Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study

Krzysztof Budzyń, Marcin Romańczyk, Diana Kitala, Paweł Kołodziej, Marek Bugajski, Hans O Adami, Johannes Blom, Marek Buszkiewicz, Natalie Halvorsen, Cesare Hassan, Tomasz Romańczyk, Øyvind Holme, Krzysztof Jarus, Shona Fielding, Melina Kunar, Maria Pellise, Nastazja Pilonis, Michał Filip Kamiński, Mette Kalager, Michael Bretthauer, Yuichi Mori

Lancet Gastroenterol Hepatol 2025; 10: 896–903

Published Online August 12, 2025 https://doi.org/10.1016/S2468-1253(25)00133-5

See Comment page 872

Department of Gastroenterology, Faculty of Medicine, Academy of Silesia, Katowice, Poland (K Budzyń MD, M Romańczyk MD, T Romańczyk MD); **Endoterapia, H-T Centrum Medyczne, Tychy, Poland** (K Budzyń, M Romańczyk, T Romańczyk); **Agency of Medical Research, Warsaw, Poland** (D Kitala PhD); **Endomed, Radom, Poland** (P Kołodziej MD); **Polish Foundation of Gastroenterology, Warsaw, Poland** (M Bugajski MD); **Clinical Effectiveness Research Group, University of Oslo and Oslo University Hospital, Oslo, Norway** (H O Adami MD, N Halvorsen MD, N Pilonis MD, Prof M F Kamiński MD, Prof M Kalager MD, Prof M Bretthauer MD, Prof Y Mori MD); **Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden** (H O Adami); **Department of Surgery, Södersjukhuset, Stockholm, Sweden** (J Blom MD); **Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden** (J Blom); **Buszkiewicz Medical Center, Gorzow Wielkopolski, Poland** (M Buszkiewicz MD); **Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy** (Prof C Hassan MD); **Endoscopy Unit, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy** (Prof C Hassan); **Institute of Health and Society, University of Oslo, Oslo, Norway** (Prof Ø Holme MD); **Research**

Summary

Background It is not known if continuous exposure to artificial intelligence (AI) changes endoscopists' behaviour when conducting colonoscopy. We assessed how endoscopists who regularly used AI performed colonoscopy when AI was not in use.

Methods We conducted a retrospective, observational study at four endoscopy centres in Poland taking part in the ACCEPT (Artificial Intelligence in Colonoscopy for Cancer Prevention) trial. These centres introduced AI tools for polyp detection at the end of 2021, after which colonoscopies had been randomly assigned to be conducted with or without AI assistance according to the date of examination. We evaluated the quality of colonoscopy by comparing two different phases: 3 months before and 3 months after AI implementation. We included all diagnostic colonoscopies, excluding those involving intensive anticoagulant use, pregnancy, or a history of colorectal resection or inflammatory bowel disease. The primary outcome was change in adenoma detection rate (ADR) of standard, non-AI assisted colonoscopy before and after AI exposure. Multivariable logistic regression was done to identify independent factors affecting ADR.

Findings Between Sept 8, 2021, and March 9, 2022, 1443 patients underwent non-AI assisted colonoscopy before (n=795) and after (n=648) the introduction of AI (median age 61 years [IQR 45–70], 847 [58·7%] female, 596 [41·3%] male). The ADR of standard colonoscopy decreased significantly from 28·4% (226 of 795) before to 22·4% (145 of 648) after exposure to AI, corresponding with an absolute difference of –6·0% (95% CI –10·5 to –1·6; p=0·0089). In multivariable logistic regression analysis, exposure to AI (odds ratio 0·69 [95% CI 0·53–0·89]), male versus female patient sex (1·78 [1·38–2·30]), and patient age ≥60 years versus <60 years (3·60 [2·74–4·72]) were the independent factors significantly associated with ADR.

Interpretation Continuous exposure to AI might reduce the ADR of standard non-AI assisted colonoscopy, suggesting a negative effect on endoscopist behaviour.

Funding European Commission and Japan Society for the Promotion of Science.

Introduction

Colorectal cancer is a major health-care problem.[1] Colonoscopy enables detection and removal of pre-cancerous lesions (ie, adenomas), leading to prevention of colorectal cancer.[2] Adenoma detection rate (ADR)—the proportion of colonoscopies in which one or more adenomas are detected—is a widely accepted indicator of colonoscopist performance, with a higher ADR associated with a greater cancer prevention effect.[3] However, around a quarter of endoscopists do not achieve the recommended ADR, according to a large-scale, randomised trial.[4] Thus, maintaining a high ADR is considered an important goal for both endoscopists, health-care systems, and researchers.

Computer-assisted polyp detection systems introduced over the last decade might increase ADR regardless of the expertise of the endoscopist; a meta-analysis of 20 randomised trials showed an absolute increase in ADR of 8·1 % with the use of artificial intelligence (AI) during colonoscopy.[5] This AI-driven increase in ADR is expected to improve prevention of colorectal cancer.[6]

Despite its promise, it is not known whether continuous exposure to AI leads to a change in endoscopist performance during standard, non-AI assisted colonoscopy. This question is crucial because the adoption of AI in medicine is spreading rapidly. Psychological studies in non-medical fields suggest that ongoing exposure to AI might change behaviour in different ways: positively, by training clinicians, or negatively, through a deskilling effect, whereby automation use leads to a decay in cognitive skills.[7] To this end, we investigated changes in the quality of standard,

Research in context

Evidence before this study

Many randomised trials show that the use of artificial intelligence (AI) for polyp detection can increase the adenoma detection rate (ADR)—a representative quality indicator of colonoscopy that is associated with colorectal cancer prevention. However, there is little research on how continuous exposure to AI affects endoscopists' behaviour and introduces the risk of deskilling, although deskilling is of concern when implementing automation devices in general. We searched PubMed from database inception to Nov 30, 2024, using the terms "((endoscopists) AND (deskilling)) AND (artificial intelligence)", without applying language restrictions. Although several guidelines and review articles cautioned against the risk of deskilling, there was no relevant original research that investigated the deskilling risk associated with the use of AI in colonoscopy.

Added value of this study

We found that routine exposure to AI in colonoscopy might reduce the ADR of standard, non-AI assisted colonoscopy. To our knowledge, this is the first study that suggests AI exposure might have a negative impact on patient-relevant endpoints in medicine in general.

Implications of all the available evidence

Despite the observational design of our study, with its sensitivity to selection bias and confounding, our results highlight the need for further high-quality research in the area of AI in gastroenterology and beyond, given the rapid adoption of AI in medicine. This emerging area of research is critical, especially considering the medical principle "do no harm", which can be extended beyond patients to include the capability of physicians. Future studies could delve deeper into physician behaviour, examining how AI affects clinical performance and identifying solutions to mitigate the risk of deskilling, including both computational solutions, such as explainable AI, and psychological interventions, such as cognitive forcing techniques.

non-AI assisted colonoscopy before and after exposure to AI in routine practice.

Methods

Study design

We conducted a multicentre, retrospective, observational study at four Polish endoscopy centres. This study is nested in the Artificial Intelligence in Colonoscopy for Cancer Prevention (ACCEPT) project, which investigates the effectiveness of AI tools in colorectal cancer screening programmes; it is registered with the UMIN-ICDS clinical trials registry, UMIN000044748. ACCEPT is a randomised trial in which individuals undergoing screening colonoscopy are randomly assigned to either AI-assisted or standard colonoscopy, depending on the date of randomisation. On AI days, the AI processor is automatically activated, whereas on non-AI days, it is automatically deactivated.

The four Polish centres were Endoterapia, H-T Centrum Medyczne, Tychy; Centre of Gastroenterology, Wodzisław Śląski; Endomed, Radom; and Buszkiewicz Medical Center, Gorzów Wielkopolski. They are referred to as Centres 1–4 for anonymity. The centres introduced AI tools for the detection of polyps towards the end of 2021 (Nov 25, 2021, at Centre 1; Dec 14, 2021, at Centre 2; Dec 7, 2021, at Centre 3; and Nov 26, 2021, at Centre 4), after which individuals were randomly assigned to colonoscopies done either with or without AI assistance according to the date of colonoscopy. We evaluated changes in the quality of all diagnostic, non-AI assisted colonoscopies between Sept 8, 2021, and March 9, 2022, by comparing two different phases: the period approximately 3 months before AI implementation in clinical practice versus the period 3 months after AI implementation in clinical practice. The first 100 colonoscopies per centre after AI introduction were excluded from analysis to allow endoscopists to become familiar with using AI. The institutional review board granted an exemption from review (number ŚIL.KB.1159.2022) for the present study as it was observational and did not use any person-identifiable information. The study was conducted in accordance with the Declaration of Helsinki.

In Poland, participant consent for the ACCEPT study is waived due to the use of a routinely approved, minimally invasive medical device, the study's focus on quality improvement in cancer screening, and its nature as a health services implementation study (ethics approval number 57/2021).[8] As the AI processor was automatically switched on or off, patients undergoing diagnostic endoscopy also had their procedures done with or without AI assistance. Offering these health services to such patients was deemed ethically feasible, as both colonoscopy with and without AI assistance are part of routine clinical practice.

Participants

We excluded data from the colonoscopies performed in patients who had an inability to undergo biopsy or polypectomy because of medication or coagulation disorders, were pregnant, had been referred for the assessment or treatment of known lesions, had a history of bowel resection, or had inflammatory bowel disease, and if the colonoscopy had been performed with a non-high definition colonoscope. Colonoscopies performed by endoscopists who did not conduct a colonoscopy in both study phases were also excluded.

Unit, Sørlandet Hospital, Kristiansand, Norway (Prof Ø Holme); **Centre of Gastroenterology, Wodzisław Śląski, Poland** (K Jarus MD); **Frontier Science (Scotland), Kincraig, UK** (S Fielding PhD); **Department of Psychology, University of Warwick, Coventry, UK** (M Kunar PhD); **Department of Gastroenterology, Hospital Clínic Barcelona, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain** (Prof M Pellise MD); **Facultat de Medicina i Ciències de la Salud, Universitat de Barcelona, Barcelona, Spain** (Prof M Pellise); **Department of Oncological Gastroenterology, National Research Institute of Oncology, Warsaw, Poland** (N Pilonis); **Division of General, Endocrine and Transplant Surgery, Medical University of Gdańsk, Gdańsk, Poland** (N Pilonis); **Department of Gastroenterology, Hepatology and Clinical Oncology, Centre of Postgraduate Medical Education, Warsaw, Poland** (Prof M F Kamiński); **Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland** (Prof M F Kamiński); **Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan** (Prof Y Mori)

Correspondence to:
Dr Marcin Romańczyk, Endoterapia, H-T Centrum Medyczne, 43–100 Tychy, Poland
marcin.romanczyk@htcentrum.pl

Procedures

We collected relevant quality indicator data for each colonoscopy, including endoscopist's expertise and specialties (either physicians or surgeons), patients' age and sex, indication for colonoscopy, use of sedation, bowel preparation status (defined as adequate when Boston Bowel Preparation Scale score was 6 or more),[9] completeness of colonoscopy, and clinicopathological characteristics of the detected polyps. Data were collected on a form by reviewing endoscopy descriptions, including data on participant's sex. Data collection was done by KB and PK between April 1, 2022, and Oct 31, 2022.

Colonoscopies were done with high-definition endoscopy processors (EVIS X1 CV 1500, Olympus Medical Systems, Tokyo, Japan) with compatible endoscopes (CF-H190L, CF-HQ190I, CF-HQ1100DI, PCF-H190I; Olympus Medical Systems, Tokyo). The AI system used was ENDO-AID CADe (OIP-1, Olympus Medical Systems, Tokyo). All the colonoscopies were performed as outpatient appointments. Patients were prepared with sodium picosulphate and magnesium citrate, low-dose polyethylene glycol, or high-dose polyethylene glycol. Sedation was administered only on patient request, with no additional clinical or procedural criteria influencing the decision. Blinding the use of AI during colonoscopy was impossible due to the system's design, as the endoscopist could see the marked target on the screen during AI use.

Outcomes

The primary outcome measure was change in ADR with standard, non-AI assisted colonoscopy before and after the endoscopists were exposed to AI in clinical practice. In addition, we measured the change in the mean number of adenomas per colonoscopy (APC) and the mean number of advanced APCs before and after AI introduction.

	Before AI introduction (n=795)	After AI introduction (n=648)	Total (n=1443)	p value
Median age, years (IQR)	62 (46–70)	59 (44–70)	61 (45–70)	0·070
Sex				0·0046
Male	302 (38·0%)	294 (45·4%)	596 (41·3%)	..
Female	493 (62·0%)	354 (54·6%)	847 (58·7%)	..
Sedation	613 (77·1%)	531 (81·9%)	1144 (79·3%)	0·024
Adequate bowel preparation*	768 (96·6%)	627 (96·8%)	1395 (96·7%)	0·87
Incomplete examination	4 (0·5%)	8 (1·2%)	12 (0·8%)	NA†
Indications				0·0090
Alarm symptoms‡	99 (12·5%)	78 (12·0%)	177 (12·3%)	..
Surveillance	119 (15·0%)	111 (17·1%)	230 (15·9%)	..
Positive faecal occult blood test	21 (2·6%)	11 (1·7%)	32 (2·2%)	..
Other§	277 (34·8%)	190 (29·3%)	467 (32·4%)	..
Unknown	279 (35·1%)	258 (39·8%)	537 (37·2%)	..

Data are n (%) unless indicated otherwise. AI=artificial intelligence. NA=not applicable. *Defined by score of at least 6 on Boston Bowel Preparation Scale.[9] †Significance assessment was not done due to too few events. ‡Weight loss, anaemia, gastrointestinal bleeding signs, and tumour seen in CT scan. §Change in bowel habits or diarrhoea.

Table 1: **Characteristics of the patients who underwent standard, non-AI assisted colonoscopies**

Statistical analysis

All data were analysed using Stata version 18. Patient characteristics were described using mean and SD for continuous and normally distributed variables, and median and IQR for skewed variables. Categorical variables were described using frequency and percentage. A p value <0·05 was considered as statistically significant.

ADR was calculated as the proportion of colonoscopy procedures in which one or more adenomas or cancers were detected. We did not consider sessile serrated lesions (SSL) in the ADR calculation, to replicate the value of ADR in the large-scale observational studies that established ADR's association with interval colorectal cancers.[10,11] However, traditional serrated adenomas were treated as adenomas in the calculation of ADR. We included all patients in the ADR calculation, including those with incomplete colonoscopy or inadequate bowel preparation. ADR of standard, non-AI assisted colonoscopy before and after the exposure to AI was compared using a χ^2 test with absolute difference and 95% CI provided. The mean number of APCs and advanced APCs were compared between groups using a *t* test, with mean difference, 95% CI, and p value provided.

Univariable mixed-effects logistic regression[12,13] with a random effect for endoscopist was performed for detection of at least one adenoma as the outcome variable and the following potential predicting variables: patient age and sex, use of sedation, bowel preparation status (Boston Bowel Preparation Scale score), completeness of colonoscopy (intubated caecum *vs* non-intubated caecum), endoscopist specialty (physicians *vs* surgeons), endoscopist experience (years after graduation from medical school), endoscopist sex, centre, and implementation of AI in clinical practice. Variables with a p value <0·05 in the univariable model were included in the multivariable model.

To account for potential imbalances in demographic characteristics, we did stratified analyses by sex and age of the patients and explored heterogenicity (female and male individuals who were younger than 60 years; female and male individuals who were 60 years or older). Regarding age stratification, although recent American guidelines have lowered the screening age to 45 years,[14] we used 60 years as the cutoff in the subgroup analysis, as many studies and risk models favour this threshold due to a marked increase in adenoma prevalence.[15]

In addition, we compared ADR in the two phases (before and after AI implementation) by subgroups of centre, specialties (physicians and surgeons), and endoscopist sex. In each subgroup, the percentage for ADR is given along with the absolute percentage change and 95% CI comparing ADR before and after AI introduction. The ADR for each endoscopist

before and after AI exposure is also presented. As a supplementary analysis, univariable and multivariable mixed-effects regression analyses were done that included patients who underwent AI-assisted colonoscopy during the study. These analyses included AI exposure, patient sex, patient age (<60 *vs* ≥60 years), adequate bowel preparation, sedation, indication for colonoscopy, endoscopist specialty, centre, endoscopist sex, and endoscopist experience.

Role of funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

In the observed period between Sept 8, 2021, and March 9, 2022, after excluding the 100 colonoscopies conducted after the introduction of AI and those done in participants who met prespecified exclusion criteria, a total of 2177 colonoscopies were conducted, including 1443 without AI use and 734 with AI. Our primary analysis focuses on the 1443 patients who underwent standard, non-AI assisted colonoscopies before (n=795) and after (n=648) AI introduction. The colonoscopies were performed by 19 endoscopists who had completed endoscopy training (16 physicians and three general surgeons), with experience of more than 2000 colonoscopies each and mean length of experience of 27·6 years (range 8–39).

Patients who underwent standard, non-AI assisted colonoscopy after AI introduction were slightly younger (median 59 years [IQR [44–70] *vs* 62 years [46–70] before AI introduction), with fewer females (354 [54·6%] of 648 compared with 493 [62·0%] of 795 before AI introduction), and a higher percentage of sedation usage (531 [81·9%] *vs* 613 [77·1%] before AI introduction; table 1). Adequate bowel preparation rates were similar before and after AI introduction (768 [96·6%] *vs* 627 [96·8%]; table 1).

A total of 177 (12·3%) of 1443 patients were referred due to alarm symptoms (weight loss, anaemia, gastrointestinal bleeding, tumour seen in CT), 230 (15·9%) were surveillance colonoscopies, 32 (2·2%) were positive for faecal occult blood test, 467 (32·4%) had other indications (change in bowel habit, diarrhoea), and 537 (37·2%) were with unknown indications. The reasons for referral were broadly similar before and after AI introduction (table 1). The proportion of patients in whom the caecum could not be reached during colonoscopy was four (0·5%) of 795 before AI introduction and eight (1·2%) of 648 after AI introduction.

ADR at standard, non-AI assisted colonoscopies decreased significantly from 28·4% (226 of 795) before AI exposure to 22·4% (145 of 648) after AI exposure, corresponding to an absolute difference of –6·0% (95% CI –10·5 to –1·6, p=0·0089; table 2, figure 1).

	Before AI implementation (n=795)	After AI implementation (n=648)	Difference (95% CI)	p value
ADR	226 (28·4%)	145 (22·4%)	–6·0 (–10·5 to –1·6)	0·0089*
Adenomas per colonoscopy	0·54 (1·23)	0·43 (1·13)	0·11 (–0·01 to 0·24)	0·071†
Advanced adenomas per colonoscopy	0·062 (0·27)	0·063 (0·33)	–0·002 (–0·03 to 0·03)	0·92†

Data are n (%) or mean (SD). AI=artificial intelligence. ADR=adenoma detection rate. *χ^2 test. †t test.

Table 2: **Quality indicators of standard, non-AI assisted colonoscopy before and after AI introduction**

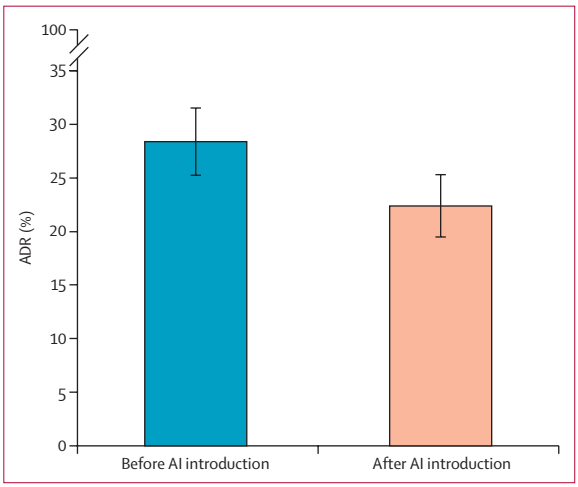

Figure 1: **Change in ADR with standard, non-AI assisted colonoscopy before and after introduction of AI for polyp detection**
The error bars represent 95% CIs. ADR=adenoma detection rate. AI=artificial intelligence.

Mean APC before AI exposure was slightly higher than after AI exposure, although this difference was not statistically significant (0·54 [SD 1·23] *vs* 0·43 [1·13]), yielding a mean difference of 0·11 (95% CI –0·01 to 0·24, p=0·071). The number of APCs in patients with at least one adenoma detected did not change significantly between the groups before and after AI exposure, with a mean of 1·91 (SD 1·64) before versus 1·92 (1·68) after AI exposure, resulting in a difference of –0·01 (95% CI –0·36 to 0·34, p=0·95). Mean advanced APC was similar between the two periods (0·062 [SD 0·27] *vs* 0·063 [0·33], mean difference –0·002 [–0·03 to 0·03]; p=0·92). Mean advanced APC on standard colonoscopy in patients with at least one adenoma detected was 0·22 (SD 0·47) before AI exposure and 0·28 (0·65) after AI exposure, with a mean difference of –0·06 (–0·18 to 0·05, p=0·26). Colorectal cancers were detected in six (0·8%) of 795 colonoscopies before AI exposure and in eight (1·2%) of 648 after AI exposure (p=0·35).

In univariable mixed-effects logistic regression, statistically significant predictors of ADR were exposure to AI, patient's male sex, age 60 years or older, and the presence of alarm symptoms. In multivariable logistic regression analysis, exposure to AI, patient's male sex,

	Adenoma, n/N (%)	Univariable analysis		Multivariable analysis*	
		Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
AI introduction in clinical practice (exposure to AI)					
Before	226/795 (28·4%)	1 (ref)	..	1 (ref)	..
After	145/648 (22·4%)	0·70 (0·55–0·90)	0·0052	0·69 (0·53–0·89)	0·0045
Sex of patient					
Female	191/847 (22·6%)	1 (ref)	..	1 (ref)	..
Male	180/596 (30·2%)	1·51 (1·19–1·93)	0·0008	1·78 (1·38–2·30)	<0·0001
Age group					
<60 years	102/687 (14·8%)	1 (ref)	..	1 (ref)	..
≥60 years	269/756 (35·6%)	3·46 (2·65–4·51)	<0·0001	3·60 (2·74–4·72)	<0·0001
Adequate bowel preparation†					
No	9/48 (18·8%)	1 (ref)	..	..	..
Yes	362/1395 (26·0%)	1·49 (0·70–3·14)	0·30	..	..
Incomplete examination					
No	369/1431 (25·8%)	..	..	..	..
Yes	2/12 (16·7%)‡	..	..	..	..
Sedation					
No	75/299 (25·1%)	1 (ref)	..	..	..
Yes	296/1144 (25·9%)	0·89 (0·63–1·27)	0·53	..	..
Indication					
Alarm symptoms	36/177 (20·3%)	1 (ref)	..	..	..
Surveillance, other, positive FOBT, or unknown	335/1266 (26·5%)	1·67 (1·12–2·51)	0·013	1·36 (0·89–2·08)	0·15
Specialty					
Physician	334/1245 (26·8%)	1 (ref)	..	..	..
Surgeon	37/198 (18·7%)	0·61 (0·35–1·08)	0·088	..	..
Centre					
1	91/354 (25·7%)	1 (ref)	..	..	..
2	66/308 (21·4%)	0·78 (0·50–1·23)	0·29	..	..
3	132/391 (33·8%)	1·37 (0·88–2·13)	0·15	..	..
4	82/390 (21·0%)	0·77 (0·47–1·1.26)	0·29	..	..
Endoscopist sex					
Male	254/1061 (23·9%)	1 (ref)	..	..	..
Female	117/382 (30·6%)	1·14 (0·71–1·85)	0·59	..	..
Endoscopist experience, per year since graduation	NA	0·98 (0·96–1·01)	0·25	..	

AI=artificial intelligence. FOBT=faecal occult blood test. NA=not applicable. *Kept in model if univariable p<0·05. All models include endoscopists as a random effect.[12,13] †Defined by score of at least 6 on Boston Bowel Preparation Scale.[9] ‡Significance assessment was not done due to the small number of events.

Table 3: **Logistic regression analysis to identify factors affecting adenoma detection rate**

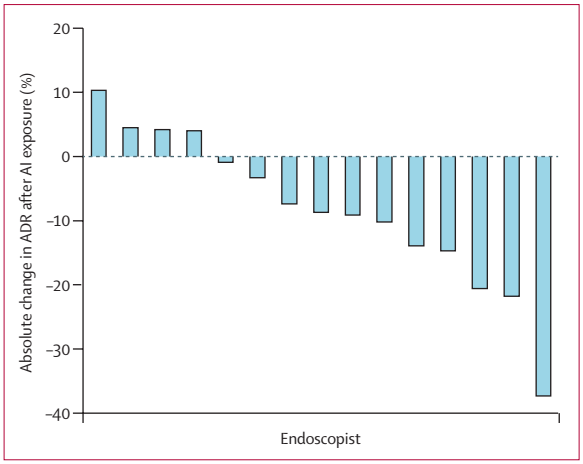

Figure 2: **Endoscopist-level change in ADR with standard, non-AI assisted colonoscopy after introduction of AI in colonoscopy at each centre**
Four endoscopists who did fewer than ten colonoscopies either before or after AI introduction are not shown. ADR=adenoma detection rate. AI=artificial intelligence.

and age 60 years or older were significantly associated with change in ADR, whereas the presence of alarm symptoms was not statistically significant (table 3).

Endoscopist-level analysis showed that most endoscopists had a lower ADR when performing standard colonoscopies after AI exposure than before, whereas only four endoscopists increased their ADR (figure 2; appendix p 2).

See Online for appendix

Among physicians, the absolute change in ADR was –6·1% (95% CI –11·0 to –1·2), from 29·7% (199 of 671) to 23·5% (135 of 574), and among surgeons ADR was –8·3% (–18·9 to –2·4), from 21·8% (27 of 124) to 13·5% (ten of 74; appendix p 1).

When focusing on sex of the endoscopists, male endoscopists (n=14) had an absolute change in their ADR of –2·9% (95% CI –8·0 to 2·3), from 25·2% (148 of 587) to 22·4% (106 of 474), whereas the absolute change in ADR in female endoscopists (n=5) was –15·1% (–24·1 to –6·0), from 37·5% (78 of 208) to 22·4% (39 of 174).

In all centres, ADR for standard, non-AI assisted colonoscopy was reduced after AI exposure, although the magnitude of ADR reduction varied greatly between centres (appendix p 3). The baseline ADR in Centres 1 and 3 (29·4% [58 of 197] in Centre 1 and 39·7% [79 of 199] in Centre 3) was higher than in Centres 2 and 4 before AI exposure (21·7% [43 of 198] in Centre 2 and 22·9% [46 of 201] in Centre 4), and Centres 1 and 3 showed greater absolute reductions in ADR compared with Centres 2 and 4 (Centre 1: –8·4% [95% CI –17·4 to 0·6], Centre 3: –12·1% [–21·4 to 2·8]; appendix p 3).

After AI exposure, the absolute change in the ADR of standard, non-AI assisted colonoscopy in female patients younger than 60 years was –6·7% (95% CI –12·8 to –0·6), and in those 60 years and older it was –8·1% (–16·5 to 0·3; appendix p 4). The absolute change in ADR in male patients younger than 60 years was –8·0% (–16·9 to 0·8) after AI exposure, whereas it was 1·6% (–9·8 to 3·0) in those 60 years and older (appendix p 4).

A comparison was also done on ADR before and after AI exposure across different indications. ADR decreased after AI exposure in all the five different indication groups, with differences ranging between –4·2% (95% CI –11·6 to 3·3) for other indications and –20·8% (–56·2 to 14·7) for a positive faecal occult blood test,

although no changes were statistically significant (appendix p 5).

In the supplementary analysis of patients who did have AI-assisted colonoscopy, the ADR was 25·3% (186 of 734). The multivariable logistic regression analysis in all 2177 colonoscopies (including those that used AI) showed that AI exposure in those without AI-assisted colonoscopy (as shown in the main analysis), female sex of the patient, and patient age younger than 60 years were independent factors that were significantly associated with ADR, when adjusted for endoscopist as a random effect. Compared with colonoscopies before the introduction of AI, use of AI was not significantly associated with a change in ADR (OR 0·80 [95% CI 0·63–1·02]); appendix p 6).

Discussion

Our primary analysis showed continuous exposure to AI for polyp detection reduced the ADR of standard, non-AI assisted colonoscopy from 28·4% to 22·4%, with a 6·0% absolute difference, suggesting a detrimental effect on endoscopist capability. Notably, during the AI introduction phase, colonoscopies involved more male patients than after AI introduction (38·0% *vs* 45·4%, p=0·0046) and greater sedation use (77·1% *vs* 81·9%, p=0·024)—both factors that would typically increase ADR. Yet ADR declined after AI introduction, emphasising the significance of this observed reduction.

Interpretation of these data is challenging. First, our observational study design necessitates cautious assessment because it is vulnerable to selection bias and confounding. To minimise the effect of confounders, we conducted stratified and multivariable analyses including risk factors of adenomas, all of which suggested that continuous exposure to AI reduces ADR. However, robustly designed prospective trials are warranted to generalise these findings.

Another challenge is to understand why the detrimental effect occurred. We assume that continuous exposure to decision support systems such as AI might lead to the natural human tendency to over-rely on their recommendations, leading to clinicians becoming less motivated, less focused, and less responsible when making cognitive decisions without AI assistance.[16] In fact, the European Society of Gastrointestinal Endoscopy advised caution about the risk of such deskilling in their guideline recommendations in 2019.[17] This interpretation was supported in an experimental study that showed reduced eye movements during colonoscopy when using AI for polyp detection, indicating a risk of overdependence.[18] In a similar study that investigated AI for breast cancer detection with mammography, physicians' detection capability decreased significantly when AI support was expected.[19] However, these early-stage experimental studies might require real-world clinical studies to validate findings using patient-important outcomes. To our knowledge, the present study is the first to examine the effect of continuous AI exposure on outcomes such as ADR in medicine.

Our study also gives a unique insight into previously published randomised trials in this area.[20,21] In these trials, AI-assisted colonoscopy led to a higher ADR than non-AI assisted colonoscopy, which was considered the standard of care. However, it is doubtful that non-AI assisted colonoscopy was truly the standard of care: our study suggests that the ADR achieved with non-AI assisted colonoscopy in these randomised trials might be negatively affected by continuous AI exposure, leading to different results to those achieved with everyday colonoscopy. This artificially reduced ADR with non-AI assisted colonoscopies might explain the significant differences in ADR observed in these randomised trials. Our results support this hypothesis. Although AI-assisted colonoscopy appeared to increase ADR from 22·4% to 25·3% within the same timeframe (appendix p 6), the actual trend from a chronological perspective suggests a decline in ADR for non-AI-assisted colonoscopy after AI implementation, decreasing from 28·4% to 22·4%. This decline indicates that the increased ADR with AI-assisted colonoscopy shown in many randomised trials might, at least in part, be due to a reduction in ADR among non-AI procedures.

Our subgroup analyses also highlight interesting findings. Centres with a relatively modest baseline ADR (21·7–22·9%) showed a less pronounced deskilling effect compared with centres with a higher baseline ADR (29·4–39·7%), suggesting that lower baseline colonoscopy quality leaves less room for further ADR reduction. A large decrease in ADR in the centres with high ADRs before AI introduction could be the result of regression to the mean. Furthermore, other subgroup analyses suggest that the detrimental effect might be more pronounced among female patients and by female endoscopists. Further research is needed to clarify these causal relationships.

Our study has several limitations. First, the patient cohorts that we compared might not have had comparable background characteristics, including indications and procedural contexts, due to the observational study design. To mitigate this limitation, we conducted multivariable logistic regression analysis. Second, only one AI system was used in this study, which might limit generalisation of the study results. However, a randomised trial showed this device led to a relative increase in ADR of about 20%,[22] which is in line with the general performance of similar AI devices.[20] Given the general nature of the AI tools and the tendency for humans to over-rely on them, we do not think the study results apply only to this specific AI. Third, our findings are based on the results with 19 experienced endoscopists who participated in the study, which might limit its generalisability. Further studies involving less-experienced operators are needed, because incorporating AI for training purposes is a major focus of the medical community.[23] Fourth, the number of

procedures per endoscopist was insufficient to reliably assess endoscopist-level ADR changes. Thus, we treated endoscopist-level analysis as a subgroup analysis, advising cautious interpretation. Fifth, a time effect could also influence endoscopists' skills, potentially hindering the interpretation of the study results. To minimise this bias, we focused on a small timeframe (approximately 3 months before and after AI implementation) rather than year-level changes. Additionally, as all the endoscopists included in our study were experienced, it is unlikely that their skills would change significantly over a short period—thereby minimising the risk of a time effect. Sixth, we primarily investigated the effect of AI exposure on human behaviour and therefore blinding was not done. However, the absence of blinding might introduce unforeseen biases and potentially distort the results. Seventh, the absence of data for endoscope withdrawal time and sample size calculation due to the retrospective observational design are also noticeable limitations. Eighth, we observed an almost two-fold increase in the number of procedures in the post-AI phase compared with the pre-AI phase (795 before *vs* 1382 after AI introduction, including 734 with AI assistance and 648 without AI assistance), even though both phases covered approximately 3 months. This finding was primarily due to the temporary suspension of the colonoscopy-based colorectal cancer screening programme in Poland on Jan 1, 2022 (lasting until December, 2022).[24] This suspension increased each centre's capacity for diagnostic colonoscopies, enabling them to accept more patients from the waiting list in the second phase, which is potentially a confounding factor. Ninth, exclusion of sessile serrated lesions from the ADR calculation might not accurately reflect colonoscopy quality, as a Dutch study showed that the sessile serrated lesion detection rate was associated with post-colonoscopy colorectal cancer incidence.[25] Tenth, although the first 100 colonoscopies per centre following AI introduction were excluded to give endoscopists time to become familiar with the technology, the number of cases performed by each endoscopist during this period might have varied. Finally, the number of examinations performed by some endoscopists was relatively low (appendix p 2) due to the observational nature of the study and could potentially affect the results.

In conclusion, we observed that continuous exposure to AI in colonoscopy might reduce the ADR of standard, non-AI assisted colonoscopy. We emphasise the urgent need for robust prospective studies, such as randomised crossover trials, to assess its generalisability and call for more behavioural research to understand the currently under-investigated mechanisms of how AI affects physician capability.

Contributors

Conceptualisation: KB, MR, PK, MBug, YM. Data curation: KB, MR, DK, PK, MBug, TR, KJ, YM. Formal analysis: KB, MR, DK, PK, MBug, JB, NH, SF, ØH, YM. Investigation: KB, MR, PK, MBug, MBus, TR, KJ, YM. Methodology: KB, MR, PK, MBug, CH, MKu, MP, NP, MFK, MKa, MBr, YM. Project administration: KB, MR, PK, MBug, CH, MKu, MP, NP, MFK, MKa, MBr, YM. Resources: KB, MR, PK, MBug, MBus, TR, KJ, CH, SF, MKu, MP, NP, MFK, MKa, MBr, YM. Supervision: MR, HOA, JB, NH, ØH, CH, SF, MKu, MP, NP, MFK, MKa, MBr, YM. Validation: MR, HOA, JB, NH, ØH, CH, SF, MKu, MP, NP, MFK, MKa, MBr, YM. Visualisation: MR, YM. Writing (original draft): KB, MR, DK, YM. Writing (review and editing): KB, MR, HOA, ØH, SF, CH, MKu, MP, NP, MFK, MKa, MBr, YM. KB, PK, MR, SF, YM had access to and verified the data of this study. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

YM has received consultancy fees, speaker's fees, and a device loan from Olympus, royalty fees from Cybernet System, and financial support from the European Commission (Horizon Europe 101057099) and the Japan Society for the Promotion of Science (22H03357). MP has received consultancy fees from Olympus Europe and Alfasigma, speaker's fees from Olympus, Casen Recordati, and Mayoly, and travel fees from Mayoly. CH has received a consultancy fee and device loan from Odin, device loan and speaker's fees from Olympus, and financial support from the European Commission (Horizon Europe 101057099), The Associazione Italiana per la Ricerca sul Cancro (IG 2022–ID. 27843 project, IG 2023–ID. 29220 project, and Bando PNRR-MCNT2–2023–12377041). SF's institution has received funds from Oslo University Hospital to cover statistical consultancy activities. MFK has received consultancy fees from Olympus, Boston Scientific, and Erbe, and speaking and teaching fees from Olympus Corp, Fujifilm, Boston Scientific, Medtronic, Erbe, Microtech, and Norgine. KB, MR, DK, PK, MBug, HOA, JB, MBus, NH, TR, ØH, KJ, MKu, NP, MKa, MBr- declare no competing interests.

Data sharing

Data and analytic codes can be shared under the following conditions: (1) approval of all the data providers or owner; (2) signed agreement for data transfer; and (3) approval of relevant ethics committees. Requests for access to data should be made to the corresponding author via the corresponding email given.

Acknowledgments

YM received financial support for this study from the European Commission (Horizon Europe: 101057099) and the Japan Society for the Promotion of Science (22H03357). CH received financial support, not related to this study, from the European Commission (Horizon Europe 101057099) and the Associazione Italiana per la Ricerca sul Cancro (IG 2022–ID. 27843 project, IG 2023–ID. 29220 project, and Bando PNRR-MCNT2–2023–12377041).

References

1 Morgan E, Arnold M, Gini A, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* 2023; **72:** 338–44.

2 Bretthauer M, Løberg M, Wieszczy P, et al, and the NordICC Study Group. Effect of colonoscopy screening on risks of colorectal cancer and related death. *N Engl J Med* 2022; **387:** 1547–56.

3 Pilonis ND, Spychalski P, Kalager M, et al. Adenoma detection rates by physicians and subsequent colorectal cancer risk. *JAMA* 2025; **333:** 400–07.

4 Bretthauer M, Kaminski MF, Løberg M, et al, and the Nordic-European Initiative on Colorectal Cancer (NordICC) Study Group. Population-based colonoscopy screening for colorectal cancer: a randomized clinical trial. *JAMA Intern Med* 2016; **176:** 894–902.

5 Hassan C, Misawa M, Rizkala T, et al, and the CADx Analysis Study Group. Computer-aided diagnosis for leaving colorectal polyps in situ: a systematic review and meta-analysis. *Ann Intern Med* 2024; **177:** 919–28.

6 Areia M, Mori Y, Correale L, et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. *Lancet Digit Health* 2022; **4:** e436–44.

7 Macnamara BN, Berber I, Çavuşoğlu MC, et al. Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cogn Res Princ Implic* 2024; **9:** 46.

8 Hakama M, Malila N, Dillner J. Randomised health services studies. *Int J Cancer* 2012; **131:** 2898–902.

9 Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc* 2009; **69:** 620–25.

10 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.

11 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.

12 Molenberghs G, Verbeke G. Models for discrete longitudinal data. Springer, 2005.

13 StataCorp. melogit—mMultilevel mixed-effects logistic regression. College Station, TX: Stata Press, 2023.

14 Issaka RB, Chan AT, Gupta S. AGA clinical practice update on risk stratification for colorectal cancer screening and post-polypectomy surveillance: expert review. *Gastroenterology* 2023; **165:** 1280–91.

15 Wieszczy P, Bugajski M, Januszewicz W, et al. Comparison of quality measures for detection of neoplasia at screening colonoscopy. *Clin Gastroenterol Hepatol* 2023; **21:** 200–209.e6.

16 Ahmad SF, Han H, Alam MM, et al. Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanit Soc Sci Commun* 2023; **10:** 311.

17 Bisschops R, East JE, Hassan C, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline—Update 2019. *Endoscopy* 2019; **51:** 1155–79.

18 Troya J, Fitting D, Brand M, et al. The influence of computer-aided polyp detection systems on reaction time for polyp detection and eye gaze. *Endoscopy* 2022; **54:** 1009–14.

19 Du-Crow E, Astley SM, Hulleman J. Is there a safety-net effect with computer-aided detection? *J Med Imaging (Bellingham)* 2020; **7:** 022405.

20 Hassan C, Spadaccini M, Mori Y, et al. Real-time computer-aided detection of colorectal neoplasia during colonoscopy: a systematic review and meta-analysis. *Ann Intern Med* 2023; **176:** 1209–20.

21 Mori Y, Bretthauer M, Kalager M. Hopes and hypes for artificial intelligence in colorectal cancer screening. *Gastroenterology* 2021; **161:** 774–77.

22 Gimeno-García AZ, Hernández Negrin D, Hernández A, et al. Usefulness of a novel computer-aided detection system for colorectal neoplasia: a randomized controlled trial. *Gastrointest Endosc* 2023; **97:** 528–36.

23 Lau LHS, Ho JCL, Lai JCT, et al, and the ENDOAID-TRAIN study group. Effect of real-time computer-aided polyp detection system (ENDO-AID) on adenoma detection in endoscopists-in-training: a randomized trial. *Clin Gastroenterol Hepatol* 2024; **22:** 630–41.

24 Polish Colorectal Cancer Screening Programme. Colonoscopy 2022. 2021. http://pbp.org.pl/2021/12/07/kolonoskopia-2022/ (accessed April 17, 2025).

25 van Toledo DEFWM, IJspeert JEG, Bossuyt PMM, et al. Serrated polyp detection and risk of interval post-colonoscopy colorectal cancer: a population-based study. *Lancet Gastroenterol Hepatol* 2022; **7:** 747–54.