Performance of artificial intelligence for colonoscopy regarding adenoma and polyp detection: a meta-analysis

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

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

*Cesare Hassan and Marco Spadaccini equally contributed to this work and are sharing first authorship position.

Corresponding author:
Marco Spadaccini, MD
Humanitas Research Hospital and University, Rozzano, Italy
Via Manzoni 56
20089 Rozzano (Milano) Italy
Tel: +39 (0)282242595
Fax: +390282242595
e-mail: marco.spadaccini@humanitas.it

Grant support: none.
Disclosure: CH, AR loan equipment and consultancy from Medtronic and Fujifilm

---

This is the author's manuscript of the article published in final edited form as:

Hassan, C., Spadaccini, M., Iannone, A., Maselli, R., Jovani, M., Chandrasekar, V. T., Antonelli, G., Yu, H., Areia, M., Dinis-Ribeiro, M., Bhandari, P., Sharma, P., Rex, D. K., Rösch, T., Wallace, M., & Repici, A. (2020). Performance of artificial intelligence for colonoscopy regarding adenoma and polyp detection: A meta-analysis. Gastrointestinal Endoscopy. https://doi.org/10.1016/j.gie.2020.06.059

ABSTRACT

BACKGROUND AND AIMS

One fourth of colorectal neoplasia is missed at screening colonoscopy, representing the main cause of interval colorectal cancer (CRC). Deep learning systems with real-time computer-aided polyp detection (CADe) showed high accuracy in artificial settings, and preliminary randomized clinical trials (RCT) reported favourable outcomes in clinical setting. Aim of this meta-analysis was to summarise 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 detection of colorectal neoplasia. Primary outcome was pooled adenoma detection rate (ADR), Secondary outcomes were adenoma per colonoscopy (APC) according to size, morphology and location, advanced APC (AAPC), as well as polyp detection rate (PDR), Polyp-per-colonoscopy (PPC), and sessile serrated lesion per colonoscopy (SPC). We calculated risk ratios (RR), performed subgroup, and sensitivity analysis, 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 groups than in the control group (791/2163, 36.6% vs 558/2191, 25.2%; RR, 1.44; 95% CI, 1.27-1.62; $p<0.01$; I^2:42%). APC was also higher in the CADe group compared with control (1249/2163, 0.58 vs 779/2191, 0.36; RR, 1.70; 95% CI, 1.53-1.89; $p<0.01$;I^2:33%). APC was higher for $\leq$5 mm (RR, 1.69; 95% CI, 1.48-1.84), 6-9 mm (RR, 1.44; 95% CI, 1.19-1.75), and $\geq$10 mm adenomas (RR, 1.46; 95% CI, 1.04-2.06), as well as for proximal (RR, 1.59; 95% CI, 1.34-1.88) and distal (RR, 1.68; 95% CI, 1.50-1.88), and for 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 SPC (RR, 1.52; 95% CI,1.14-2.02), whereas a nonsignificant trend for AADR was found (RR, 1.35; 95% CI, 0.74 – 2.47; $p = 0.33$; I^2:69%). Level of evidence for RCTs was graded moderate.

CONCLUSIONS

According to available evidence, the incorporation of Artificial Intelligence as aid for detection of colorectal neoplasia results in a significant increase of the detection of colorectal neoplasia, and such effect is independent from main adenoma characteristics.

Key Words: Artificial intelligence; Screening; Prevention; Quality.

INTRODUCTION

Interval colorectal cancer (CRC) represents one of the most dismal consequences of screening colonoscopy with an incidence of 0.5 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 CRC[7,8]. In detail, the adoption of convoluted neural networks (CNN) 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 adenomas per colonoscopy (APC), as well as possible harms such as deskilling of the endoscopist or time wasting due to 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 adenoma detection rate (ADR), is also associated with an increase in the detection of advanced adenomas, defined as either $\geq$10 mm or unfavorable histology[13,14].

The aim of our systematic review and meta-analysis is to assess the relationship between the increased detection led by CADe and the main features of the detected lesions.

MATERIALS AND METHODS

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

Data sources and search strategy

We performed a comprehensive literature search in MEDLINE, Cochrane Central Register of Controlled Trials (CENTRAL) and Embase (up to March 31, 2020) electronic databases to identify Randomized Clinical Trials (RCT) 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 1A-C). 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. When there were 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, adenoma detection rate, polyp detection rate, number and characteristics (size, 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:

(1) Population: all adults (>18 years old) undergoing colonoscopy in nonemergency setting.

(2) Intervention: colonoscopy with high-definition endoscopes implemented with real-time CADe systems.

(3) Comparison: colonoscopy with high-definition endoscopes.

(4) Outcome: adenoma and polyp detection rate.

(5) Study design: only randomized controlled trials were considered.

Exclusion criteria were as follows:

(1) Essential information not available;

(2) Studies not published as full text article;

(3) Studies including less than 50 patients in each group.

Outcomes

Primary outcome

- Adenoma detection rate (ADR): proportion of individuals undergoing a complete colonoscopy who had at least one adenoma detected (and removed).

Secondary outcomes

- Adenomas per colonoscopy (APC): number of adenomas per colonoscopy, 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 (PDR): proportion of individuals undergoing a complete colonoscopy who had at least one polyp detected.
- Polyps per colonoscopy (PPC): number of polyps per colonoscopy, calculated by dividing the total number of polyps detected by the total number of colonoscopies. PPC was investigated according to lesion characteristics, namely size, morphology, location, and histology.
- Sessile serrated lesions per colonoscopy (SPC): number of sessile serrated lesions (SSLs) 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 adenomas per colonoscopy (AAPC): number of advanced adenomas per colonoscopy, 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 or treatment time were 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.

Data Synthesis and Analysis

In individual trials, we estimated risk ratios (RRs) for dichotomous outcomes and mean differences (MDs) for continuous ones, together with their 95% confidence intervals (95% 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 Chi2 (Cochran Q) and I2 statistics. 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 adenomas per colonoscopy and polyps per colonoscopy by colonic segment (right colon segement, transverse colon, left colon segment, or rectum), colonic site (proximal to splenic flexure or distal colon), size ($\leq$5 mm, 6-9 mm, $\geq$10 mm; and <10 mm, $\geq$10 mm), and morphology (polypoid or nonpolypoid). We estimated differences among subgroups by the Mantel-Haenszel test and heterogeneity using the Chi2 and I^2 statistics.

We planned sensitivity analyses by the leave-one-out approach for the primary outcomes (ie, adenoma detection rate 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 which 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 (GRADE) approach[20].

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

RESULTS

Study characteristics and quality

The initial literature search resulted in 554 articles (Figure 1). A total of 6 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 using a non-CADe CNN-based system focusing on withdrawal time and scope speed. Of the 5 included CADe-based RCTs,[10,11,21–23] one coupled the CADe algorithm to CNN models developed to assess quality indicators (withdrawal time, withdrawal stability, and bowel preparation).[22] Different CADe systems characteristics are summarized in Supplementary Table 1.

Most of the 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.

The total number of participants included in the analysis was 4354 (2163 in the CNN 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.

Adenoma and Polyp detection rate

Based on the data reported by all the 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<0.01). All RCT reported a significant ADR increase. However, there was moderate heterogeneity (I^2: 42%) in the level of the effect (Figure 2). Leave-one-out sensitivity analysis on ADR is reported in Supplementary Figure 2.

The overall PDR was also significantly improved in CADe group compared with control group with 1089 and 758 patients with at least one polyp out of 2163 and 2191 patients, respectively (50.3% vs 34.6%; RR, 1.43; 95% CI, 1.34-1.53; p<0.01) with low level of heterogeneity (I^2: 0%) across the 5 studies (Supplementary Fig. 3).

We explored publication bias of both the outcomes using funnel plots resulting in no small-study effect (Supplementary Fig. 4).

Adenoma, Sessile serrated lesions, and Polyps per colonoscopy

The number of adenomas detected per colonoscopy (APC) was significantly higher in the CADe group compared with the control group (1249/2163, 0.58 vs 779/2191, 0.36, RR: 1.70; 95% CI, 1.53-1.89; p<0.01) with moderate level of heterogeneity (I^2: 33%). The performance of CADe systems in significantly improving APC was confirmed irrespectively from adenoma size (RR $\leq$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 $\geq$10 mm: 1.46, 95% CI, 1.04-2.06, Figure 3), location (RR

proximal: 1.59 95% CI, 1.34-1.88 vs RR distal: 1.68; 95% CI, 1.50-1.88; Figure 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) (Table 2).

Based on data of 3 studies [10,11,21], only a not statistically significant trend in the number of advanced adenomas detected was found in favor of CADe as compared with control group (116/1347, 0.09 vs 71/1358, 0.05 RR: 1.35; 95% CI, 0.74 – 2.47; p: 0.33, I^2: 69%).

The number of sessile serrated lesions detected per colonoscopy (SPC) was significantly improved in the CADe group compared with the control group (109/1855, 0.06 vs 73/1876, 0.04, RR: 1.52; 95% CI, 1.14-2.02; p<0.01) with low level of heterogeneity (I^2: 0%) (Supplementary Fig. 6) across 4 studies[10,21,21,23].

Results of polyps per colonoscopy (PPC) and other per-polyps analysis were extensively reported in Supplementary Table 2.

Withdrawal time

No statistically significant difference between CADe and control groups was observed in term of mean withdrawal time (mean difference CADe vs control: 0.34 minutes,95% CI,-0.10 – 0.78; p = 0.13) with high level of heterogeneity (I^2: 97%). However, the heterogeneity level turn to be low (I^2: 0%), if excluding in a sensitivity analysis the study in which the CADe systems play a direct role in influencing the withdrawal time[22] (mean difference: 0.10; 95% CI, 0.02 – 0.18; p: 0.02) (Supplementary Fig. 7).

Quality of evidence

The quality of evidence was assessed by applying the GRADE methodology. The level of evidence for RCTs was downgraded due to 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 the imprecision due to possible differences in advanced adenoma definitions across studies. Details can be found in Supplementary Table 3.

DISCUSSION

The 44% and 70% relative increase in adenoma detection rate and adenoma per colonoscopy, consistently shown across the 5 included 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 subanalysis that were performed. In detail, CADe led to a statistically significant increase detection of both diminutive, small, and large adenomas, of those located in the proximal as well as in the distal colon, and of those flat and polypoid. In

addition, there was an increase detection of SSL 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 one or the other mechanism[3,6]. On the other hand, the CADe-RCT included in our meta-analysis represents a parallel methodology where only failure to recognize the lesion contributes to the additional detection.

The resilience of CADe efficacy from the traditional classifications of colorectal neoplasia is far from being unexpected as CADe and human perceptions are based on completely different mechanisms, namely a probabilistic analysis of the image based on training-acquired parameters vs 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 missing suggests that characteristics other than size and morphology are exploited by the machine to recognize the lesion. Speculatively, these factors may be represented by the texture, color and shape factors other than those perceived by the human mind.

Despite the additional detection of adenomas $\geq$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 advanced adenoma pool is represented by $\geq$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 of the analysis were performed at 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. Secondly, one of the adopted technology included both a 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 to discriminate the possible impact of each of the 2 components, leaving uncertainty on the possible synergistic effect between CADe technology, on one side, 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 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. As 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 with the Western setting, when considering the very low ADR in the control group of some of the 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 ADR in the control group could suggest an efficacy of CADe for low-detectors. However, dedicated studies are needed. More in general, 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 AI 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, Polkowski M, Wojciechowska U, Didkowska J, et al. Quality Indicators for Colonoscopy and the Risk of Interval Cancer. New England Journal of Medicine. 2010;362:1795–803.

2. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma Detection Rate and Risk of Colorectal Cancer and Death. New England Journal of Medicine. 2014;370:1298–306.

3. Zhao S, Wang S, Pan P, Xia T, Chang X, Yang X, et al. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. Gastroenterology. 2019;156:1661-1674.e11.

4. Dj R, Da L, Sj W, Dj A, Ja B, A S, 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;158:1287-1299.e2.

6. Rex D, Cutler C, Lemmel G, Rahmani E, Clark D, Helper D, 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, Antonino M, Berzin T, Byrne M, et al. Proceedings from the First Global Artificial Intelligence in Gastroenterology and Endoscopy Summit. Gastrointest Endosc. Epub 2020 Apr 25.

8. Bisschops R, East JE, Hassan C, Hazewinkel Y, Kamiński MF, Neumann H, 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, Maselli R, Craviotto V, Spadaccini M, 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, Correale L, Radaelli F, Rondonotti E, et al. Real-Time Computer Aided Diagnosis for Detection of Colorectal Neoplasia: a randomized controlled trial. Gastroenterology. 2020

11. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, 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, Saito Y, Ikematsu H, Hotta K, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med. 2018 18;169:357–66.

13. Greenspan M, Rajan KB, Baig A, Beck T, Mobarhan S, Melson J. 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, Irina G, Daniel P, Asaturi A, 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, Altman DG, PRISMA Group. 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, Tong T. 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. X W, W W, J L, T T. Estimating the Sample Mean and Standard Deviation From the Sample Size, Median, Range and/or Interquartile Range. Vol. 14, BMC medical research methodology. BMC Med Res Methodol; 2014 [cited 2020 May 25]. Available at: https://pubmed.ncbi.nlm.nih.gov/25524443/

18. R D, N L. Meta-analysis in Clinical Trials. Vol. 7, Controlled clinical trials. Control Clin Trials; 1986 [cited 2020 May 25]. Available at: https://pubmed.ncbi.nlm.nih.gov/3802833/

19. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.

20. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill 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, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. The Lancet Gastroenterology Hepatology. 2020;5:343–51.

22. Su J-R, Li Z, Shao X-J, Ji C-R, Ji R, Zhou R-C, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointestinal Endoscopy. 2020;91:415-424.e4.

23. Liu W-N, Zhang Y-Y, Bian X-Q, Wang L-J, Yang Q, Zhang X-D, 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, Mu G, Shen L, Liu J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. The Lancet Gastroenterology Hepatology. 2020;5:352–61.

Search number	Search query	Results
1	exp Colorectal Neoplasms/	197032
2	exp Cecal Neoplasms/	5470
3	exp Adenomatous Polyps/	7933
4	exp Intestinal Polyps/	14519
5	colo* cancer*.tw,kw.	138200
6	colo* adenoma*.tw,kw.	5652
7	colo* polyp*.tw,kw.	6375
8	or/1-7	250532
9	exp Artificial Intelligence/	92581
10	exp Diagnosis, Computer-Assisted/	82109
11	exp Image Processing, Computer-Assisted/	228955
12	artificial intelligence.tw,kw.	6715
13	cad*.tw,kw.	187891
14	computer aided detection.tw,kw.	1367
15	computer aided system*.tw,kw.	287
16	deep learning.tw,kw.	7875
17	or/9-16	543850
18	8 and 17	5555
19	randomized controlled trial.pt.	500805
20	controlled clinical trial.pt.	93552
21	pragmatic clinical trial.pt.	1301
22	randomi#ed.tw.	606532
23	placebo.ab.	205210
24	clinical trials as topic/	190226
25	randomly.ab.	327736
26	(crossover or cross-over).tw.	83501
27	Cross-over Studies/	47190
28	trial.tw.	578824
29	or/19-28	1508898
30	animals/ not (humans/ and animals/)	4640785
31	29 not 30	1383304
32	18 and 31	347

Appendix 1A: Medline 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	1,841
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

Appendix 1B: EMBASE 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	15848
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	21207
12	#5 AND #11	70

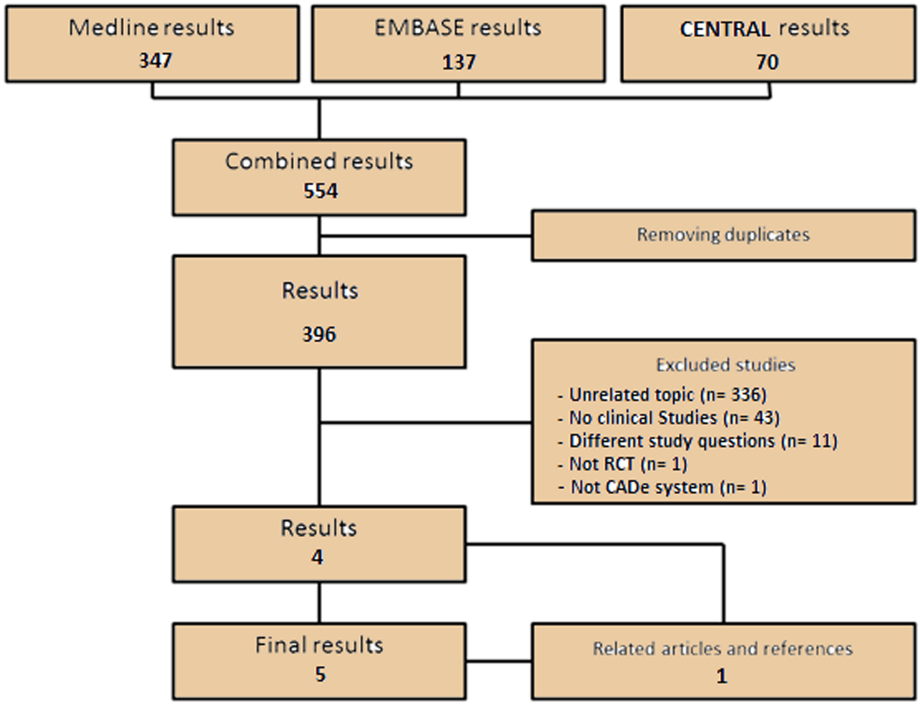
Appendix 1C: Cochrane Central Register of Controlled Trials (CENTRAL) search strategy.

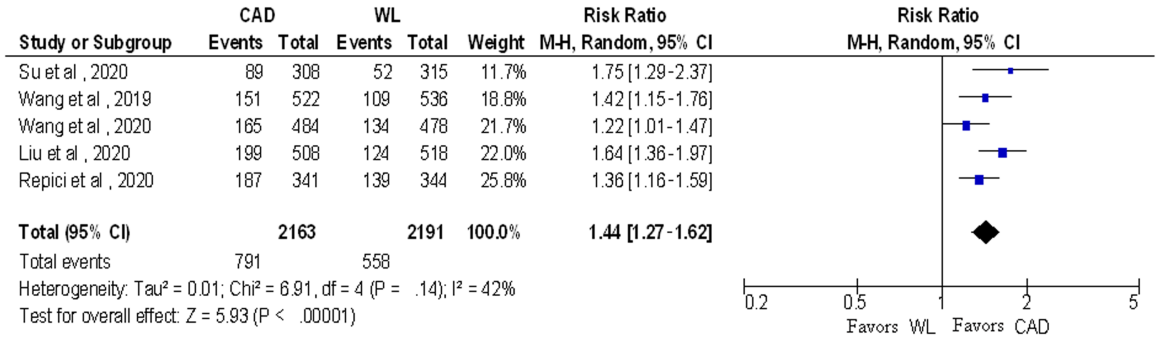
Reference	Publication year	Country	Type of system	Colonoscopy Indications	Endoscopists	Patients		Gender (male)		Age (mean years)		Withdrawal time		
						Control	CAD	Control	CAD	Control	CAD	Control	CAD	*P* value
Wang et al[11]	2019	China	CADe	Symptomatic: 974 Screening or surveillance: 84	2 senior endoscopists (>20 000 colonoscopies) 2 midlevel endoscopists (3000 to 10 000) 4 junior endoscopists (100 to 500)	536	522	249 (46%)	263 (50%)	49.9 ±13.8	51.1 ±13.2	6.1±1.1	6.2±1.4	ns
Wang et al[21]	2020	China	CADe	Symptomatic: 804 Screening or surveillance: 158	4 senior endoscopists (> 5 years' experience and > 1000 colonoscopies per year)	478	484	254 (53%)	241 (50%)	49.0 (40-56)	49.0 (39-60)	6.4±1.1	6.5±1.3	ns
Repici et al[10]	2020	Italy	CADe	Symptomatic: 161 Screening or surveillance: 524	6 experienced endoscopists (>2000 screening colonoscopies)	344	341	165 (48%)	172 (50%)	61.1 ±10.6	61.5 ±9.7	7.0±1.5	7.1±1.5	ns
Liu et al[23]	2020	China	CADe	Symptomatic: 960 Screening or surveillance: 66	\	518	508	287 (55%)	264 (52%)	50.1 ±12.7	51 ±12.3	6.1±1.0	6.2±1.3	ns
Su et al[22]	2020	China	CADe + Quality	Symptomatic: 407 Screening or surveillance: 216	6 endoscopists (5000 to 8000 colonoscopies)	315	308	148 (47%)	159 (49%)	51.6 ±9.0	50.5 ±10.3	5.7±1.1	7.0±1.0	ns

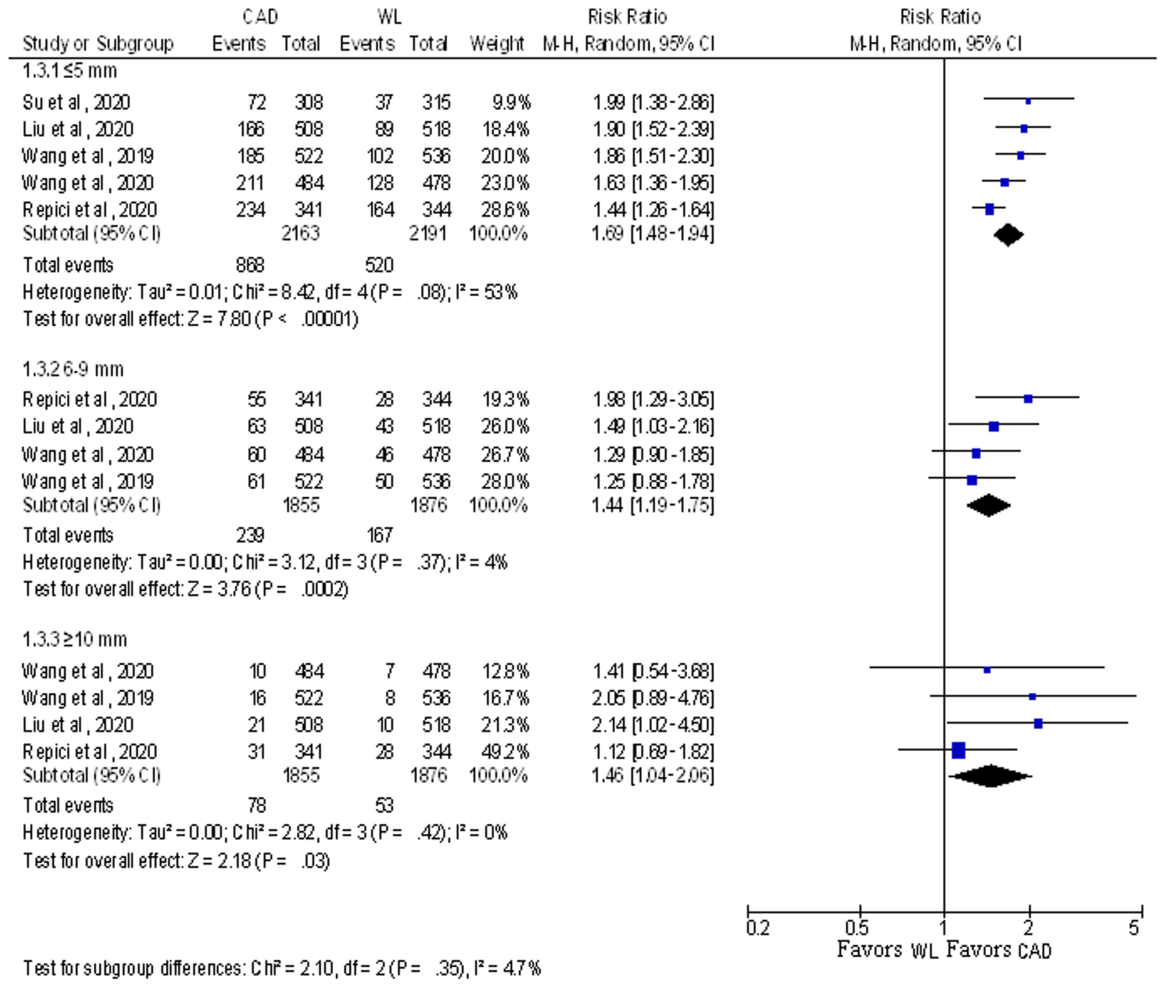
Table 1: Study characteristics. CAD: Computer-aided diagnosis. CADe: Computer-aided polyp detection; ns: not statistically significant

Reference	Adenoma <5 mm (%)			Adenoma 6-9 mm (%)			Adenoma ≥10 mm (%)			Adenoma proximal (%)			Adenoma distal (%)			Adenoma polypoid (%)			Adenoma flat (%)		
	Control	**CAD**	***P* value**	**Control**	**CAD**	***P* value**	**Control**	**CAD**	***P* value**	**Control**	**CAD**	***P* value**	**Control**	**CAD**	***P* value**	**Control**	**CAD**	***P* value**	**Control**	**CAD**	***P* value**
Wang et al[11]	102(63.8)	185(70.6)	<0.05	50(31.6)	61(23.3)	ns	8(5.0)	16(6.1)	ns	76(47.5)	122(46.6)	<0.05	84(52.5)	140(53.4)	<0.05	97(60.6)	144(55.0)	<0.05	63(39.4)	118(45.0)	<0.05
Wang et al[21]	128(71)	211(75)	<0.05	46(25)	60(21)	ns	7(4)	10(4)	ns	85(47.0)	132(47.0)	<0.05	96(53.0)	149(53.0)	<0.05	180(99)	276(98)	<0.05	1(1)	5(2)	<0.05
Repici et al[10]	164(74.5)	234(73.1)	<0.05	28(12.7)	55(17.2)	<0.05	28(12.7)	31(9.7)	ns	151(63.0)	200(58.3)	<0.05	89(37.0)	143(41.7)	<0.05	143(59.6)	204(68.1)	<0.05	97(40.4)	139(31.9)	<0.05
Liu et al[23]	89(62.7)	166(66.4)	<0.05	43(30.3)	63(25.2)	ns	10(7.0)	21(8.4)	ns	81(57.0)	131(52.4)	<0.05	61(42.9)	119(47.6)	<0.05	82(57.7)	128(51.2)	<0.05	60(42,3)	122(48.8)	<0.05
Su et al[22]	37(66.1)	72(63.7)	<0.05	\	\	\	\	\	\	18(32.1)	48(42.5)	<0.05	38(67.9)	65(57.5)	<0.05	35(62.5)	75(66.4)	<0.05	21(37.5)	38(33.6)	<0.05

Table 2: Adenoma detection subgrouped according to size, location, and morphology. CAD: Computer-aided diagnosis


Medline results
347
EMBASE results
137
CENTRAL results
70
Combined results
554
Removing duplicates
Results
396
Excluded studies
- Unrelated topic (n= 336)
- No clinical Studies (n= 43)
- Different study questions (n= 11)
- Not RCT (n= 1)
- Not CADe system (n= 1)
Results
4
Final results
5
Related articles and references
1


CAD
WL
Risk Ratio
Risk Ratio
Study or Subgroup
Events
Total
Events
Total
Weight
M-H, Random, 95% CI
M-H, Random, 95% CI
Su et al , 2020 89 308 52 315 11.7% 1.75 [1.29-2.37]
Wang et al , 2019 151 522 109 536 18.8% 1.42 [1.15-1.76]
Wang et al , 2020 165 484 134 478 21.7% 1.22 [1.01-1.47]
Liu et al , 2020 199 508 124 518 22.0% 1.64 [1.36-1.97]
Repici et al , 2020 187 341 139 344 25.8% 1.36 [1.16-1.59]
Total (95% CI) 2163 2191 100.0% 1.44 [1.27-1.62]
Total events 791 558
Heterogeneity: Tau² = 0.01; Chi² = 6.91, df = 4 (P = .14); I² = 42%
Test for overall effect: Z = 5.93 (P < .00001)
0.2
0.5
1
2
5
Favors WL
Favors CAD


CAD
WL
Risk Ratio
Risk Ratio
Study or Subgroup Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
1.3.1 ≤5 mm
Su et al , 2020 72 308 37 315 9.9% 1.99 [1.38 - 2.86]
Liu et al , 2020 166 508 89 518 18.4% 1.90 [1.52 - 2.39]
Wang et al , 2019 185 522 102 536 20.0% 1.86 [1.51 - 2.30]
Wang et al , 2020 211 484 128 478 23.0% 1.63 [1.36 - 1.95]
Repici et al , 2020 234 341 164 344 28.6% 1.44 [1.26 - 1.64]
Subtotal (95% CI) 2163 2191 100.0% 1.69 [1.48 - 1.94]
Total events 868 520
Heterogeneity: Tau² = 0.01; Chi² = 8.42, df = 4 (P = .08); I² = 53%
Test for overall effect: Z = 7.80 (P < .00001)
1.3.2 6-9 mm
Repici et al , 2020 55 341 28 344 19.3% 1.98 [1.29 - 3.05]
Liu et al , 2020 63 508 43 518 26.0% 1.49 [1.03 - 2.16]
Wang et al , 2020 60 484 46 478 26.7% 1.29 [0.90 - 1.85]
Wang et al , 2019 61 522 50 536 28.0% 1.25 [0.88 - 1.78]
Subtotal (95% CI) 1855 1876 100.0% 1.44 [1.19 - 1.75]
Total events 239 167
Heterogeneity: Tau² = 0.00; Chi² = 3.12, df = 3 (P = .37); I² = 4%
Test for overall effect: Z = 3.76 (P = .0002)
1.3.3 ≥10 mm
Wang et al , 2020 10 484 7 478 12.8% 1.41 [0.54 - 3.68]
Wang et al , 2019 16 522 8 536 16.7% 2.05 [0.89 - 4.76]
Liu et al , 2020 21 508 10 518 21.3% 2.14 [1.02 - 4.50]
Repici et al , 2020 31 341 28 344 49.2% 1.12 [0.69 - 1.82]
Subtotal (95% CI) 1855 1876 100.0% 1.46 [1.04 - 2.06]
Total events 78 53
Heterogeneity: Tau² = 0.00; Chi² = 2.82, df = 3 (P = .42); I² = 0%
Test for overall effect: Z = 2.18 (P = .03)
0.2
0.5
1
2
5
Favors WL
Favors CAD
Test for subgroup differences: Chi² = 2.10, df = 2 (P = .35), I² = 4.7%

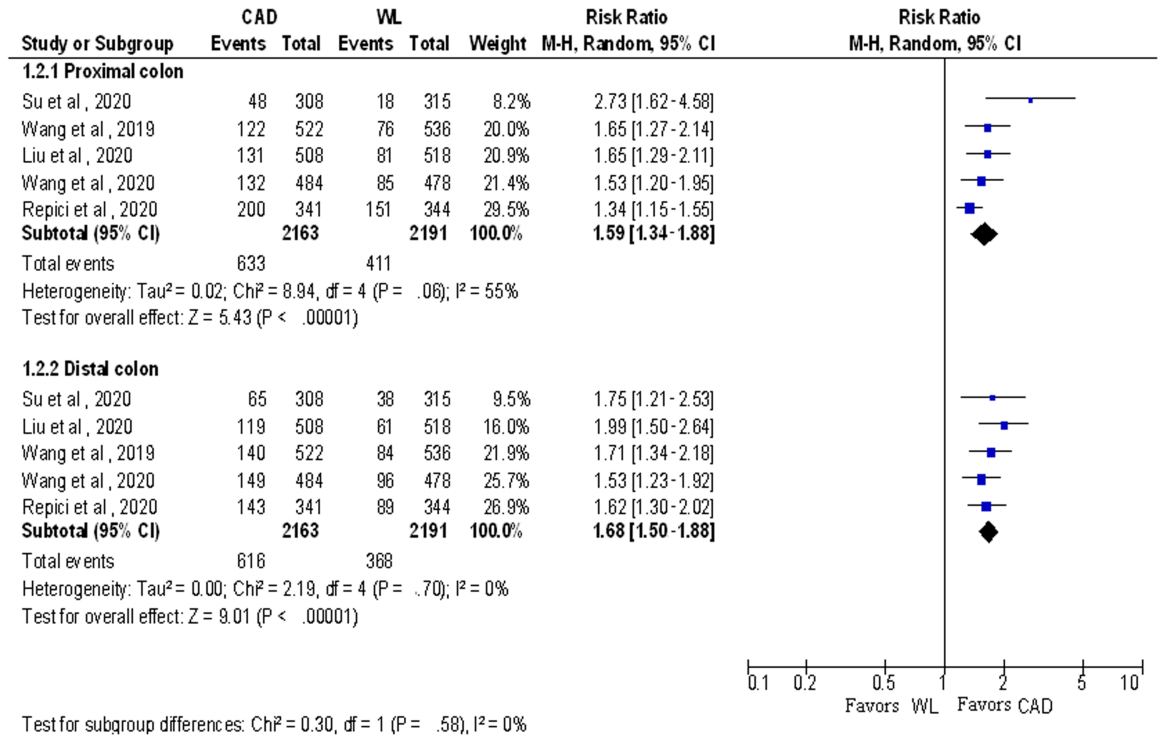

CAD
WL
Risk Ratio
Risk Ratio
Study or Subgroup Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
1.2.1 Proximal colon
Su et al , 2020 48 308 18 315 8.2% 2.73 [1.62 - 4.58]
Wang et al , 2019 122 522 76 536 20.0% 1.65 [1.27 - 2.14]
Liu et al , 2020 131 508 81 518 20.9% 1.65 [1.29 - 2.11]
Wang et al , 2020 132 484 85 478 21.4% 1.53 [1.20 - 1.95]
Repici et al , 2020 200 341 151 344 29.5% 1.34 [1.15 - 1.55]
Subtotal (95% CI) 2163 2191 100.0% 1.59 [1.34 - 1.88]
Total events 633 411
Heterogeneity: Tau² = 0.02; Chi² = 8.94, df = 4 (P = .06); I² = 55%
Test for overall effect: Z = 5.43 (P < .00001)
1.2.2 Distal colon
Su et al , 2020 65 308 38 315 9.5% 1.75 [1.21 - 2.53]
Liu et al , 2020 119 508 61 518 16.0% 1.99 [1.50 - 2.64]
Wang et al , 2019 140 522 84 536 21.9% 1.71 [1.34 - 2.18]
Wang et al , 2020 149 484 96 478 25.7% 1.53 [1.23 - 1.92]
Repici et al , 2020 143 341 89 344 26.9% 1.62 [1.30 - 2.02]
Subtotal (95% CI) 2163 2191 100.0% 1.68 [1.50 - 1.88]
Total events 616 368
Heterogeneity: Tau² = 0.00; Chi² = 2.19, df = 4 (P = .70); I² = 0%
Test for overall effect: Z = 9.01 (P < .00001)
0.1 0.2 0.5 1 2 5 10
Favors WL Favors CAD
Test for subgroup differences: Chi² = 0.30, df = 1 (P = .58), I² = 0%

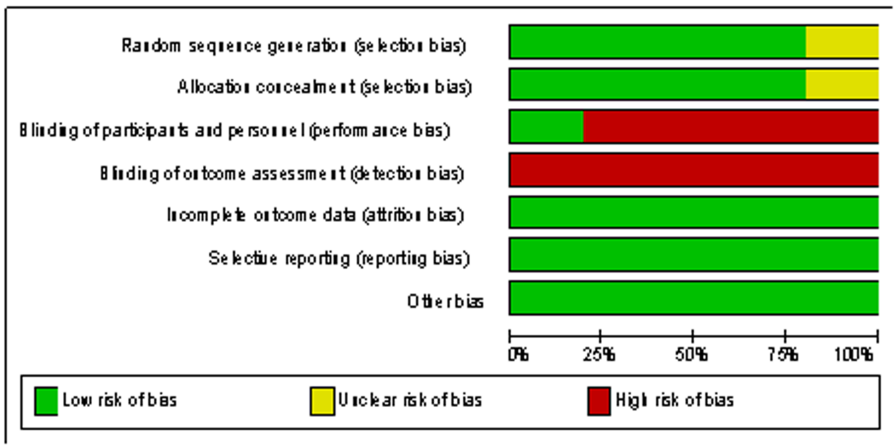
Figure Captions

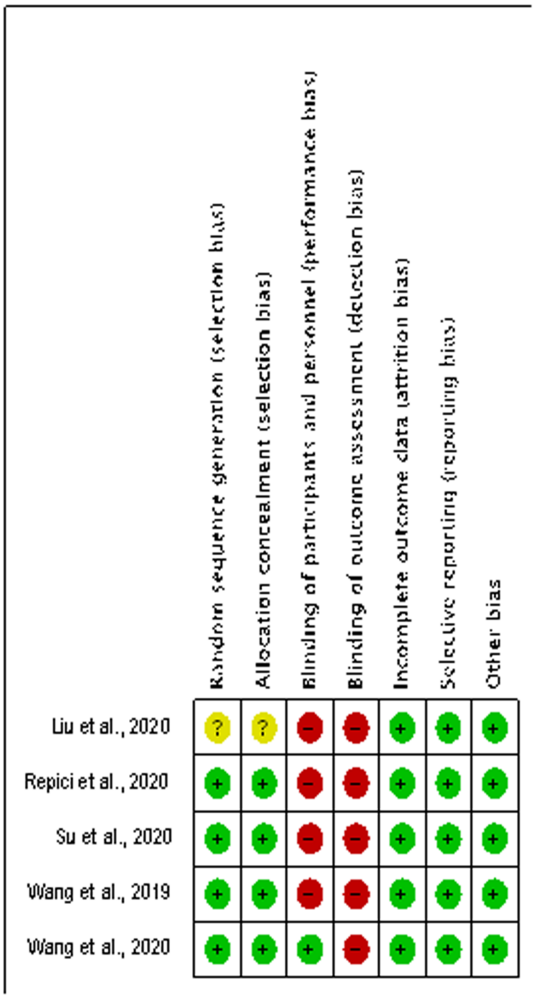
Figure 1: Study selection flow chart. CADe: Computer-aided polyp detection.

Figure 2: Comparative effectiveness of CAD versus control group on ADR. CAD: Computer-aided diagnosis. ADR: adenoma detection rate.

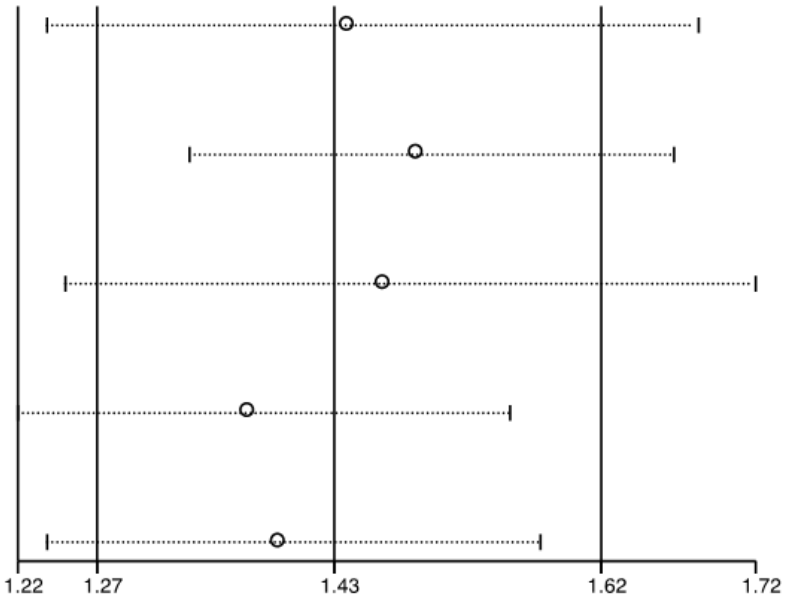
Figure 3: Comparative effectiveness of CAD versus control group on APC subgrouped according to size. CAD: Computer-aided diagnosis. APC: adenoma per colonoscopy.

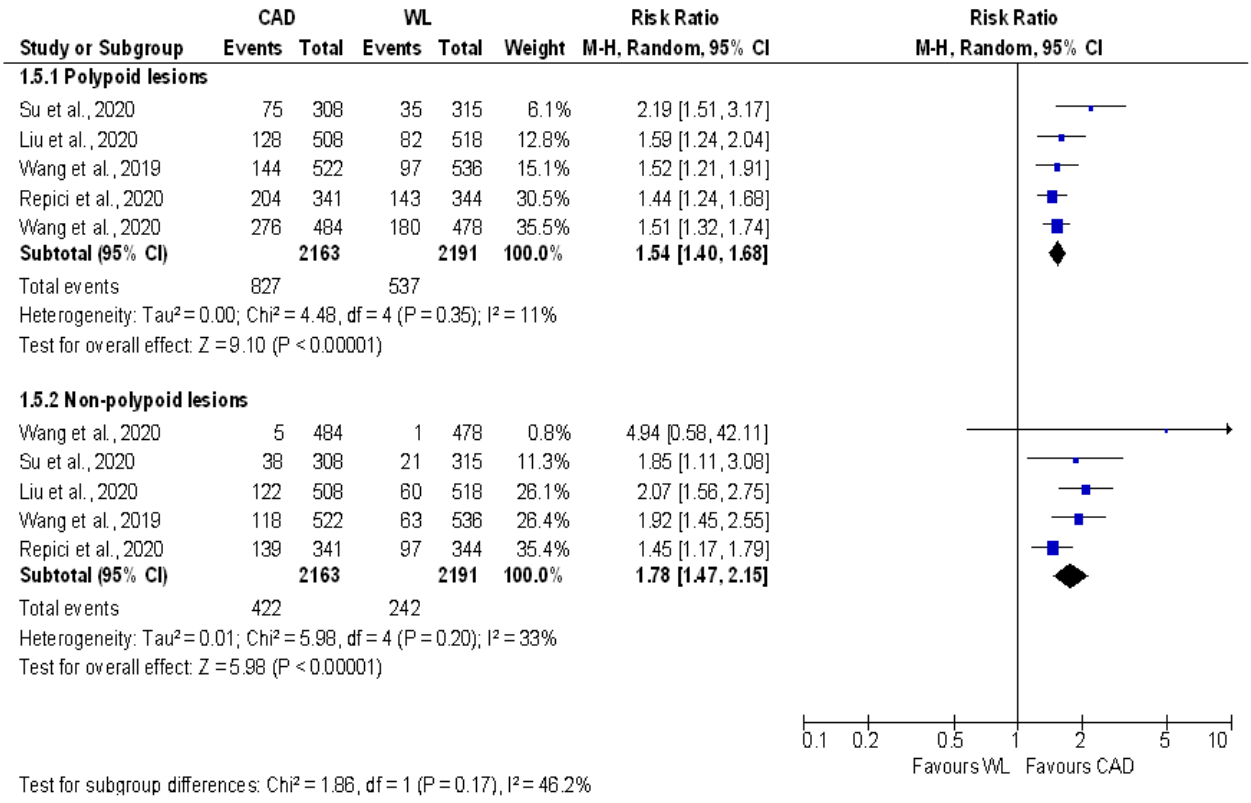
Figure 4: Comparative effectiveness of CAD versus control group on APC subgrouped according to location. CAD: Computer-aided diagnosis. APC: adenoma per colonoscopy.

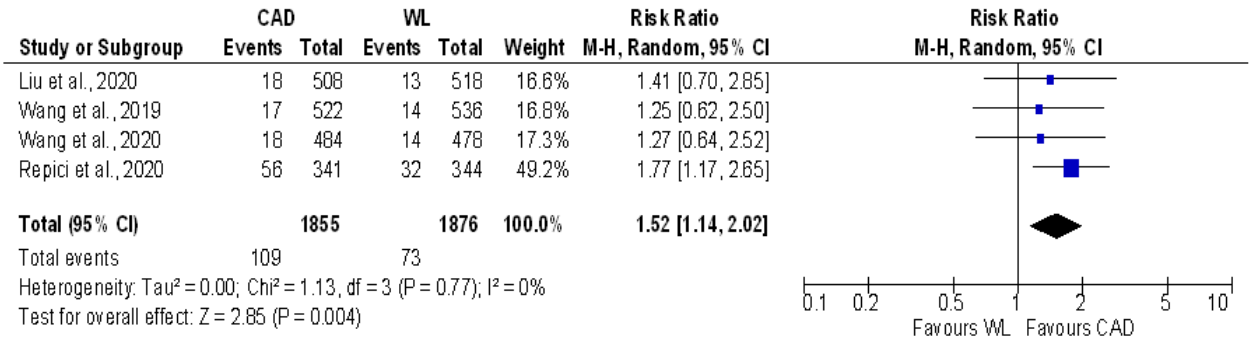

Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias)
Blinding of outcome assessment (detection bias)
Incomplete outcome data (attrition bias)
Selective reporting (reporting bias)
Other bias
0%
25%
50%
75%
100%
Low risk of bias
Unclear risk of bias
High risk of bias

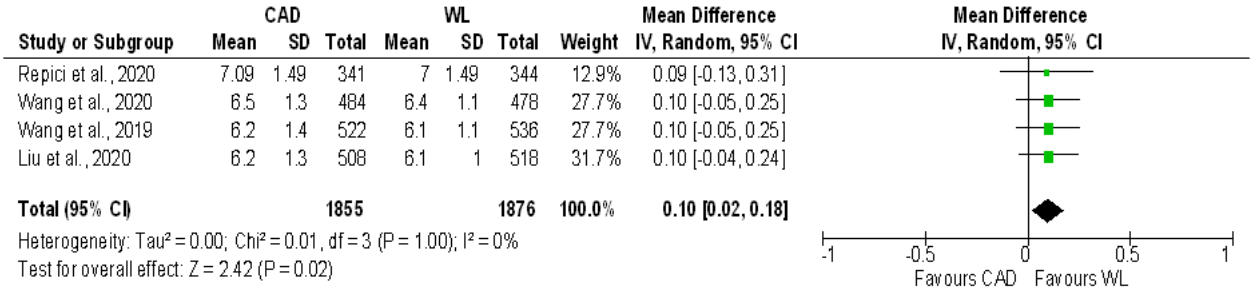

Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias)
Blinding of outcome assessment (detection bias)
Incomplete outcome data (attrition bias)
Selective reporting (reporting bias)
Other bias
Liu et al., 2020
Repici et al., 2020
Su et al., 2020
Wang et al., 2019
Wang et al., 2020

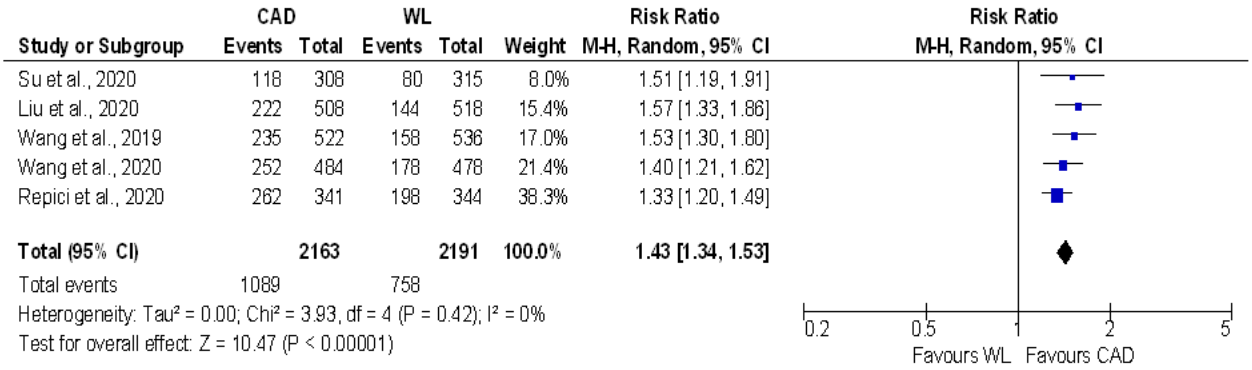
Excluded study	Risk Ratio (95%CI)
Wang et al., 2019	1.44 (1.24-1.68)
Wang et al., 2020	1.49 (1.34-1.67)
Repici et al., 2020	1.47 (1.25-1.72)
Liu et al., 2020	1.38 (1.22-1.55)
Su et al., 2020	1.40 (1.24-1.58)




CAD
WL
Risk Ratio
Risk Ratio
Study or Subgroup Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
1.5.1 Polypoid lesions
Su et al., 2020 75 308 35 315 6.1% 2.19 [1.51, 3.17]
Liu et al., 2020 128 508 82 518 12.8% 1.59 [1.24, 2.04]
Wang et al., 2019 144 522 97 536 15.1% 1.52 [1.21, 1.91]
Repici et al., 2020 204 341 143 344 30.5% 1.44 [1.24, 1.68]
Wang et al., 2020 276 484 180 478 35.5% 1.51 [1.32, 1.74]
Subtotal (95% CI) 2163 2191 100.0% 1.54 [1.40, 1.68]
Total events 827 537
Heterogeneity: Tau² = 0.00; Chi² = 4.48, df = 4 (P = 0.35); I² = 11%
Test for overall effect: Z = 9.10 (P < 0.00001)
1.5.2 Non-polypoid lesions
Wang et al., 2020 5 484 1 478 0.8% 4.94 [0.58, 42.11]
Su et al., 2020 38 308 21 315 11.3% 1.85 [1.11, 3.08]
Liu et al., 2020 122 508 60 518 26.1% 2.07 [1.56, 2.75]
Wang et al., 2019 118 522 63 536 26.4% 1.92 [1.45, 2.55]
Repici et al., 2020 139 341 97 344 35.4% 1.45 [1.17, 1.79]
Subtotal (95% CI) 2163 2191 100.0% 1.78 [1.47, 2.15]
Total events 422 242
Heterogeneity: Tau² = 0.01; Chi² = 5.98, df = 4 (P = 0.20); I² = 33%
Test for overall effect: Z = 5.98 (P < 0.00001)
0.1 0.2 0.5 1 2 5 10
Favours WL Favours CAD
Test for subgroup differences: Chi² = 1.86, df = 1 (P = 0.17), I² = 46.2%


CAD
WL
Risk Ratio
Risk Ratio
Study or Subgroup
Events
Total
Events
Total
Weight
M-H, Random, 95% CI
M-H, Random, 95% CI
Liu et al., 2020 18 508 13 518 16.6% 1.41 [0.70, 2.85]
Wang et al., 2019 17 522 14 536 16.8% 1.25 [0.62, 2.50]
Wang et al., 2020 18 484 14 478 17.3% 1.27 [0.64, 2.52]
Repici et al., 2020 56 341 32 344 49.2% 1.77 [1.17, 2.65]
Total (95% CI) 1855 1876 100.0% 1.52 [1.14, 2.02]
Total events 109 73
Heterogeneity: Tau² = 0.00; Chi² = 1.13, df = 3 (P = 0.77); I² = 0%
Test for overall effect: Z = 2.85 (P = 0.004)
0.1
0.2
0.5
1
2
5
10
Favours WL
Favours CAD


CAD
WL
Mean Difference
Mean Difference
Study or Subgroup
Mean
SD
Total
Mean
SD
Total
Weight
IV, Random, 95% CI
IV, Random, 95% CI
Repici et al., 2020 7.09 1.49 341 7 1.49 344 12.9% 0.09 [-0.13, 0.31]
Wang et al., 2020 6.5 1.3 484 6.4 1.1 478 27.7% 0.10 [-0.05, 0.25]
Wang et al., 2019 6.2 1.4 522 6.1 1.1 536 27.7% 0.10 [-0.05, 0.25]
Liu et al., 2020 6.2 1.3 508 6.1 1 518 31.7% 0.10 [-0.04, 0.24]
Total (95% CI) 1855 1876 100.0% 0.10 [0.02, 0.18]
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 3 (P = 1.00); I² = 0%
Test for overall effect: Z = 2.42 (P = 0.02)
-1
-0.5
0
0.5
1
Favours CAD
Favours WL


CAD
WL
Risk Ratio
Risk Ratio
Study or Subgroup
Events Total Events Total Weight M-H, Random, 95% CI
M-H, Random, 95% CI
Su et al., 2020 118 308 80 315 8.0% 1.51 [1.19, 1.91]
Liu et al., 2020 222 508 144 518 15.4% 1.57 [1.33, 1.86]
Wang et al., 2019 235 522 158 536 17.0% 1.53 [1.30, 1.80]
Wang et al., 2020 252 484 178 478 21.4% 1.40 [1.21, 1.62]
Repici et al., 2020 262 341 198 344 38.3% 1.33 [1.20, 1.49]
Total (95% CI) 2163 2191 100.0% 1.43 [1.34, 1.53]
Total events 1089 758
Heterogeneity: Tau² = 0.00; Chi² = 3.93, df = 4 (P = 0.42); I² = 0%
Test for overall effect: Z = 10.47 (P < 0.00001)
0.2
0.5
1
2
5
Favours WL
Favours CAD

ACRONYMS and ABBREVIATIONS:

Colorectal Cancer: CRC.

Computer-Aided Polyp Detection: CADe.

Randomized Clinical Trials: RCT.

adenoma detection rate: ADR.

Adenoma per colonoscopy: APC.

Polyp per colonoscopy: PPC.

Advanced Adenoma per colonoscopy: AAPC.

Polyp detection rate: PDR.

Sessile serrated lesion per colonoscopy: SPC.

Risk ratio: RR.

Convoluted Neural Networks: CNN.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses: PRISMA.

Sessile serrated lesions: SSL.

Mean differences: MDs.

Grades of Recommendation, Assessment, Development and Evaluation: GRADE.