341	165 (48)	172 (50)	61.1 ± 10.6	61.5 ± 9.7	7.0 ± 1.5	7.1 ± 1.5	ns
518	508	287 (55)	264 (52)	50.1 ± 12.7	51 ± 12.3	6.1 ± 1.0	6.2 ± 1.3	ns
315	308	148 (47)	159 (49)	51.6 ± 9.0	50.5 ± 10.3	5.7 ± 1.1	7.0 ± 1.0	ns

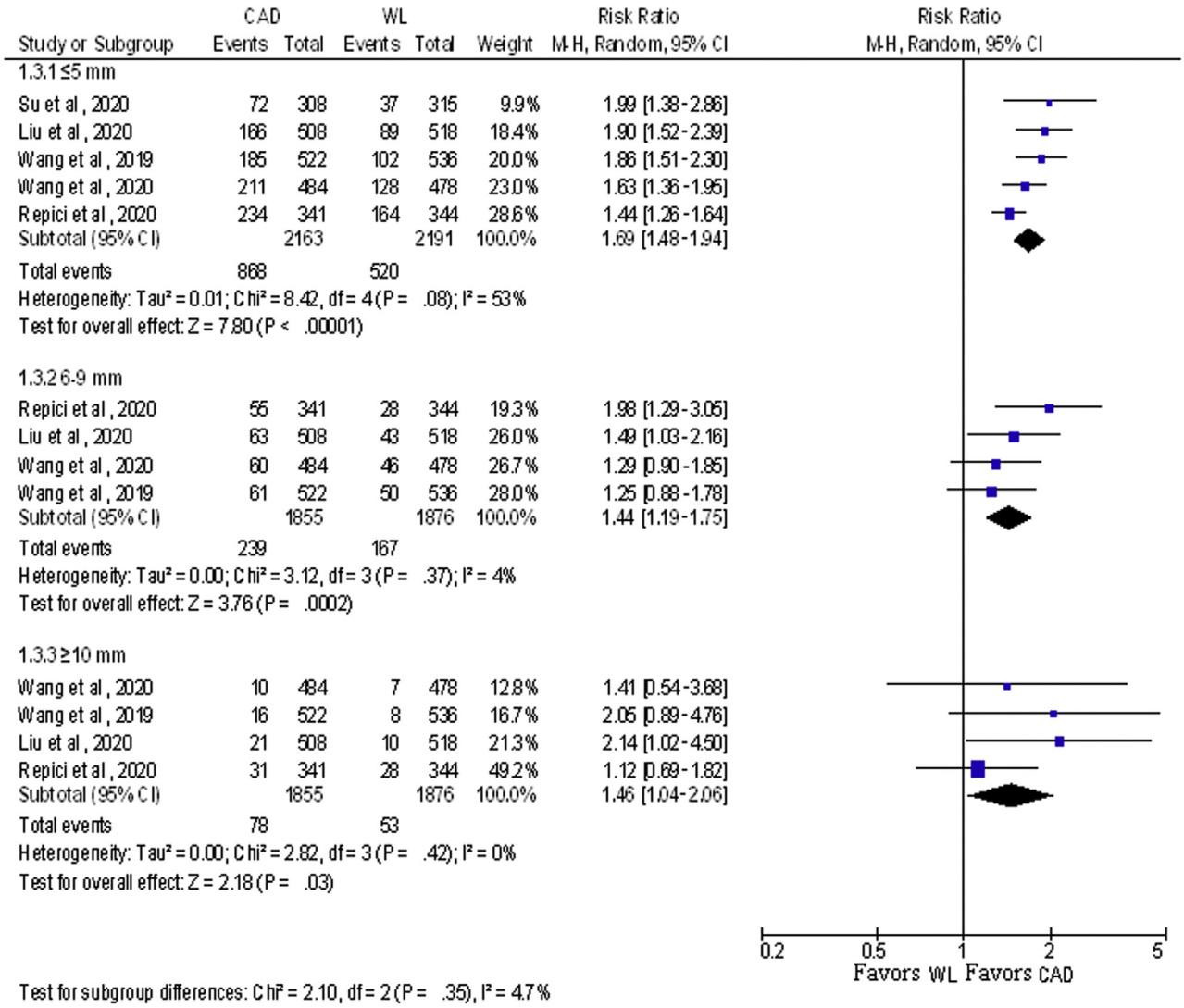


Figure 3. Comparative effectiveness of CAD versus control group on adenoma per colonoscopy subgrouped according to size. *CAD*, Computer-aided diagnosis; *WL*, white light; *CI*, confidence interval.

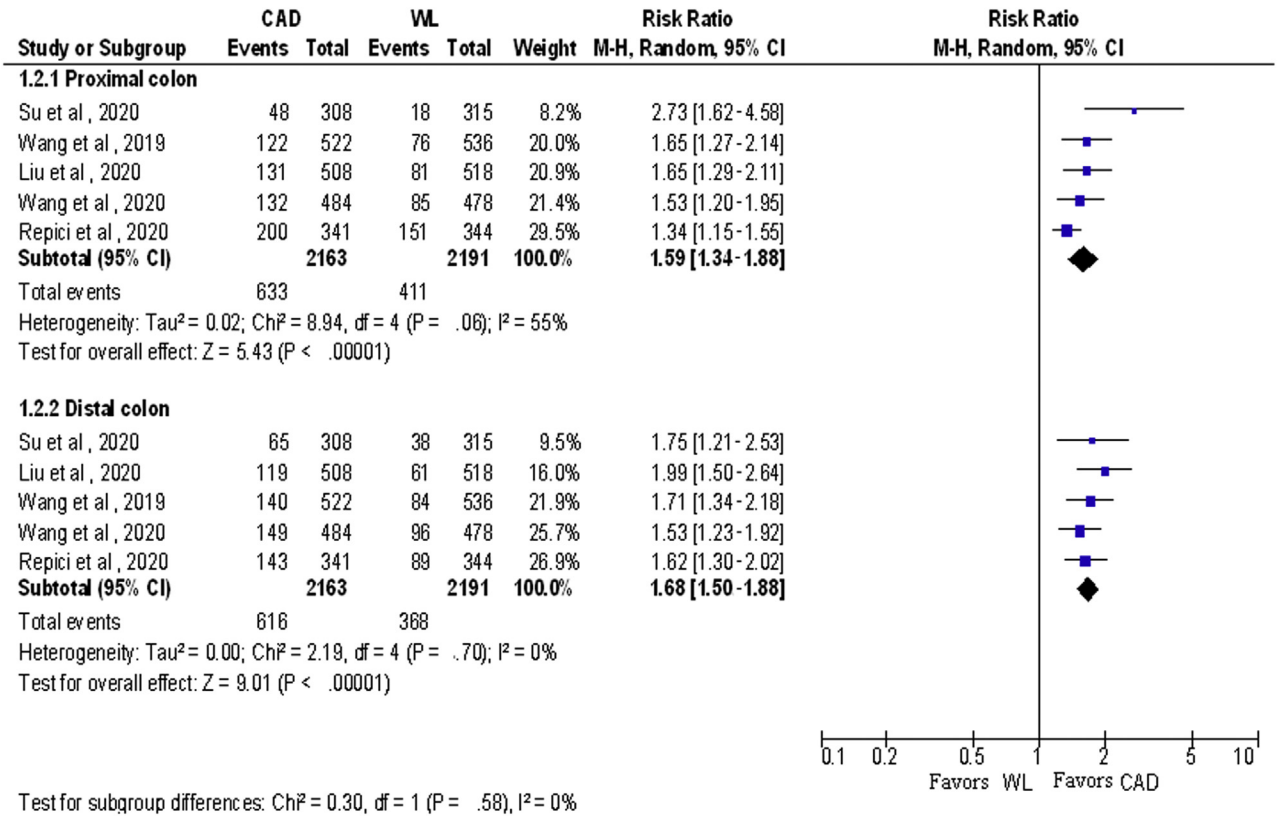


Figure 4. Comparative effectiveness of CAD versus control group on adenoma per colonoscopy subgrouped according to location. *CAD*, Computer-aided diagnosis; *WL*, white light; *CI*, confidence interval.

RESULTS

Study characteristics and quality

The initial literature search resulted in 554 articles (Fig. 1). Six RCTs,[10,11,21-24] all published between 2019 and 2020, tested the impact of CNN-based systems on detection of colorectal neoplasia. One study[24] was excluded because it used a non-CADe CNN-based system focusing on withdrawal time and endoscope speed. Of the 5 included CADe-based RCTs,[10,11,21-23] 1 coupled the CADe algorithm to CNN models developed to assess quality indicators (withdrawal time, withdrawal stability, and bowel preparation).[22] Different CADe system characteristics are summarized in Supplementary Table 1 (available online at www.giejournal.org).

Most studies were conducted in China (n = 4),[11,21-23] and only 1 study came from Western countries (Italy).[10] All but 1 study were single-center experiences. The objective assessment of risk of bias is reported in Supplementary Figure 1 (available online at www.giejournal.org).

The total number of participants included in the analysis was 4354 (2163 in the CNN group and 2191 in the control group), and the individual study sample size ranged from 623 to 1058 patients. Study characteristics are comprehensively shown in Table 1.

TABLE 2. Adenoma detection subgrouped according to size, location, and morphology

Reference	Adenoma <5 mm			Adenoma 6-9 mm			Adenoma ≥10 mm		
	Control	CAD	*P* value	Control	CAD	*P* value	Control	CAD	*P* value
Wang et al[11]	102 (63.8)	185 (70.6)	<.05	50 (31.6)	61 (23.3)	ns	8 (5.0)	16 (6.1)	ns
Wang et al[21]	128 (71)	211 (75)	<.05	46 (25)	60 (21)	ns	7 (4)	10 (4)	ns
Repici et al[10]	164 (74.5)	234 (73.1)	<.05	28 (12.7)	55 (17.2)	<.05	28 (12.7)	31 (9.7)	ns
Liu et al[23]	89 (62.7)	166 (66.4)	<.05	43 (30.3)	63 (25.2)	ns	10 (7.0)	21 (8.4)	ns
Su et al[22]	37 (66.1)	72 (63.7)	<.05	\	\	\	\	\	\

Values are n (%).
CAD, Computer-aided diagnosis; *ns*, not statistically significant; \, not available.

ADR and polyp detection rate

Based on data reported by all 5 studies, the overall ADR was significantly higher for the CADe group compared with the control group, respectively (791/2163 [36.6%] vs 558/2191 [25.2%]; RR, 1.44; 95% CI, 1.27-1.62; $P < .01$). All RCTs reported a significant ADR increase. However, there was moderate heterogeneity ($I^2 = 42\%$) in the level of the effect (Fig. 2). Leave-1-out sensitivity analysis on ADR is reported in Supplementary Figure 2 (available online at www.giejournal.org).

The overall polyp detection rate was also significantly improved in the CADe group compared with the control group, with 1089 of 2163 patients and 758 of 2191 patients with at least 1 polyp out (50.3% vs 34.6%; RR, 1.43; 95% CI, 1.34-1.53; $P < .01$) with a low level of heterogeneity ($I^2 = 0\%$) across the 5 studies (Supplementary Fig. 3, available online at www.giejournal.org). We explored publication bias of both outcomes using funnel plots, resulting in no small study effect (Supplementary Fig. 4, available online at www.giejournal.org).

APCs, sessile serrated lesions, and polyps per colonoscopy

The number of APCs was significantly higher in the CADe group compared with the control group (1249/2163 [.58] vs 779/2191 [.36]; RR, 1.70; 95% CI, 1.53-1.89; $P < .01$) with a moderate level of heterogeneity ($I^2 = 33\%$). The performance of CADe systems in significantly improving APC was confirmed irrespective of adenoma size (RR ≤5 mm, 1.69 [95% CI, 1.48-1.84] vs RR 6-9 mm, 1.44 [95% CI, 1.19-1.75] vs RR ≥10 mm, 1.46 [95% CI, 1.04-2.06]) (Fig. 3), location (RR proximal, 1.59 [95% CI, 1.34-1.88] vs RR distal, 1.68 [95% CI, 1.50-1.88]) (Fig. 4), and morphology (RR flat, 1.78 [95% CI, 1.47-2.15] vs RR polypoid, 1.54 [95% CI, 1.40-1.68]) (Supplementary Fig. 5 [available online at www.giejournal.org], Table 2).

Based on data from 3 studies,[10,11,21] a statistically significant trend in the number of advanced adenomas detected was not found in favor of CADe as compared with the control group (116/1347 [.09] vs 71/1358 [.05]; RR, 1.35; 95% CI, .74-2.47; $P = .33$; $I^2 = 69\%$). The number of sessile serrated lesions detected per colonoscopy was significantly improved in the CADe group compared with the control group (109/1855 [.06] vs 73/1876 [.04]; RR, 1.52; 95% CI, 1.14-2.02; $P < .01$) with a low level of heterogeneity ($I^2 = 0\%$) (Supplementary Fig. 6, available online at www.giejournal.org) across 4 studies.[10,21,21,23] The results of polyps per colonoscopy and other per-polyp analyses are extensively reported in Supplementary Table 2 (available online at www.giejournal.org).

Withdrawal time

No statistically significant difference between the CADe and control groups was observed in terms of mean withdrawal time (mean difference of CADe vs control, .34 minutes; 95% CI, –.10 to .78; $P = .13$) with a high level of heterogeneity ($I^2 = 97\%$). However, the heterogeneity level was low ($I^2 = 0\%$) when excluding a sensitivity analysis in the study in which the CADe systems played a direct role in influencing withdrawal time[22] (mean difference, .10; 95% CI, .02-.18; $P = .02$) (Supplementary Fig. 7, available online at www.giejournal.org).

Quality of evidence

The quality of evidence was assessed by applying the Grades of Recommendation, Assessment, Development and Evaluation methodology. The level of evidence for RCTs was downgraded because of moderate quality of the included RCTs (assessed by Cochrane Risk Bias Tool for randomized studies), inconsistency attributed to endoscopists (eg, subjective assessments of lesion location and size) and patients (ie, different indications for colonoscopy), and possible differences in advanced adenoma definitions across studies. Details can be found in Supplementary Table 3 (available online at www.giejournal.org).

DISCUSSION

The 44% and 70% relative increase in ADR and APC, respectively, consistently shown across the 5 included

TABLE 2. Continued

Adenoma proximal			Adenoma distal			Adenoma polypoid			Adenoma flat		
Control	CAD	*P* value	Control	CAD	*P* value	Control	CAD	*P* value	Control	CAD	*P* value
76 (47.5)	122 (46.6)	<.05	84 (52.5)	140 (53.4)	<.05	97 (60.6)	144 (55.0)	<.05	63 (39.4)	118 (45.0)	<.05
85 (47.0)	132 (47.0)	<.05	96 (53.0)	149 (53.0)	<.05	180 (99)	276 (98)	<.05	1 (1)	5 (2)	<.05
151 (63.0)	200 (58.3)	<.05	89 (37.0)	143 (41.7)	<.05	143 (59.6)	204 (68.1)	<.05	97 (40.4)	139 (31.9)	<.05
81 (57.0)	131 (52.4)	<.05	61 (42.9)	119 (47.6)	<.05	82 (57.7)	128 (51.2)	<.05	60 (42.3)	122 (48.8)	<.05
18 (32.1)	48 (42.5)	<.05	38 (67.9)	65 (57.5)	<.05	35 (62.5)	75 (66.4)	<.05	21 (37.5)	38 (33.6)	<.05

studies in nearly 4300 randomized patients supports the benefit when adding CADe to colonoscopy with no meaningful effect on the efficiency of colonoscopy as shown by the similar withdrawal time between the 2 arms. The main result of our study is the independence between the additional benefit of CADe and the traditional features of colorectal neoplasia in the adequately powered subanalyses performed. In detail, CADe led to a statistically significant increased detection of diminutive, small, and large adenomas; of those located in the proximal and distal colon; and of flat and polypoid adenomas. In addition, there was an increased detection of sessile serrated lesions and a trend for a nearly 2-fold increase in advanced neoplasia. It could be argued that in previous pre-CADe studies neoplasia miss rate at colonoscopy was selectively associated with some of these features, such as flat morphology, proximal location, and diminutive to small size.[3] However, such evidence came from back-to-back studies, representing a markedly different methodology as the one adopted in the included RCTs.[3,6] In the tandem setting, miss rates from failure in lesion recognition and incomplete exposure of the mucosa are mixed, preventing a clear attribution of such miss rates to either mechanism.[3,6] On the other hand, the CADe RCTs included in our meta-analysis represent a parallel methodology where only failure to recognize the lesion contributed to the additional detection.

The resilience of CADe efficacy from the traditional classifications of colorectal neoplasia is far from being unexpected because CADe and human perceptions are based on completely different mechanisms, namely a probabilistic analysis of the image based on training-acquired parameters versus human cognitive perception that may be highly variable depending on training, experience, visual acuity, gaze patterns, and endoscopist personality. Thus, the finding that the additional CADe-driven detection is independent of features traditionally associated with miss rate suggests that characteristics other than size and morphology are exploited by the machine to recognize the lesion. Speculatively, these factors may be represented by texture, color, and shape factors other than those perceived by the human mind.

Despite the additional detection of adenomas ≥10 mm, our study failed to show an increase in the detection of advanced adenomas at a per-patient or per-polyp level. This discrepancy is unexpected considering that most of the advanced adenoma pool is represented by ≥10-mm lesions. The authors of the Chinese studies did not respond to our requests for details regarding their definitions of advanced adenomas. Uncertainties regarding these definitions may have influenced the result we report here regarding advanced adenomas.

There are limitations to our meta-analysis. Most analyses were performed at the per-polyp level, because per-patient data according to different features of lesions were not available. However, there is no reason to assume that the conversion between per-patient and per-polyp analysis would be different in the 2 arms. Second, an adopted technology included both CADe and an algorithm to improve other factors of quality of colonoscopy, such as the withdrawal technique or the level of cleansing. Unfortunately, such study did not allow the discrimination of the possible impact of each of the 2 components, leaving uncertainty on the possible synergistic effect between CADe technology on the one hand and algorithms focusing on optimization of mucosa exposure during withdrawal on the other. This limitation also applies to possible synergism between CADe and devices aiming to expose more mucosa, such as the cup or endocuff, when considering that such devices were not used in the included studies. Thus, new studies specifically addressing the possible synergism between CADe and CAD technologies alarming the endoscopist when poorly exploring the mucosa are required.

In addition, we showed the additional efficacy of CADe for a variety of polyp categories according to size, morphology, and histology. However, it is unclear how many of these categories were adequately represented in the training database. Because the training database is the only clinical information that is fully transparent to endoscopist, any model should clearly report how many lesions for each category were included. Finally, the Chinese setting cannot be immediately translated to the Western setting when considering the very low ADR in the control group of some studies that could depend on both a different prevalence of disease and operator skill. The fact that CADe was successful in these studies with very low ADRs in the control group could suggest an efficacy of CADe for low detectors. However, dedicated studies are needed. More generally, we could not properly assess the relationship between baseline ADR and CADe benefit. Thus, specific studies on both low and high detectors are needed, as well as additional studies testing CADe in Western populations.

In conclusion, lesion detection by artificial intelligence is not impacted by factors such as size and morphology that are known to affect detection by human observers. According to the current evidence, there is substantial and convergent evidence for the incorporation of artificial intelligence to increase detection of colorectal neoplasia during colonoscopy.

REFERENCES

1. 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.
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. 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.
4. Robertson DJ, Lieberman DA, Winawer SJ, et al. Colorectal cancers soon after colonoscopy: a pooled multicohort analysis. Gut 2014;63:949-56.
5. Anderson R, Burr NE, Valori R. Causes of post-colonoscopy colorectal cancers based on world endoscopy organization system of analysis. Gastroenterology 2020;158:1287-99.
6. Rex D, Cutler C, Lemmel G, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology 1997;112:24-8.
7. Parasa S, Wallace M, Bagci U, et al. Proceedings from the First Global Artificial Intelligence in Gastroenterology and Endoscopy Summit. Gastrointest Endosc 2020;92:938-45.
8. Bisschops R, East JE, Hassan C, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) guideline—update 2019. Endoscopy 2019;51:1155-79.
9. Hassan C, Wallace MB, Sharma P, et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. Gut 2020;69:799-800.
10. 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.
11. 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.
12. Mori Y, Kudo S-E, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med 2018;169:357-66.
13. Greenspan M, Rajan KB, Baig A, et al. Advanced adenoma detection rate is independent of nonadvanced adenoma detection rate. Am J Gastroenterol 2013;108:1286-92.
14. Penz D, Ferlitsch A, Waldmann E, et al. Impact of adenoma detection rate on detection of advanced adenomas and endoscopic adverse events in a study of over 200,000 screening colonoscopies. Gastrointest Endosc 2020;91:135-41.
15. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006-12.
16. Luo D, Wan X, Liu J, et al. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res 2018;27:1785-805.
17. Wan X, Wang W, Liu J, et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014;14:135. Available at: https://pubmed.ncbi.nlm.nih.gov/25524443/. Accessed May 25, 2020.
18. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.
19. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
20. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res 2004;4:38.
21. 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.
22. Su J-R, Li Z, Shao X-J, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020;91:415-24.
23. Liu W-N, Zhang Y-Y, Bian X-Q, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26:13.
24. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. Lancet Gastroenterol Hepatol 2020;5:352-61.

Abbreviations: ADR, adenoma detection rate; APC, adenoma per colonoscopy; CADe, computer-aided polyp detection; CNN, convoluted neural network; RR, risk ratio; RCT, randomized controlled trials.

DISCLOSURE: *The following authors disclosed financial relationships: C. Hassan, A. Repici: Consultant for and equipment loan from Medtronic and Fujifilm. P. Sharma: equipment loan from Medtronic Italy; consultant for and grant support from Olympus, Medtronic USA, and Fujifilm; consultant for Lumendi, Boston Scientific, and Bausch; grant support from US Endoscopy, Ironwood, Erbe, Docbot, Cosmo Pharmaceuticals, and CDx Labs; D.K. Rex: owner of Satisfai Health, consultant for Medtronic, Boston Scientific, Aries Pharmaceutical, Lumenid Ltd, Braintree Laboratories, Norgine, Endokey, and GI Supply; consultant for and research support from Olympus Corporation; research support from EndoAid, Medivators, and Erbe USA, Inc; M. Wallace: consultant for Virgo Inc, Cosmo/Aries Pharmaceuticals, Anx Robotica (2019), Covidien, and GI Supply; research grants from Fujifilm, Boston Scientific, Olympus, Medtronic, Ninepoint Medical, and Cosmo/Aries Pharmaceuticals; stock options from Virgo; consulting on behalf of Mayo Clinic, GI Supply (2018), Endokey, Endostart, Boston Scientific, and Microtek; minor food/beverage from Synergy Pharmaceuticals, Boston Scientific, and Cook Medical. All other authors disclosed no financial relationships.*

*Drs Hassan and Spadaccini contributed equally to this article.

0016-5107/$36.00
https://doi.org/10.1016/j.gie.2020.06.059

Received May 25, 2020. Accepted June 18, 2020.

Current affiliations: Digestive Endoscopy Unit, Nuovo Regina Margherita Hospital, Rome, Italy (1), Endoscopy Unit, Humanitas Clinical and Research Center–IRCCS, Rozzano, Italy (2), Department of Biomedical Sciences, Humanitas University, Rozzano, Italy (3), Section of Gastroenterology, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy (4), Division of Gastroenterology and Hepatology, Johns Hopkins Hospital, Baltimore, Maryland, USA (5), Division of Gastroenterology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA (6), Gastroenterology and Hepatology, Kansas City VA Medical Center, Kansas City, Missouri, USA (7), Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China (8), Department of Gastroenterology, Portuguese Oncology Institute of Coimbra, Coimbra, Portugal (9), MEDCIDS-Department of Community Medicine, Information and Decision in Health, Faculty of Porto, University of Medicine, Porto, Portugal (10), Department of Gastroenterology, Queen Alexandra Hospital, Portsmouth, UK (11), Division of Gastroenterology/Hepatology, Indiana University School of Medicine, Indianapolis, Indiana, USA (12), Department of Interdisciplinary Endoscopy, University Hospital Hamburg-Eppendorf, Hamburg, Germany (13), Department of Gastroenterology, Mayo Clinic, Jacksonville, Florida, USA (14).

Reprint requests: Marco Spadaccini, MD, Humanitas Research Hospital and University, Via Manzoni 56, 20089 Rozzano (Milano) Italy.

If you would like to chat with an author of this article, you may contact Dr Spadaccini at marco.spadaccini@humanitas.it.

APPENDIX 1

Search strategies

Medline search strategy

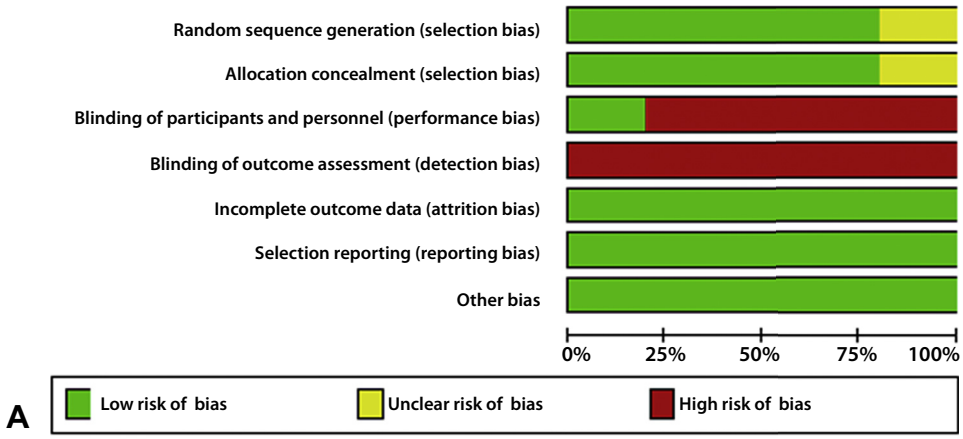
Search number	Search query	Results
1	exp Colorectal Neoplasms/	197,032
2	exp Cecal Neoplasms/	5470
3	exp Adenomatous Polyps/	7933
4	exp Intestinal Polyps/	14,519
5	colo* cancer*.tw,kw.	138,200
6	colo* adenoma*.tw,kw.	5652
7	colo* polyp*.tw,kw.	6375
8	or/1-7	250,532
9	exp Artificial Intelligence/	92,581
10	exp Diagnosis, Computer-Assisted/	82,109
11	exp Image Processing, Computer-Assisted/	228,955
12	artificial intelligence.tw,kw.	6715
13	cad*.tw,kw.	187,891
14	computer aided detection.tw,kw.	1367
15	computer aided system*.tw,kw.	287
16	deep learning.tw,kw.	7875
17	or/9-16	543,850
18	8 and 17	5555
19	randomized controlled trial.pt.	500,805
20	controlled clinical trial.pt.	93,552
21	pragmatic clinical trial.pt.	1301
22	randomi#ed.tw.	606,532
23	placebo.ab.	205,210
24	clinical trials as topic/	190,226
25	randomly.ab.	327,736
26	(crossover or cross-over).tw.	83,501
27	Cross-over Studies/	47,190
28	trial.tw.	578,824
29	or/19-28	1,508,898
30	animals/ not (humans/ and animals/)	4,640,785
31	29 not 30	1,383,304
32	18 and 31	347

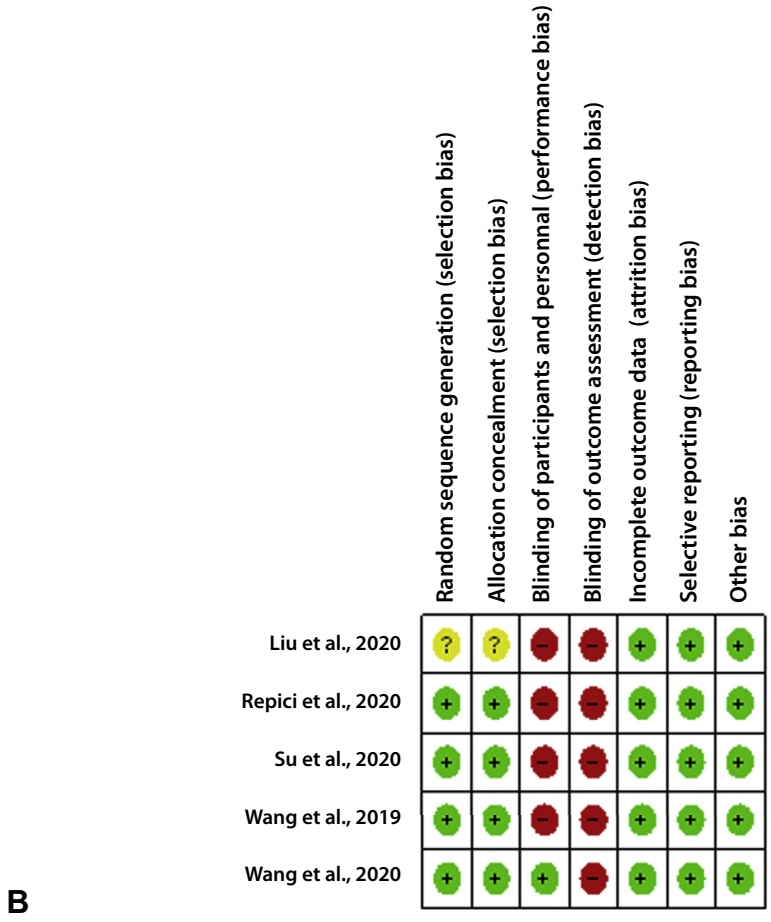
EMBASE search strategy

Search number	Search query	Results
1	colorectal AND ('adenoma'/exp OR adenoma)	24,082
2	colorectal AND polyp$	21,242
3	adenomatous AND polyp$	13,725
4	#1 OR #2 OR #3	39,242
5	'artificial intelligence'	29,751
6	'computer assisted diagnosis'	38,416
7	'computer-assisted'	836,274
8	'cad':ab,ti	65,216
9	'computer aided system'	283
10	'deep learning'	10,705
11	#5 OR #6 OR #7 OR #8 OR #9 OR #10	927,483
12	#4 AND #11	1841
13	'randomized controlled trial'	778,348
14	'crossover procedure'	62,131
15	'double blind procedure'	170,104
16	'single blind procedure'	37,978
17	random*	1,721,357
18	factorial*	38,617
19	placebo*	455,566
20	assign*	387,430
21	allocat*	163,890
22	#13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21	2,283,452
23	#12 AND #22	137

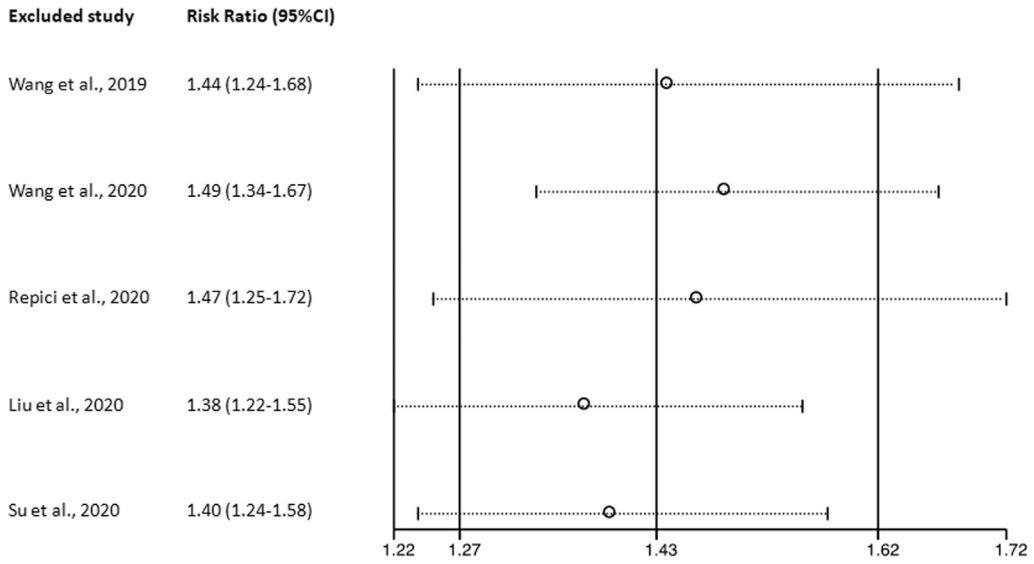
Cochrane Central Register of Controlled Trials search strategy

Search number	Search query	Results
1	colorectal polyp in Trials (Word variations have been searched)	1563
2	colorectal adenoma in Trials (Word variations have been searched)	1545
3	colon adenoma in Trials (Word variations have been searched)	1291
4	adenomatous polyp in Trials (Word variations have been searched)/	546
5	#1 OR #2 OR #3 OR #4	2296
6	artificial intelligence in Trials	412
7	computer assisted in Trials	15,848
8	CAD in Trials	4775
9	computer aided system in Trials	265
10	deep learning in Trials	438
11	#6 OR #7 OR #8 OR #9 OR #10	21,207
12	#5 AND #11	70

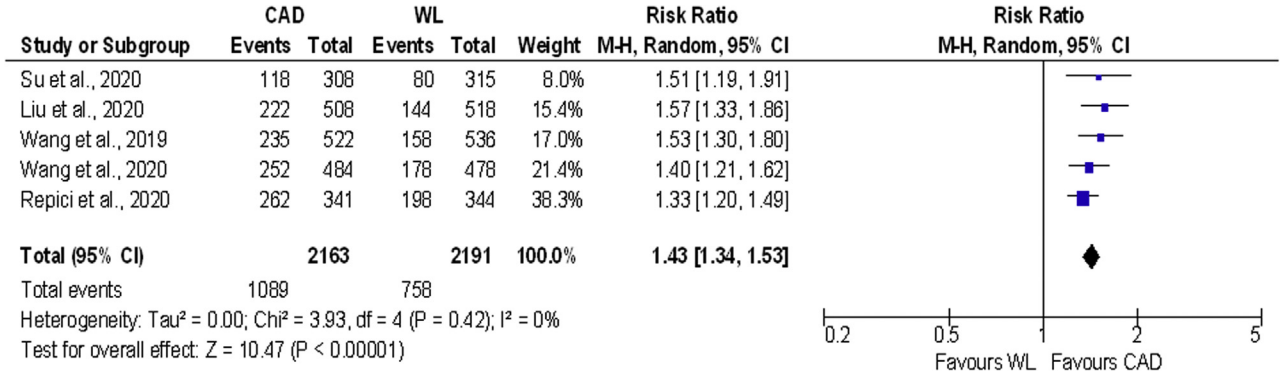




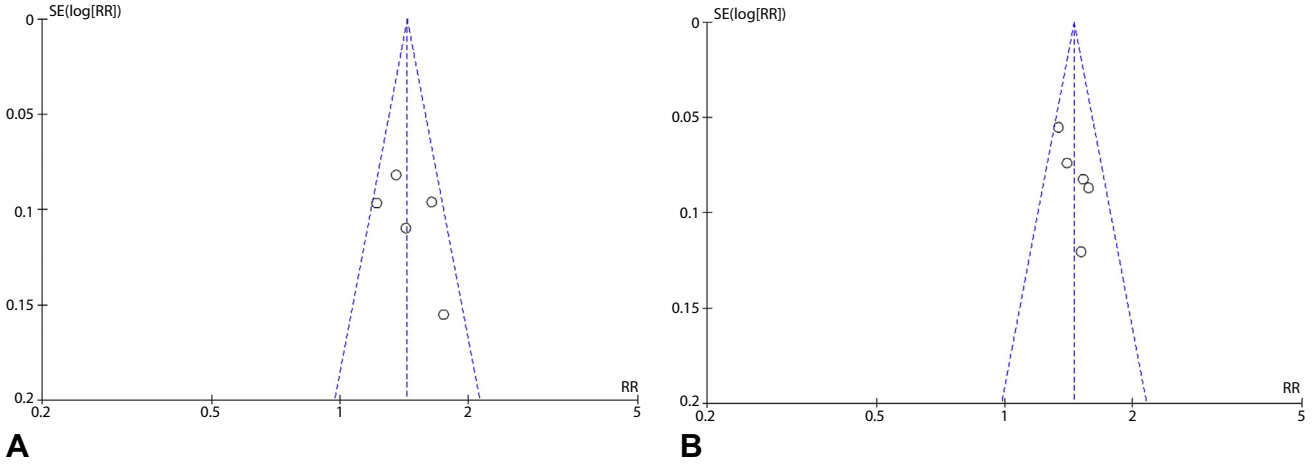
Supplementary Figure 1. A, Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies. **B,** Risk of bias summary: review authors' judgements about each risk of bias item for each included study.



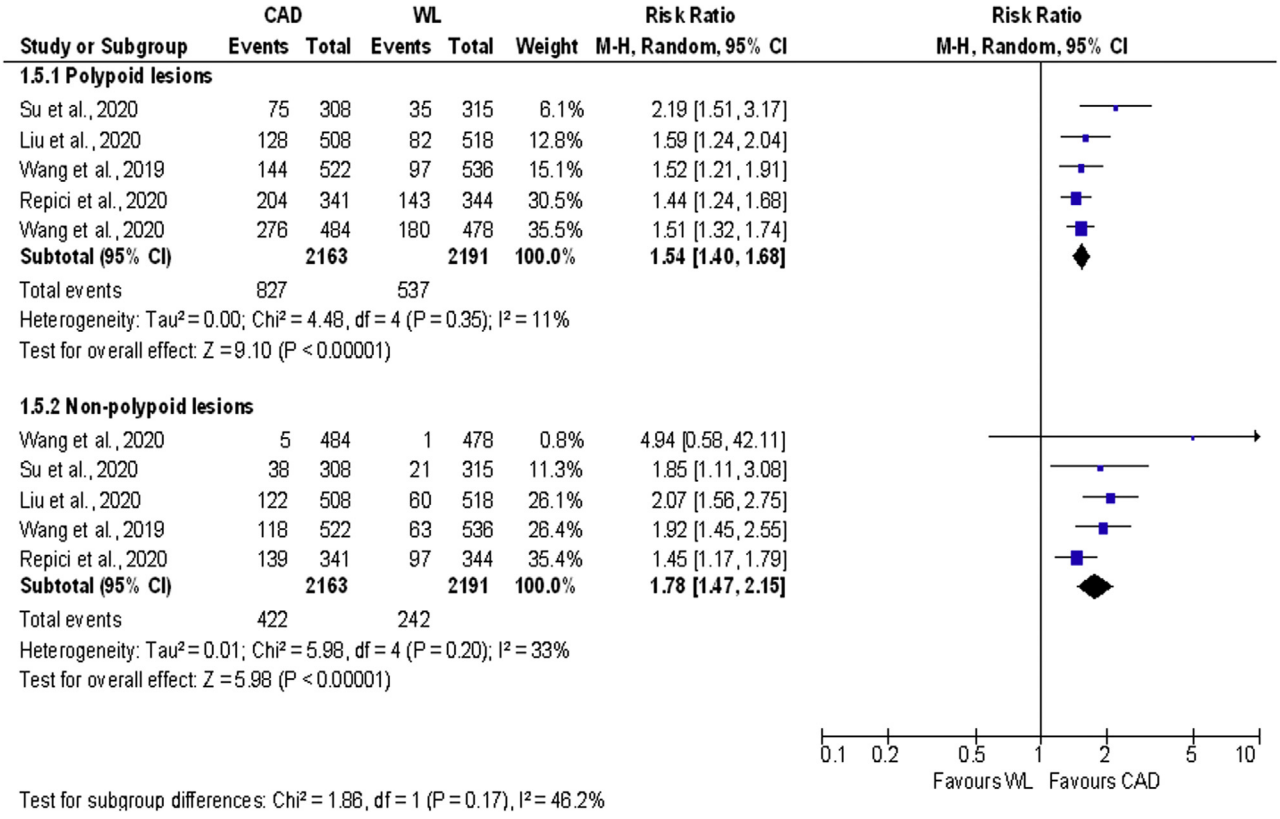
Supplementary Figure 2. Leave-1-out sensitivity analysis for adenoma detection rate. *CI*, Confidence interval.



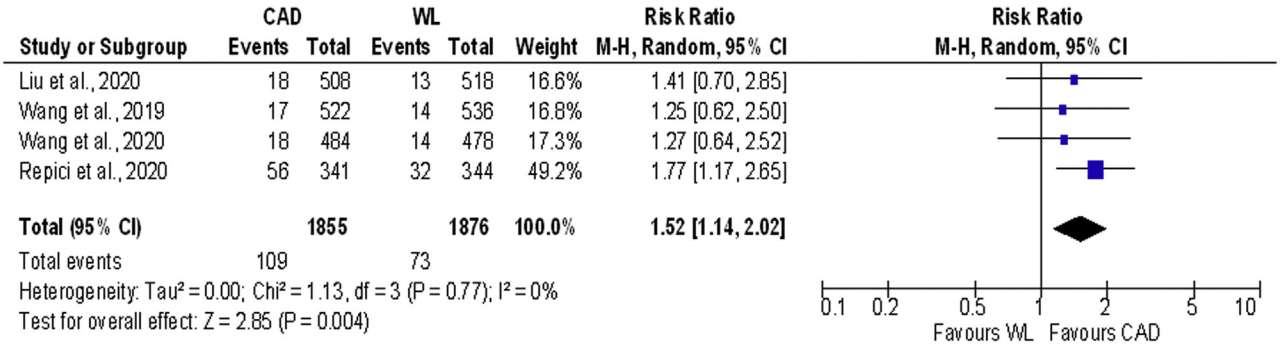
Supplementary Figure 3. Comparative effectiveness of CAD versus control group on polyp detection rate. *CAD*, Computer-aided diagnosis; *WL*, white light; *CI*, confidence interval.



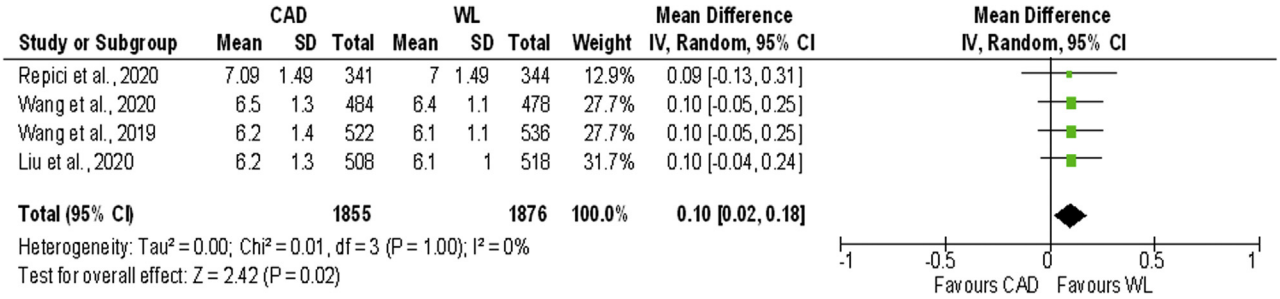
Supplementary Figure 4. Funnel plots for (**A**) adenoma detection rate and (**B**) polyp detection rate. *SE*, Standard error; *RR*, relative risk.



Supplementary Figure 5. Comparative effectiveness of CAD versus control group on adenoma per colonoscopy subgrouped according to morphology. *CAD*, Computer-aided diagnosis; *WL*, white light; *CI*, confidence interval.



Supplementary Figure 6. Comparative effectiveness of CAD versus control group on sessile serrated lesions per colonoscopy. *CAD*, Computer-aided diagnosis; *WL*, white light; *CI*, confidence interval.



Supplementary Figure 7. Sensitivity analysis for withdrawal time by excluding studies investigating systems that provide control on the endoscope withdrawal technique. *CAD*, Computer-aided diagnosis; *WL*, white light; *SD*, standard deviation; *CI*, confidence interval.

SUPPLEMENTARY TABLE 1. Computer-aided diagnosis system characteristics

Author and year	Scopes	Computer-aided diagnosis system	Component of convoluted neural networks model
Wang et al 2019[11]	Olympus	EndoScreener	RT polyp detection
Wang et al 2020[21]	Fuji	EndoScreener	RT polyp detection
Repici et al 2020[10]	Olympus, Fuji	GI-genius	RT polyp detection
Liu et al 2020[23]	Fuji	Henan Xuanweitang Medical Information Technology Co Ltd	RT polyp detection
Su et al 2020[22]	Pentax	\	RT polyp detection, withdrawal time, withdrawal stability, bowel preparation

RT, Real time; \, not available.

SUPPLEMENTARY TABLE 2. Per-polyp analysis

Outcome	Computer-aided diagnosis vs control: relative risk	95% Confidence interval	Heterogeneity	*P* value
Overall polyps per colonoscopy	1.94	1.83-2.06	I^2 = 0%	<.05
Size (subgroup)				
≤5 mm	1.98	1.74-2.25	I^2 = 72%	<.05
6-9 mm	1.47	1.18-1.84	I^2 = 54%	<.05
≥10 mm	1.24	.95-1.62	I^2 = 0%	.12
Location (subgroup)				
Proximal	2.22	1.72-2.86	I^2 = 80%	<.05
Distal	1.68	1.57-1.80	I^2 = 0%	<.05
Morphology (subgroup)				
Polypoid	1.71	1.53-1.92	I^2 = 0%	<.05
Flat	1.92	1.60-2.30	I^2 = 71%	<.05

SUPPLEMENTARY TABLE 3. Grades of Recommendation, Assessment, Development and Evaluation evidence profile for efficacy of computer-aided diagnosis systems in polyp detection

	Quality assessment					
Outcomes	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Quality**
ADR	Serious*	Not serious	Not serious	Not serious	Not serious	⊕⊕⊕O Moderate
Polyp detection rate	Serious*	Not serious	Not serious	Not serious	Not serious	⊕⊕⊕O Moderate
APC, all lesions	Serious*	Not serious	Not serious	Not serious	Not serious	⊕⊕⊕O Moderate
APC, >10 mm	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
APC, <10 mm	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
APC, proximal	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
APC, distal	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
PPC, all lesions	Serious*	Not serious	Not serious	Not serious	Not serious	⊕⊕⊕O Moderate
APC, >10 mm	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
APC, <10 mm	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
APC, proximal	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
APC, distal	Serious*	Serious†	Not serious	Not serious	Not serious	⊕⊕OO Low
Advanced Adenoma per colonoscopy	Serious*	Serious†	Not serious	Serious‡	Not serious	⊕OOO Very Low
Sessile serrated lesion per colonoscopy	Serious*	Not serious	Not serious	Not serious	Not serious	⊕⊕⊕O Moderate

ADR, Adenoma detection rate; *APC*, adenoma per colonoscopy; *PPC*, polyp per colonoscopy.

*Risk of bias was judged as serious because of the moderate quality of the included randomized controlled trials.

†Inconsistency risk was judged as serious because of heterogeneity among endoscopists (ie, risk of subjective influence in evaluating lesion location and size) and patients (ie, different indications for undergoing a colonoscopy).

‡Imprecision risk was judge as serious because of possible differences in term of advanced adenoma definitions across the studies.