Clinical Gastroenterology and Hepatology 2023;21:949–959

ENDOSCOPY

Impact of Artificial Intelligence on Colonoscopy Surveillance After Polyp Removal: A Pooled Analysis of Randomized Trials

Yuichi Mori,[1,2,3,*] Pu Wang,[4,*] Magnus Løberg,[1,2] Masashi Misawa,[3] Alessandro Repici,[5,6] Marco Spadaccini,[5] Loredana Correale,[5] Giulio Antonelli,[7,8] Honggang Yu,[9,10,11] Dexin Gong,[9,10,11] Misaki Ishiyama,[3] Shin-ei Kudo,[3] Shunsuke Kamba,[12] Kazuki Sumiyama,[12] Yutaka Saito,[13] Haruo Nishino,[14] Peixi Liu,[4] Jeremy R. Glissen Brown,[15] Nabil M. Mansour,[16] Seth A. Gross,[17] Mette Kalager,[1,2] Michael Bretthauer,[1,2] Douglas K. Rex,[18] Prateek Sharma,[19] Tyler M. Berzin,[20] and Cesare Hassan[5,6]

[1]Clinical Effectiveness Research Group, University of Oslo, Oslo, Norway; [2]Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway; [3]Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan; [4]Department of Gastroenterology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Sichuan, China; [5]Endoscopy Unit, Humanitas Clinical and Research Center-IRCCS, Rozzano, Italy; [6]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Italy; [7]Gastroenterology and Digestive Endoscopy Unit, Ospedale dei Castelli Hospital, Ariccia, Rome, Italy; [8]Department of Anatomical, Histological, Forensic Medicine and Orthopedics Sciences, Sapienza University of Rome, Italy; [9]Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China; [10]Key Laboratory of Hubei Province for Digestive System Disease, Renmin Hospital of Wuhan University, Wuhan, China; [11]Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Wuhan University Renmin Hospital, Wuhan, China; [12]Department of Endoscopy, the Jikei University School of Medicine, Tokyo, Japan; [13]Endoscopy Division, National Cancer Center Hospital, Tokyo, Japan; [14]Coloproctology Center, Matsushima Hospital, Yokohama, Japan; [15]Division of Gastroenterology, Duke University Medical Center, Durham, North Carolina; [16]Section of Gastroenterology and Hepatology, Baylor College of Medicine, Houston, Texas; [17]Division of Gastroenterology and Hepatology, NYU Langone Health, New York, New York; [18]Division of Gastroenterology/Hepatology, Indiana University School of Medicine, Indianapolis, Indiana; [19]Department of Gastroenterology and Hepatology, Kansas City VA Medical Center and University of Kansas School of Medicine, Kansas City, Kansas; and [20]Center for Advanced Endoscopy, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts

BACKGROUND AND AIMS: **Artificial intelligence (AI) tools aimed at improving polyp detection have been shown to increase the adenoma detection rate during colonoscopy. However, it is unknown how increased polyp detection rates by AI affect the burden of patient surveillance after polyp removal.**

METHODS: **We conducted a pooled analysis of 9 randomized controlled trials (5 in China, 2 in Italy, 1 in Japan, and 1 in the United States) comparing colonoscopy with or without AI detection aids. The primary outcome was the proportion of patients recommended to undergo intensive surveillance (ie, 3-year interval). We analyzed intervals for AI and non-AI colonoscopies for the U.S. and European recommendations separately. We estimated proportions by calculating relative risks using the Mantel-Haenszel method.**

RESULTS: **A total of 5796 patients (51% male, mean 53 years of age) were included; 2894 underwent AI-assisted colonoscopy and 2902 non-AI colonoscopy. When following U.S. guidelines, the proportion of patients recommended intensive surveillance increased from 8.4% (95% CI, 7.4%–9.5%) in the non-AI group to 11.3% (95% CI, 10.2%–12.6%) in the AI group (absolute difference, 2.9% [95% CI, 1.4%–4.4%]; risk ratio, 1.35 [95% CI, 1.16–1.57]). When following European guidelines, it increased from 6.1% (95% CI, 5.3%–7.0%) to 7.4% (95% CI, 6.5%–8.4%) (absolute difference, 1.3% [95% CI, 0.01%–2.6%]; risk ratio, 1.22 [95% CI, 1.01–1.47]).**

*Authors share co-first authorship.

Abbreviations used in this paper: ADR, adenoma detection rate; AI, artificial intelligence; CRC, colorectal cancer; ESGE, European Society of Gastrointestinal Endoscopy; SSL, sessile serrated lesions; USMSTF, U.S. Multi-Society Task Force on Colorectal Cancer.

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2022.08.022

CONCLUSIONS: **The use of AI during colonoscopy increased the proportion of patients requiring intensive colonoscopy surveillance by approximately 35% in the United States and 20% in Europe (absolute increases of 2.9% and 1.3%, respectively). While this may contribute to improved cancer prevention, it significantly adds patient burden and healthcare costs.**

Keywords: Computer-Aided Diagnosis; Surveillance Interval; Machine Learning.

Colorectal cancer (CRC) is the second leading cause of cancer death worldwide. Screening colonoscopy is expected to reduce incidence and mortality from CRC through early detection of cancer and removal of precancerous adenomas.[1] It is well known that colonoscopy efficacy to prevent CRC varies according to endoscopists' adenoma detection rate (ADR). Patients examined by endoscopists with high ADR have lower CRC mortality as compared with those examined by endoscopists with low ADR.[2]

The exact relationship between ADR and future cancer prevention is still unknown. There could be a linear relationship,[2] but there remains the possibility of a threshold effect, so that above a certain level of ADR there may be no further improvement in cancer prevention.[3] If there is a threshold between ADR and CRC prevention, increased ADR beyond a threshold may result in additional cost and burden for patients and health systems without significant additional benefit.[4,5]

An important part of costs and burden for patients with polyps is colonoscopy surveillance after polyp removal. Current guidelines recommend frequent surveillance colonoscopy for patients with polyps.[6–8] As patient surveillance after colonoscopy is based on the number of polyps removed (in addition to size and histological features), more patients examined by endoscopists with high ADR are recommended surveillance colonoscopies. Recently, the use of artificial intelligence (AI) for polyp detection has been shown to increase ADR of individual endoscopists by about 12% (ie, from 25% to 37%).[9] While there is benefit of increased ADR, there is also increased burden associated more intensive surveillance colonoscopy.

The present pooled analysis of randomized trials of AI colonoscopy aims at quantifying how the use of AI affects postpolypectomy surveillance and thus affects patient burden.

Materials and Methods

Included Studies

We conducted a pooled analysis of randomized trials of colonoscopy with or without AI tools aimed at improving polyp detection. We identified randomized trials comparing AI and non-AI colonoscopy that had published results by May 2021 through a PubMed search with search terms of "colonoscopy," "randomized controlled trial," "computer aided," and "artificial intelligence," and invited the lead investigators of all trials to participate in the study. Among 11 available trials, 9 agreed to participate and provided data for analyses[10–17] while the remaining 2 did not respond.[18,19]

Among the included trials—5 in China,[10–14] 2 in Italy,[15,16] 1 in Japan,[17] and 1 in the United States[20]—6 were parallel randomized trials (AI-assisted colonoscopy vs standard non-AI colonoscopy) and 3 were randomized tandem trials. In the randomized tandem trials, we used the first-pass data in which each patient was randomized to either AI-assisted colonoscopy or standard non-AI colonoscopy, while we did not include data from the second pass in which the patient had the comparator procedure. All polyps detected during colonoscopy were removed or biopsied and included in the analyses, except for tiny hyperplastic polyp-like lesions in the rectosigmoid colon. Primary endpoints included ADR or adenoma missed rate in all trials.

Specifically, we asked the authors of the included studies to provide the number of patients of the AI group and non-AI group, respectively, in accordance with the 3-class, adenoma-based risk (low risk, intermediate risk, and high risk). The definition of this risk classification is written in the following section. We also obtained information on the colonoscopy indications of each of the participants (screening colonoscopy or non–screening colonoscopy).

Outcome Measures

The primary outcome measure of the present study is the proportion of patients advised to undergo intensive colonoscopy surveillance, defined as colonoscopy surveillance after 3 years based on the following guidelines of 3 major endoscopy societies[6–8]: the American Society of Gastrointestinal Endoscopy, part of the U.S. Multi-Society Task Force on Colorectal Cancer (USMSTF)[7]; the European Society of Gastrointestinal Endoscopy (ESGE)[8]; and the Japan Gastroenterological Endoscopy Society.[6]

All 3 guidelines recommend colonoscopy surveillance after 3 years for patients with 5 or more nonadvanced adenomas, or 1 or more advanced adenomas. The U.S. and Japanese guidelines also recommend 3-year surveillance for patients with 3–4 nonadvanced adenomas, while the European guidelines recommends surveillance for these patients after 10 years.

We evaluated the proportions of patients recommended surveillance after 3 years comparing AI-assisted colonoscopy with non-AI colonoscopy with pooled data from the 9 randomized trials. We classified all patients into 3 categories based on the adenoma-based risk: low risk (no adenoma or 1–2 nonadvanced adenomas), intermediate risk (3–4 nonadvanced adenomas), and high risk (≥5 nonadvanced adenomas or any advanced adenoma)[8]; advanced adenomas were defined as adenomas with villous components, size ≥10 mm, or high-grade dysplasia.[7] Subsequently, we determined the surveillance interval of all patients based on the assigned risk levels. The length of surveillance interval had some range in the USMSTF guidelines according to patients' risks, and thus we simplified the surveillance interval slightly; we assumed 10-year surveillance interval instead of 7- to 10-year interval for the low-risk patients, and 3-year surveillance interval instead of 3- to 5-year interval for the intermediate-risk patients (Table 1). We did not take sessile serrated lesions (SSLs) into consideration in the present study because the contribution of SSLs to future cancer risk is uncertain and all current AI polyp detection software mainly targets adenomas.[8] We also performed subgroup analyses including only colonoscopy done for CRC screening.

AI Tools Used in the Randomized Trials

All AI tools were designed to improve polyp detection. Most tools used in the trials had similar functions in terms of identifying areas suspicious for a polyp with a bounding frame alert on the monitor during colonoscopy. The exception was the study by Gong et al,[14] who used a quality improvement system that showed blind spots to endoscopists instead of showing a bounding frame during colonoscopy. The tools used a similar deep learning architecture with small difference in terms of algorithm development and performance. Preceding pilot studies showed that most of the AI tools for polyp detection provided over 95% sensitivity for polyp recognition in still image and video analysis.[21–23]

Ethics

The investigators of the 9 trials provided aggregated data for pooled analyses. No person-identifiable data were transferred to us for analyses. Data were saved and analyzed in a secure computer at the University of Oslo. The protocols of all included trials have been approved by the local ethics committees and were registered in clinical trial registries. All participating patients provided informed consent in the context of each trial. No new consent was obtained for the present pooled analysis. The present study was conducted in accordance with the Declaration of Helsinki.

What You Need to Know

Background

It is unknown how artificial intelligence tools aimed at improving polyp detection changes colonoscopy surveillance after polyp removal.

Findings

The use of artificial intelligence during colonoscopy increased the proportion of patients requiring intensive colonoscopy surveillance by approximately 35% in the United States and 20% in Europe (absolute increases of 2.9% and 1.3%, respectively).

Implications for patient care

While this may contribute to improved cancer prevention, it significantly adds patient burden and healthcare costs.

Statistical Analysis

Because all 9 included trials adopted similar study design, inclusion and exclusion criteria, intervention, and outcome measures, we applied fixed effect models for the pooled analyses. Absolute difference of the proportion was expressed with 95% CI. We performed pairwise comparisons of AI-assisted colonoscopy vs non-AI colonoscopy using a Mantel-Haenszel fixed effects model and reported relative risks with 95% CIs and forest plots. As a sensitivity analysis, we also estimated values based on a random effects model with the restricted maximum likelihood approach. Statistical heterogeneity among studies was quantified using the I^2 statistic. We used Stata SE 17.0 (StataCorp, College Station, TX) for all statistical analyses.

Table 1. Simplified Guideline Recommendations on Surveillance Interval, in Accordance With the Detected Adenoma-Based Risk Stratification

	USMSTF	ESGE	JGES
Low risk (0–2 nonadvanced adenomas)	10 y	10 y	5 y
Intermediate risk (3–4 nonadvanced adenomas)	3 y	10 y	3 y
High risk (5 nonadvanced adenomas or any advanced adenomas)	3 y	3 y	3 y

ESGE, European Society of Gastrointestinal Endoscopy; JGES, Japan Gastroenterological Endoscopy Society; USMSTF, US Multi-Society Task Force on Colorectal Cancer.

Results

Overview of the Pooled Data

A total of 5796 patients (51% male, mean 53 years of age) were included; 2894 underwent AI-assisted colonoscopy and 2902 underwent standard colonoscopy. Among them, 1299 were done as a primary screening examination. Higher ADRs in the AI-assisted colonoscopy groups were observed in all included trials. Demographic data and ADRs for AI and non-AI colonoscopies from the 9 included trials are presented in Table 2.

Polyp Risk Stratification

Among the 2894 patients who underwent AI-assisted colonoscopy, 2565 (88.7%) were classified as being in the low-risk group, 114 (3.9%) in the intermediate-risk group, and 215 (7.4%) in the high-risk group (Table 3). Among the 2902 patients who underwent non-AI colonoscopy, 2658 (91.6%) were classified as being in the low-risk group, 67 (2.3%) in the intermediate-risk group, and 177 (6.1%) in the high-risk group.

A similar trend was confirmed in which screening was the indication for colonoscopy; in the AI-assisted colonoscopy group (666 patients), 89.2%, 4.2%, and 6.6% of patients were classified as low, intermediate, and high risk, respectively. In the non-AI colonoscopy group (633 patients), 91.9%, 2.1%, and 6.0% of patients were classified as low, intermediate, and high risk, respectively.

Surveillance Recommendation

When following the U.S. and Japanese guidelines, the proportion of patients recommended intensive surveillance increased from 8.4% (95% CI, 7.4% to 9.5%) in the non-AI group to 11.3% (95% CI, 10.2% to 12.6%) in the AI group, with an absolute difference of 2.9% (95% CI, 1.4% to 4.4%) and a risk ratio of 1.35 (95% CI, 1.16 to 1.57; $I^2 = 29\%$) (Table 4). When following the European guidelines, it increased from 6.1% (95% CI, 5.3 to 7.0%) to 7.4% (95% CI, 6.5% to 8.4%), with an absolute difference of 1.3% (95% CI, 0.01% to 2.6%) and a risk ratio of 1.22 (95% CI, 1.01 to 1.47; $I^2 = 21\%$).

Among those undergoing colonoscopy for CRC screening, the proportion of patients recommended intensive surveillance increased from 8.1% (95% CI, 6.1% to 10.5%) to 10.8% (95% CI, 8.6% to 13.4%), with an absolute difference of 2.7% (95% CI, −0.5% to 5.9%) and a risk ratio of 1.32 (95% CI, 0.95 to 1.84; $I^2 = 0\%$), in accordance with the U.S. and Japanese guidelines. On the other hand, it increased from 6.0% (95% CI, 4.3% to 8.1%) to 6.6% (95% CI, 4.8% to 8.8%), with an absolute difference of 0.6% (95% CI, −2.0% to 3.2%) and a risk ratio of 1.09 (95% CI, 0.72 to 1.64; $I^2 = 0\%$), in accordance with the European guidelines. Forest plots of the analyzed data are shown in Figure 1 and Supplementary Figures 1 and 2.

The number needed to scope with AI-assisted colonoscopy to detect 1 more patient who is recommended intensive surveillance as compared with non-AI colonoscopy was 37 when we referred to the American and Japanese guidelines.

Proportion of Low-Risk Patients

When compared with standard colonoscopy, there was a slight reduction in patients being in the low-risk group after AI-assisted colonoscopy, from 91.6% (95% CI, 90.5% to 92.6%) to 88.6% (95% CI, 87.4% to 89.8%). This represents an absolute reduction of −3.0% (95% CI, −4.5% to −1.5%), and a risk ratio of 0.97 (95% CI, 0.95 to 0.98; $I^2 = 5\%$) according to the Mantel-Haenszel fixed effects model. For patients who underwent screening colonoscopy, the absolute reduction was −2.7% (95% CI, −5.9% to −0.5%), for a relative risk of 0.97 (95% CI, 0.94 to 1.01; $I^2 = 13\%$) (see Figure 2).

Heterogeneity of the Analyzed Data and Application of the Random Effects Model

Heterogeneity among studies was moderate based on the fixed effects model ($I^2 = 29\%$ when following the U.S. and Japanese guidelines; 21% when following the European guidelines). When applying a random effects model, the results were similar (see Supplementary Figure 3).

Sensitivity Analysis Focused on Computer-Aided Detection for Polyps

We also conducted a sensitivity analysis focused on computer-aided detection for polyps by excluding Gong et al's study.[14] The analysis showed a slight decrease in the proportion of patients for whom intensive surveillance is recommended, with the risk ratio decreasing from 1.35 to 1.34. On the other hand, the heterogeneity of the included studies was reduced from 29% to 7% (Supplementary Figure 4).

Discussion

According to our pooled analysis, utilization of AI tools aimed at improving polyp detection results in a 35% increase (ranging from a 16% increase to as high as 57% increase) in the proportion of patients allocated to an intensive postpolypectomy surveillance in the United States and Japan (absolute increase of 2.9%). This is primarily because of the reallocation of patients into intermediate- or high-risk categories. Such AI-driven enrichment of the higher-risk category is likely to be explained by the AI-related increase in the adenomas per colonoscopy that has been described in these AI-based trials.[9,24]

The main clinical relevance of such AI-related change in risk stratification of adenoma-bearing patients is that patients with high-risk adenomas are known to be at a

Table 2. Overview of the Pooled Trial Data

Publication	Design	Country	Number of Patients			Gender		Mean Age (Years)	Indication of Colonoscopy		ADR (%)[a]		AI System
			Total	AI-Assisted Colonoscopy	Standard Colonoscopy	Men	Women		Screening	Nonscreening	AI-Assisted Colonoscopy	Standard Colonoscopy	
Wang et al, 2019[12]	Parallel RCT	China	1058	522	536	512	546	50	84	974	29	20	EndoScreener
Wang et al, 2020[10]	Parallel RCT	China	962	484	478	495	467	49	158	804	34	28	EndoScreener
Liu et al, 2020[13]	Parallel RCT	China	790	393	397	374	416	48	182	608	29	21	EndoScreener
Gong et al, 2020[14]	Parallel RCT	China	704	355	349	345	359	50	123	581	16	8	ENDOANGEL
Repici et al, 2020[16]	Parallel RCT	Italy	684	341	343	336	348	61	152	532	55	40	GI GENIUS
Repici et al, 2022[15]	Parallel RCT	Italy	660	330	330	330	330	62	192	468	53	45	GI GENIUS
Wang et al, 2020[11]	Tandem RCT[b]	China	369	184	185	179	190	47	113	256	35	26	EndoScreener
Kamba et al, 2021[17]	Tandem RCT[b]	Japan	346	172	174	265	81	61	162	184	65	54	YOLOv3
Glissen Brown et al, 2022[20]	Tandem RCT[b]	United States	223	113	110	122	101	61	133	90	50	44	EndoScreener

ADR, adenoma detection rate; AI, artificial intelligence; RCT, randomized controlled trial.
[a]The proportion of patients who had adenomas detected during colonoscopy.
[b]ADRs of the tandem colonoscopy trials were calculated based on the first pass of colonoscopy.

Table 3. Patients Number for Each Risk Category Based on the Detected Adenomas

Indication	AI or Non-AI	Risk Based on Detected Adenomas	Number of Patients	Percentage
All	AI-assisted colonoscopy (n = 2894)	High risk	215	7.4
		Intermediate risk	114	3.9
		Low risk	2565	88.7
	Standard colonoscopy (n = 2902)	High risk	177	6.1
		Intermediate risk	67	2.3
		Low risk	2658	91.6
Screening colonoscopy	AI-assisted colonoscopy (n = 666)	High risk	44	6.6
		Intermediate risk	28	4.2
		Low risk	594	89.2
	Standard colonoscopy (n = 633)	High risk	38	6.0
		Intermediate risk	13	2.1
		Low risk	582	91.9

Low risk: no adenoma or 1–2 nonadvanced adenomas; intermediate risk: 3–4 nonadvanced adenomas; high risk: 5 nonadvanced adenomas or any advanced adenomas.
AI, artificial intelligence.

higher risk of developing metachronous cancer and dying from it. In detail, CRC incidence and mortality are nearly 2 times more likely in patients with high-risk adenomas as compared with patients with no adenomas, and more than 2 times more likely than patients with low-risk adenomas.[25] When considering the proven efficacy of an intensive endoscopic surveillance in mitigating such excess CRC risk, the 35% increase in 3-year surveillance interval colonoscopies could be a relevant mechanism for the additional CRC prevention expected on the basis of the AI-related ADR increase. In other words, some of these AI-upshifted high-risk adenoma patients would have been erroneously allocated in the low-risk adenoma group by standard colonoscopy.

By increasing the proportion of patients who fall into the intermediate- and high-risk categories by approximately 35% (which may range between 16% and 57%), it could be argued that the financial costs of the additional AI-related intensive surveillance represent a possible drawback of AI implementation in CRC screening programs. Such excess surveillance may not result in a higher CRC prevention, which is well known as the overdiagnosis bias. Such bias could mainly apply to the intermediate-risk category that is represented by those with 3–4 nonadvanced adenomas.[8] While previous studies based on metachronous advanced adenomas showed an excess risk for this intermediate category, more recent studies with metachronous CRC incidence or mortality as endpoints tended to exclude such excess risk.[8] Of note, the consequences of potential AI-related overdiagnosis depend on not only the baseline risk category, but also the adopted guidelines. For instance, the intermediate-risk category triggers an intensive surveillance according to the U.S. and Japanese guidelines but not according to the European guidelines. This results in a substantial difference in the excess of AI-related intensive surveillance that is roughly halved when passing from the U.S. and Japanese guidelines to the European guidelines. Thus, the consequences of possible overdiagnosis may be mitigated by more conservative surveillance recommendations, especially in those health systems that are already experiencing suboptimal capacity with the population-based CRC screening programs in the first place.

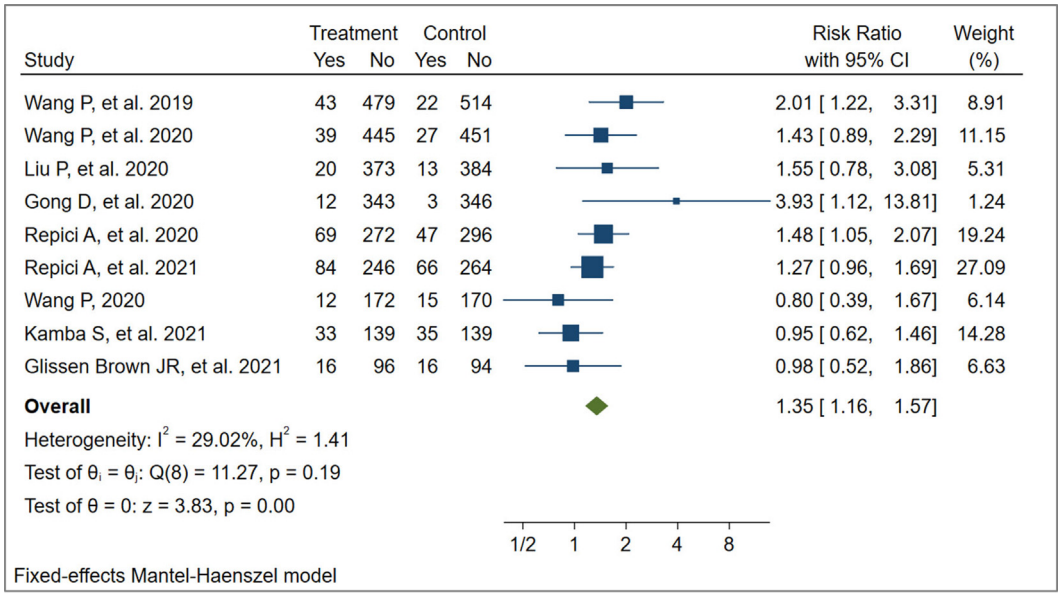
In a recent cost-effectiveness modelling analysis on postpolypectomy surveillance estimates with the cost of a surveillance colonoscopy at approximately $1000, its cost-effectiveness was less than the recommended $100,000 per quality-adjusted life-years, irrespective of the low vs high intensity of surveillance.[26] Thus, assuming a 3% absolute increase of the intensive postpolypectomy surveillance as shown by our analysis for the U.S. scenario, and the need of approximately 2 more colonoscopies in the next 10 years, we expect an increase in cost of 3% × 2 × $1000 that equals an undiscounted $60 per screening examination. According to a recently published microsimulation study, the cost for AI per screening examination is $20. Thus, the cost of additional surveillance would become the main determinant of the cost of AI in colonoscopy. Of course, this is not cost-saving intervention in the short term. On the other hand, a recently published microsimulation study suggests that CRC prevention effect due to the increased ADR with AI could reduce this increased short-term cost.[27]

The present study measures the actual proportion of intensive surveillance colonoscopies, but it is not a microsimulation study that estimates the number of surveillance colonoscopies in a specific period. In fact,

Table 4. Relative Risk on the Intensive Surveillance Recommendation (ie, 3-Year Interval), Comparing Artificial Intelligence–Assisted Colonoscopy With Standard Colonoscopy

	USMSTF[7]		ESGE[8]		JGES[6]	
Indication	Relative Risk	95% CI	Relative Risk	95%CI	Relative Risk	95%CI
All	1.35	1.16–1.57	1.22	1.01–1.47	1.35	1.16–1.57
Screening colonoscopy	1.32	0.95–1.84	1.09	0.72–1.64	1.32	0.95–1.84

The values were calculated based on the American, European, and Japanese guidelines.
CI, confidence interval; ESGE, European Society of Gastrointestinal Endoscopy; JGES, Japan Gastroenterological Endoscopy Society; USMSTF, U.S. Multi-Society Task Force on Colorectal Cancer.



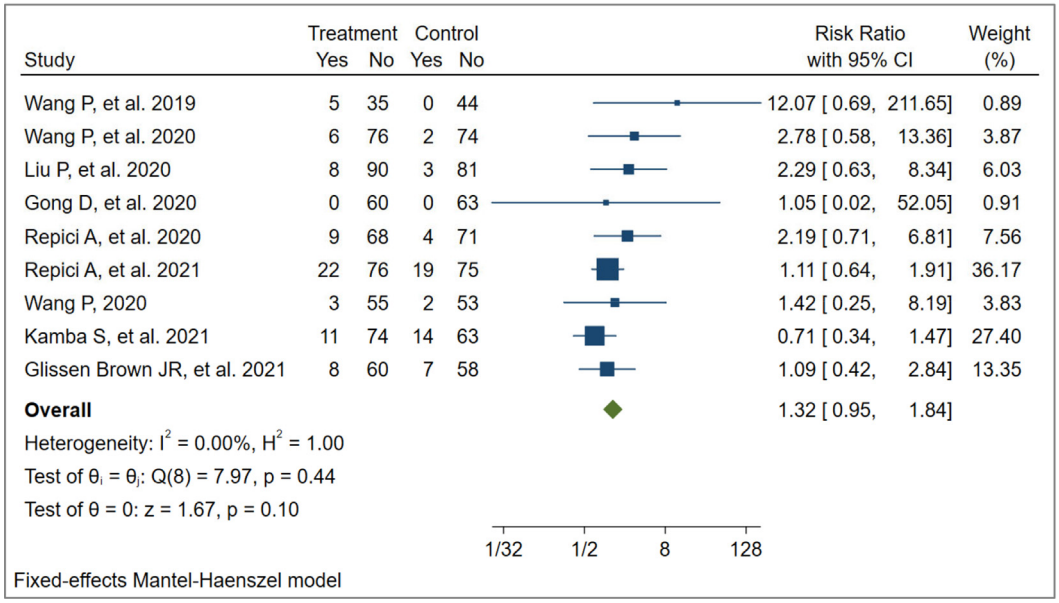


Figure 1. (Top) Relative risk on the intensive surveillance recommendation according to the U.S. guidelines, comparing AI-assisted colonoscopy with standard colonoscopy (all indication; the fixed effect model). (Bottom) Relative risk on the intensive surveillance recommendation according to the U.S. guidelines, comparing AI-assisted colonoscopy with standard colonoscopy (only screening colonoscopy; the fixed effect model).

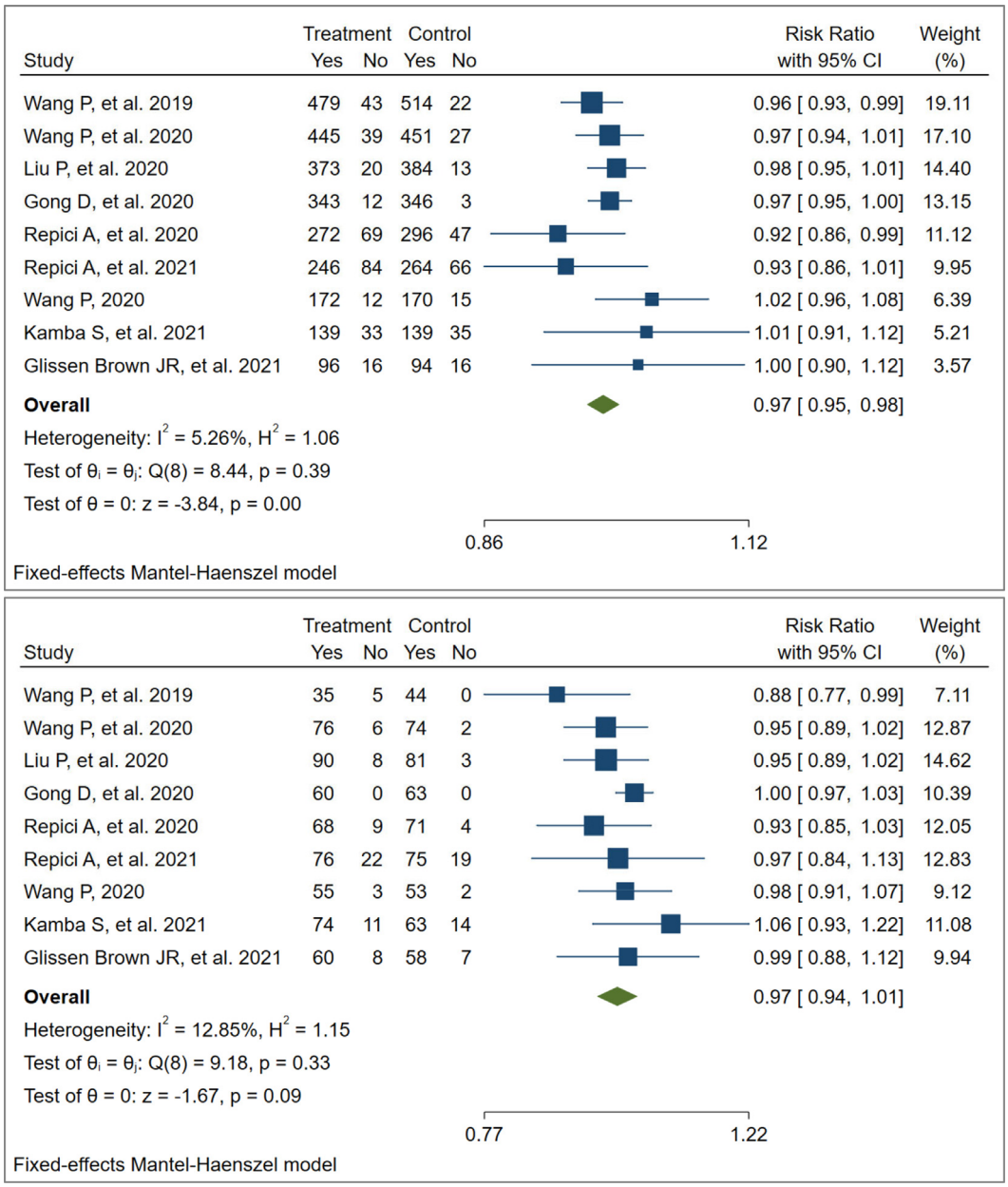


Figure 2. (Top) Relative risk on being classified into the low-risk group, comparing AI-assisted colonoscopy with standard colonoscopy (all indication). (Bottom) Relative risk on being classified into the low-risk group, comparing AI-assisted colonoscopy with standard colonoscopy (only screening colonoscopy).

increased burden of the patients and society may be assessed more properly in a microsimulation study with the number of surveillance colonoscopies. We recently conducted another simulation study, which showed that the number of surveillance colonoscopies increased from 4123 to 4706 per 100,000 in the first 10 years after introduction of AI in colonoscopy.[27] However, results of microsimulation modeling are always subject to considerable uncertainty because of many assumptions. Therefore, we conducted this real-world data-driven study to provide more reliable information about the burden of patients in the short term.

As far as we know, this is the first pooled data analysis on the impact of a new technology for polyp detection on the interval of surveillance. Thus, our analysis may be used as reference standard for future studies addressing the same surveillance endpoint for non-AI technologies, such as devices to increase mucosal exposure or chromoendoscopy series. Furthermore, the present data contribute to robust estimate of cost-effectiveness of AI in colonoscopy, which is a crucially important topic given the growing use of AI in clinical practice.[5]

The main limitation of our study is the lack of inclusion of SSLs in our pooled cohorts. There are 3 reasons

for this: (1) because we thought it to be important to simplify the study outcome measures, (2) because the prevalence of SSLs in the included studies was small, and (3) because there is ongoing discussion on the risk of SSLs for metachronous CRC. The USMSTF emphasizes earlier surveillance of patients with SSLs, while the ESGE guidelines do not recommend any surveillance for those with <10-mm serrated polyps. Nevertheless, given the uncertain performance of AI in increasing the detection of SSLs,[24] inclusion of the SSLs may not have very much influenced the study outcomes. Second, most of the included trials were performed in highly controlled research settings, preventing its generalization to community-based centers. Third, we assumed a complete patient compliance with the recommended surveillance intervals. However, both under- and overuse of endoscopic surveillance has been shown.[28] Fourth, the length of surveillance intervals has some range in the USMSTF guidelines according to patients' risks, and thus we simplified the surveillance interval slightly; we assumed 10-year surveillance interval instead of 7- to 10-year interval for the low-risk patients, and 3-year surveillance interval instead of 3- to 5-year interval for the intermediate-risk patients. Therefore, our interpretation that the use of AI increases the burden of surveillance colonoscopy may be overestimated. Fifth, surveillance guidelines of each region were developed and fine-tuned in accordance with region-specific background information such as capacity of colonoscopy, disease rate, and health economic consideration. In this regard, superimposing surveillance guidelines of a specific region over the population of another region might provide less relevant results. Sixth, we did not do systematic review to identify the randomized studies to be analyzed, which might lead to selection bias.

Furthermore, to better understand the difference of AI benefits between the low detectors and high detectors, we should classify participating endoscopists into those with lower ADR in the screening colonoscopy and those with higher ADR in the screening colonoscopy. Unfortunately, there has been no randomized controlled trial that has focused only on the benefits of using AI in screening colonoscopy and provided subgroup analysis data according to the endoscopists' expertise. On the other hand, a recently published interesting comparison between 2 Italian randomized controlled trials (one trial included only expert endoscopists, while the other trial included only nonexperts)[15] showed that the use of AI, but not the level of endoscopist experience, was associated with increased ADR. Thus, similar benefit and potential burden may be expected from the use of AI for polyp detection even between the different groups of endoscopists.

In conclusion, our study showed an impact by AI on baseline risk stratification, which shifted a considerable proportion of patients to higher-risk categories with little influence on the proportion of patients in low-risk categories. This in turn prompts a more intensive post-polypectomy surveillance that may lead to a more effective cancer prevention. Surveillance strategies should take such an increase into account balancing between higher efficacy on the one hand and endoscopy capacity and risk of overdiagnosis on the other hand. Large-scale population-based trials with long-term follow-up will bring clear answers to these important questions.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at http://doi.org/10.1016/j.cgh.2022.08.022.

Reference:

1. Kaminski MF, Wieszczy P, Rupinski M, et al. Increased Rate of adenoma detection associates with reduced risk of colorectal cancer and death. Gastroenterology 2017; 153:98–105.
2. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014; 370:1298–1306.
3. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010;362:1795–1803.
4. Kalager M, Wieszczy P, Lansdorp-Vogelaar I, et al. Overdiagnosis in colorectal cancer screening: time to acknowledge a blind spot. Gastroenterology 2018;155:592–595.
5. Mori Y, Bretthauer M, Kalager M. Hopes and hypes for artificial intelligence in colorectal cancer screening. Gastroenterology 2021;161:774–777.
6. Saito Y, Oka S, Kawamura T, et al. Colonoscopy screening and surveillance guidelines. Dig Endosc 2021;33:486–519.
7. Gupta S, Lieberman D, Anderson JC, et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2020;158:1131–1153.e5.
8. Hassan C, Antonelli G, Dumonceau JM, et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) guideline - update 2020. Endoscopy 2020; 52:687–700.
9. Hassan C, Spadaccini M, Iannone A, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021;93:77–85.e6.
10. Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020;5:343–351.
11. Wang P, Liu P, Glissen Brown JR, et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. Gastroenterology 2020;159:1252–1261.e5.
12. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and

adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–1819.
13. Liu P, Wang P, Glissen Brown JR, et al. The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study. Therap Adv Gastroenterol 2020;13:1756284820979165.
14. 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;5:352–561.
15. Repici A, Spadaccini M, Antonelli G, et al. Artificial intelligence and colonoscopy experience: lessons from 2 randomised trials. Gut 2022;71:757–765.
16. Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020;159:512–520.e7.
17. Kamba S, Tamai N, Saitoh I, et al. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. J Gastroenterol 2021; 56:746–757.
18. 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 2020;91:415–424.e4.
19. 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;26:13–19.
20. Glissen Brown JR, Mansour NM, Wang P, et al. Deep learning computer-aided polyp detection reduces adenoma miss rate: a U.S. multi-center randomized tandem colonoscopy study (CADeT-CS Trial). Clin Gastroenterol Hepatol 2022;20:1499–1507.e4.
21. Misawa M, Kudo SE, Mori Y, et al. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). Gastrointest Endosc 2022;93:960–967.e3.
22. 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.
23. Wang P, Xiao X, Glissen Brown JR, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nature Biomedical Engineering 2018;2:741–748.
24. Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy 2021;53:277–284.
25. Duvvuri A, Chandrasekar VT, Srinivasan S, et al. Risk of colorectal cancer and cancer related mortality after detection of low-risk or high-risk adenomas, compared with no adenoma, at index colonoscopy: a systematic review and meta-analysis. Gastroenterology 2021;160:1986–1996.e3.
26. Meester RGS, Lansdorp-Vogelaar I, Winawer SJ, et al. High-intensity versus low-intensity surveillance for patients with colorectal adenomas: a cost-effectiveness analysis. Ann Intern Med 2019;171:612–622.
27. Areia M, Mori Y, Correale L, et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. Lancet Dig Health 2022;4:e436–e444.
28. van Heijningen EM, Lansdorp-Vogelaar I, Steyerberg EW, et al. Adherence to surveillance guidelines after removal of colorectal adenomas: a large, community-based study. Gut 2015; 64:1584–1592.

Correspondence
Address correspondence to: Yuichi Mori, MD, PhD, Clinical Effectiveness Research Group, University of Oslo, Gaustad Sykehus, Bygg 20, Sognsvannsveien 21, 0372 Oslo, Norway. e-mail: yuichi.mori@medisin.uio.no.

Acknowledgments
Study data, analytic methods, and study materials can be available upon request to the corresponding author. Conclusion of the agreement between the data provider and recipients, and submission of the study protocol is the prerequisite for the data sharing.

CRediT Authorship Contributions

Yuichi Mori (Conceptualization: Lead; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Project administration: Equal; Resources: Equal; Software: Equal; Supervision: Equal; Validation: Equal; Visualization: Equal; Writing – original draft: Equal)

Pu Wang (Conceptualization: Lead; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Project administration: Equal; Resources: Equal; Software: Equal; Supervision: Equal; Validation: Equal; Visualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Magnus Løberg (Formal analysis: Lead; Investigation: Equal; Methodology: Equal; Writing – review & editing: Equal)

Masashi Misawa (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Alessandro Repici (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Marco Spadaccini (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Loredana Correale (Formal analysis: Equal)

Giulio Antonelli (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Honggang Yu (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Dexin Gong (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Misaki Ishiyama (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Shin-ei Kudo (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Shunsuke Kamba (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Kazuki Sumiyama (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Yutaka Saito (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Haruo Nishino (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Peixi Liu (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Jeremy Brown (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Nabil Mansour (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Seth Gross (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Mette Kalager (Methodology: Equal; Writing – review & editing: Equal)

Michael Brethauer (Investigation: Equal; Methodology: Equal; Writing – review & editing: Equal)

Douglas Rex (Investigation: Equal; Methodology: Equal; Writing – review & editing: Equal)

Prateek Sharma (Investigation: Equal; Methodology: Equal; Writing – review & editing: Equal)

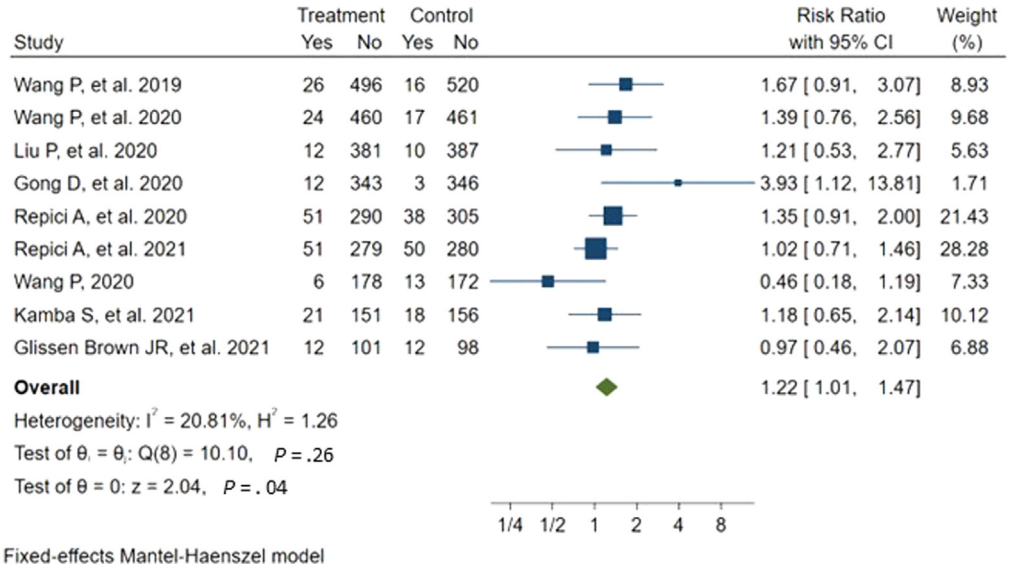
Tyler Berzin (Data curation: Equal; Investigation: Equal; Writing – review & editing: Equal)

Cesare Hassan (Conceptualization: Lead; Data curation: Equal; Investigation: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

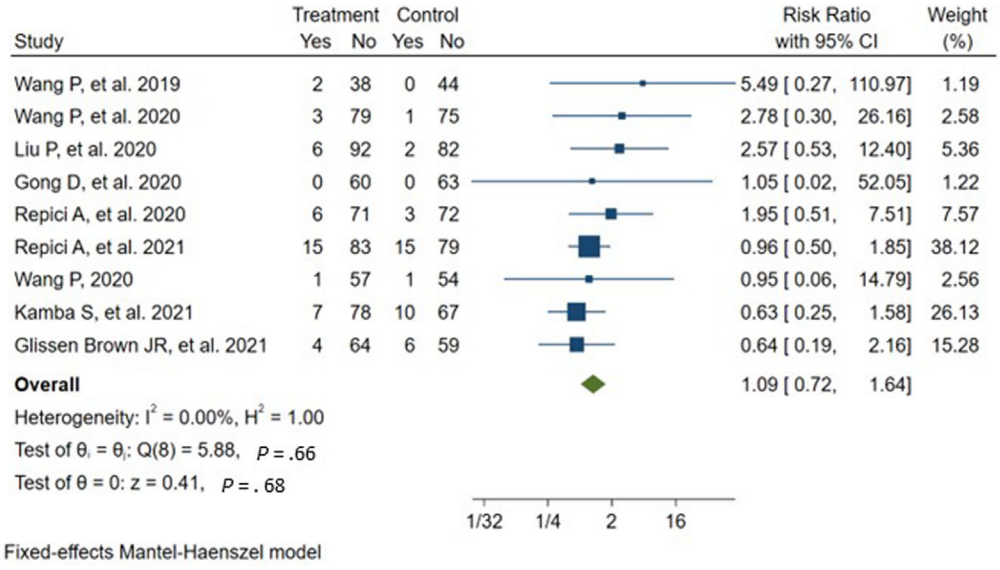
Conflicts of Interest
These authors disclose the following: Yuichi Mori and Masashi Misawa have served as consultants and received equipment on loan from Olympus; and

have ownership interest in Cybernet System. Alessandro Repici has served as a consultant for Fujifilm and Cosmo Pharmaceuticals; received research grant support from Fujifilm and Boston Scientific; has served on the advisory board for Medtronic; and has received speaker fees from Medtronic and Boston Scientific. Michael Bretthauer has served as a consultant for Cybernet System. Seth A. Gross has served as a consultant for Olympus, Cook, Cook, Pentax, Ambu, and Iterative Scopes; and served on the advisory board for Docbot. Douglas K. Rex has an ownership interest in Satisfai Health; and has served as a consultant for Olympus. Prateek Sharma has served as a consultant for Medtronic, Olympus, Boston Scientific, Fujifilm, Salix Pharmaceuticals, and Lumendi; and received research grant support from Ironwood, Erbe, Docbot, Cosmo Pharmaceuticals, and CDx Labs. Tyler M. Berzin has served as a consultant for Medtronic, Boston Scientific, Wision AI, and Magnetiq AI; and served on the advisory board for Docbot AI. Cesare Hassan has served as a consultant for Medtronic, Fujifilm, and Pentax; and received equipment on loan from Medtronic and Fujifilm. All remaining authors disclose no conflicts.

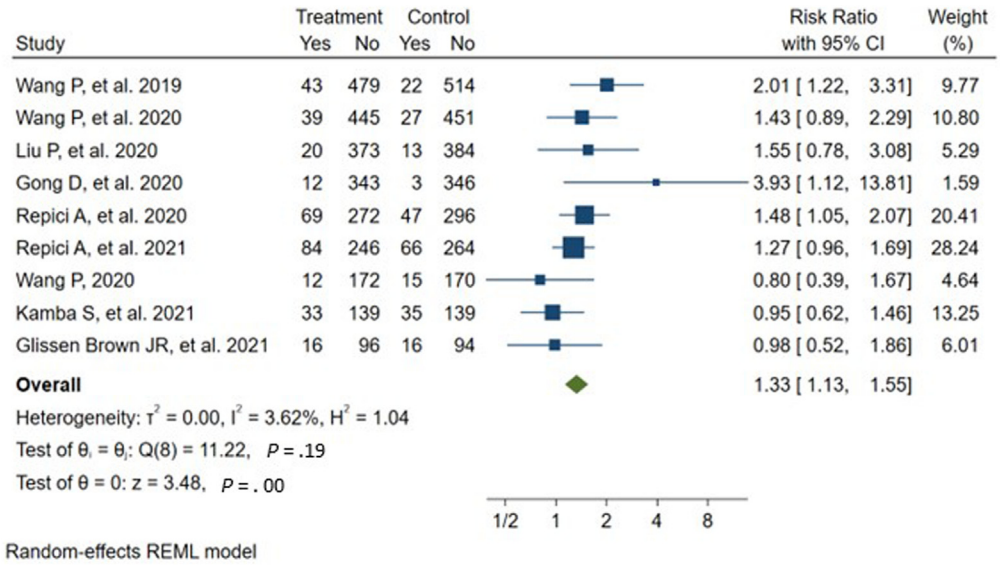
Funding
Yuichi Mori was supported by the European Commission (Horizon2020 MSCA-IF No. 101026196) and the Japan Society for the Promotion of Science (No. 19KK0421). Shunsuke Kamba was supported by the Japan Agency for Medical Research and Development (18ck0106272h0002).



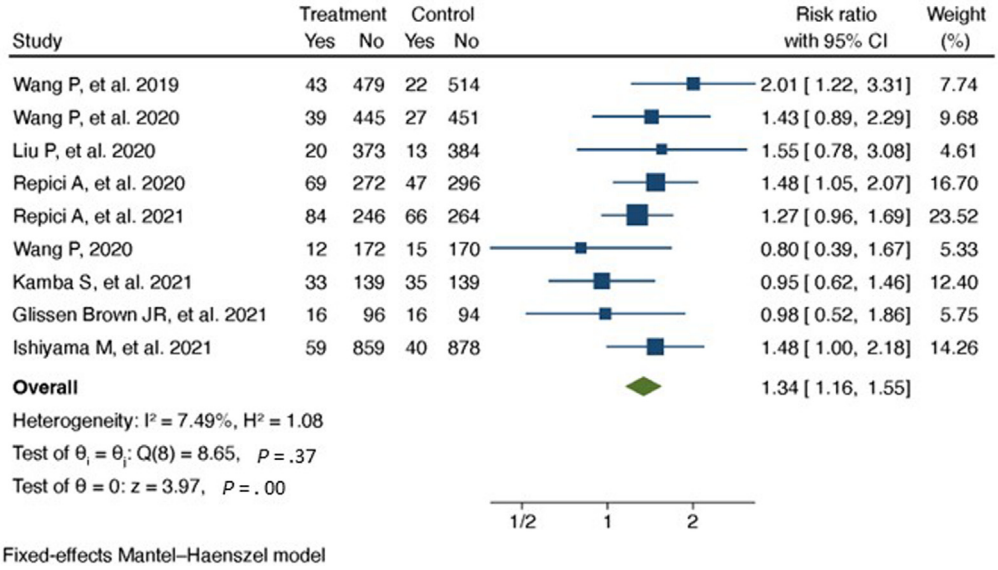
Supplementary Figure 1. Relative risk on the intensive surveillance recommendation according to the European guideline, comparing artificial intelligence–assisted colonoscopy with standard colonoscopy (all indication; the fixed effect model).



Supplementary Figure 2. Relative risk on the intensive surveillance recommendation according to the European guideline, comparing artificial intelligence–assisted colonoscopy with standard colonoscopy (only screening colonoscopy; the fixed effect model).



Supplementary Figure 3. Relative risk on the intensive surveillance recommendation according to the U.S. guideline, comparing artificial intelligence–assisted colonoscopy with standard colonoscopy (all indication; the random effect model). REML, restricted maximum likelihood.



Supplementary Figure 4. A sensitivity analysis focused on computer-aided detection for polyps by excluding Gong's study. Relative risk on the intensive surveillance recommendation according to the U.S. guideline, comparing artificial intelligence–assisted colonoscopy with standard colonoscopy (all indication; the fixed effect model).