Journal of Gastroenterology and Hepatology

JGHF

doi:10.1111/jgh.15070

META ANALYSIS AND SYSTEMATIC REVIEW

The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: A systematic review with meta-analysis

Muhammad Aziz,* Rawish Fatima,* Charles Dong,* Wade Lee-Smith† and Ali Nawras‡

*Department of Internal Medicine, University of Toledo Medical Center, Toledo, Ohio USA †University of Toledo Libraries, University of Toledo Medical Center, Toledo, Ohio USA ‡Department of Gastroenterology, University of Toledo Medical Center, Toledo, Ohio USA

Key words
artificial intelligence, colonoscopy, deep convolutional neural network, high-definition.

Accepted for publication 5 April 2020.

Correspondence
Muhammad Aziz, Department of Internal Medicine, University of Toledo Medical Center, 3000 Arlington Avenue, Toledo, OH 43614, USA.
Email: marajani@hotmail.com

Declaration of conflict of interest: The authors declare no conflict of interest. No human subjects/animals were involved in this systematic review and meta-analysis.
Author contribution: Muhammad Aziz planned and conducted the study, collected and interpreted the data, conducted the statistical analysis, and drafted the manuscript. Rawish Fatima collected the data, interpreted the data, and drafted the manuscript. Dong Charles interpreted the data and drafted the manuscript. Wade Lee-Smith created the search strategy and collected the data, and drafted the manuscript. Ali Nawras was involved in study design/conception and critical revision of the manuscript. All authors have approved the submission of this manuscript to the journal and are accountable for all aspects of the work as relevant.
Financial support: No funding was received for the preparation of this manuscript.
Guarantor of the article: Muhammad Aziz, MD

Abstract

Background and Aim: The utility of artificial intelligence (AI) in colonoscopy has gained popularity in current times. Recent trials have evaluated the efficacy of deep convolutional neural network (DCNN)-based AI system in colonoscopy for improving adenoma detection rate (ADR) and polyp detection rate (PDR). We performed a systematic review and meta-analysis of the available studies to assess the impact of DCNN-based AI-assisted colonoscopy in improving the ADR and PDR.

Methods: We queried the following database for this study: PubMed, Embase, Cochrane Library, Web of Sciences, and Computers and Applied Sciences. We only included randomized controlled trials that compared AI colonoscopy to standard colonoscopy (SC). Our outcomes included ADR and PDR. Risk ratios (RR) with 95% confidence interval (CI) were calculated using random effects model and DerSimonian–Laird approach for each outcome.

Results: A total of three studies with 2815 patients (1415 in SC group and 1400 in AI group) were included. AI colonoscopy resulted in significantly improved ADR (32.9% *vs* 20.8%, RR: 1.58, 95% CI 1.39–1.80, $P = < 0.001$) and PDR (43.0% *vs* 27.8%, RR: 1.55, 95% CI 1.39–1.72, $P = < 0.001$) compared with SC.

Conclusion: Given the results and limitations, the utility of AI colonoscopy holds promise and should be evaluated in more randomized controlled trials across different population, especially in patients solely undergoing colonoscopy for screening purpose as improved ADR will ultimately help in reducing incident colorectal cancer.

Introduction

Since the advent of colonoscopy, colorectal screening has been demonstrated to decrease mortality and morbidity of colorectal cancer (CRC) in adults.[1] Colonoscopy in its current state is effective at detecting adenomas, a common precursor for CRC.[2] Improving adenoma detection rate (ADR), an important quality metric and colonoscopy outcome, has shown to substantially decrease both incidence of CRC and mortality.[3]

Numerous interventions have been introduced to improve the ADR of colonoscopy. These include the use of add-on devices (endocap, endoring, endocuff, G-EYE, and amplifEYE), electronic chromoendoscopy (narrow band imaging, linked color imaging, blue laser imaging, and flexible spectral imaging color enhancement), water-aided techniques (water immersion and water exchange), use of second observer, antispasmodic administration, retroflexion in the right colon or second forward view,

quality control initiatives, and improved bowel preparation.[4–12] However, there are still a number of adenomas that are missed during colonoscopy, and current efforts are being directed to prevent this.

Recent studies in literature have evaluated the impact of deep convolutional neural network (DCNN), a branch of artificial intelligence (AI) that allows detailed image analysis in real-time by computer algorithms during colonoscopy.[13] The current efforts and trials are aimed at improving these AI algorithms to reflect the ADR of expert endoscopists that would assist endoscopists with low ADR in the clinical setting. The algorithms utilized in newer trials have demonstrated excellent per-image accuracy and sensitivity for polyp detection (> 95% and > 90%).[14,15]

Although the results of these individual trials look promising, the applicability is limited because of overall low number of patients. We performed a systematic review and meta-analysis of the available trials to summarize and compare the outcomes for AI and standard colonoscopy (SC), ADR, and polyp detection rate (PDR).

Methods

Study definitions. Adenoma detection rate is defined as the proportion of patients with detection of at least one adenoma. PDR is defined as the proportion of patients with detection of at least one polyp. Cecal intubation time (CIT) is defined as the time for the colonoscope tip to reach proximal to the ileocecal valve. Withdrawal time (WT) is defined as the examination time of the colon while the colonoscope is being withdrawn after achieving cecal intubation. Total procedural time is defined as the time from the start to end of colonoscopy. Polyp per subject (PPS) is defined as the total number of polyps detected divided by the total number of subjects examined. Adenoma per subject (APS) is defined as the total number of adenomas detected divided by the total number of subjects examined. SC is defined as patients undergoing high-definition colonoscopy as the norm in clinical practice without any added intervention. AI colonoscopy is defined as patients undergoing colonoscopy with computer-aided real-time automated quality control or polyp/adenoma detection based on DCNN algorithms. False alarm refers to any lesion detected by AI colonoscopy that was anything other than a polyp/adenoma.

Search strategy. We conducted a detailed search of the following databases through January 8, 2020: PubMed/Medline, Embase, Cochrane Register of Controlled Trials, Web of Sciences Core Collection, and Computers and Applied Sciences Complete. The search strategy was formulated by a librarian (W. L. S.) and cross checked by another reviewer (M. A.). Article screening was performed by two independent reviewers (M. A. and R. F.), and any discrepancy was resolved through mutual discussion. Bibliography of the selected articles were also screened to broaden the literature search. PRISMA guidelines were adhered for the purpose of this manuscript. An example of detailed search strategy on EMBASE is highlighted in Table S1.

Inclusion and exclusion criteria. Our search strategy was restricted to include only randomized controlled trial (RCT). We excluded all other study designs (cohort, case–control, case series, case reports, reviews, and guidelines) as well as studies published as abstracts only to generate the highest level of evidence. We did not restrict our search based on language. The studies were finalized based on the following: (i) use of DCNN-based artificial intelligence system during colonoscopy in the intervention group; (ii) standard high-definition colonoscopy as control group; (iii) patients undergoing colonoscopy based on average risk screening or symptoms; and (iv) outcome of interest, that is, ADR and PDR reported.

Data collection. Baseline demographics data (age, sex, and indication for colonoscopy), quality indicators (CIT and WT), and outcomes (ADR, PDR, APS, and PPS) were extracted by two independent reviewers (M. A. and R. F.). Discrepancy in data extraction process was resolved through mutual discussion.

Primary and secondary outcomes. The primary outcome of interest was ADR, and secondary outcomes included PDR, APS, PPS, CIT, and WT. Additional outcomes for PPS and APS on the basis of location in colon and morphology were attempted if data were available.

Data synthesis and sensitivity analysis. Extracted data were tabulated in Microsoft Excel (Microsoft, Redmond, WA, USA). Proportional outcomes (PDR and ADR) and outcomes with mean difference (CIT and WT) were compared using weighted random effects and DerSimonian–Laird method. The fixed effect using Mantel–Haenzel method was used as a sensitivity tool. Summary estimates, that is, risk ratio (RR) and mean difference (MD) with 95% confidence interval (CI) and *P*-values were calculated and displayed using forest plots. We evaluated the APS and PPS using the generic inverse variance meta-analysis, and MD with 95% CI and *P*-values was calculated. Study heterogeneity was assessed using the I^2 statistic, and a value of > 50% was considered as significant heterogeneity.[16] The meta-analysis was performed on Open Meta Analyst (CEBM, University of Oxford, Oxford, UK) and Comprehensive Meta-Analysis (BioStat, Englewood, NJ, USA).

Non-meta-analytical outcomes, that is, APS and PPS were assessed using two sample *t*-test and computed using SPSS v26 (IBM, Armonk, NY, USA). A P-value < 0.05 was considered to be statistically significant for all assessed outcomes. The outcomes were assessed on per-protocol basis, that is, patients who completed the study and were not excluded for reasons such as inadequate bowel prep, failed procedure, and findings of cancer/inflammatory bowel disease/previous surgery.

Bias assessment and quality of evidence. Study quality was assessed using the Cochrane risk of bias tools for RCTs.[17] Publication bias was assessed qualitatively by visually assessing the funnel plots and quantitively using Egger's regression analysis. Quality of evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation approach.[18]

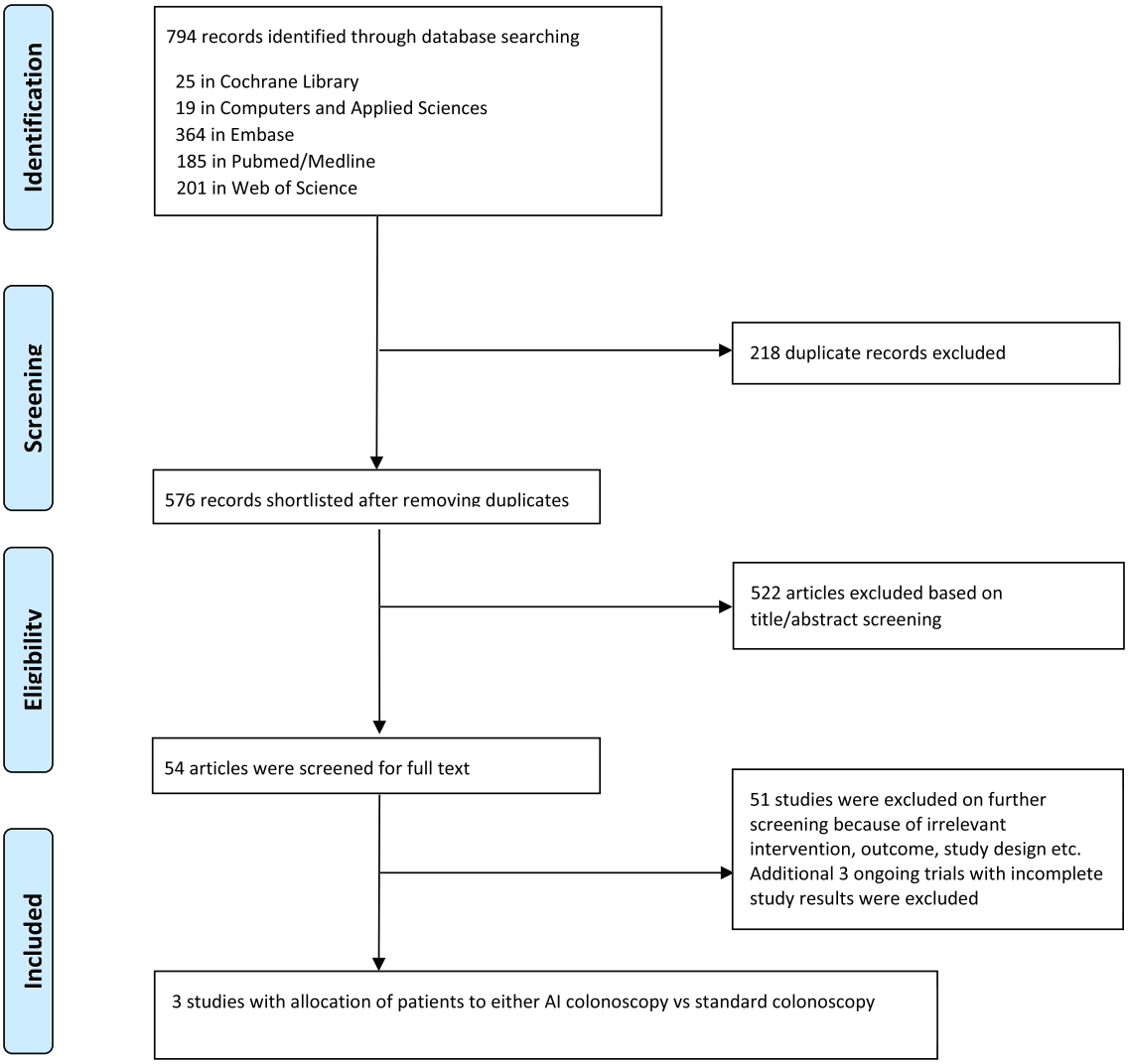

Figure 1 Preferred reporting items for systematic review and meta-analysis (PRISMA) flow diagram. [Color figure can be viewed at wileyonlinelibrary.com]

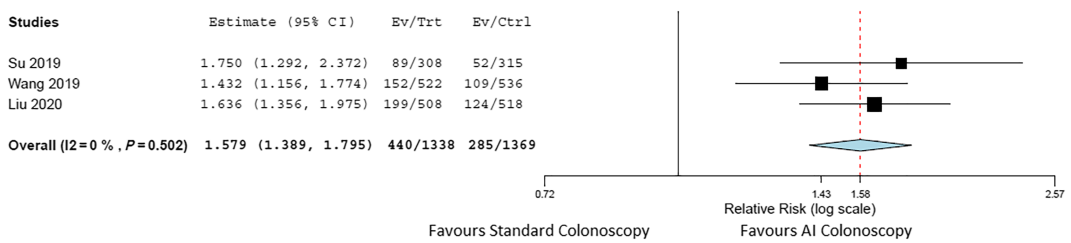

Figure 2 Forest plot comparing adenoma detection rate using artificial intelligence (AI) *versus* standard colonoscopy (SC) group. CI, confidence interval; Ctrl, SC; Trt, AI. [Color figure can be viewed at wileyonlinelibrary.com]

Results

A total of three RCTs were finalized based on the search strategy (Fig. 1).[19–21] All studies were published after 2019 and originated from China. For AI group, all three studies utilized DCNN models with capability of detecting polyp in real time. One study further supplemented the DCNN models to assess additional quality indicators (withdrawal time, quality of bowel preparation, and withdrawal stability).[21] Two studies reported the name of

manufacturer for AI system, that is, Shagai Wision AI co. Ltd[19] and Henan Xuanweitang Medical Information technology Co. Ltd.[20] One study reported developing and testing of the DCNN algorithm models by study investigators, and name of manufacturer was not reported.[21] Pretesting of the diagnostic algorithm was conducted by two studies, and a diagnostic accuracy of 98.4%[19] and 93.33%–98.30%[21] was reported. Further details regarding AI system in individual studies are given in Data S1.

The risk of bias assessment of included RCTs are highlighted in Table S2. The risk of bias was high for all studies particularly because the colonoscopists were not blinded to treatment allocation because of practical reasons. Publication bias was difficult to assess based on funnel plots because of low number of studies. Egger's regression analysis did not demonstrate significant publication bias ($P = 0.76$).

Study details and demographics of patients are summarized in Tables 1 and S3. The total number of patients included in three studies were 2815 (1415 in SC group and 1400 in AI group). Study completion rate for all studies was 96.2% (96.7% *vs* 95.6% for SC and AI group, respectively). No baseline significant difference was observed in age range (49.9–51.6 *vs* 50.5–51.0), male gender (50.0% *vs* 51.3%), and screening/surveillance indication (13.2% *vs* 13.8%).

Detection rate. The outcomes ADR and PDR are summarized in Table 2. The overall ADR was significantly higher for AI group compared with SC group (32.9% *vs* 20.8%, RR: 1.58, 95% CI 1.39–1.80, $P = < 0.001$, $I^2 = 0\%$; Fig. 2). A subgroup analysis of two studies that only used polyp detection DCNN system also showed significantly increased ADR for AI group (34.1% *vs* 22.1%, RR: 1.54, 95% CI 1.34–1.78, $P = < 0.001$, $I^2 = 0\%$).

The overall PDR was also significantly elevated for AI group compared with SC group (43.0% *vs* 27.8%, RR: 1.55, 95% CI 1.39–1.72, $P = < 0.001$, $I^2 = 0\%$; Fig. 3). Subgroup analysis of two studies that only used polyp detection DCNN system also showed significantly increased PDR for AI group (44.4% *vs* 28.5%, RR: 1.56, 95% CI 1.39–1.75, $P = < 0.001$, $I^2 = 0\%$).

No polyps were missed by the intervention group. The details of false alarms, that is, AI suggesting the possibility of a lesion that was later proven to be false positive is highlighted in Table 2. The quality of evidence was low for both ADR and PDR (rated down because of significant risk of bias as noted in Table S2 and possible confounding factor as diagnostic colonoscopy was the indication in majority of the patients).

Polyp/adenoma per subject. The mean PPS and APS were significantly improved when colonoscopy was performed using AI compared with SC (0.87 *vs* 0.45, $P < 0.001$ and 0.46 *vs* 0.26, $P < 0.001$ for polyps and adenomas, respectively). The results were consistent when meta-analysis using inverse variance method was applied (MD: 0.356, 95% CI 0.220–0.493, $P < 0.001$, $I^2 = 0\%$ and MD: 0.202, 95% CI 0.133–0.271, $P < 0.001$, $I^2 = 0\%$) for both PPS and APS (Figs S1 and S2). The quality of evidence was very low for both PPS and APS because of factors mentioned earlier for ADR and PDR as well as the indirectness of the overall results. The results were also consistent on further analysis using *t*-test based on location of polyps

Table 1 Study details

S. No	Study, year	PMID	Type of study	Study site, Country	Study Period	Main components of DCNN model	Total patients, *n*	Patients in control group, *n*	Patients in intervention group, *n*	Study completion control (%)	Study completion intervention (%)
1.	Wang, 2019	30814121	RCT	Sichuan Provincial People's Hospital, China	September 2017–February 2018	Polyp detection	1130	567	563	94.5	92.7
2.	Su, 2019	31454493	RCT	Qilu Hospital, China	October 2018–May 2019	Polyp detection, withdrawal time, withdrawal stability, quality of bowel preparation	659	330	329	95.5	93.6
3.	Liu, 2020	31898644	RCT	988th Hospital of Joint Logistic Support Force of PLA, China	October 2018–March 2019	Polyp detection	1026	518	508	100	100

AI, artificial intelligence colonoscopy; *n*, no. of patients; PMID, PubMed Index number; RCT, randomized controlled trial; SC, standard colonoscopy.

Table 2 Detection rates and procedure time compared using meta-analysis

Study	Wang 2019		Su 2019		Liu 2019		*P*-value*
	SC	AI	SC	AI	SC	AI	
ADR, *n*/*N* (%)	109/536 (20.3%)	152/522 (29.12%)	52/315 (16.5%)	89/308 (28.9%)	124/518 (23.90%)	199/508 (39.20%)	< 0.001
PDR, *n*/*N* (%)	156/536 (29.1%)	235/522 (45.0%)	80/315 (25.4%)	118/308 (38.3%)	144/518 (27.80%)	222/508 (43.70%)	< 0.001
False alarm (*n*)	—	39	—	62	—	36	NA
Missed polyps (*n*)	—	0	—	0	—	0	NA
Cecal intubation time in minutes, mean (*SD*)	5.71 (3.90)	5.63 (4.03)	6.38 (2.25)	6.27 (2.17)	5.96 (4.06)	5.68 (4.09)	0.25
Withdrawal time in minutes, mean (*SD*)	6.39 (1.21)	6.89 (1.79)	5.68 (1.26)	7.03 (1.01)	6.74 (1.62)	6.82 (1.78)	0.03
Total procedure time in minutes, mean (*SD*)	12.10 (4.08)	12.52 (4.38)	NR	NR	12.70 (4.16)	12.41 (4.25)	0.853
PPS (*n*)	0.502	0.954	0.305	0.575	0.479	0.957	< 0.001
APS (*n*)	0.299	0.502	0.178	0.367	0.274	0.492	< 0.001

**P*-value is assessed using meta-analysis.

ADR, adenoma detection rate; AI, artificial intelligence colonoscopy; *n*, number event; *N*, total patients; PDR, polyp detection rate; SC, standard colonoscopy.

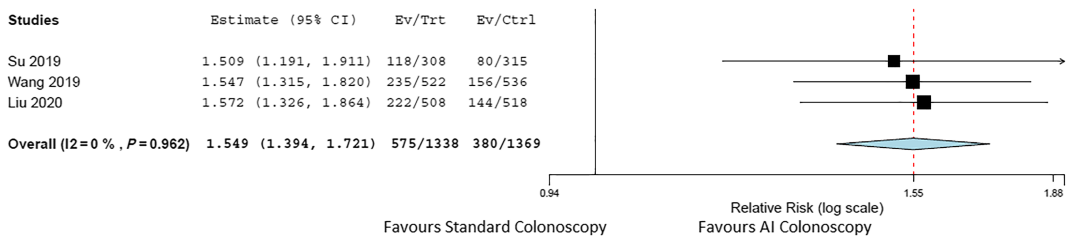

Figure 3 Forest plot comparing polyp detection rate using artificial intelligence (AI) *versus* standard colonoscopy (SC) group. CI, confidence interval; Ctrl, SC; Trt, AI. [Color figure can be viewed at wileyonlinelibrary.com]

(right sided or left sided), location of adenomas (right sided or left sided), and morphology of adenomas (flat/sessile or pedunculated; Table 3). The mean sessile serrated adenoma per subject (SAPS) was reported by two studies. Improved SAPS was noted with AI colonoscopy compared with SC colonoscopy (0.03 *vs* 0.02, $P < 0.001$).

Cecal intubation time/withdrawal time. No significant difference in mean cecal intubation time was noted for either arm (MD: −0.14, 95% CI −0.39–0.10, $P = 0.25$; Fig. 4a). Significantly increased mean withdrawal time was observed for patients undergoing colonoscopy in AI group (MD: 0.90, 95% CI 0.09–1.71, $P = 0.03$; Fig. 4b). Only two studies reported mean total procedural time, and no significant difference was observed (MD: 0.066, 95% CI −0.630–0.762, $P = 0.853$).

Discussion

This systematic review and meta-analysis demonstrated statistically significant results for ADR and PDR using AI colonoscopy compared with SC colonoscopy. The results were also consistent when mean APS and PPS were compared. Further subgroup analysis demonstrated consistent results based on location, size, and morphology of polyps and adenomas. Collectively, these results support the utility of AI colonoscopy for improving ADR in clinical setting.

The efficacy of colonoscopy lies in the detection of precancerous lesions, that is, adenomas and serrated lesions. The detection rate of these lesions is significantly affected by several factors including endoscopist experience, bowel preparation, colonoscopy indication, physician fatigue, and examination time.[22–24] Substantial variations exist when collectively looking at these factors, and hence, the need for a more standardized and systematic approach is needed to circumvent this. DCNN-based AI colonoscopy might be the answer to limit the human error and subjective bias associated with this issue. This brings into play the study by Su *et al*. that used DCNN models targeting not only polyp detection but also other quality metrics such as bowel preparation, withdrawal time, and withdrawal stability.[21] The increased detection rates, that is, ADR and PDR can also be attributed to increased detection of flat and small lesions as shown by a previous study.[25] Our study demonstrated significant improvement in both flat/sessile APS and adenomas < 10 mm using AI colonoscopy that may have resulted in overall improved ADR and PDR.

The attractiveness of the AI system is based on real-time automated detection of polyps. However, for the system to effectively work overall in clinical setting, high diagnostic sensitivity,

Table 3 Results of detected polyps/adenomas

Study	Wang 2019		Su 2019		Liu 2019		Total across all studies		*P*-value
	SC	AI	SC	AI	SC	AI	SC	AI	
Total polyps (*n*)	269	498	96	177	248	486	613	1161	—
PPS (*n*)	0.502	0.954	0.305	0.575	0.479	0.957	0.448	0.868	< 0.001
Location of polyp (*n*)									
Right	95	212	34	75	96	218	215	505	—
Left	174	286	62	102	152	268	388	656	—
RPPS (*n*)	0.178	0.406	0.108	0.243	0.185	0.429	0.157	0.377	< 0.001
LPPS (*n*)	0.325	0.548	0.197	0.331	0.293	0.528	0.283	0.490	< 0.001
Size of the polyp, *n*									
< 10 mm (PPS)	—	—	NR	NR	—	—	—	—	—
≥ 10 mm (PPS)	259 (0.457)	482 (0.856)	—	—	232 (0.448)	464 (0.913)	491 (0.453)	946 (0.883)	< 0.001
	10 (0.018)	16 (0.031)	—	—	16 (0.031)	22 (0.043)	26 (0.024)	38 (0.035)	< 0.001
Total adenomas (*n*)	160	262	56	113	142	250	358	625	—
APS (*n*)	0.299	0.502	0.178	0.367	0.274	0.492	0.262	0.467	< 0.001
Location of adenomas (*n*)									
Right	76	122	18	48	81	131	175	301	—
Left	84	140	38	65	61	119	183	324	—
RAPS (*n*)	0.142	0.234	0.057	0.156	0.156	0.258	0.128	0.225	< 0.001
LAPS (*n*)	0.157	0.268	0.121	0.211	0.118	0.234	0.134	0.242	< 0.001
Size of the adenomas, *n*									
< 10 mm (APS)	152 (0.268)	246 (0.437)	NR	NR	132 (0.255)	229 (0.451)	284 (0.262)	475 (0.444)	< 0.001
≥ 10 mm (APS)	8 (0.014)	16 (0.031)	—	—	10 (0.019)	21 (0.041)	18 (0.017)	37 (0.035)	< 0.001
Advanced adenomas (*n*)	16	17	NR	NR	16	14	32 (2 studies)	31 (2 studies)	—
AAPS (*n*)	0.029	0.033	—	—	0.031	0.028	0.030	0.030	1.00
Sessile Serrated Adenomas (*n*)	14	17	NR	NR	13	18	17 (2 studies)	35 (2 studies)	—
SAPS (*n*)	0.026	0.033	—	—	0.025	0.035	0.02	0.03	< 0.001
Morphology of adenomas (*n*)									
Flat/Sessile	127	223	43	98	106	210	435	531	—
Pedunculated	33	39	13	15	36	40	82	94	—
FAPS (*n*)	0.237	0.427	0.137	0.318	0.205	0.413	0.318	0.397	< 0.001
PAPS (*n*)	0.062	0.075	0.041	0.049	0.069	0.079	0.060	0.070	< 0.001
Cancer	0	0	NR	NR	0	0	0	0	NA

P-values for mean per subject calculated using two sample *t*-test.

AI, artificial intelligence colonoscopy; AAPS, advanced adenoma per subject; APS, adenoma per subject; FAPS, flat/sessile APS; LAPS, left colon APS; LPPS, left colon PPS; *n*, no./mean; NA, not applicable; NR, not reported; PAPS, pedunculated APS; PPS, polyp per subject; RAPS, right colon APS; RPPS, right colon PPS; SAPS, sessile serrated adenoma per subject; SC, standard colonoscopy.

specificity, and accuracy as well as efficiency is required. An inadequate specificity would result in large number of false positives (as noted in our study), while an inadequate sensitivity would result in large number of false negatives. The significantly increased WT in AI group can be partially attributed to the number of alarms (false positives) in our study. A previous study showed significant improvement in ADR with increased withdrawal time or repeated/longer examination.[26] The current study results with improved ADR and PDR can also be somewhat attributed to increased WT, that is, longer examination time. The significant heterogeneity noted with WT can be due to several factors including different DCNN models for AI group, difference in endoscopic experience/practice, and patient-related factors.

Interestingly, we found all AI to significantly improve SAPS as well. Serrated lesions (sessile serrated adenomas and traditional serrated adenomas) are notorious for causing CRC through serrated adenoma pathway, accounting for roughly 20% of all CRC.[27] Hence, there is an interest in improving the detection rates of these lesions.[8] Although our results favor AI, the clinical relevance is low as the actual mean SAPS detected across two studies was small. There is certainly a need for modalities to improve the detection of these lesions, and optimizing AI might be one potential solution.

Another important concern in our study was the indication of colonoscopy as the majority of patients were included for diagnostic purpose. This is problematic as true ADR should be accounted for screening colonoscopies only.[28,29] The low ADR seen in the SC group across all studies can be explained because of this issue. The studies also did not mention the screening method, that is, direct colonoscopy *versus* stool-based testing leading to colonoscopy. With < 15% of included patients with screening indication, our study results should be interpreted with caution taking this important limitation into account.

Several other important limitations exist with our study. The biggest one is the low number of available studies as it can undermine all outcomes of interest including ADR, PDR, APS, PPS, and false alarms. Another important limitation is the high risk of bias in included RCTs as it was impossible to blind

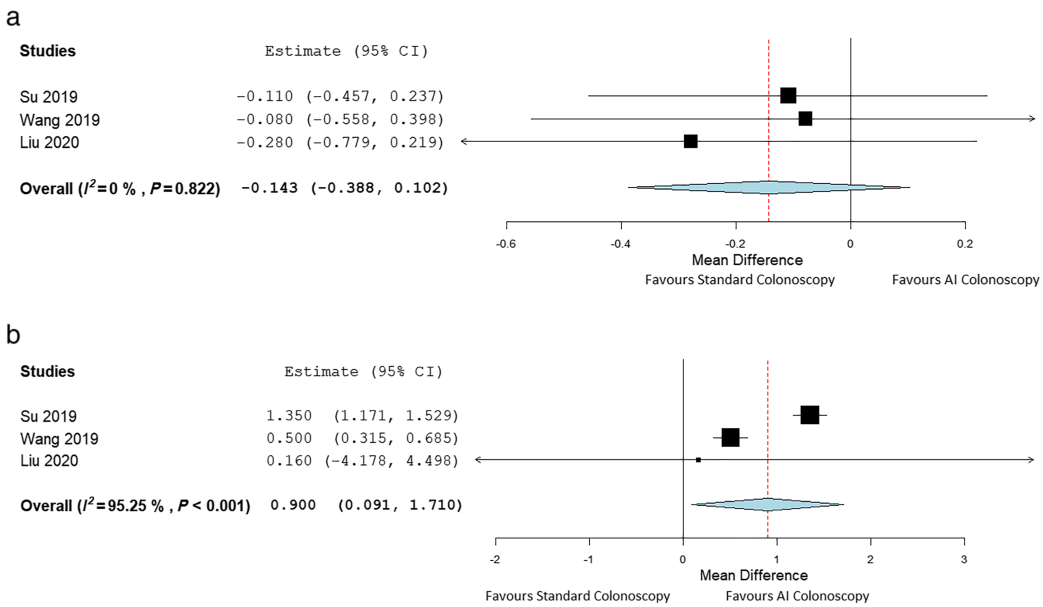

Figure 4 Forest plot comparing artificial intelligence (AI) *versus* standard colonoscopy (SC) for (a) CIT (b) WT. CI, confidence interval; CIT, cecal intubation time; Ctrl, SC; Trt, AI; WT, withdrawal time. [Color figure can be viewed at wileyonlinelibrary.com]

endoscopist to patient's treatment allocation. Further, one of the studies utilized additional quality controls in their DCNN algorithms,[21] while the other two did not. We were also not able to account for factors such as bowel preparation, physician factors (timing of colonoscopy, fatigue, endoscopy experience, and level of training), and patient factors (ethnicity, smoking, family history, and genetic syndrome). All the studies originated from a single country, and hence, the generalizability of the results is still questionable. Lastly, these studies did not perform tandem colonoscopies to assess the true impact of missed lesions/polyps.

Despite these limitations, all the included studies were RCTs with a robust number of patients. In addition to our strict inclusion and exclusion criteria for selecting RCTs, each included RCT had further strict criteria for enrolling patients. Moreover, our results remained consistent on all sensitivity and subgroup analysis.

In conclusion, the use of AI-assisted colonoscopy resulted in improved colonoscopy outcome metrics, that is, ADR and PDR. Although the results look promising, there is a need for more RCTs evaluating the efficacy of AI colonoscopy in patients with screening indication only to truly unravel the potential of this technique. There is a need for more studies from diverse and multiple population (for generalizability) as well as evaluation using tandem colonoscopies (i.e. AI followed by SC and SC followed by AI) to assess the true impact of AI on ADR, PDR, and missed lesions. Further, studies are also needed to assess the efficacy of AI colonoscopy compared with other interventions such as add-on devices, enhanced imaging modalities, and chromoendoscopy.

References

1 Aziz M, Thoguluva CV. Improving the ADR: narrow band, broad benefits. *Dig Dis Sci* 2019: 1–2. https://doi.org/10.1007/s10620-019-05997-6.
2 Rex D. Colonoscopy: the current king of the hill in the USA. *Dig Dis Sci* 2014; **60**: 639–46.
3 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.
4 Aziz M, Weissman S, Khan Z *et al*. Use of 2 observers increases adenoma detection rate during colonoscopy: systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2019. https://doi.org/10.1016/j.cgh.2019.07.033.
5 Facciorusso A, Triantafyllou K, Murad MH *et al*. Compared abilities of endoscopic techniques to increase colon adenoma detection rates: a network meta-analysis. *Clin Gastroenterol Hepatol* 2019; **17**: 2439–54.
6 Desai M, Bilal M, Hamade N *et al*. Increasing adenoma detection rates in the right side of the colon comparing retroflexion with a second forward view: a systematic review. *Gastrointest Endosc* 2019; **89**: 453–9.
7 Seo JY, Jin EH, Bae JH *et al*. Multidirectional colonoscopy quality improvement increases adenoma detection rate: results of the Seoul National University Hospital Healthcare System Gangnam Center Colonoscopy Quality Upgrade Project (Gangnam-CUP). *Dig Dis Sci*. https://doi.org/10.1007/s10620-019-05944-5.
8 Aziz M, Desai M, Hassan S *et al*. Improving serrated adenoma detection rate in the colon by electronic chromoendoscopy and distal attachment: systematic review and meta-analysis. *Gastrointest Endosc* 2019; **90**: 721–31.
9 Sulz MC, Kröger A, Prakash M, Manser CN, Heinrich H, Misselwitz B. Meta-analysis of the effect of bowel preparation on adenoma detection: early adenomas affected stronger than advanced adenomas. *PLoS ONE* 2016; **11**: e0154149.
10 Fuccio L, Frazzoni L, Hassan C *et al*. Water exchange colonoscopy increases adenoma detection rate: a systematic review with network meta-analysis of randomized controlled studies. *Gastrointest Endosc* 2018; **88**: 589–97.
11 Aziz M, Sharma S, Fatima R *et al*. How to increase proximal adenoma detection rate: a meta-analysis comparing water exchange, water immersion and air/CO2 insufflation methods for colonoscopy. *Ann Gastroenterol* 2020; **33**: 1–9.

12 Aziz M, Sharma S, Ghazaleh S *et al*. The anti-spasmodic effect of peppermint oil during colonoscopy: a systematic review and meta-analysis. *Minerva Gastroenterol Dietol* 2020: 1–8. https://doi.org/10.23736/S1121-421X.20.02652-5.
13 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.
14 Urban G, Tripathi P, Alkayali T *et al*. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018; **155**: 1069–78.
15 Byrne MF, Chapados N, Soudan F *et al*. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019; **68**: 94–100.
16 Shuster J. Review: Cochrane handbook for systematic reviews for interventions, Version 5.1.0, published 3/2011. Julian P.T. Higgins and Sally Green, Editors. *Res Synth Methods* 2011; **2**: 126–30.
17 Higgins JP, Altman DG, Gøtzsche PC *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; **343**: d5928–8.
18 Puhan MA, Schünemann HJ, Murad MH *et al*. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014; **349**: g5630.
19 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.
20 Liu W, Zhang Y, Bian X *et al*. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol* 2020; **26**: 13. https://doi.org/10.4103/sjg.SJG_377_19.
21 Su JR, Li Z, Shao XJ *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* 2019; **91**: 415–24. https://doi.org/10.1016/j.gie.2019.08.026.
22 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.
23 Froehlich F, Wietlisbach V, Gonvers JJ, Burnand B, Vader JP. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. *Gastrointest Endosc* 2005; **61**: 378–84.
24 Atkins L, Hunkeler EM, Jensen CD *et al*. Factors influencing variation in physician adenoma detection rates: a theory-based approach for performance improvement. *Gastrointest Endosc* 2016; **83**: 617–26.
25 Buchner AM, Shahid MW, Heckman MG *et al*. Trainee participation is associated with increased small adenoma detection. *Gastrointest Endosc* 2011; **73**: 1223–31.
26 Ai X, Qiao W, Han Z *et al*. Results of a second examination of the right side of the colon in screening and surveillance colonoscopy: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol* 2018; **30**: 181–6.
27 Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* 2010; **138**: 2088–100.
28 Hassan C, Repici A, Rex DK. FITting ADR to colonoscopy indication. *United European Gastroenterol J* 2017; **5**: 149–52.
29 Anderson JC, Butterly LF, Goodrich M, Robinson CM, Weiss JE. Differences in detection rates of adenomas and serrated polyps in screening versus surveillance colonoscopies, based on the New Hampshire colonoscopy registry. *Clin Gastroenterol Hepatol* 2013; **11**: 1308–12.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. EMBASE search strategy.
Table S2. Risk of bias in the included RCTs.
Table S3. Demographics and characteristics of patients included in the study.

Data S1. Supporting Information.

Figure S1. Forest plot comparing PPS using AI vs SC group (AI: Artificial intelligence colonoscopy, CI: Confidence interval, SC: Standard colonoscopy).

Figure S2. Forest plot comparing APS using AI vs SC group (AI: Artificial intelligence colonoscopy, CI: Confidence interval, SC: Standard colonoscopy).