Clinical Validation of Artificial Intelligence for Colorectal Cancer Screening with Colonoscopy

A Doctoral Thesis by

Ishita Barua

Institute of Health and Society

Faculty of Medicine

University of Oslo

&

Department of Transplantation Medicine

Oslo University Hospital

2022

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0142-6

Print production: Graphics Center, University of Oslo.

Acknowledgements

This doctoral thesis is the result of research conducted at the Department of Health Management and Health Economics, Institute of Health and Society, University of Oslo, at the Department of Transplantation Medicine, Oslo University Hospital, at Baerum Hospital, Vestre Viken, at Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, USA. The research was made possible through internal financing from our research group The Clinical Effectiveness Research Group, Caroline Musæus Aarsvold's Foundation for Young Internists and the U.S.-Norway Fulbright Foundation for Educational Exchange.

I have had the great honor of being supervised by Professor Michael Bretthauer, Associate Professor Magnus Løberg, Associate Professor Øyvind Holme and Associate Professor Yuichi Mori.

After completing medical studies and junior doctor internship, I wanted to be involved in medical research and learn that craftmanship. I had heard about this innovative and world-class research group doing groundbreaking research on clinical effectiveness, that was led by Professor Mette Kalager. At this point, I along with many other young female physicians, was already a huge fan of Mette's brilliant mind, resilience and grit. I knew I wanted to learn from her and be part of this specific research group. I contacted Mette and Michael and had a very inspiring meeting with them to talk about research possibilities in the group. They even invited me to social gatherings with the whole group, and I became even more certain that this was the kind of colleagues and research I wanted to be a part of. However, the timing was off since there were no new PhD-projects to be involved in, but we decided to keep in touch. Michael generously put me in contact with Baerum Hospital where I soon started a dual residency in Internal Medicine and Gastroenterology and was trained as an endoscopist in the colorectal screening pilot program. This accelerated my interest in colorectal cancer epidemiology and screening. From the very beginning of my endoscopy training, I was very

aware of the vulnerability in doing operator-dependent procedures such as colonoscopy and had reflected on different approaches to address this. So, when the opportunity arose to become a researcher in the group and evaluate the clinical performance of an AI-based tool in colonoscopy, I immediately said yes. At this point, the position was only temporary and not a full PhD position. Many thought this was a gamble. However, knowing Mette and Michael's track record and personal integrity, I did not hesitate to say yes. This leap of faith, has proven to be life changing for me and I would like to sincerely thank them both for all the opportunities they have presented and all the doors they have opened.

Michael has been my main supervisor on this thesis, and with his extraordinary analytical mind and unparalleled ability to put complicated medical concepts into clear and concise writing, he has pushed me towards the highest possible academic standards and thus been the greatest mentor I could wish for. I feel honored and privileged to have been able to learn from such a scholarly esteemed mentor like you. Magnus, has been one of my three co-supervisors. Needless to say, his outstanding methodology abilities have saved me numerous times, and the combination with his genuine compassion and calming effect on colleagues makes him truly exceptional. Øyvind, has also been one my three co-supervisors, and one I can always call whenever I am stressed and need advice from a calm and generous mind. I can always rely on Øyvind's support in solving any academic challenge, but also to deliver the funniest jokes; both of which are invaluable in order to get through the PhD alive. Yuichi, has also been one of my three co-authors, and the person I have been working closest with from day to day alongside Michael. It is Yuichi, with his strong and unique academic background and network within the field of artificial intelligence in gastroenterology, that has enabled me to pursue a thesis focusing on artificial intelligence in colonoscopy practice. Without his support and guidance, this thesis would not have been possible.

I am forever grateful to the whole supervisor team with their different academic backgrounds and areas of expertise perfectly complementing each other.

For me family always comes first. I would not have been a researcher today, if I had not been a physician first. And I would not have been a physician if it were not for the dedication of my parents, their relentless support and unconditional love. What they gave up in order to provide for me and my siblings, can never be repaid. I am the daughter of two of the most hard-working and selfless people I know and it is because of them that I will be the first in my family to become both a doctor and hold a PhD. Imon, my younger brother, I am forever grateful to you for all the moral and intellectual support you provide and for telling me about this research group to begin with. Simon, my youngest brother, whose opinion and counselling I always seek, completes my support system from home. My in-laws Anne Liv, Michael and Carly, thank you for all the support and for taking such good care of both me and Philip. I would like to sincerely thank my colleagues Lise, Frederik Emil and Erle for their tremendous support and always providing fun and thought-provoking conversations at the office. A very special thanks to Henriette and Anita who always have my back, keep me sane and listen to my long rants. I would like to thank Sofie, Ragnhild, Ida and Inga for their endless support and whose advice I sought during most of my PhD period. I am lucky to have friends like you.

My darling husband Philip, you are the love of my life, my best friend and my favorite person. We met during the early days of my PhD and since then you have been my strongest supporter and caretaker, even when I was away for my research stay in Boston alone. It's the greatest privilege of my life that we got married and I would never be able to complete this PhD or any other side projects without your unconditional love and support. Thank you, love.

Oslo, May 2022
Ishita Barua

1 Abbreviations

AADR	Advanced adenoma detection rate
ADR	Adenoma detection rate
AI	Artificial Intelligence
ASGE	American Society for Gastrointestinal Endoscopy
CAD	Computer-aided diagnosis
CADe	Computer-aided diagnosis for detection
CADx	Computer-aided diagnosis for characterization
CI	Confidence interval
CRC	Colorectal cancer
ESGE	The European Society of Gastrointestinal Endoscopy
FIT	Fecal Immunochemical Test
GI	Gastrointestinal
ICC	Intra-cluster correlation coefficient
IQR	Interquartile ranges
ITT	Intention-to-treat analyses
MAAP	Mean number of advanced adenomas detected per colonoscopy
MAP	Mean number of adenomas detected per colonoscopy
MPP	Mean number of polyps detected per colonoscopy
NPV	Negative predictive value
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RCT	Randomized controlled trial
REML	Restricted maximum likelihood
RR	Risk ratio
SSL	Sessile serrated lesion
TSA	Traditional serrated adenomas

2 Thesis Summary

Background

Colorectal cancer is considered a growing health burden and a preventable disease. It is the third most common cancer and the second leading cause of cancer death worldwide. Norway has one of the world's highest rates of colorectal cancer. In 2020, approximately 4,500 Norwegians were diagnosed with colorectal cancer, while 1,500 died of colorectal cancer. In order to reduce the risk of colorectal cancer incidence and mortality, many countries have implemented colorectal cancer screening. Such a national screening program is scheduled to be implemented in Norway in 2022. Colonoscopy is the gold standard of colorectal cancer screening, but it is dependent on endoscopist performance and the technology used. Novel technologies such as artificial intelligence targeting improved performance and standardization is expected to play a bigger role in colorectal cancer screening with colonoscopy in the future. In this thesis, I aim to investigate the clinical performance of artificial intelligence (AI) to optimize colonoscopy for colorectal cancer screening.

Methods

We performed a meta-analysis to determine if colonoscopy with AI-based polyp detection systems increased adenoma, polyp and colorectal cancer detection rates compared to colonoscopy without AI-assistance. We also performed an international multicenter clinical trial to investigate whether an AI-based device for optical diagnosis can increase sensitivity in identification of small, rectosigmoid adenomas compared with visual inspection by the endoscopist alone. We further evaluated the specificity to differentiate adenomas and assessed the confidence level of endoscopists' optical diagnosis of polyps, comparing colonoscopies performed with and without AI-assistance. Lastly, we performed a clinical, implementation

trial to evaluate the performance of an AI-based speedometer that monitors the withdrawal speed. We evaluated withdrawal time difference, adenoma detection rate and proportion of colonoscopies with withdrawal time ≥ 6 minutes, comparing colonoscopies performed with and without an AI-based speedometer.

Results

The meta-analysis revealed five eligible RCTs on AI-based polyp detection. We found that ADR had a relative increase of 50% with the use of AI-based polyp detection systems. These systems only increased detection of small polyps and non-advanced adenomas, but did not show any difference in detection of advanced adenomas known to have the highest potential for malignancy.

The multicenter clinical trial evaluating an AI-based device for optical diagnosis did not show any significant increase in sensitivity for small, neoplastic polyps during colonoscopy with the use of an AI-based device for optical diagnosis compared to standard method. The study did suggest an increase in specificity for optical diagnosis of small neoplastic polyps and an increase in confidence level for the endoscopists in optical diagnosis of polyps.

The clinical implementation trial evaluating an AI-based speedometer, showed no benefit from using the AI-speedometer during colonoscopy in order to increase withdrawal time. We found no significant withdrawal time difference, no increase in ADR or proportion of colonoscopies with withdrawal time ≥ 6 minutes when using the AI-speedometer.

Conclusion

The studies we conducted and included in this thesis do not show that AI-based tools add clinical value to colonoscopy for colorectal cancer screening and thus show no clear benefit, but may cause harm.

3 Sammendrag av avhandlingen

Bakgrunn og mål

Tarmkreft (tykk- og endetarmskreft) regnes som en økende helsebyrde og er en sykdom det er mulig å forebygge. Det er den tredje vanligste kreftformen og den kreftformen som står for det nest høyeste antallet døde på verdensbasis. Norge har en av verdens høyeste forekomster av tarmkreft. I 2020 ble cirka 4500 nordmenn diagnostisert med tarmkreft, og 1500 døde som følge av sykdommen. For å redusere risikoen for forekomst og dødelighet av tarmkreft, har mange land innført screening for tarmkreft. Et slikt nasjonalt screeningprogram er planlagt implementert i Norge i løpet av 2022. Koloskopi er gullstandarden for tarmkreftscreening, men det er avhengig av endoskopørens ferdigheter og teknologien som brukes. Ny teknologi som kunstig intelligens som tar sikte på å forbedre og standardisere utførelsen av koloskopier, er forventet å spille en større rolle innen tarmkreftscreening i fremtiden. I denne avhandlingen, har jeg som mål å undersøke i hvilken grad kunstig intelligens kan bidra til å optimalisere tarmkreftscreening med koloskopi.

Metoder

Vi utførte en metaanalyse for å undersøke om koloskopi med kunstig intelligens-baserte polyppdeteksjonssystemer økte adenom-, polypp- og tarmkreftdeteksjonsraten sammenlignet med koloskopi uten støtte fra kunstig intelligens-baserte systemer. Vi gjennomførte i tillegg en internasjonal, klinisk, multisenterstudie for å undersøke om et kunstig intelligens-basert system for optisk diagnose kan øke sensitiviteten ved identifisering av små adenomer i rektum og sigmoideum, sammenlignet med visuell inspeksjon av endoskopøren alene. Vi evaluerte videre spesifisiteten for å differensiere adenomer og vurderte hvor trygge endoskopørene var ved optisk diagnose av polypper, og sammenlignet koloskopier utført med og uten assistanse fra kunstig intelligens. Til slutt gjennomførte vi en klinisk, implementeringsstudie der vi

vurderte den kliniske effekten av et kunstig intelligens-basert system for måling av fart ved koloskopi-tilbaketrekning. Vi evaluerte tidsforskjellen i koloskopi-tilbaketrekning, adenomdeteksjonsrate og andelen koloskopier som hadde koloskopi-tilbaketrekningstid på ≥ 6 minutter, og sammenlignet koloskopier utført med og uten kunstig intelligens-basert fartsmåler.

Resultater

Metaanalysen viste at fem randomiserte kontrollerte studier om kunstig intelligens-baserte polyppdeteksjonssystemer var kvalifiserte. Vi fant at adenomdeteksjonsraten hadde en relativ økning på 50% ved bruk av kunstig intelligens-baserte polyppdeteksjonssystemer. Disse systemene økte bare deteksjonen av små polypper og ikke-avanserte adenomer, men viste ingen forskjell i deteksjonen av avanserte adenomer som er kjent for å ha det høyeste potensialet for malignitet. Den kliniske multisenterstudien som vurderte et kunstig intelligens-basert system for optisk diagnose, viste ingen signifikant økning i sensitivitet for små neoplastiske polypper ved koloskopi med bruk av et kunstig intelligens-basert system for optisk diagnose sammenlignet med standardmetoden. Studien antydet en økning i spesifisitet for optisk diagnose av små neoplastiske polypper og en økt selvsikkerhet for endoskopørene i optisk diagnose av polypper. Den kliniske implementeringsstudien som vurderte et kunstig intelligens-basert system for måling av fart ved koloskopi-tilbaketrekning, viste ingen fordeler ved bruk under koloskopi for å øke koloskopi-tilbaketrekningstiden. Vi fant ingen signifikant tidsforskjell i tilbaketrekning, ingen økning i adenomdeteksjonsrate eller økt andel av koloskopier med tilbaketrekningstid ≥ 6 minutter ved bruk av kunstig intelligens.

Konklusjon: Studiene vi har utført og inkludert i denne avhandlingen viser ikke at kunstig intelligens-baserte verktøy forbedrer koloskopi for tarmkreftscreening og viser dermed ingen klar fordel, men kan medføre ulemper.

4 Articles in the Thesis

Article I:

Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis

Ishita Barua, Daniela Guerrero Vinsard, Henriette C. Jodal, Magnus Løberg, Mette Kalager, Øyvind Holme, Masashi Misawa, Michael Bretthauer, Yuichi Mori.

Endoscopy 2021; 53:277-284

Article II:

Real-Time AI-Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy

Ishita Barua MD, Paulina Wieszczy PhD, Shin-ei Kudo MD, Masashi Misawa MD, Øyvind Holme MD, Shraddha Gulati MD, Sophie Williams MD, Kensaku Mori PhD, Hayato Itoh PhD, Kazumi Takishima MD, Kenichi Mochizuki MD, Yuki Miyata MD, Kentaro Mochida MD, Yoshika Akimoto MD, Takanori Kuroki MD, Yuriko Morita MD, Osamu Shiina MD, Shun Kato MD, Tetsuo Nemoto MD, Bu Hayee MD, Mehul Patel MD, Nishmi Gunasingam MD, Alexandra Kent MD, Andrew Emmanuel MD, Carl Munck MD, Jens Aksel Nilsen MD, Stine Astrup Hvattum MD, Svein Oskar Frigstad MD, Petter Tandberg MD, Magnus Løberg MD, Mette Kalager MD, Amyn Haji MD, Michael Bretthauer MD and Yuichi Mori MD.

New England Journal of Medicine Evidence 2022; DOI:10.1056/EVIDoa2200003

Article III:

Speedometer for withdrawal time monitoring during colonoscopy: A clinical implementation trial

Ishita Barua, Masashi Misawa, Jeremy R Glissen Brown, Trent Walradt, Shin-ei Kudo, Sunil G Sheth, Judy Nee, Johanna Iturrino, Rupa Mukherjee, Catherine P Cheney, Mandeep S Sawhney, Douglas K Pleskow, Kensaku Mori, Magnus Løberg, Mette Kalager, Paulina Wieszczy, Michael Bretthauer, Tyler M Berzin and Yuichi Mori.

Manuscript

5 Background

5.1 Colorectal cancer epidemiology

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer death worldwide, accounting for more than 1.9 million new CRC cases and 935,000 deaths, representing approximately 10% of all cancer deaths.[1] The lifetime risk of CRC in the Western world is around 5%, and there are geographical variations in CRC incidence and mortality.[2] Incidence rates are 4-fold higher in developed countries compared with developing countries, but with less variation in the mortality rates because of higher fatality in developing countries. The incidence rates are highest in European regions, Australia, New Zealand and Northern America. Rates of CRC incidence is lowest in Africa and in South Central Asia.[3]

These geographic differences and disparities in CRC incidence and outcomes can be attributed to differences in access to high-quality health care services (CRC screening availability, diagnostic resources and treatment), dietary and environmental exposures, low socioeconomic status, and genetic predisposition.[4]

The cumulative risk of developing CRC before the age of 75 is 1,8% for women and 2,7% for men globally. Men are on average 5 to 10 years younger when they are diagnosed with CRC, and also have a higher risk of CRC than women.[5] Hungary rank first in CRC incidence for men and Norway rank first in CRC incidence for women. In Norway approximately 4,500 individuals were diagnosed with CRC and 1,500 individuals died of CRC in 2020.[6] Since the establishment of the Cancer Registry of Norway for more than 70 years ago, Norway has had one of the highest CRC rates and the incidence has increased three-fold, which is significantly higher than neighboring Nordic countries for reasons that are not fully understood.[7–9]

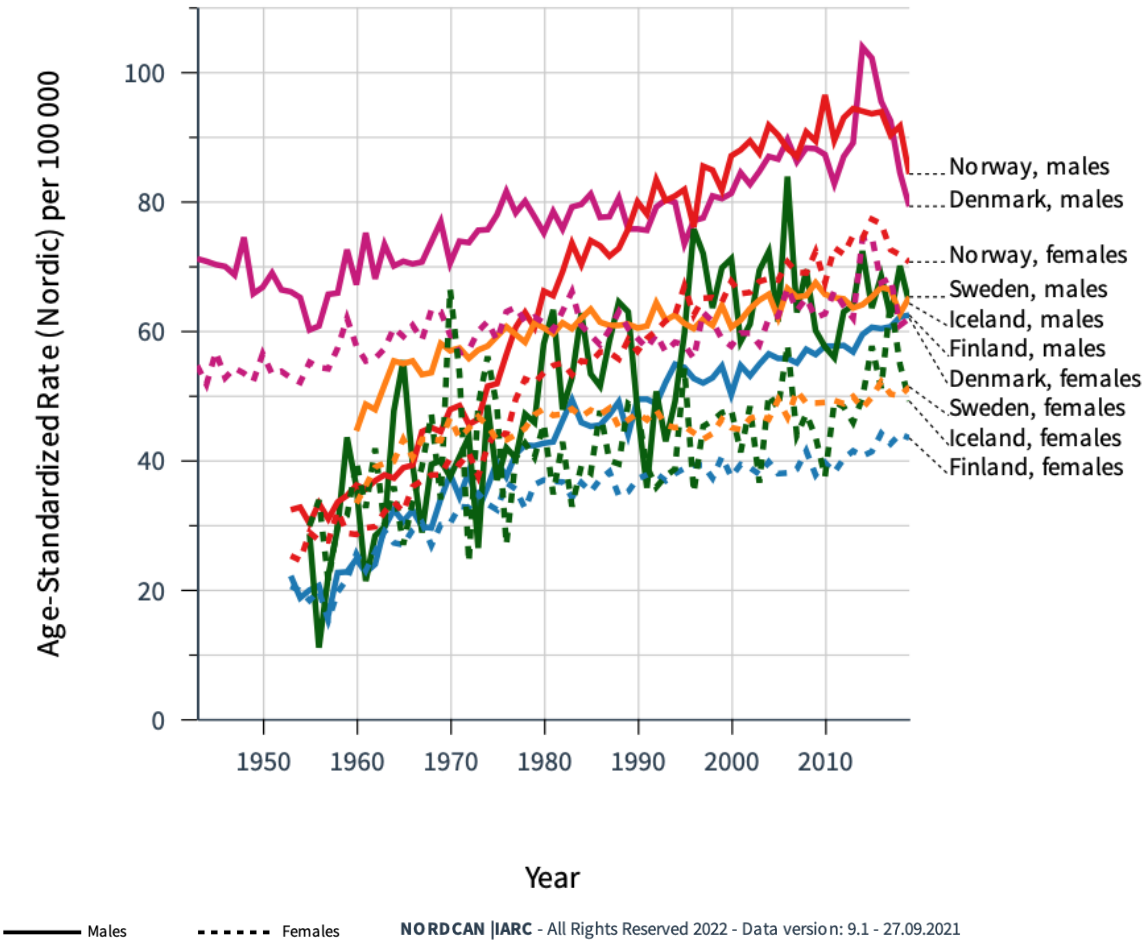


Figure 1: Age-standardized incidence rates (1943-2019) for colorectal cancer in the Nordic countries (Norway, Denmark, Finland, Iceland and Sweden) per 100 000, among males and females. Tabulated data are made available and graphics made with permission from NORDCAN and the Association of the Nordic Cancer Registries (ANCR).

What is better understood is that CRC is associated with both non-modifiable risk factors such as sex, age and family history,[10] and with modifiable risk factors such as excess body weight, unhealthy diet, physical inactivity, smoking, high alcohol consumption.[11,12] CRC incidence rates tend to rise uniformly with increasing socioeconomic growth in countries undergoing major societal and economic transitions that are strongly linked to a western lifestyle and with many risk factors predominant in developed countries.[13] This suggest a causal relationship and CRC can thus be considered a determinant of socioeconomic progress.

5.2 Colorectal cancer pathogenesis

CRC is the development of cancer from the colon and rectum. Normal cells grow and divide in an orderly way, but somatic gene mutations can result in unregulated growth and cause visible wart-like lesions on the colonic mucosa protruding into the gastrointestinal (GI) lumen. Polyps may occur anywhere in the GI tract but are most common in the colorectal region and with a predominantly left-sided anatomic distribution.[14] They can be classified according their macroscopic appearance; depressed, flat, sessile (broad-based with no stalk) or pedunculated (with a stalk).[15]

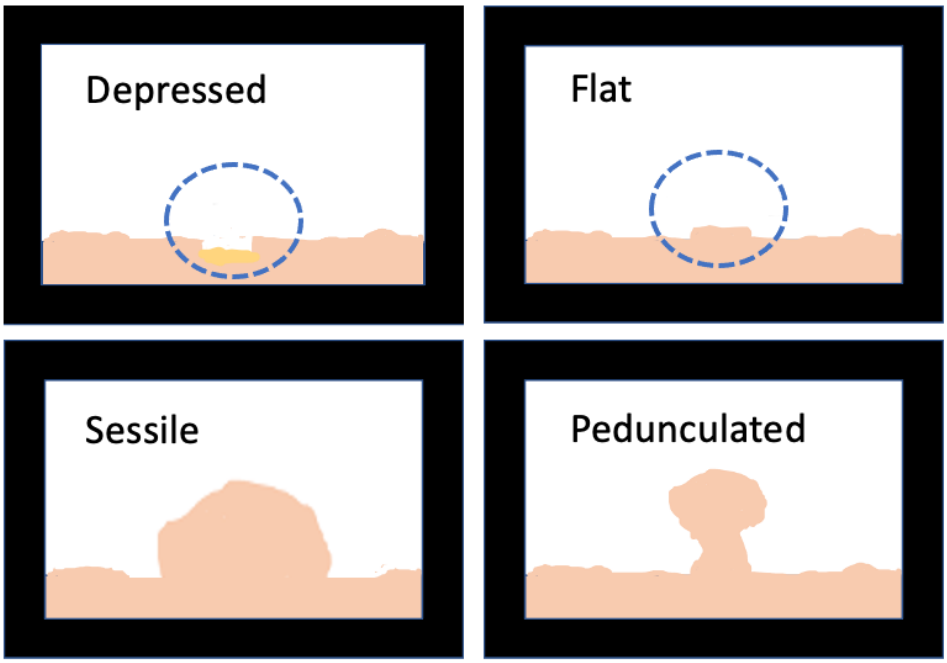


Figure 2: Different types of polyps based on macroscopic appearance.

Colonic polyps can also be subdivided according to their potential for cancer development into two main histological categories: non-neoplastic and neoplastic. Non-neoplastic polyps harbor no malignant potential. This group includes hyperplastic polyps, hamartomatous polyps, inflammatory polyps and mucosal polyps. Neoplastic polyps harbor malignant potential and have epithelial dysplasia. This group includes adenomas and carcinomas.

Adenomas can be tubular, tubulovillous, or villous based on the glandular architecture and further classified according to their grade of dysplasia being either low or high.[14–16] Historically, only colorectal polyps classified as adenomas were considered to be a precursor lesion having malignant potential recognized through histopathological examination. The adenoma-carcinoma sequence is the stepwise mutational transformation from normal colonic epithelium to adenoma, and then progressing to CRC.[17,18] This malignant transformation from a detectable adenoma to cancer is estimated to take about 10-15 years.[19,20] It is not yet possible to determine which adenomas will progress, but certain histopathological features of the adenomas correlate with the risk of diagnosing cancer in those polyps. Increased polyp size, villous histology, and severe dysplasia are all associated with an increased risk of malignancy in an adenoma.[14,15]

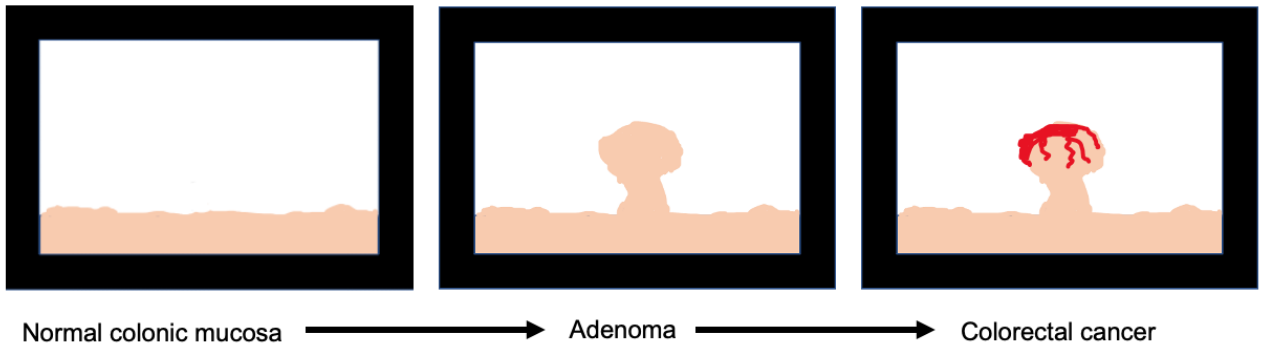


Figure 3: The adenoma-carcinoma sequence

In the last two decades, a second precursor of CRC and an alternative pathway for colorectal carcinogenesis has been widely accepted and is called the serrated pathway.[21] Through this pathway normal colonic epithelium evolves into a serrated polyps, and then progresses into CRC. Serrated polyps have a saw-tooth appearance under the microscope and include hyperplastic polyps, sessile serrated lesions (SSL), and traditional serrated adenomas (TSA). These polyps have different cancer risks and are sometimes difficult to discriminate. SSLs and TSAs are generally considered to have malignant potential, while hyperplastic

polyps are not. It is estimated that 20-25% of all CRCs develop through the serrated pathway and 5% are attributed to Lynch syndrome.[22–24] However, around two-thirds of all colonic polyps are adenomatous and approximately 60-75% of all CRCs are considered to have developed through the adenoma-carcinoma sequence.[23,25–27]

The CRCs evolving from the different pathways may be genetically different[28] and thus have different potential for prevention by screening. Whether genetic differences in colorectal tumors (that also varies among different ethnicities) can be used to predict disease progression or treatment selection is still unknown and the complete natural history of untreated colonic polyps is still not fully understood. Thus, knowledge gaps regarding colorectal cancer development still exist, but the detectable precursor lesions serve as a prerequisite for the cancer preventive effect of colorectal cancer screening.

5.3 Screening

CRC is considered a growing global public health challenge and a preventable disease. According to demographic projections it is expected that CRC will increase by 60% to more than 2.2 million new cases and 1.1 million deaths by 2030.[13] In order to mitigate this challenge many countries have implemented a strategy for reducing disparities in CRC incidence and mortality through screening.

Cancer screening is the process of identifying cancer in individuals who don't have any symptoms or signs. The aims of a cancer screening program is to reduce incidence and mortality, and are based on two principles; prevention and early detection. The rationale for screening effect through prevention is that detection and removal of pre-cancerous lesions and cancers may result in reduced incidence and mortality. The rationale behind early detection is that cancers detected at an early and curable stage usually have a better prognosis than

cancers detected at a later stage. Early detection can only reduce mortality and does not reduce cancer incidence, while prevention reduces both incidence and mortality.[5,29,30]

There are four different outcomes of screening with different implications as shown in the following figure.

	True	False
Individuals with a negative screening test	True negative (healthy individuals without the disease)	False negative (presumably healthy individuals, but are actually at risk or have the disease)
Individuals with a positive screening test	True positive (individual with the disease)	False positive (presumably individuals at risk or with the disease, but are actually healthy)

Figure 4: Individuals with a negative screening test or a positive screening test can either be true or false.

The outcomes of screening that are the least desirable are the false outcomes such as false-positive and false-negative tests. False outcomes can result in overtreatment or undertreatment of the disease, in addition to psychological distress due to the test result.

5.3.1 Overdiagnosis

Screening based on early detection may increase the incidence rates and screening may also cause harmful effects due to overdiagnosis. Overdiagnosis is when a disease is detected in presumably healthy individuals with no clinical signs or symptoms, and would not have been clinically identified within the individual's lifetime without the screening test.[5,29,30] This can be the result of either a spontaneous regression or that the individual dies from other causes and before the disease clinically manifested due to slow progression.

In the case of a cancerous tumor identified during screening, there is no certain way to distinguish between an overdiagnosed cancer and an aggressive, deadly cancer.[31] They will both nonetheless trigger treatment and surveillance, but the ones that are overdiagnosed will experience no real benefits from the screening and potentially be subjugated to harmful effects from complications, side effects and emotional distress associated with the screening outcome.[31–33] Overdiagnosis was mostly disregarded 10 years ago, but is now acknowledged as a significant harm of cancer screening.[31–33] A misconception worth mentioning and emphasize is that overdiagnosis is not the same as a false positive, since overdiagnosis occurs when a disease or cancer is *correctly* diagnosed albeit would not result in symptoms, while a false positive result occurs when an *incorrectly* diagnosis or cancer is identified.

5.3.2 Colorectal cancer screening

Most western countries have introduced CRC screening programs since the early 2000s with different screening modalities.[2] The most frequently used screening modalities in CRC screening programs are FIT (Fecal Immunochemical Test) which identifies traces of blood in stool samples (potentially a early symptom of CRC) and colonoscopy which is inspection of the whole colon and removal of lesions.[30,34] FIT is considered an early detection screening test for colorectal cancer, while colonoscopy is considered both a preventive screening test and an early detection screening test for colorectal cancer.

There is currently a paucity for randomized controlled trials examining the effect of CRC screening with FIT and colonoscopy on incidence and mortality, especially those looking at a follow-up time of more than 5 years.[30] Results from such trials are expected to be ready in the coming years ahead.[35] CRC screening is managed and provided as either an opportunistic screening through referrals from physicians or as an organized population-wide screening program.

The screening approach in the USA is largely opportunistic, but in Europe organized population-based screening programs are more predominant. Population-wide screening for CRC is now being planned in Norway and expected to be implemented in 2022.[36] This cancer screening program will be the third national screening program implemented (breast cancer and cervix cancer already established), but the first screening program for both sexes. In the Norwegian CRC screening program, FIT will be used for primary screening and positive screening tests will be followed by colonoscopy for definitive diagnosis. Gradually, the program will adopt to primary colonoscopy screening.

5.4 Colonoscopy

Colonoscopy is the gold standard of colorectal cancer screening and provides approximately 50% reduction in CRC incidence and mortality.[37–40] During a colonoscopy the endoscopist does an inspection of the entire colon using a flexible colonoscope allowing a direct assessment of the colonic mucosa. The primary aim of colonoscopy is detection of precancerous lesions and CRCs, whilst the secondary aim is removal of precancerous lesions and CRCs if feasible according to size, location and level of experience with the endoscopist. During a single procedure, colonoscopy can achieve both aims.

The effectiveness of colonoscopy screening to prevent colorectal cancer is dependent on high adenoma detection rates (ADR), i.e. the proportion of colonoscopies where at least one adenoma is detected.[41] ADR is considered a quality indicator of colonoscopy because it has shown that it is inversely associated with the risk of CRC and each 1% increase in ADR is associated with a 3% decrease in the risk for CRC.[41–43] There is considerable variation in ADR between individual endoscopists. Patients examined by an endoscopist with a high ADR, experience significantly lower risk of CRC diagnosed after screening (so-called interval

cancer that occur between screening and post-screening surveillance) as compared to endoscopists with a lower ADR. ADR of less than 20% is related to a higher risk of developing interval cancer.[41,42] It is important to both increase ADR and reduce variation in colonoscopist ADR to reduce interval cancer risk for patients after colonoscopy.[44] Interval cancers are the result of one of the three following reasons; missed adenomas, previously incomplete resection and new lesions. Missed adenomas represent the majority of these.

5.4.1 Missed adenomas

Although detection of neoplastic polyps is key for CRC prevention, approximately one fourth of these polyps are missed at colonoscopy screening and responsible for 50-60% of interval cancers, which appear after a negative screening test or examination.[45,46] 80-89% of all interval cancers are deemed avoidable.[47,48] Missed adenomas can be caused by two types of errors. One way of missing adenomas during colonoscopy is through recognition failure when the adenomas are present and visual on the screen, but the endoscopist fail to recognize them. These errors are called cognitive errors. The other way of missing adenomas is through incomplete mucosal exposure and blind spots that may be connected to withdrawal speed, endoscopists' skill, degree of bowel cleansing, and other factors. These errors are called exposure errors.[45,47,49]

5.4.2 Withdrawal time

Withdrawal time is defined as time spent examining the colon during withdrawal of the colonoscope. It is measured as the duration from identification of the cecum base and until exit from the rectum.[50] Withdrawal time serves as a surrogate quality indicator of colonoscopy that impacts the ADR through mucosal inspection.[41] A withdrawal time of at least 6 minutes is recommended and may indicate a more careful inspection of the colorectal mucosa during screening colonoscopy.[50–52] However, the recommended withdrawal time is

sometimes ignored or not feasible to achieve due to the busyness of clinical practice or other reasons that result in considerable variation in endoscopist withdrawal time.[50,51,53] It is important to maintain a stable maneuvering and uniform speed throughout withdrawal. This is crucial when passing colonic flexures where the endoscopists can experience endoscope slipping, resulting in quickly changing frames and blurry images, reduced mucosal exposure and increasing blind spots.

5.4.3 Optical diagnosis

Colonoscopy activity is steadily increasing around the world, mainly due to the introduction of CRC screening programs. This results in more polyp removals and histopathological assessments that together represent an increasing burden for the health care system in many countries. The current practice at most colonoscopy centers around the world is to remove all detected polyps during colonoscopy followed by submission to histopathological assessment. This is due to the difficulty for endoscopists to do a procedure known as “optical diagnosis”, where they reliably distinguish between non-neoplastic and neoplastic polyps. However, as many polyps never grow into cancer, many non-neoplastic polyps are removed today with considerable waste of resources. Reliable real-time optical diagnosis of polyps during colonoscopy could enable targeted removal only of polyps classified as neoplastic (e.g., adenomas), while non-neoplastic polyps (e.g., hyperplastic polyps) could be left behind. For rectosigmoid small (≤ 5mm in diameter) polyps, this strategy is called “diagnose-and-leave”.

There are several gastroenterology societies with guidelines recommending the diagnose-and-leave strategy due to the overall low prevalence of malignancy in rectosigmoid small polyps, combined with the potential cost-effectiveness stemming from reduction in procedure time and costs related to histopathology.[54,55] However, despite guidelines providing

quality standards and benchmarks that need to be achieved, the clinical uptake and the widespread use of optical diagnosis is low and there are several reasons for this.

On an individual endoscopist level there are reasons such as concerns about their capacity to make accurate diagnoses, assignment of inappropriate surveillance intervals and the perceived risk of legal consequences in the case of an incorrect diagnosis.[56,57] On a group level there are concerns regarding the potential variability in endoscopist performance with studies showing contradictory data with endoscopists both achieving the recommended accuracy thresholds[58,59] and falling short of these thresholds.[60]

5.4.4 Areas of improvement for CRC screening with colonoscopy

Colonoscopy is a cost-effective procedure that obviates the need for surgery in patients with identified CRCs or adenomas not invading deeper than the superficial submucosa. However, the quality of colonoscopy procedures is judged on the basis of several factors such as the level of endoscopists experience, endoscopists performance and technology used. Establishing practical and feasible methods that will both increase the ADR, increase withdrawal time and mitigate other technical variations in daily endoscopy practice is essential. This accentuates the need for operator-independent performance, increasing procedure precision and standardization in colonoscopy practice with two goals; improving overall endoscopist performance and reducing endoscopist variability during colonoscopy.

New research and technology developments such as artificial intelligence focusing on tackling these challenges and attaining standardization is already here and expected to play a bigger role in colonoscopy practice and CRC screening in the coming years. Clinical validation of the efficacy of these new tools will play an important role in the early adoption of new technological trends.

6.1 Artificial intelligence in medicine

Digital transformation is disrupting the whole society, including the health care sector. Health care data amassed from electronic medical records, clinical registries, biosensors, smart devices etc. is generated in such large volumes that we call it big data.[61] Big data in healthcare relies on the ability to detect and recognize patterns, and then convert large data sets into actionable knowledge for clinicians. Big data coupled with artificial intelligence have the potential to improve medical outcomes and population health. The term artificial intelligence (AI), is defined by the ability of computers to perform tasks that usually require human intelligence and that mimic human cognitive function. [62] Machine learning is a subcategory of AI that uses statistical methods to produce an algorithm founded on data, for pattern recognition and making decisions or predictions.[63,64] Deep learning is a subset of machine learning that uses feature extraction from multiple layers of data to make meaningful outputs.[63,64]

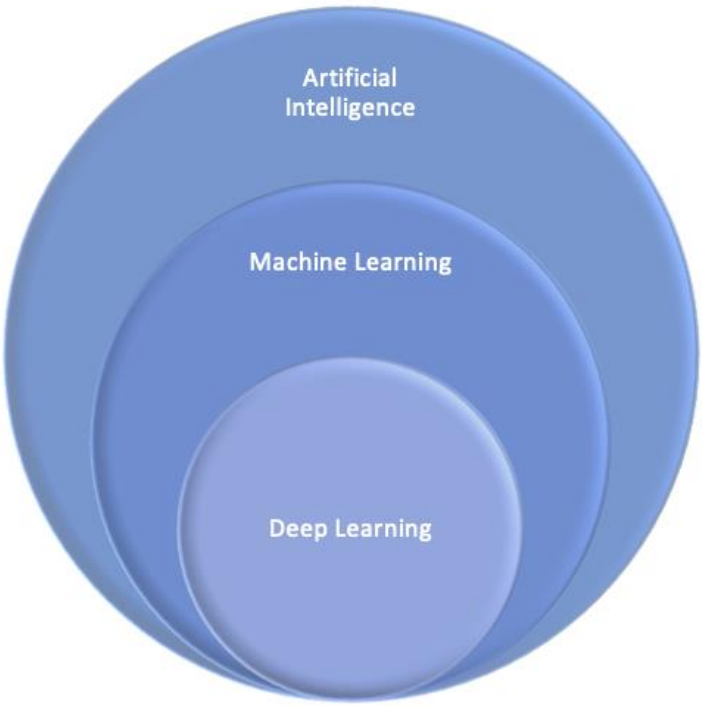


Figure 5: The different subfields of artificial intelligence.

Artificial intelligence has already started to permeate society in other fields than medicine, but AI applications for clinical medicine is advancing rapidly. In the future, AI may help us with some of the tasks presented in the following figure.

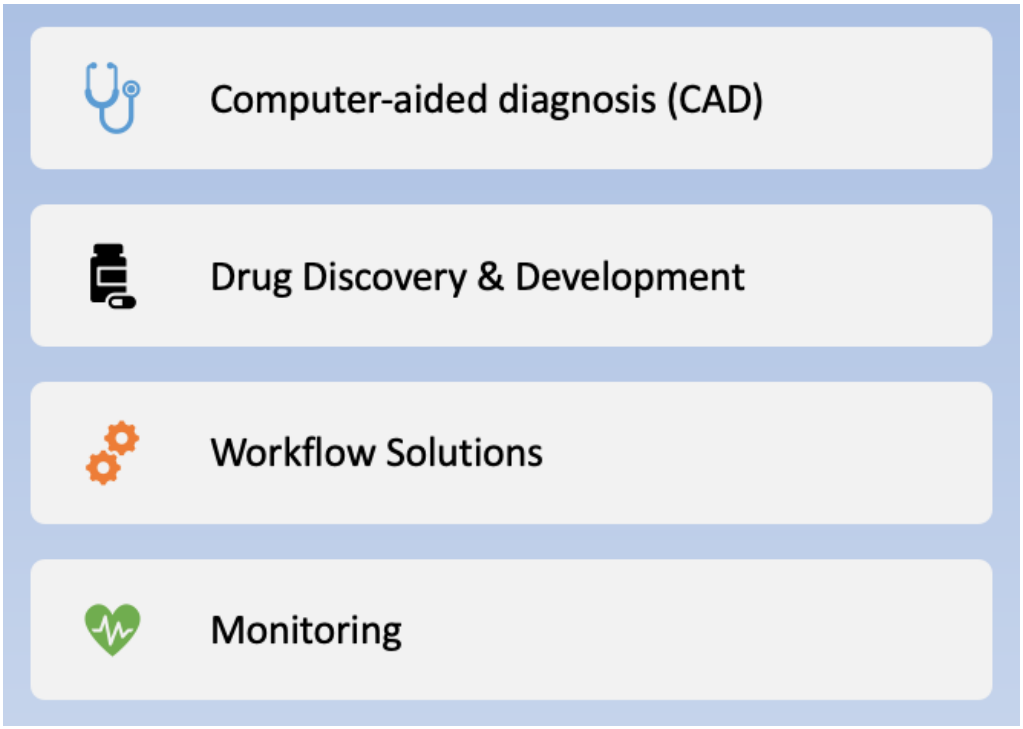


Figure 6: Key tasks that AI is expected to solve through recent progress and expected in the future.

One of the most important tasks for AI in medicine is computer-aided diagnosis (CAD). By processing vast amount of available information from large data sets or medical images, CAD-systems are able to augment the performance of clinicians or supplement and assist in decision making. CAD can be based on both machine learning and deep learning, but especially deep learning has become more frequent as basis for medical AI applications.[64] One of the more recent developments in AI application is drug discovery and development through automated drug design and compound selection that are more clinically meaningful and effective.[65]

AI has also proven its value in healthcare automation by improving clinical workflows to manage high-volume, repetitive tasks and workflows that machines do best and thus freeing up the clinicians' time to do direct patient work.[65–67] And the last two-three years during the COVID-19 pandemic have also highlighted the importance of monitoring devices. Wearables and embedded biosensors based on AI can now reduce the burden on health care professionals and provide monitoring – such as continuous glucose monitors or pulse oximeters, that potentially can prevent unwanted medical outcomes while simultaneously producing a vast amount of unprocessed data available to clinicians for more informed decision-making.

6.2 Clinical validation of AI-based tools

In medicine, we have a long tradition of practicing "evidence-based medicine", a term coined 31 years ago and defined as making clinical decisions based on the most current and available evidence.[68,69] In order for AI to be utilized and widely implemented into clinical practice, these novel auxiliary tools must prove clinical utility meaning effect on health outcomes,[70] through clinical validation and deliver evidence on efficiency and performance. However, there is a paucity for clinical validation of AI algorithms and clinical AI-based tools, in terms of prospective studies. Most validation studies for AI are retrospective studies done in silico (computer simulation) and not prospective studies performed in a real-world clinical setting.[65,66,71,72]

The retrospective studies are important for validation of AI algorithms, however they must be followed by prospective studies to establish clinical validation. Radiology is the medical field that many considers to be leading in terms of number of approved AI applications. There is currently 198 CE-marked (safe for sale and use in EU)[73] and commercially available AI-softwares.[74] However, research shows that even in this leading field the clinical validation is sparse and peer-reviewed evidence of efficacy is limited.[71,75]

Figure 7: Stepwise validation approach into clinical implementation. Adapted from the original by Topol, E.J.[66]

6.3 Applied AI-based tools in colonoscopy

Recent developments of AI applications for colonoscopy have shown great potential to improve the quality of colonoscopy. Thus, its imminent arrival and clinical implementation in colonoscopy practice is expected to increase in the coming years. AI-based tools in colonoscopy consist of integrated data sets from colonoscopy images and recordings, combined with high-volume computational power. Gastroenterology has for the last couple of years been one of the early adaptors in clinical medicine and a leading field when it comes to clinical validation.[44]

AI have attained two major roles in colonoscopy; polyp detection and polyp characterization. A third role is also expected to attain significance. This third group will in this thesis be referred to as "other AI-based systems" and focuses on the indirect quality assessments of colonoscopy in terms of withdrawal time, bowel preparation etc. What is unique for all types of AI application in colonoscopy compared to other medical applications, is the need for real-time assistance in order for AI to be of clinical relevance and support clinical decision-making.

6.3.1 Computer-aided diagnosis for detection (CADe)

One subset of CAD is computer-aided diagnosis for detection (CADe). CADe systems for colonoscopy is primarily developed through “annotation” or labelling of polyp images collected from colonoscopy recordings. Domain experts such as board-certified gastroenterologists will do the annotations and outline lesions that are considered to be a polyp, and this supervised learning process will ultimately lead to an algorithmic output where the CADe system can correctly detect and identify a polyp.[63] CADe in colonoscopy aims at improving ADR by reducing the rate of missed polyps. CADe can reduce the rate of missed polyps through targeting cognitive errors made by endoscopists. Cognitive errors during colonoscopy can be the failure of recognizing a polyp that is in fact visualized on the monitor.

6.3.2 Computer-aided diagnosis for characterization (CADx)

Another subset of CAD is computer-aided diagnosis for characterization (CADx). CADx systems for colonoscopy is developed through a similar process of supervised learning and annotations made by domain experts, as CADe. Machine learning was the basis for both CADe and CADx previously, but deep learning has for the last couple of years accelerated the performance of these CAD subfields.[64] CADx in colonoscopy aims at improving optical biopsy of polyps to obviate the need for histology and save time spent on waiting for the results.

6.3.3 Other AI-based systems

While CADe targets cognitive errors leading to missed polyps, other AI-based systems focuses on measures to overcome exposure errors that can help to increase ADR. These other AI-based systems can be deployed to increase mucosal exposure during colonoscopy and result in a more thorough visualization of the colon.

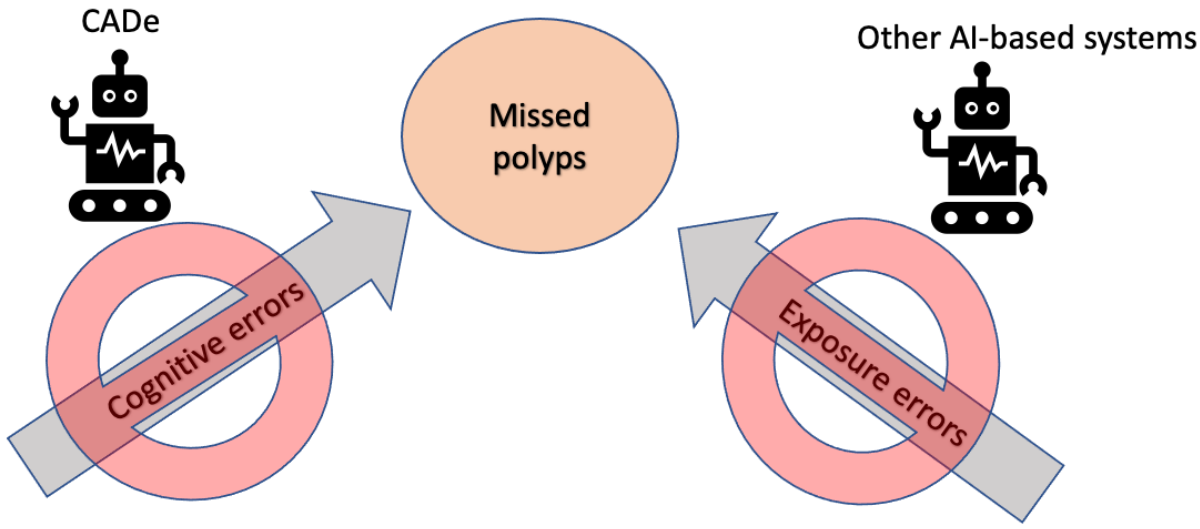


Figure 8: Missed polyps are caused by cognitive errors and exposure errors. CADe can mitigate cognitive errors, and other AI-based systems can mitigate exposure errors.

These other AI-systems are thus used for quality assessments related to colonoscopy such as withdrawal time monitoring or assessments of bowel preparation. One example is an AI-based speedometer that monitors the withdrawal time by alerting the endoscopist whenever the withdrawal speed exceeds a predefined threshold. This is an indirect way of achieving slower and more attentive withdrawal. Based on the results from two studies,[78,80] AI-based withdrawal time monitoring can help increase withdrawal time significantly.

Table 1: List of randomized controlled trials investigating CADe performance

Publication	**Year**	**Study design**	**Country**	**No of study centers**	**Scope company**	**CADe system**	**Patients**
Wang et al.[76]	2019	Parallel	China	1	Olympus	EndoScreener	1058
Repici et al.[77]	2020	Parallel	Italy	3	Olympus, Fujifilm	GI genius® (Medtronic)	685
Gong et al.[78]	2020	Parallel	China	1	Olympus, Fujifilm, Sonoscape	Endoangel	704
Liu et al.[79]	2020	Parallel	China	1	Fujifilm	Henan Tongyu	1026
Su et al.[80]	2020	Parallel	China	1	Pentax	Self-developed	623
Wang et al.[81]	2020	Parallel	China	1	Fujifilm	EndoScreener	962
Luo et al.[82]	2021	Parallel	China	1	Olympus	YOLO	150
Repici et al.[83]	2021	Parallel	Italy, Switzerland	5	Olympus, Fujifilm	GI genius® (Medtronic)	660
Xu et al.[84]	2021	Parallel	China	6	Olympus	Self-developed	2488
Yao et al.[85]	2021	Parallel	China	1	Olympus, Fujifilm	Endoangel	539
Glissen Brown et al.[86]	2021	Tandem	US	4	Olympus	EndoScreener	223
Kamba et al.[87]	2021	Tandem	Japan	4	Olympus	YOLOv3	344
Wang et al.[88]	2021	Tandem	China	1	Fujifilm	EndoScreener	369
Wallace et al.[89]	2022	Tandem	US, UK, Italy	8	Fujifilm	GI genius® (Medtronic)	230
Rondonotti et al.[90]	2022	Parallel	Italy	5	Fujifilm	CAD-EYE® (Fujifilm)	800

Table 2: List of prospective trials investigating CADx performance

Publication	Year	Country	No of study centers	Modality	Patients
Aihara et al.[91]	2013	Japan	1	AFI	32
Kuiper et al.[92]	2015	The Netherlands	1	WavStat4	87
Rath et al.[93]	2016	Germany	1	WavStat4	27
Kominami et al.[94]	2016	Japan	1	Magnifying NBI	41
Mori et al.[95]	2018	Japan	1	Endocytoscopy	791
Horiuchi et al.[96]	2019	Japan	1	AFI	95
Barua et al.[97]	2022	Norway, UK, Japan	3	Endocytoscopy	1,289
Minegishi et al.[98]	2022	Japan	1	NBI	186

Table 3: List of prospective trials investigating AI-based tools for withdrawal time monitoring.

Publication	Study design	Withdrawal time in standard colonoscopy	Withdrawal time with only a speedometer	Withdrawal time with CADe and speedometer combined	*P*-value
Gong *et al*. (2020)[78]	Randomized controlled trial	4.76 min	N/A	6.38 min	$P < 0.0001$
Su *et al*. (2020)[80]	Randomized controlled trial	5.68 min	N/A	7.03 min	$P < 0.001$
Yao *et al*. (2021)[85]	Randomized tandem, four-group, parallel, controlled study	9.71 min	10.14 min	10.17 min	* $P = 0.302$ ** $P = 0.413$

*: Standard colonoscopy vs. speedometer only.
**: CADe and speedometer combined vs speedometer only.

7 Aims and objectives of the thesis

Over the past decade, numerous AI-tools have been developed to improve clinical performance during colonoscopy using endoscopic still images and video data. Clinical validation on the efficacy of these new AI-tools will play an important role in the potential early adoption of new technological trends.

The overall objective for this thesis is to investigate the clinical performance of novel technology to optimize colonoscopy for CRC screening. This thesis includes three articles with the following aims:

Article I:

Investigate if recently developed AI-tools can increase the detection of polyps and colorectal cancer during colonoscopy compared to colonoscopy performed with standard method.

Article II:

Compare the clinical performance of an AI-based device for optical diagnosis in distinguishing neoplastic and non-neoplastic small polyps in the rectosigmoid colon during colonoscopy, with standard visual inspection-based optical diagnosis in routine clinical practice.

Article III:

Investigate if an AI-based speedometer for withdrawal time monitoring can improve suboptimal withdrawal time during colonoscopy, compared to colonoscopy performed with standard method.

8 Material & Methods

8.1 Article I

Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis

Endoscopy 2021; 53:277-284

The protocol for this systematic review and meta-analysis is registered with PROSPERO (CRD42020171860).

8.1.1 Study aim

The aim of this meta-analysis was to determine if colonoscopy with AI-based polyp detection systems increased adenoma, polyp and CRC detection rates compared to colonoscopy without AI-assistance. Additionally, the aim was to determine if colonoscopy with AI-based polyp detection systems increased the mean number of adenomas detected per colonoscopy (MAP), mean number of polyps detected per colonoscopy (MPP) and mean number of advanced adenomas detected per colonoscopy (MAAP), compared to colonoscopy without AI-assistance. Lastly, the aim was to perform stratification of mean number of adenomas by polyp size; rate of false positives and false negative AI diagnoses; colorectal cancer detection rates; and withdrawal time during colonoscopy.

8.1.2 Study population

In this meta-analysis a total of 4,311 patients between the mean ages of 49 and 51.6 were included between 2019 or 2020. All the patients included were from studies performed in China. The indication for colonoscopy was screening, symptomatic, or surveillance. Patients who had inflammatory bowel disease, history of colorectal cancer, history of radiotherapy and/or chemotherapy, or biopsy contraindications, were excluded.

8.1.3 Study intervention

In this meta-analysis we included five prospective studies, all five of them were RCTs. The study intervention was the use of an AI-based polyp detection system during colonoscopy. From these five studies we evaluated the performance of four different AI-based polyp detection systems. All the five RCTs used AI-based systems developed with deep learning models and with input from endoscopists and modelers. The comparator was colonoscopy without AI-assistance for four studies and with a sham control for one study.

8.1.4 Data extraction and rating of evidence

With the guidance from a medical librarian, we performed a search in MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) for the period between 1946 to February 2020. Two reviewers independently screened all search results, titles, abstracts and full texts to check further eligibility. Additionally, two reviewers searched ClinicalTrials.gov for ongoing studies not published yet, and contacted the relevant investigators for information. Data extraction was performed by two reviewers independently using a standardized form. Risk of bias was evaluated by two reviewers independently by using a modified Cochrane tool proven to increase reliability and validity in such assessments.[99] Consensus was achieved on all levels.

8.1.6 Statistical analysis

We used Stata version 16 (StataCorp, College Station, Texas, USA) for all data analyses and followed the reporting standards by Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA).[100] We based all analyses on per-protocol analyses of the trials. We used random-effects modelling and restricted maximum likelihood (REML) for analyzing summary measures.

The dichotomous outcomes were calculated with risk ratio (RR) with 95 % confidence intervals (CI) and mean differences were calculated with 95 % CI's. Absolute risks are the difference in overall detection rates and mean number between non-AI and AI groups.

8.1.6 Ethical approval

There was no need for ethical approval since this was a study summarizing already published evidence and no new data was collected.

8.2 Article II

Real-Time AI-Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy

New England Journal of Medicine Evidence 2022; DOI:10.1056/EVIDoa2200003

Article II in this thesis is based on the EndoBRAIN International study and was registered with UMIN Clinical Trials Registry (UMIN00003521).

8.2.1 Study aim

The aim of this study was to investigate whether an AI-based device for optical diagnosis can increase sensitivity in identification of small (≤5 mm in diameter), rectosigmoid adenomas compared with visual inspection by the endoscopist alone. Additionally, the aim was to evaluate the specificity to differentiate adenomas. Lastly, the aim was to assess the confidence level of endoscopists' optical diagnosis of polyps.

8.2.2 Study population

In this clinical trial we recruited 1,289 patients all whom were 18 years old or older in the period between May 2019 to May 2021. Out of these, 518 patients had 892 small and detected polyps (≤5 mm) in the rectosigmoid colon. The patients were from three different countries; Norway, Japan and the United Kingdom. The indication for colonoscopy was colorectal

cancer screening, surveillance, or diagnostic. In Norway, only those patients enrolled in the national CRC screening program pilot were eligible. Patients who had inflammatory bowel disease, polyposis syndrome, history of or current chemotherapy or radiation for rectosigmoid tumors, inability to undergo polypectomy (e.g., anticoagulants, comorbidities) pregnancy, and referral for removal of polyps with known histology. were excluded.

8.2.3 Study intervention

In this international, multicenter clinical trial we compared the sensitivity of identifying small (≤5 mm in diameter) polyps in the rectosigmoid colon as adenomas during colonoscopy with the combination of standard visual inspection and the CADx system, and of standard visual inspection alone, compared with gold-standard histopathology. All polyps were removed and submitted for histopathological analysis and all colonoscopies were performed according to routine standards at the participating centers, including pre-procedure assessment, bowel preparation, sedation practices, and post-procedure recovery and care.

8.2.4 Statistical Analysis

All statistical analyses were performed by using R version 3.4.1 and Stata version 17 software. We estimated the sensitivity, specificity, positive predictive value, and negative predictive value for the standard method and the CADx method compared with histopathology, respectively. The unit of analysis was the polyp rather than the study participant, with each participant providing 0 to 5 polyps for each analysis. For the primary analyses of sensitivity and specificity, SSLs were classified as neoplastic. For the secondary analyses, SSLs were classified as non-neoplastic.

8.2.5 Ethical approval

The EndoBRAIN International study was approved by the Regional Ethics Committee of South-Eastern Norway (2011/1272). In Norway, only participants enrolled in the national screening program pilot at Baerum Hospital were eligible for participation and written informed patient consent was included in the consent of the screening program.

8.3 Article III

Speedometer for withdrawal time monitoring during colonoscopy: A clinical implementation trial

Manuscript

Article III in this thesis is based on the Speedy-AI trial and was registered at ClinicalTrials.gov (NCT04710251).

8.3.1 Study aim

The aim if this study was to investigate the implementation of an AI-based speedometer that monitors the withdrawal speed and warns the colonoscopist whenever the speed limit is exceeded in real-time during colonoscopy. We wanted to evaluate the withdrawal time difference between colonoscopies performed with and without the speedometer. Additionally, the aim was to evaluate ADR of the participating endoscopists. Lastly, the aim was to assess the proportion of colonoscopies with withdrawal time ≥ 6 minutes.

8.3.2 Study population

All patients from the age of 18 years or older undergoing colonoscopy at a single study center were eligible for enrollment. Exclusion criteria were known colorectal cancer present before colonoscopy, hereditary colorectal polyposis, inflammatory bowel disease, or history of colorectal resection.

8.3.3 Study intervention

We performed a prospective, clinical implementation trial where patients were examined in one of the two study periods; either during the pre-implementation period with no use of the AI-based speedometer during colonoscopy or during the post-implementation period with AI-based speedometer used during colonoscopy.

All colonoscopy procedures in the trial were video-recorded and edited by a research assistant to mask whether or not the AI-speedometer was used, before being reviewed by two independent assessors. The independent research assistant kept the scrambling key that contained information about which video recording corresponded to which study period. This information was unavailable to the research team.

8.3.4 Statistical analysis

All analyzes were performed using Stata version 16.0 (Texas, USA). All analyses done are based on patients with a complete colonoscopy (cecum intubated) and their video recordings technically assessable. For the main endpoint withdrawal time, we reported mean and 95% confidence intervals. We defined statistical significance if $p<0.05$ and all p-values are two-sided.

8.3.5 Ethical approval

This trial was approved by the local institutional review board at Beth Israel Deaconess Medical Center, and registered at ClinicalTrials.gov (NCT04710251). We obtained written informed consent from all participants before enrolment.

9 Results and summary of the articles

9.1 Article I

Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis

Endoscopy 2021; 53:277-284

Our search yielded 1548 potentially relevant a documents and after completing the reviewing process, we ended up including five prospective studies which were all RCTs in our meta-analysis. These five included trials all conducted in China enrolled a total of 4,311 patients. Three out of five trials were single-blinded RCTs, one was a non-blinded RCT and one a double-blinded RCT. We found that four out of five trials had high risk of bias in terms of blinding, and one trial had no criterion for high risk of bias.

Compared to standard colonoscopy without AI-assistance, we found high-certainty evidence that AI-based polyp detection systems deployed during colonoscopy increases adenoma detection rates and polyp detection rates. Patients who had colorectal cancer were excluded in all the included trials, and thus we were not able to report on this outcome.

We also found high-certainty evidence that there was no difference in the mean number of advanced adenomas detected per colonoscopy between AI and non-AI assisted colonoscopy. Adenoma detection rate with AI was 29.6% (95%CI 22.2-37.0) as compared to 19.3% (95%CI 12.7-25.9) without AI (RR 1.52, 95%CI 1.31-1.77). Polyp detection rate with AI was 45.4% (95%CI 41.1-49.8) as compared to 30.6% (95%CI 26.5-34.6) without AI (RR 1.48, 95%CI 1.37-1.60). The mean number of advanced adenomas detected per colonoscopy between AI and non-AI assisted colonoscopy was 0.03 in each group, equal to a mean difference of 0 (95%CI -0.01-0.01). We further found moderate evidence that mean number of polyps detected per colonoscopy increases when comparing AI and non-AI assisted

colonoscopy; 0.93 vs. 0.51 (mean difference 0.42, 95%CI 0.33-0.50) and moderate evidence that mean number of adenomas detected per colonoscopy increases when comparing AI and non-AI assisted colonoscopy; 0.41 vs. 0.23 (mean difference 0.18, 95%CI 0.13-0.22).

Lastly, we found with high certainty that mean number of adenomas per colonoscopy increases when comparing AI and non-AI assisted colonoscopy for small adenomas ≤5 mm (mean difference 0.15, 95%CI 0.12-0.18), but not for larger adenomas >5 mm and ≤10 mm (mean difference 0.03, 95%CI 0.01-0.05) or larger adenomas >10 mm (mean difference 0.01, 95%CI 0.00-0.02). This study shows that AI-based polyp detection systems increased the adenoma detection rates, polyp detection rates, mean number of polyps detected during colonoscopy and mean number of adenomas detected during colonoscopy compared to non-AI assisted colonoscopy.

The ADR had a relative increase of 50% with the use of AI-based polyp detection systems. The AI-based polyp detection systems seem to augment the detection of more small polyps and non-advanced adenomas, but does not show any difference in detection of advanced adenomas known to have the highest potential for malignancy. The clinical added value of detecting more small polyps and non-advanced adenomas, need to be weighed against the pitfalls of overdiagnosis and overtreatment. Thus, CADe or AI-based polyp detection systems need to be supplemented by CADx, also known as AI-based tools for real time optical diagnosis of polyps.

9.2 Article II

Real-Time AI-Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy

New England Journal of Medicine Evidence 2022; DOI:10.1056/EVIDoa2200003

We recruited 1,289 individuals undergoing colonoscopy and assessed them for eligibility at three colonoscopy centers, in Norway, the United Kingdom and Japan. Patients included in the analyses had the median age of 67 years, and most patients had colonoscopy performed with CRC screening or polyp surveillance as indication. In our analyses we included 892 eligible polyps collected from 518 patients, constituting 359 neoplastic and 533 non-neoplastic polyps. Polyps identified in the rectosigmoid colon as 5 mm or less were assessed through standard visual inspection followed by standard visual inspection combined with CADx assessment. These polyps were subsequently removed and submitted for gold-standard histopathological evaluation. Participating study endoscopists were non-experts with average experience.

The difference in sensitivity for diagnosis of small, neoplastic polyps with the standard method vs. with the CADx method was not significant (p=0.33). The sensitivity with the standard method was 88.4% (95%CI 84.3-91.5) versus 90.4% (95%CI 86.8-93.1) with the CADx method. The specificity with the standard method was 83.1% (95%CI 79.2-86.4) versus 85.9% (95%CI 82.3-88.8) with the CADx method.

The percentage of polyp assessments made with high confidence for classifying into neoplastic or non-neoplastic polyps was 74.2% (95%CI 70.9-77.3) vs. 92.6% (95%CI 90.6-94.3) when comparing the standard method with the CADx method. The negative predictive value (NPV) with the standard method compared to the CADx method was 91.5% (95%CI 88.5-93.8) vs. 92.8% (95%CI 90.1-94.9). Our second analysis classifying SSLs as non-neoplastic polyps showed no significant differences in CADx performance.

This study did not show any significant increase in sensitivity for small, neoplastic polyps during colonoscopy with the use of a specific CADx system compared to standard method. The study also suggested an increase in specificity for optical diagnosis of small

neoplastic polyps and an increase in confidence level for the endoscopists in optical diagnosis of polyps. These results imply that CADx for characterization of polyps cannot mitigate the burden of overdiagnosis and overtreatment stemming from an increase in detected polyps with the use of CADe. Thus, the added clinical value of using CADx for polyp characterization versus the standard method, might be limited to increased confidence level for the endoscopists.

9.3 Article III

Speedometer for withdrawal time monitoring during colonoscopy: A clinical implementation trial

Manuscript

We enrolled 352 patients scheduled for colonoscopy into one of two study periods; pre-implementation period or post-implementation period. In our analysis, 332 patients were included; 166 patients underwent a colonoscopy without the speedometer and 166 patients underwent a colonoscopy with the speedometer. Median patient age was 61 years, and 53% were women.

Withdrawal time was measured as the duration between the cecum base identification and until exit from the rectum. Biopsy and polypectomy time were subtracted to reflect the true withdrawal times. We found that the withdrawal time difference was 2.3 seconds and not significant, comparing colonoscopy without the use of the speedometer and colonoscopy with the use of the speedometer (95%CI -42.3-37.7 p=0.91). There was no significant increase in ADR (p=0.91) comparing colonoscopy done without the speedometer (95%CI 37.6-52.8) and with the speedometer (95%CI 38.2-53.4). There were no significant differences in the proportion of colonoscopies with withdrawal time ≥ 6 minutes comparing colonoscopies without use of the speedometer and colonscopies with the use of the speedometer (p=0.75). Lastly, there were no differences in the mean number of adenomas per colonoscopy

comparing colonoscopies done without use of the speedometer and colonscopies done with the use of the speedometer (p=0.83).

This study shows no benefit from using the speedometer during colonoscopy in order to increase withdrawal time. We found no significant withdrawal time difference, no increased ADR or proportion of colonoscopies with withdrawal time ≥ 6 minutes when using the speedometer. The potential added clinical value of using a speedometer to supplement CADe systems and further increase the ADR by reducing the number of missed adenomas, was not supported by our findings.

10 Discussion of main findings

The purpose of all innovation and new inventions is to solve an existing problem, challenge status quo and enhance existing solutions. In medicine this translates into whether AI can improve clinical outcomes and produce clinical utility. AI has a great potential to transform our healthcare system by enhancing and equipping clinicians with powerful tools, such as AI-tools for colonoscopy and colorectal cancer screening. Thus, the overall objective for this thesis was to investigate the clinical performance of novel technology to optimize colonoscopy for CRC screening. In order to achieve that objective, we focused on the performance of three different AI-based tools each targeting different challenges in colonoscopy practice; CADe (polyp detection), CADx (polyp characterization) and other AI-tools (increasing withdrawal time).

We have found in our meta-analysis, that CADe or AI-tools for polyp detection results in a relative increase of 50% in ADR and 10.3 percentage points in absolute increase. However, the increment in detection is only valid for small polyps and non-advanced adenomas as we also found that CADe showed no difference in the detection of advanced

adenomas known to have the highest potential for malignancy. These findings largely echoes the results from other similar meta-analyses published.[101–107] Based on these significant and promising results on clinical validation of CADe for polyp detection, the implementation into real-world colonoscopy practice may seem imminent.

Clinical implementation will however beg the question of who the CADe will be the most advantageous for; non-expert endoscopists or all endoscopists? The trials included in our meta-analysis did not provide information regarding the effect of using AI on individual levels, thus preventing further assessment. Clinical application of AI in general is considered beneficial especially for non-experts because it can level the field by standardization. However, there might be a levelling-off effect between the ADR and the CRC risk.[41] The association between the ADR and the CRC risk is most significant for endoscopists with low ADRs, for instance 20% in ADR, and less for those with already high ADRs. This association implies that any additional increase in ADR over this threshold, have limited or no benefit in CRC prevention and could potentially add to the risk of overdiagnosis and overtreatment.

This raises the question of whether chasing the highest possible ADR for everyone, truly contributes the most to reducing overall CRC risk. We might also have to rethink what purpose and clinical utility a CADe actually represents. It may seem more cost-effective to focus on the endoscopists that have low ADRs, and one way of doing that is to utilize AI. CADe can standardize and make sure that endoscopists perform with less variability and always above a pre-defined threshold.

In addition to improving ADR for those with low ADRs, we may also focus more on AADR (advanced adenoma detection rate). Our meta-analysis and several others could not show any increments in the detection of those lesions that have the most malign potential, namely larger adenomas and advanced adenomas.[101–107] One possible explanation as to why

CADe currently fails to increase the detection of larger adenomas and advanced adenomas, could be the number of patients included in the studies for the meta-analyses. Superiority of using CADe for detection of advanced adenomas might be feasible if a larger number of patients were included in these studies.

In our study investigating the performance of an AI-based speedometer for withdrawal time monitoring we found no significant increase in withdrawal time difference, ADR or proportion of colonoscopies with withdrawal time ≥ 6 minutes when compared to standard colonoscopy without AI-assistance. Other studies have investigated the combined use of CADe and other AI-based systems for withdrawal time monitoring, and found a significant increase in ADR.[78,80,85] However, for two of these studies,[75,77] the AI-based system for withdrawal time monitoring also included a blind spot detector, which might have influenced the ADR more than the increment in withdrawal time itself. The sole function of an AI-based speedometer does also seem to increase ADR in one study, and the combination of CADe and an AI-based speedometer can increase ADR without prolonging the withdrawal time.[89] However, there have been little focus on reducing withdrawal time, and most focus is on prolonging it in order to increase ADR. More research is needed in order to access the cost-effectiveness of both reducing polypectomies and not increasing withdrawal time beyond recommended threshold.

If CADe becomes widely implemented in clinical practice, it may result in overdiagnosis and overtreatment, which again can lead to more waste of resources both in terms of time and costs. Thus, the need for supplementary CADx that can distinguish between neoplastic and non-neoplastic polyps arises and becomes more urgent. Unfortunately, our multicenter, prospective, trial did not show support for this mitigation of overdiagnosis since it showed no significant increase in sensitivity for optical diagnosis during AI-assisted

colonoscopy compared to non-AI assisted colonoscopy. This corroborates results from one other clinical trial investigating the diagnostic performance of CADx.[92] But the majority of other studies have been showing superiority of using CADe for real-time assessments of colorectal polyps.[91,93–96,98]

Our clinical study is nonetheless the most high-quality study among these previously prospective trials performed (Table 2), due to having the highest number of subjects (1289 included) and is the only multicenter trial to date. Two more prospective, multicenter studies on CADx for colonic polyp characterization are on the way according to data on ClinicalTrials.gov; one from Singapore which is still recruiting (NCT05034185), and one from Italy that has completed recruitment (NCT04607083). Since these studies are not international, the external validity will be lower than our study. Nonetheless, it will be interesting to see if these other multicenter studies also show the same results as we had, or if they show a significant improvement with the use of CADx.

We planned to do a clinical validation of AI for polyp characterization, but we inadvertently also ended up validating ourselves as highly competent endoscopists. We found a sensitivity level of 88.4% without any involvement from CADx. However, this may not reflect the true baseline and could be attributed to the existence of endoscopy training programs at all three study centers, which were all teaching hospitals. Furthermore, the study showed a NPV of 91.5% (95%CI 88.5-93.8) with the standard method and no use of CADx. With the additional use of CADx, NPV increased with 1.3% and performed well above the recommended NPV threshold of >90%, which could support a more widespread use of the diagnose-and-leave strategy.

The clinical impact of these CAD-based systems is still unknown and there is currently few guidelines on the application of AI in colonoscopy. The European Society of

Gastrointestinal Endoscopy (ESGE) published its first guideline endorsing the use of AI during colonoscopy in 2019,[108] and the American Society for Gastrointestinal Endoscopy (ASGE) has published a position statement to accelerate the implementation of AI in endoscopy practice.[109] The recommendation from ESGE was based an expert opinion-based recommendation, and thus can be categorized as a weak and of low-quality evidence. The recommendation to implement CAD-systems in clinical practice should according to this statement require high-quality multicenter in vivo clinical studies.

11 Methodological considerations

We performed a meta-analysis in article I, and had to choose between two main statistical models used for meta-analyses; The fixed-effects modelling or the random-effects modelling. The choice between these methods depends on the underlying assumptions and will impact the conclusions of the meta-analysis.[110] The random-effects modelling assumes that all studies included in a meta-analysis have different true underlying effect sizes that are distributed around a mean. The fixed-effects modelling assumes that all studies included in a meta-analysis have the same true effect size.[110] For our meta-analysis we assumed that the different studies included would have different true effect sizes, e.g., different inclusion or exclusion criteria, different AI-tools for polyp detection used etc. However, since our meta-analysis ended up only including studies from China as opposed to different countries, there is a possibility that fixed-effects modelling assuming same true effect size could also be used. We selected random-effects modelling for our meta-analysis and decided on doing the subtype of random-effects modelling called restricted maximum likelihood (REML) method to aggregate data.

The REML method is preferred when the included studies in a meta-analysis are small, outcomes are rare or the heterogeneity is large.[111] The included studies in our meta-analysis were not so small and the outcomes were not so rare, but some of the studies had outcomes showing a heterogeneity that was considered moderate to large. A more commonly used method than REML is the DerSimonian-Laird method which is the simplest version of random-effects modelling. We could have used the DerSimonian-Laird method and expected similar results. However, since some of the studies included had outcomes showing a heterogeneity that was considered moderate to large, this could negatively bias the estimates using the DerSimonian-Laird method. Simulation studies have also shown that the DerSimonian-Laird method can be negatively biased in scenarios with small studies and in scenarios with a rare outcome. As we from previous research were familiar with the REML-method, it is less biased in many other situations, and therefore recommended over other methods,[111] we chose to use this method in our meta-analysis. We have also analyzed the data using the DerSimonian-Laird method, and found that the point estimates and confidence intervals only changed in the third decimal place.

All analyses are based on per-protocol analyses of the five trials selected and found eligible for our meta-analysis. Although it is known that such studies are often prone to biases as the original randomization is broken, we chose to only include per protocol and not intention-to-treat (ITT) analyses. This was mainly because only one study provided both ITT analysis and per-protocol analysis[112], while the other four studies provided data based solely on the per-protocol analyses. While we consider ITT the best analysis when it comes to results that occur after some follow-up, we do not think this is of importance for the purpose of this study.

In article II, we performed a prospective, open-label, single-arm study. We considered doing a RCT and using positive predictive value (PPV) and negative predictive value (NPV) as primary outcomes. However, we ended up choosing a single-arm study and sensitivity as a primary outcome instead. Sensitivity is the most relevant and generalizable outcome to assess the safety and efficacy of the "diagnose-and-leave" strategy for non-neoplastic polyps which may be achieved with an AI-tool. High sensitivity means low false negative rate, and thus safe diagnose-and-leave polyp strategy. Even though guidelines from ASGE uses NPV to define thresholds for test performance, we feel that sensitivity and also specificity are clinically more relevant. Contrary to the NPV and PPV, the sensitivity and specificity of a diagnostic test are not dependent on the prevalence of the target condition (here adenomas), and gives thus the preferred primary and secondary outcome in this trial.

The rationale for this choice of study design lies in which improvement we are interested in measuring. We wanted to measure whether the use of an AI-tool increased the diagnostic sensitivity in the same endoscopist before and after the use of an AI-tool, which we thought was clinically more relevant and favors the single-arm study design.

This was preferred over measuring the difference in diagnostic performance between an endoscopist without an AI-tool vs. an endoscopist with an AI-tool, and requiring a RCT as study design. However, one disadvantage connected to the prospective study design is the learning curve experienced by participating endoscopists in our study. The learning curve may have resulted in CADx underperformance.

In article III, we chose to do a prospective, clinical implementation trial investigating the performance of an AI-speedometer. When planning our sample size, the rationale was based on two sources of information; firstly, a randomized trial from China[78] which showed a withdrawal time difference between standard colonoscopies and colonoscopies with both a

speedometer and a blind spot detector was 1.6 minutes, and secondly unpublished data collected from 125 colonoscopy cases performed by expert colonoscopists at the same study center. Because the Chinese trial[78] used a non-parametric test to compare withdrawal time between the groups, we assumed that time distribution was not normal, and chose to adjust for that with +/- 5% to our sample size estimates. Thus, with a statistical power of 80%, an alpha of 5% and no intra-cluster correlation coefficient (ICC) among colonoscopists, the required sample size was 299 patients in total. Expecting a 10% drop out, we planned to enroll 332 patients; 166 patients in the pre-implementation period and 166 patients in the post-implementation period, totaling 332 patients enrolled over the course of 4-5 months.

Due to the COVID-19 pandemic, elective procedures such as colonoscopies were cancelled on short notice, and activity levels did not pick up for several months. As a result, our study was postponed for several months as well. However, if we had been able to start enrolling patients as early as planned and allocate more resources during the study periods, we might have selected a different study design, such as a RCT comparing two groups of patients: patients undergoing colonoscopy with an AI-speedometer vs. patients undergoing a colonoscopy with standard method. This would be a more powerful study design, but also one that would require a larger sample size. Given the limitations in study period and cancelled or postponed colonoscopy procedures, this was not realistic or feasible during an ongoing pandemic. Furthermore, one can also argue that a RCT might not have changed the results we got from our study. We had a difference in gender-distribution between our two study periods with one having significantly more women than the other. We adjusted for this gender-difference, but found no difference in our primary outcome or result. On the other hand, since our study design was non-blinded, operational bias and the Hawthorne effect[113] might have influenced the end result and more research evaluating the efficiency of AI-tools with a blinded design could countereffect this bias.

12 Ethical considerations

There are some unique ethical considerations and challenges connected to the application of clinical AI-based tools. The reason for that lies in nature of how AI can serve as clinical decision support and influence physicians' opinions and decisions. If our decisions as clinicians are being supported or influenced, then who will be held accountable for decisions that are wrong, cause harm or fatal mistakes? Who is liable for AI-based misdiagnosis and to what degree? In the case of a CRC screening program that uses AI applications, who will be held accountable if AI-based tools support the strategy of diagnose-and-leave but the patient is identified with an interval cancer years later? Or what if the endoscopist disagree with the AI-based polyp detection system that recommends polypectomy of a lesion that the endoscopists themselves feel is not a lesion. Should they resect anyway and possibly subject the patient to unnecessary risk of bleeding and perforation? These examples justify why many clinicians might feel critical if there is a lack of explainability to AI algorithms.

A relevant academic subfield in AI has emerged the last couple of years focusing on the need for transparency in clinical decision support, called "Explainable AI" (XAI).[114] The purpose of XAI is to provide transparent explanations of how AI-based systems make recommendations and thus avoid "black boxes", which are AI-models without transparency or explanation of how an AI-algorithm performs or delivers prognosis.[115]

In the future, clinicians will need tailored education targeting ethical challenges specific to AI applications in medicine, and learn more about how to interpret outputs. Future physicians may need to recognize underpowered and biased datasets, but at the very least critically evaluate recommendations given by the AI-based tools and not readily accept them.

A clear consensus on medical liability in the case of errors made by an AI algorithm or AI-based system, has not been established but is on the way after the EU announced their proposal for regulation of artificial intelligence in April 2021.[116,117] Today, physicians together with their patients decide on the correct course of treatment and follow-up. In the future, this process might involve clinical AI-systems as well. Thus, it remains to be seen if the physicians will keep their status as "gold standard" and having the final say in clinical decisions, or if AI will challenge this position by outperforming physicians. Much like automation in the automobile industry, where fully automated cars may replace human drivers if they can prove to be safer and more trustworthy. Thus, the future of the physicians' role in patient care may rely on patient trust and explainability of clinical decisions assisted by AI.

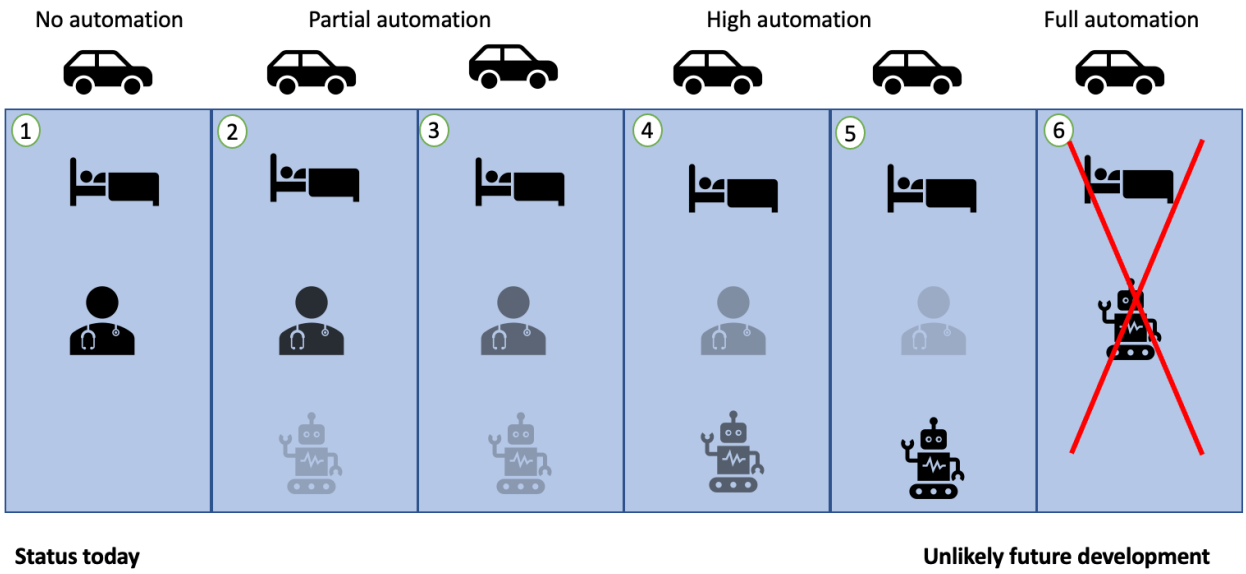


Figure 7: The analogy between self-driving cars and medicine, showing how clinicians' decisions may be more and more diluted by AI until it is fully automated with no physicians involved in medical decision-making. That scenario is today regarded as highly unlikely to occur. Adapted from the original by Topol, E.J.[66]

13 Conclusion and Implications

13.1 The hype

All three articles included in this thesis paint a picture of AI-based tools for colonoscopy, as a technology still in its infancy. For CADe the expectations of increased polyp detection was met. However, due to the result of finding only more small polyps and non-advanced adenomas, the need for additional solutions to handle potential overdiagnosis and overtreatment from CADe also emerged. The CADx did not deliver on our pre-defined performance goal and the AI-based speedometer did not show any significant difference during colonoscopy. More similar studies are currently ongoing and may contradict our findings, but the overall impression of AI underdelivering on its promise will remain as long as there is a paucity for clinical validation of AI applications.

13.1 The hurdles

For AI-based tools to be implemented in colonoscopy and CRC screening in the future, they must show clinical utility and proof of benefit. Although there is an added clinical value of significantly increasing the ADR for cancer prevention, especially up until a certain threshold before seeing a levelling-off effect,[41] that increase must be weighed against the harms from overdiagnosis and overtreatment. Most of the harm and resource use take place during the screening test itself and the surveillance that follows if the screening test is positive. The studies we conducted do not show that AI-based tools add clinical value to colonoscopy for colorectal cancer screening, without simultaneously causing harm. Harm in this context, can be the result of several conditions. One is the increased risk of adverse events such as bleeding and perforation due to a higher number of polypectomies with the use of AI-based tools for polyp detection. Another risk of harm is caused by prolonged procedure time both with and without polypectomies. Our studies on CADe and CADx have shown that the

procedure time increases with about 30-40 seconds due to the application of AI itself. Future cost-effectiveness studies must explore whether prolonging the colonoscopy procedure time pays off with the benefit of an improved health outcome, which potentially can reduce costs. In order to show that AI can improve health outcomes such as CRC incidence and mortality, we need to conduct research within population-based cancer screening programs with 10-15 years of follow-up time. One such large-scale study was recently established and has already started enrolling patients in several European countries and Japan (UMIN000044748).

13.1 The hope

The hope for the future is to clinically prove that AI can lessen the burden of CRC screening with minimal to no harm. This is the general expectation of clinical AI applications, that AI can produce equal access to healthcare and democratize healthcare.[118] As with all technological advancements, there might even be the possibility of doing leapfrogging with AI. Leapfrogging with AI in CRC screening would enable developing countries to go directly from having no CRC screening programs, to having full-scale screening facilities where less experienced endoscopists can utilize AI to compensate for the lack of procedural experience. In other words, they can bypass all the intermediary hurdles and steps that Western countries with functioning CRC programs learned from. Just like leapfrogging with 5G networks.

This thesis does not add to the hype, but rather adds to the current understanding of where AI falls short and what we need in order to go forward and improve. In search of the best technology applications available to solve future challenges, we must keep an open mind and be able to do both of the following two things; not dismiss AI simply because it underdelivers on promises today and not try to make AI become the universal solution to all our problems and thus blocking the idea or possibility that there might be other ways to solve existing and future medical challenges.

13 References

1. Ferlay, J. *et al.* (2020). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [19 April 2022].
2. Schreuders, E. H. *et al.* Colorectal cancer screening: A global overview of existing programmes. *Gut* **64**, 1637–1649 (2015).
3. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
4. Carethers, J. M. & Doubeni, C. A. Causes of Socioeconomic Disparities in Colorectal Cancer and Intervention Framework and Strategies. *Gastroenterology* **158**, 354–367 (2020).
5. Helsingen, L. M. & Kalager, M. Colorectal Cancer Screening-Approach, Evidence, and Future Directions. (2022) doi:10.1056/EVIDra2100035.
6. Cancer in Norway 2020 Cancer incidence, mortality, survival and prevalence in Norway Kreft i Norge-hva sier tallene?
7. Malila, N. & Hakulinen, T. Epidemiological trends of colorectal cancer in the Nordic countries. *Scand. J. Surg.* **92**, 5–9 (2003).
8. Engholm, G. *et al.* (2010) NORDCAN – a Nordic tool for cancer information, planning, quality control and research, Acta Oncologica. **49:5**, 725–736, DOI: 10.3109/02841861003782017.
9. Larønningen, S. *et al.* NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 9.1 (27.09.2021). *Association of the Nordic Cancer Registries. Cancer Registry of Norway. Available from: https://nordcan.iarc.fr/, accessed on 01/05/2022*.
10. Wang, X. *et al.* Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. *BMJ Open Gastroenterol.* **6**, (2019).
11. Islami, F. *et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA. Cancer J. Clin.* **68**, 31–54 (2018).
12. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA. Cancer J. Clin.* **70**, 145–164 (2020).
13. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
14. Markowitz, A. J. & Winawer, S. J. Management of colorectal polyps. *CA. Cancer J. Clin.* **47**, 93–112 (1997).
15. Galuppini, F. *et al.* The histomorphological and molecular landscape of colorectal adenomas and serrated lesions. *Pathol. - J. Ital. Soc. Anat. Pathol. Diagnostic Cytopathol.* **113**, 218–229 (2021).
16. Atkin, W. S. *et al.* European guidelines for quality assurance in colorectal cancer screening and diagnosisFirst Edition Colonoscopic surveillance following adenoma removal. *Endoscopy* **44**, (2012).
17. Vogelstein, B. *et al.* Genetic Alterations during Colorectal-Tumor Development. *N. Engl. J. Med.* **319**, 525–532 (1988).
18. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
19. Morson, B. C. The evolution of colorectal carcinoma. *Clin. Radiol.* **35**, 425–431 (1984).

20. Binefa, G., Rodríguez-Moranta, F., Teule, À. & Medina-Hayas, M. Colorectal cancer: From prevention to personalized medicine. *World J Gastroenterol* **20**, 6786–6808 (2014).
21. Torlakovic, E., Skovlund, E., Snover, D. C., Torlakovic, G. & Nesland, J. M. Morphologic reappraisal of serrated colorectal polyps. *Am. J. Surg. Pathol.* **27**, 65–81 (2003).
22. Patai, Á. V., Molnár, B., Tulassay, Z. & Sipos, F. Serrated pathway: Alternative route to colorectal cancer. *World J. Gastroenterol.* **19**, 607 (2013).
23. Jaravaza, D. R. & Rigby, J. M. Hyperplastic polyp or sessile serrated lesion? The contribution of serial sections to reclassification. *Diagn. Pathol.* **15**, 1–9 (2020).
24. Bettington, M. *et al.* Clinicopathological and molecular features of sessile serrated adenomas with dysplasia or carcinoma. *Gut* **66**, 97–106 (2017).
25. Snover, D. C. Update on the serrated pathway to colorectal carcinoma. *Hum. Pathol.* **42**, 1–10 (2011).
26. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
27. Burnett-Hartman, A. N. *et al.* Original Contribution Differences in Epidemiologic Risk Factors for Colorectal Adenomas and Serrated Polyps by Lesion Severity and Anatomical Site. (2013) doi:10.1093/aje/kws282.
28. Pancione, M., Remo, A. & Colantuoni, V. Genetic and epigenetic events generate multiple pathways in colorectal cancer progression. *Patholog. Res. Int.* **2012**, (2012).
29. Bretthauer, M. & Kalager, M. Principles, effectiveness and caveats in screening for cancer. doi:10.1002/bjs.8995.
30. Jodal, H. C. *et al.* Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. *BMJ Open* **9**, (2019).
31. Kalager, M. *et al.* Overdiagnosis in Colorectal Cancer Screening: Time to Acknowledge a Blind Spot. *Gastroenterology* vol. 155 592–595 (2018).
32. Carter, J. L., Coletti, R. J. & Harris, R. P. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* **350**, (2015).
33. Wieszczy, P. *et al.* Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models. *BMJ Open* **11**, e042158 (2021).
34. Helsingen, L. M. *et al.* Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ* **367**, 12 (2019).
35. Robertson, D. J., Kaminski, M. F. & Bretthauer, M. Effectiveness, training and quality assurance of colonoscopy screening for colorectal cancer. *Gut* (2015) doi:10.1136/gutjnl-2014-308076.
36. Kreftregisteret. Når kommer tarmscreeningtilbudet til deg? *Available from: https://www.kreftregisteret.no/screening/tarmscreening/utrulling/ (Accessed 02.05.2022).*
37. Zauber, A. G. *et al. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths*. *New England Journal of Medicine* vol. 366 687–696 (Massachussetts Medical Society, 2012).
38. Holme, Ø. *et al.* Effect of Flexible Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality: A Randomized Clinical Trial. *JAMA* **312**, 606–615 (2014).
39. Shaukat, A. *et al.* Long-Term Mortality after Screening for Colorectal Cancer. *N. Engl. J. Med.* **369**, 1106–1114 (2013).
40. R. Nishihara, K. Wu, P. Lochhead, et al. Long-Term Colorectal-Cancer Incidence and Mortality after Lower Endoscopy. *N. Engl. J. Med.* **12**, 1095–105 (2013).

41. Kaminski, M. F. *et al.* Quality Indicators for Colonoscopy and the Risk of Interval Cancer. *N. Engl. J. Med.* **362**, 1795–1803 (2010).
42. Corley, D. A. *et al.* Adenoma Detection Rate and Risk of Colorectal Cancer and Death. *N. Engl. J. Med.* **370**, 1298–1306 (2014).
43. Kaminski, M. F. *et al.* Increased Rate of Adenoma Detection Associates With Reduced Risk of Colorectal Cancer and Death. *Gastroenterology* **153**, 98–105 (2017).
44. Berzin, T. M. & Topol, E. J. Adding artificial intelligence to gastrointestinal endoscopy. *Lancet* **395**, 485 (2020).
45. Zhao, S. *et al.* Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. *Gastroenterology* **156**, 1661-1674.e11 (2019).
46. Robertson, D. J. *et al.* Colorectal cancers soon after colonoscopy: A pooled multicohort analysis. *Gut* **63**, 949–956 (2014).
47. Anderson, R., Burr, N. E. & Valori, R. Causes of Post-Colonoscopy Colorectal Cancers Based on World Endoscopy Organization System of Analysis. *Gastroenterology* **158**, 1287-1299.e2 (2020).
48. Lui, T. K. L. *et al.* New insights on missed colonic lesions during colonoscopy through artificial intelligence–assisted real-time detection (with video). *Gastrointest. Endosc.* **93**, 193-200.e1 (2021).
49. Rex, D. K. *et al.* Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* **112**, 24–28 (1997).
50. Barclay, R. L., Vicari, J. J., Doughty, A. S., Johanson, J. F. & Greenlaw, R. L. Colonoscopic Withdrawal Times and Adenoma Detection during Screening Colonoscopy. *N. Engl. J. Med.* **355**, 2533–2541 (2006).
51. Vavricka, S. R. *et al.* Monitoring colonoscopy withdrawal time significantly improves the adenoma detection rate and the performance of endoscopists. *Endoscopy* **48**, 256–262 (2016).
52. Sawhney, M. S. *et al.* Effect of Institution-Wide Policy of Colonoscopy Withdrawal Time ≥7 Minutes on Polyp Detection. *Gastroenterology* **135**, 1892–1898 (2008).
53. Shaukat, A. *et al.* Longer Withdrawal Time Is Associated With a Reduced Incidence of Interval Cancer After Screening Colonoscopy. *Gastroenterology* **149**, 952–957 (2015).
54. Rex, D. K. *et al.* The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest. Endosc.* **73**, 419–422 (2011).
55. Monika Ferlitsch, A. *et al.* Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. | *Endosc.* **49**, 270–297 (2017).
56. Willems, P. *et al.* Uptake and barriers for implementation of the resect and discard strategy: an international survey. *Endosc. Int. open* **8**, E684–E692 (2020).
57. Taghiakbari, M. *et al.* The location-based resect and discard strategy for diminutive colorectal polyps: A prospective clinical study. *Endoscopy* **54**, 354–363 (2022).
58. Abu Dayyeh, B. K. *et al.* ASGE technology committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest. Endosc.* **81**, 502.e1-502.e16 (2015).
59. Vleugels, J. L. A. *et al.* Effects of Training and Feedback on Accuracy of Predicting Rectosigmoid Neoplastic Lesions and Selection of Surveillance Intervals by Endoscopists Performing Optical Diagnosis of Diminutive Polyps. *Gastroenterology* **154**, 1682-1693.e1 (2018).

60. Rees, C. J. *et al.* Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. *Gut* **66**, 887–895 (2017).
61. Raghupathi, W. & Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Heal. Inf. Sci. Syst.* **2**, 1–10 (2014).
62. Ruffle, J. K., Farmer, A. D. & Aziz, Q. Artificial Intelligence-Assisted Gastroenterology - Promises and Pitfalls. *Am. J. Gastroenterol.* **114**, 422–428 (2019).
63. Taghiakbari, M., Mori, Y. & von Renteln, D. Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. *World J. Gastroenterol.* **27**, 8103 (2021).
64. Glissen Brown, J. R., Waljee, A. K., Mori, Y., Sharma, P. & Berzin, T. M. Charting a path forward for clinical research in artificial intelligence and gastroenterology. *Dig. Endosc.* **34**, 4–12 (2022).
65. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med. 2022 281* **28**, 31–38 (2022).
66. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* vol. 25 44–56 (2019).
67. Bohr, A. & Memarzadeh, K. The rise of artificial intelligence in healthcare applications. *Artif. Intell. Healthc.* 25 (2020) doi:10.1016/B978-0-12-818438-7.00002-2.
68. GH., G. Evidence-based medicine. *Ann Intern Med.* **114**, (1991).
69. Guyatt, G. *et al.* Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA* **268**, 2420–2425 (1992).
70. Bossuyt, P. M. M., Reitsma, J. B., Linnet, K. & Moons, K. G. M. Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clin. Chem.* **58**, 12–1636 (2012).
71. Aristidou, A., Jena, R. & Topol, E. J. Bridging the chasm between AI and clinical implementation. *Lancet* **399**, 620 (2022).
72. Nagendran, M. *et al.* Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies in medical imaging. *BMJ* **368**, m689 (2020).
73. CE marking. https://ec.europa.eu/growth/single-market/ce-marking_en.
74. AI for radiology. https://grand-challenge.org/aiforradiology/.
75. Van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M., Van Ginneken, B. & De Rooij, M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. doi:10.1007/s00330-021-07892-z/Published.
76. Wang, P. *et al.* Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* **68**, 1813–1819 (2019).
77. Repici, A. *et al.* Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* **159**, 512-520.e7 (2020).
78. Gong, D. *et al.* Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol. Hepatol.* (2020) doi:10.1016/s2468-1253(19)30413-3.
79. Liu, W. N. *et al.* Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J. Gastroenterol.* **26**, 13–19 (2020).
80. Su, J. R. *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.* **91**, 415-424.e4 (2020).
81. Wang, P. *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. **5**, 343–351 (2020).
82. Luo, Y. *et al.* Artificial Intelligence-Assisted Colonoscopy for Detection of Colon Polyps: a Prospective, Randomized Cohort Study. *J. Gastrointest. Surg.* **25**, 2011 (2021).
83. Repici, A. *et al.* Artificial intelligence and colonoscopy experience: lessons from two randomised trials. *Gut* **71**, 757–765 (2021).
84. Xu, L. *et al.* Artificial intelligence-assisted colonoscopy: A prospective, multicenter, randomized controlled trial of polyp detection. *Cancer Med.* **10**, 7184 (2021).
85. Yao, L. *et al.* An artificial intelligence-based quality improvement system significantly improved the efficacy of computer-aided detection system in colonoscopy: A four group parallel study. *Endoscopy* (2021) doi:10.1055/A-1706-6174.
86. Glissen Brown, J. R. *et al.* Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial). *Clin. Gastroenterol. Hepatol.* (2022) doi:10.1016/J.CGH.2021.09.009.
87. Kamba, S. *et al.* Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. *J. Gastroenterol.* **56**, 746–757 (2021).
88. Wang, P. *et al.* Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study. *Gastroenterology* **159**, 1252-1261.e5 (2020).
89. Wallace, M. B. *et al.* Impact of Artificial Intelligence on Miss Rate of Colorectal Neoplasia. *Gastroenterology* **0**, (2022).
90. Rondonotti, E. *et al.* Efficacy of a Computer Aided Detection (CADe) system in a FIT-based organized colorectal cancer screening program: a randomized controlled trial (AIFIT study). *Endoscopy* (2022) doi:10.1055/A-1849-6878.
91. Aihara, H. *et al.* Computer-aided diagnosis of neoplastic colorectal lesions using 'real-time' numerical color analysis during autofluorescence endoscopy. *Eur. J. Gastroenterol. Hepatol.* **25**, 488–494 (2013).
92. Kuiper, T. *et al.* Automatic optical diagnosis of small colorectal lesions by laser-induced autofluorescence. *Endoscopy* **47**, 56–62 (2015).
93. Rath, T. *et al.* In vivo real-time assessment of colorectal polyp histology using an optical biopsy forceps system based on laser-induced fluorescence spectroscopy. *Endoscopy* **48**, 557–562 (2016).
94. Kominami, Y. *et al.* Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest. Endosc.* **83**, 643–649 (2016).
95. Mori, Y. *et al.* Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Ann. Intern. Med.* **169**, 357–366 (2018).
96. Horiuchi, H. *et al.* Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software. *Scand. J. Gastroenterol.* **54**, 800–805 (2019).
97. Barua, I. *et al.* Real-Time Artificial Intelligence–Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy. *NEJM Evid.* (2022) doi:10.1056/EVIDOA2200003.
98. Minegishi, Y. *et al.* Comprehensive Diagnostic Performance of Real-Time Characterization of Colorectal Lesions Using an Artificial Intelligence–Assisted System: A Prospective Study. *Gastroenterology* **0**, (2022).
99. Akl, E. A. *et al.* Specific instructions for estimating unclearly reported blinding status

in randomized trials were reliable and valid. *J. Clin. Epidemiol.* **65**, 262–267 (2012).
100. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* **8**, 336–341 (2010).
101. Hassan, C. *et al.* Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest. Endosc.* **93**, 77-85.e6 (2021).
102. Deliwala, S. S. *et al.* Artificial intelligence (AI) real-time detection vs. routine colonoscopy for colorectal neoplasia: a meta-analysis and trial sequential analysis. *Int. J. Colorectal Dis.* **36**, 2291–2303 (2021).
103. Mohan, B. P. *et al.* Real-time computer aided colonoscopy versus standard colonoscopy for improving adenoma detection rate: A meta-analysis of randomized-controlled trials. *EClinicalMedicine* **29**–**30**, (2020).
104. Aziz, M. *et al.* The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: A systematic review with meta-analysis. (2020) doi:10.1111/jgh.15070.
105. Li, J., Lu, J., Yan, J., Tan, Y. & Liu, D. Artificial intelligence can increase the detection rate of colorectal polyps and adenomas: A systematic review and meta-analysis. *Eur. J. Gastroenterol. Hepatol.* 1041–1048 (2021) doi:10.1097/MEG.0000000000001906.
106. Zhang, Y., Zhang, X., Wu, Q., Gu, C. & Wang, Z. Artificial Intelligence-Aided Colonoscopy for Polyp Detection: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *J. Laparoendosc. Adv. Surg. Tech. A* **31**, 1143–1149 (2021).
107. Ashat, M. *et al.* Impact of real-time use of artificial intelligence in improving adenoma detection during colonoscopy: A systematic review and meta-analysis. *Endosc. Int. Open* **09**, E513–E521 (2021).
108. Bisschops, R. *et al.* Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy* **51**, 1155–1179 (2019).
109. Berzin, T. M. *et al.* Position statement on priorities for artificial intelligence in GI endoscopy: a report by the ASGE Task Force. *Gastrointest. Endosc.* **92**, 951–959 (2020).
110. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* **1**, 97–111 (2010).
111. Langan, D. *et al.* A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res. Synth. Methods* **10**, 83–98 (2019).
112. Gong, D. *et al.* Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet. Gastroenterol. Hepatol.* **22**, 22 (2020).
113. McCambridge, J., Witton, J. & Elbourne, D. R. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *J. Clin. Epidemiol.* **67**, 267 (2014).
114. Holzinger, A., Langs, G., Denk, | Helmut, Zatloukal, K. & Müller, H. Causability and explainability of artificial intelligence in medicine. (2019) doi:10.1002/widm.1312.
115. Watson, D. S. *et al.* Clinical applications of machine learning algorithms: beyond the black box. *BMJ* **364**, (2019).
116. The European Medical Devices Regulation (EU) 2017/745 (EU MDR) . https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02017R0745-20170505.

117. Beckers, R., Kwade, Z. & Zanca, F. The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. *Phys. Medica* **83**, 1–8 (2021).
118. Topol, E. J. *Deep medicine : How artificial intelligence can make healthcare human again*. (Basic Books, Inc., USA., 2019).

14 Articles and Supplements

Article I

Article II

NEJM Evidence

Published April 13, 2022

DOI: 10.1056/EVIDoa2200003

ORIGINAL ARTICLE

Real-Time Artificial Intelligence–Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy

Ishita Barua, M.D.,[1,2] Paulina Wieszczy, Ph.D.,[1,2,3] Shin-ei Kudo, M.D.,[4] Masashi Misawa, M.D.,[4] Øyvind Holme, M.D.,[1,2,5] Shraddha Gulati, M.D.,[6] Sophie Williams, M.D.,[6] Kensaku Mori, Ph.D.,[7] Hayato Itoh, Ph.D.,[7] Kazumi Takishima, M.D.,[4] Kenichi Mochizuki, M.D.,[4] Yuki Miyata, M.D.,[4] Kentaro Mochida, M.D.,[4] Yoshika Akimoto, M.D.,[4] Takanori Kuroki, M.D.,[4] Yuriko Morita, M.D.,[4] Osamu Shiina, M.D.,[4] Shun Kato, M.D.,[4] Tetsuo Nemoto, M.D.,[8] Bu Hayee, M.D.,[6] Mehul Patel, M.D.,[6] Nishmi Gunasingam, M.D.,[6] Alexandra Kent, M.D.,[6] Andrew Emmanuel, M.D.,[6] Carl Munck, M.D.,[9] Jens Aksel Nilsen, M.D.,[9] Stine Astrup Hvattum, M.D.,[9] Svein Oskar Frigstad, M.D.,[9] Petter Tandberg, M.D.,[9] Magnus Løberg, M.D.,[1,2] Mette Kalager, M.D.,[1,2] Amyn Haji, M.D.,[6] Michael Bretthauer, M.D.,[1,2] and Yuichi Mori, M.D.[1,2,4]

Abstract

BACKGROUND Artificial intelligence using computer-aided diagnosis (CADx) in real time with images acquired during colonoscopy may help colonoscopists distinguish between neoplastic polyps requiring removal and nonneoplastic polyps not requiring removal. In this study, we tested whether CADx analyzed images helped in this decision-making process.

METHODS We performed a multicenter clinical study comparing a novel CADx-system that uses real-time ultra-magnifying polyp visualization during colonoscopy with standard visual inspection of small (≤5 mm in diameter) polyps in the sigmoid colon and the rectum for optical diagnosis of neoplastic histology. After committing to a diagnosis (i.e., neoplastic, uncertain, or nonneoplastic), all imaged polyps were removed. The primary end point was sensitivity for neoplastic polyps by CADx and visual inspection, compared with histopathology. Secondary end points were specificity and colonoscopist confidence level in unaided optical diagnosis.

RESULTS We assessed 1289 individuals for eligibility at colonoscopy centers in Norway, the United Kingdom, and Japan. We detected 892 eligible polyps in 518 patients and included them in analyses: 359 were neoplastic and 533 were nonneoplastic. Sensitivity for the diagnosis of neoplastic polyps with standard visual inspection was 88.4% (95% confidence interval [CI], 84.3 to 91.5) compared with 90.4% (95% CI, 86.8 to 93.1) with CADx (P=0.33). Specificity was 83.1% (95% CI, 79.2 to 86.4) with standard visual inspection and 85.9% (95% CI, 82.3 to 88.8) with CADx. The proportion of polyp assessment with high confidence was 74.2% (95% CI, 70.9 to 77.3) with standard visual inspection versus 92.6% (95% CI, 90.6 to 94.3) with CADx.

Drs. Barua, Wieszczy, and Kudo contributed equally as co-first authors, and Drs. Haji, Bretthauer, and Mori contributed equally as co-last authors.

The author affiliations are listed at the end of the article.
Dr. Bretthauer can be contacted at michael.bretthauer@medisin.uio.no or at Clinical Effectiveness Research Group, Institute of Health and Society, University of Oslo, Oslo 0027, Norway.

CONCLUSIONS Real-time polyp assessment with CADx did not significantly increase the diagnostic sensitivity of neoplastic polyps during a colonoscopy compared with optical evaluation without CADx. (Funded by the Research Council of Norway [Norges Forskningsråd], the Norwegian Cancer Society [Kreftforeningen], and the Japan Society for the Promotion of Science; UMIN number, UMIN000035213.)

Introduction

Colorectal cancer is the third most common cancer and the second leading cause of cancer deaths worldwide.[1] Removal of precancerous polyps during colonoscopy is the cornerstone of colorectal cancer screening. Most colorectal polyps are small (≤5 mm in diameter) and located in the sigmoid colon and the rectum. Although most colorectal cancers develop from polyps, many small polyps are not neoplastic and do not have any malignant potential.[2]

With current standard colonoscopy equipment, many endoscopists, especially those with less experience, cannot reliably distinguish between neoplastic and nonneoplastic polyps on visual inspection, a procedure known as "optical diagnosis."[3,4] Therefore, the current standard of care is to remove all polyps and submit them for histopathologic diagnosis. Reliable real-time optical diagnosis of small polyps during colonoscopy could enable targeted removal only of polyps classified as neoplastic, while small, nonneoplastic polyps could be left behind.[5]

In a recent single-center, proof-of-concept study of a novel artificial intelligence (AI) system for computer-aided polyp diagnosis (CADx), we achieved a reliable distinction between small neoplastic and nonneoplastic polyps in the distal colon and the rectum.[6] The CADx system combines colonoscopes with 520× magnification of polyp surfaces during colonoscopy in real time, and it enables AI-derived automated optical diagnosis of neoplastic and nonneoplastic polyps in about 40 seconds. The automated diagnosis is signaled to the colonoscopist by an acoustic and optical alarm during each polyp assessment.[6-8]

The current multicenter clinical study was designed to compare the clinical performance of AI CADx-based optical diagnosis in distinguishing neoplastic from nonneoplastic small polyps in the sigmoid colon and the rectum during colonoscopy with standard visual inspection–based optical diagnosis in routine clinical colonoscopy practice.

Methods

STUDY DESIGN AND OVERSIGHT

We performed a multicenter clinical study of AI CADx polyp classification and visual inspection versus standard visual inspection alone. Study procedures were performed at three participating endoscopy centers: Baerum Hospital (Norway), King's College Hospital London (United Kingdom), and Showa University Northern Yokohama Hospital (Japan).

The institutional review board (IRB) at each of the three participating centers approved the conduct of the study. The study protocol and statistical analysis plan are available with the full text of this article at evidence.nejm.org. Patient consent was implemented at the three study sites according to local IRB practice; In Norway, only participants enrolled in the national screening program pilot were eligible for participation and written informed patient consent was included in the consent of the screening program. In Japan, the IRB approved an opt-out consent approach because of the low risk related to the study intervention (standard treatment was performed for all polyps detected). In London, all patients provided informed consent.

All co-authors agreed on publishing the article and vouch for the completeness and accuracy of the data and the adherence to the protocol.

PATIENTS

Eligible patients were individuals 18 years of age or older who were scheduled for colonoscopy for colorectal cancer screening, polyp surveillance, or evaluation of clinical signs or symptoms at the participating centers between May 2019 and May 2021. Exclusion criteria were inflammatory bowel disease, polyposis syndrome (familial adenomatous polyposis, serrated polyposis), history of or current chemotherapy or radiation for rectosigmoid tumors, inability to undergo polypectomy (e.g., anticoagulants, comorbidities), pregnancy, and referral for removal of polyps with known histology.

All patients with small polyps (≤5 mm in diameter) in the sigmoid colon or the rectum (jointly called rectosigmoid

colon) detected during colonoscopy were included in this study. For patients with more than five eligible polyps, the first five polyps were included and evaluated according to the study interventions (described next).

COLONOSCOPY PROCEDURES

All colonoscopies were performed according to routine standards at the participating centers, including preprocedure assessment, bowel preparation, sedation practices, and postprocedure recovery and care.

The following information was assessed and was registered in the study database immediately during and after each procedure: indication for colonoscopy, quality of bowel preparation assessed by the Boston Bowel Preparation Score (a 9-point assessment scale for cleaning quality during colonoscopy, with higher numbers indicating better preparation)[9]; most proximal segment of the colon reached during colonoscopy; insertion and withdrawal duration; and size, shape, and location of all detected polyps. All detected polyps were removed for histologic assessment for final diagnosis. By study design, study colonoscopists were nonexperts, defined as having between 1 and 5 years of colonoscopy experience or having independently performed between 200 and 1000 procedures before joining the study as an endoscopist. This aspect of the study design was included because we wanted to determine whether CADx improved the performance of reasonably trained, but nonexpert, endoscopists and thus shortened the learning curve in endoscopy training so the study colonoscopists behaved like experts. The study endoscopists were accredited for standard colonoscopy in the participating countries, but they did not have additional training in optical polyp diagnosis before the study. For the purpose of this study, study endoscopists received training on handling the study colonoscopes and devices and image interpretation. Novice endoscopists were not included because they are unlikely to make optical diagnoses independently from supervisors in clinical practice.

EQUIPMENT

The study centers were provided with high-resolution magnification colonoscopes (CF-H290ECI; Olympus Corp., Tokyo, Japan). These appear to be standard instruments by design, feel, and function, including narrow band imaging. In addition, the study colonoscope featured a light microscopy system integrated into the distal tip of the colonoscope. The extra feature provided 520-fold magnification at a focusing depth of 35 μm, and a field of view of 570 × 500 μm, for high-resolution magnified images on demand, which the colonoscopist controlled with a hand-operated lever.[6] This feature enabled real-time, in vivo evaluation of polyp microvascular morphology.

AI SYSTEM

The study centers were also provided with a real-time polyp classification CADx device (EndoBRAIN; Cybernet Systems Corp., Tokyo, Japan), connected to a standard colonoscopy processor unit (EVIS LUCERA ELITE, CV-290; Olympus Corp.). As noted earlier, the CADx system provides an automated diagnosis of rectosigmoid polyps by analyzing images obtained in the magnification mode of the colonoscopes for detected polyps, as previously described.[6-8]

Briefly, the CADx algorithm comprises three steps. The first is feature extraction, which is the analysis of textures characterized by differences in contrast for polyp vessels and lumens, quantified in 312 validated variables. Second is classification, which comprises support-vector machine classification of polyps as nonneoplastic or neoplastic on the basis of the 312 variables through machine learning. For the system training and validation, more than 35,000 polyp images were used which were collected from five Japanese endoscopy centers, as described previously.[10] Finally, in the diagnostic output step, the predicted diagnosis is displayed (Fig. 1) for the colonoscopist as "neoplastic" or "nonneoplastic" with a confidence probability for neoplasia (0 to 100%).

If the CADx diagnosis has a confidence probability of less than 70%, the system flags it as "low confidence," on the basis of a previous preclinical study.[10] If the quality of the captured image is not appropriate for system diagnosis (e.g., artifacts caused by mucus, low image quality), the analysis is flagged as "not a good sample," and no diagnosis is provided.

The nonneoplastic category comprises polyps with no neoplastic features, such as hyperplastic polyps, inflammatory polyps, and juvenile polyps. The neoplastic category comprises polyps with neoplastic features, such as adenomas and cancers.

POLYP HANDLING

For each detected polyp, four consecutive steps were applied. Step 1 comprised the standard endoscopic assessment. First, colonoscopists assessed the size, shape, and appearance of each detected polyp 5 mm or less in diameter in the rectosigmoid colon. Morphology was categorized

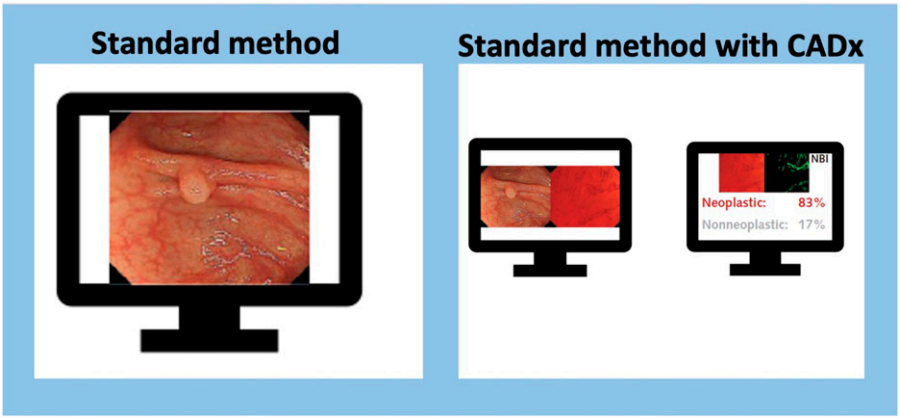


Figure 1. The Standard Method and the Combined Use of the Standard Method and the CADx System.
The Cybernet Systems EndoBRAIN system was used in this study. CADx denotes computer-aided diagnosis.

according to the Paris classification.[11] The endoscopists then classified polyps as either neoplastic (adenoma) or nonneoplastic (nonadenoma) using a binary scale (i.e., low or high confidence level in a nonneoplastic diagnosis, following recommendations in current guidelines[12-14]). Once the endoscopist registered their optical diagnosis, the CADx predicted classification was reported immediately for each polyp and registered in the study database.

Step 2 was the CADx assessment. After the standard assessment as described earlier, colonoscopists captured at least five images from each polyp using narrow band imaging and magnification mode to feed the CADx system. The CADx system then immediately provided the suggested diagnosis of the polyp as either neoplastic or nonneoplastic according to the algorithms described earlier (Fig. 1).

Step 3 was performed after standard clinical assessment and after CADx assessment, respectively. The colonoscopist again scored the confidence level of classification prediction of each polyp as either "high" or "low" and relayed it to the study nurse for immediate capture in the study database.

In step 4, all polyps were removed by snare polypectomy, biopsy forceps, or endoscopic mucosal resection and submitted for histopathologic evaluation. All polyps were evaluated by board-certified (the local board for each country of practice) gastrointestinal pathologists at each center. All pathologists were blinded to colonoscopic diagnoses of the polyps.

All polyps that were diagnosed histopathologically as nonneoplastic but had been considered by the colonoscopist as neoplastic with high confidence after standard assessment were submitted for a second histopathologic review by a different pathologist. The second pathologist was blinded to the first histopathologic diagnosis. See Supplementary Appendix, Section 2 for details.

STUDY END POINTS

The primary endpoint of the study was to compare the sensitivity of identifying small (≤5 mm in diameter) polyps in the rectosigmoid colon as adenomas during colonoscopy with the combination of standard visual inspection and the CADx system, and of standard visual inspection alone, compared with gold-standard histopathology.

Secondary outcome measures included specificity, positive predictive value (PPV), negative predictive value (NPV), rate of high-confidence optical diagnosis, and rate of rectosigmoid polyps of 5 mm or less with adequate images captured for CADx analysis.

Polyps that were not removed, those that were nonepithelial (neuroendocrine polyps, lymphoid aggregates), and those with unsuccessful image capturing were excluded from analyses.

SAMPLE SIZE CALCULATION

On the basis of a pilot study in Japan, we assumed a 6.7-percentage-point increase in sensitivity with the CADx system compared with the standard method, assuming discordance between the two methods of 14.4 percentage points (see the study protocol at evidence.nejm.org). We considered this difference to be clinically meaningful

to uncover. With a statistical power of 90%, the required sample size using a two-sided 5% significance level was 345 neoplastic polyps. We estimated that we needed to enroll 767 patients on the basis of a 25% prevalence of neoplastic eligible polyps, a mean of two eligible polyps per patient, and 90% of polyps with satisfactory prediction by the CADx system. The 90% threshold was motivated by U.S. guidelines recommending an NPV of 90% or greater for optical diagnosis of small neoplastic polyps.[5]

STATISTICAL ANALYSES

Sensitivity, specificity, PPV, and NPV for the standard method and the CADx method compared with histopathology, respectively, were estimated using generalized estimating equation analyses with exchangeable correlation accounting for correlation between multiple polyps within one patient. We did not account for clustering within colonoscopist, site, or country. We calculated 95% confidence intervals (CIs) using sandwich estimates of the variance. Sensitivity and specificity of the two interventions were compared using an exact version of the McNemar test. We did not adjust for multiple comparisons. Polyps that were not removed, from which specimens were lost after removal, or that had nonepithelial histology were excluded from analyses. All tests were performed in relation to the 0.05 significance level and used R version 3.4.1 and Stata version 17 software.

In primary analyses of sensitivity and specificity, sessile serrated lesions were classified as neoplastic (similar to adenomas). In secondary analyses, sessile serrated lesions were classified as nonneoplastic (no adenomas).

No interim analysis was planned at the study start in 2019. Because of slow recruitment during the Covid-19 pandemic, the study team decided to amend the protocol and performed a blinded interim analysis in April 2020. The interim analysis applied an a priori stopping rule for futility (see details in the study protocol on evidence.nejm.org), which was not met. Thus, the study was continued until preplanned recruitment was fulfilled. Because of the blinded nature of the interim analysis, we did not adjust for it in the final analysis.

Results

PATIENTS

The median age of patients included in analyses was 67 years (interquartile range [IQR], 60 to 74), and 63.1% were men (Table 1). Of the 1242 patients who underwent study colonoscopy, 525 had 903 eligible rectosigmoid polyps that received visual inspection.

Of the 903 eligible polyps, 11 were not included in analyses. Of these, 5 were not removed, 3 were lost after removal, and 3 were nonepithelial (two neuroendocrine tumors and one leiomyoma). Consequently, 892 polyps (359 neoplastic polyps and 533 nonneoplastic polyps) from 518 patients were included in the analyses (Fig. 2). The distribution of sex and age of the participants reflects real-world clinical practice (Table S2). We did not register the race and ethnicity of participants.

Twenty-two colonoscopists, including 20 physicians and two nurse endoscopists, performed the study procedures.

COLONOSCOPY PERFORMANCE AND COMPLICATIONS

Baseline characteristics of patients and colonoscopy performance are shown in Table 2. Most colonoscopies were for colorectal cancer screening or polyp surveillance. The median colonoscopy insertion time was 12 minutes (IQR, 8 to 19), and the median withdrawal time with polyp assessments and polypectomies was 28 minutes (IQR, 20 to 40). We did not observe any complications or adverse events related to the colonoscopy or to polyp assessment or removal.

POLYP CHARACTERISTICS

The 518 eligible patients had 892 detected and removed polyps that were 5 mm or less in the rectosigmoid colon. On the basis of the histopathologic examination of the removed polyps, 359 were neoplastic. Of these, 319 were tubular adenomas with low-grade dysplasia, 2 were tubular adenomas with high-grade dysplasia, 9 were tubulovillous adenomas with low-grade dysplasia, and 3 were tubulovillous adenomas with high-grade dysplasia. Of the 26 remaining polyps that were categorized as neoplastic, 7 were traditional serrated adenomas with low-grade dysplasia and 19 were sessile serrated lesions without dysplasia. On the basis of histopathologic examination, 533 polyps were found to be nonneoplastic. Of these, 485 were hyperplastic polyps, 8 were inflammatory polyps, and 40 had other nonneoplastic histology.

PERFORMANCE OF OPTICAL DIAGNOSIS

In primary analyses, the sensitivity for neoplastic polyps was 88.4% (95% CI, 84.3 to 91.5) with the standard

Table 1. Baseline Characteristics of 518 Included Patients and Their Colonoscopies.*

Characteristic	Value
Median age — yr	67 (60 to 74)
Sex	
Men	327 (63.1)
Women	191 (36.9)
Colonoscopy Indication	
Screening colonoscopy (primary screening or fecal testing)	266 (51.4)
Polyp surveillance colonoscopy	161(31.1)
Clinical signs or symptoms	67 (12.9)
Therapy of large polyps	23 (4.4)
Other	1 (0.2)
Median insertion time — min	12 (8 to 19)
Median withdrawal time — min	28 (20 to 40)
Preparation quality good or very good†	481 (92.9)

* Data are presented as the median (interquartile range) or no. (%).

† The Boston Bowel Preparation Scale is a 9-point assessment scale for cleaning quality during colonoscopy. The colon is divided into three segments: proximal, transverse, and distal. Each segment is classified from 0 to 3 depending on the degree of soiling. The sum total of the three segments represents the degree of soiling (≤5 points indicates poor bowel preparation; 6–7 good bowel preparation, and ≥8 very good bowel preparation).[9]

method and 90.4% (95% CI, 86.8 to 93.1) with the CADx method (P=0.33). The percentage of discordant pairs between the standard method and the CADx method was 7.2% (Fig. 3).

The specificity for neoplastic polyps was 83.1% (95% CI, 79.2 to 86.4) with the standard method and 85.9% (95% CI, 82.3 to 88.8) with the CADx method. The discordance between the standard method and the CADx method was 7.9%.

The percentage of polyp assessments with high confidence for categorization into neoplastic or nonneoplastic polyp increased from 74.2% (95% CI, 70.9 to 77.3) with the standard method to 92.6% (95% CI, 90.6 to 94.3) with the CADx method.

In secondary analyses classifying sessile serrated lesions as nonneoplastic, the sensitivity for neoplastic polyps was 91.2% (95% CI, 87.5 to 93.9) with the standard method and 94.1% (95% CI, 91.2 to 96.2) with the CADx method. The specificity for neoplastic polyps was 82.3% (95% CI, 78.4 to 85.6) with the standard method and 85.5% (95% CI, 81.9 to 88.5) with the CADx method. For separate center analyses, see Tables S3 through S8.

Discussion

Implementation of AI in cancer screening and clinical diagnosis requires proof of benefits from high-quality clinical studies. Our international multicenter study assessed the incremental gain of a specific CADx AI system for real-time polyp assessment during colonoscopy. Our study indicates that real-time AI with CADx may not significantly increase the sensitivity for small neoplastic polyps. However, CADx may improve specificity for optical diagnosis of small neoplastic polyps and increase colonoscopist confidence with visual diagnosis of polyps.

AI polyp detection tools (so-called computer-aided polyp detection) during colonoscopy could potentially increase detection of small polyps by up to 50%.[15] While this potentially could increase screening benefit, it also increases health care costs, risk of overtreatment, and patient burden.[16] Most additionally detected polyps are small ones in the distal colon and the rectum, and many of these are nonneoplastic; that is, they do not need to be removed if reliable, real-time classification were possible. One may further argue that removal of small polyps contributes little in terms of cancer prevention.[17]

The "diagnose-and-leave" strategy recently proposed by the American Society for Gastrointestinal Endoscopy (ASGE) suggests not to remove small polyps during colonoscopy if they can be reliably classified (defined as NPV of ≥90%) by optical diagnosis as nonneoplastic.[5] This strategy is not easy to apply because such reliable diagnosis is difficult to achieve with standard colonoscopy systems. Our study provides high-quality data to address this critical issue.

Our main outcome did not reach the prespecified increase of 6.7% in sensitivity with CADx, which was based on preclinical testing, observational studies, and a single-center study. Our study thus emphasizes the importance of rigorous clinical studies to assess AI performance and quantifies the added value and the limitation of CADx in colonoscopy.

According to our results, CADx may not reduce overlooking adenomas during visual inspection of polyps. However, our study showed a potential improvement in specificity for neoplastic polyps, albeit one in which we cannot declare statistical significance because our primary outcome failed to reach that level with the CADx system. There was also a trend toward improved confidence in

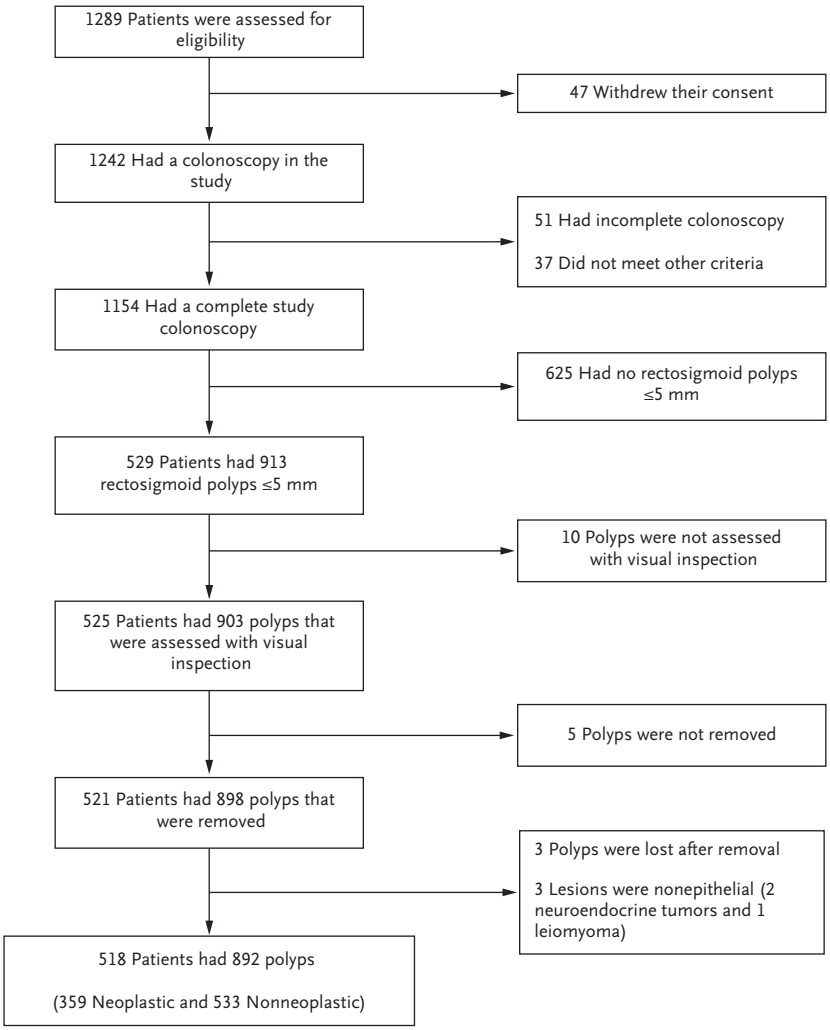


Figure 2. Study Flow Chart.

optical diagnosis of polyps. If this can be established through additional clinical trials, it could potentially contribute to a clinically important reduction in the unnecessary removal of small nonneoplastic polyps by giving the operator the ability to make a high-confidence prediction during a procedure.[5]

PPVs and NPVs are influenced by the prevalence of disease (polyps) and do not adequately assess tools or devices as such. Therefore, our primary outcomes of interest were sensitivity and specificity. However, we also analyzed the predictive values of CADx and observed increments of 1.3% for NPV and 3.1% for PPV with CADx (Table 3). Our results are consistent with the hypothesis that CADx can fulfill the criteria for the diagnose-and-leave strategy with 95% CIs above the NPV threshold of 90%.

The strengths of the current study are the comparison with both non-AI optical diagnosis and gold-standard histopathology for all included polyps; the inclusion of centers from different countries and continents; and the focus on endoscopists with average experience and workload, mimicking real-world colonoscopy practice. A limitation of this study is the inability of the CADx tool to identify sessile serrated polyps, a recently recognized polyp type with likely neoplastic potential. To alleviate this challenge, we conducted two analyses (one classifying sessile serrated polyps as neoplastic and the other classifying them as nonneoplastic) without significant differences in the performance of the CADx tool. Another limitation is the learning curve of the colonoscopists during the study period due to the prospective study design, which may contribute to underestimation of the CADx performance. However, we

Table 2. Characteristics of the 892 Small Polyps (≤5 mm in diameter) in the Distal Colon and the Rectum.*		
Characteristic	**Neoplastic Polyps (n=359)**	**Nonneoplastic Polyps (n=533)**
Median size — mm	4 (3 to 5)	3 (2 to 3)
Location		
Sigmoid colon	274 (76.3)	260 (48.8)
Rectum	85 (23.7)	273 (51.2)
Morphology†		
Polypoid (type Is or Ip)	175 (48.7)	109 (20.5)
Nonpolypoid (type IIa)	184 (51.3)	424 (79.5)
Removal method		
Snare polypectomy	247 (68.8)	265 (49.7)
Forceps	65 (18.1)	258 (48.4)
Endoscopic mucosal resection	46 (12.8)	10 (1.9)

* Data are presented as the median (interquartile range) or no. (%). Sessile serrated lesions were classified as neoplastic polyps in the primary analysis.

† The Paris classification was used. Morphologic classification systems for polyps during colonoscopy classify polyps into polypoid and nonpolypoid, with six different subtypes.[12]

may also have overestimated nonexpert endoscopists' performance because the sensitivity we found to predict adenomas, without the aid of CADx, was 88.4%, which is slightly higher than that reported in previous studies.[18,19] This may be related to the fact that our study was conducted at teaching hospitals with endoscopy training programs.

Finally, the colonoscopes used in the current study are not widely used today, although they are commercially available in Europe, the Middle East, and Asia. Provided that colonoscopes with surface enhancement functions facilitating CADx systems like the one we tested prove to be useful, they would likely become used more widely.

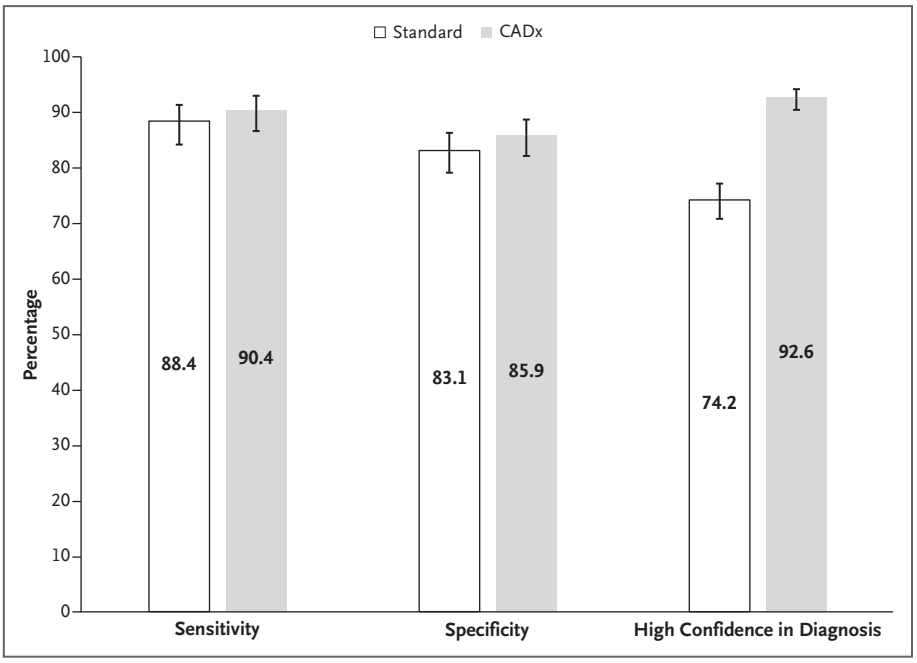


Figure 3. Sensitivity, Specificity, and Confidence of Diagnosis of Standard and AI-Derived CADx Optical Diagnosis of Small Rectosigmoid Polyps during Colonoscopy Compared with Histopathology.

All bars are represented with corresponding 95% confidence intervals. AI denotes artificial intelligence and CADx computer-aided diagnosis.

Table 3. Performance of Standard and AI-Derived CADx Optical Diagnosis of Small Rectosigmoid Polyps during Colonoscopy Compared with Histopathology.*

Parameter	Standard Diagnosis	CADx Diagnosis
Sensitivity	88.4 (84.3 to 91.5)	90.4 (86.8 to 93.1)
Specificity	83.1 (79.2 to 86.4)	85.9 (82.3 to 88.8)
Positive predictive value	78.9 (74.3 to 82.9)	82.0 (77.6 to 85.6)
Negative predictive value	91.5 (88.5 to 93.8)	92.8 (90.1 to 94.9)
High confidence in optical diagnosis	74.2 (70.9 to 77.3)	92.6 (90.6 to 94.3)

* Sessile serrated lesions were classified as neoplastic polyps according to the primary analysis plan. Values are presented as percentages (95% confidence intervals). AI denotes artificial intelligence and CADx denotes computer-aided diagnosis.

Our study suggests that the use of CADx helped the provider have higher confidence in optical diagnosis. If this can be replicated, it could contribute to cost reduction because more polyps could be left in situ. Better confidence comes at a cost; CADx assessment prolongs colonoscopy procedure time, which increases health care cost. In previous studies, we demonstrated that the time necessary for CADx assessment of one small polyp, as applied in this study, is about 40 seconds.[6] We consider this additional time well spent with regard to the gain in terms of reduction of unnecessary removal of polyps and histopathologic assessment. Future cost-effectiveness studies may explore whether the prolonged procedure time pays off with the benefit of reduced polypectomies.

In conclusion, real-time assessment with CADx did not significantly increase sensitivity for neoplastic polyps during colonoscopy. There are promising signals for increased specificity and improved confidence of optical diagnosis, but our statistical approach precludes us from making any definitive statements about the identification and removal of small rectosigmoid polyps using the colonoscopy system we employed.

Disclosures

Author disclosures and other supplementary materials are available at evidence.nejm.org.

This study was investigator initiated and funded by public research grants from the Research Council of Norway (Norges Forskningsråd), the Norwegian Cancer Society (Kreftforeningen), and the Japan Society for the Promotion of Science. Cybernet Systems and Olympus loaned the AI software and endoscopes, respectively, to the study at no cost. The companies and funders did not have any influence on and did not play any role in the design, conduct, analysis of the study, or interpretation of the results and were not involved in the writing of the manuscript. Olympus provided a research grant for the study.

Author Affiliations

[1] Clinical Effectiveness Research Group, University of Oslo, Oslo
[2] Clinical Effectiveness Research Group, Department of Transplantation Medicine, Oslo University Hospital, Oslo
[3] Department of Gastroenterology, Hepatology and Clinical Oncology, Centre of Postgraduate Medical Education, Warsaw, Poland
[4] Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan
[5] Department of Medicine, Sørlandet Hospital Kristiansand, Kristiansand, Norway
[6] King's Institute of Therapeutic Endoscopy, King's College Hospital NHS Foundation Trust, London
[7] Graduate School of Informatics, Nagoya University, Nagoya, Japan
[8] Department of Diagnostic Pathology, School of Medicine, Showa University Northern Yokohama Hospital, Kanagawa, Japan
[9] Department of Medicine, Baerum Hospital, Vestre Viken Hospital Trust, Gjettum, Norway

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-249. DOI: 10.3322/CAAC.21660.
2. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366:687-696. DOI: 10.1056/NEJMoa1100370.
3. Abu Dayyeh BK, Thosani N, Konda V, et al. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2015;81:502.e1-502.e16. DOI: 10.1016/j.gie.2014.12.022.
4. Wadhwa V, Alagappan M, Gonzalez A, et al. Physician sentiment toward artificial intelligence (AI) in colonoscopic practice: a survey of US gastroenterologists. Endosc Int Open 2020;8:E1379-E1384. DOI: 10.1055/a-1223-1926.
5. Rex DK, Kahi C, O'Brien M, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2011;73:419-422. DOI: 10.1016/J.GIE.2011.01.023.

6. Mori Y, Kudo SE, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med 2018;169:357-366. DOI: 10.7326/M18-0249.

7. Misawa M, Kudo SE, Mori Y, et al. Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy. Gastroenterology 2016;150:1531-1532.e3. DOI: 10.1053/j.gastro.2016.04.004.

8. Mori Y, Kudo SE, Chiu PW, et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study. Endoscopy 2016;48:1110-1118. DOI: 10.1055/s-0042-113609.

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;69:620-625. DOI: 10.1016/J.GIE.2008.05.057.

10. Kudo SE, Misawa M, Mori Y, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol 2020;18:1874-1881. DOI: 10.1016/j.cgh.2019.09.009.

11. Participants in the Paris Workshop. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon. Gastrointest Endosc 2003;58(Suppl):S3-S4. DOI: 10.1016/s0016-5107(03)02159-x.

12. Hewett DG, Kaltenbach T, Sano Y, et al. Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. Gastroenterology 2012;143:599-607.e1. DOI: 10.1053/J.GASTRO.2012.05.006.

13. IJspeert JEG, Bastiaansen BAJ, van Leerdam ME, et al. Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. Gut 2016;65:963-970. DOI: 10.1136/GUTJNL-2014-308411.

14. Sumimoto K, Tanaka S, Shigita K, et al. Diagnostic performance of Japan NBI Expert Team classification for differentiation among noninvasive, superficially invasive, and deeply invasive colorectal neoplasia. Gastrointest Endosc 2017;86:700-709. DOI: 10.1016/j.gie.2017.02.018.

15. 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. DOI: 10.1055/a-1201-7165.

16. Hassan C, Pickhardt PJ, Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. Clin Gastroenterol Hepatol 2010;8:865-869, 869.e1-869.e3. DOI: 10.1016/J.CGH.2010.05.018.

17. Wieszczy P, Kaminski MF, Løberg M, et al. Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models. BMJ Open 2021;11:e042158. DOI: 10.1136/bmjopen-2020-042158.

18. Rees CJ, Rajasekhar PT, Wilson A, et al. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. Gut 2017;66:887-895. DOI: 10.1136/gutjnl-2015-310584.

19. Ladabaum U, Fioritto A, Mitani A, et al. Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions. Gastroenterology 2013;144:81-91. DOI: 10.1053/j.gastro.2012.09.054.

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Ishita Barua, Paulina Wieszczy, Shin-ei Kudo, et al. Real-Time Artificial Intelligence–Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy.
NEJM Evidence. DOI: 10.1056/EVIDoa2200003.

Supplementary Appendix Barua et al.

Table of Contents

1. Authors contributions

The investigators IB, PW, SKu, MM, OH, KMori, HI, BH, ML, MK, AH, MB, and YMori designed the study.

IB, SG, SW, KT, KMochizuki, YMiyata, KMochizuki, YA, TK, YMorita, OS, SKa, TN, MP, NG, AK, AE, CM, JAN, SAH, SOF, and PT contributed to collecting the data.

IB, PW, and YMori analyzed the data.

IB, PW, MB, and YMori drafted the manuscript.

2. Polyp histology re-assessment

Forty-eight polyps were reviewed in a second histopathological evaluation; among these, 43 polyps remained unchanged after second assessment while diagnosis was changed in 5 polyps: 2 changed from hyperplastic polyp to adenoma; and 3 from hyperplastic to traditional serrated adenoma with low-grade dysplasia.

3. Sessile serrated lesions

The CADx tool used in this study was not designed to differentiate sessile serrated lesions (SSLs) from other types of polyps. Although the colonoscopic appearance of SSLs resembles hyperplastic polyps, SSL's are considered to have neoplastic potential and thus should be removed. Given there is no available CADx tools that can identify SSLs, risk of misidentifying SSLs under the aid of CADx is unavoidable. Among the pathologically proven 19 SSLs in the present study, endoscopists predicted 5 as neoplastic, while 14 were predicted being non-neoplastic with the aid of CADx. In this regard, development of CADx systems that can differentiate SSL is desirable. On the other hand, this issue is not likely to be very clinically relevant, because the prevalence of SSL among small polyps (≤5mm) is low. In our study, it was only 2% (19 out of 892 polyps).

Table S1: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy and sessile serrated polyps considered non-neoplastic (all centers)

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	91.2 (87.5-93.9)	94.1 (91.2-96.2)
Specificity - % (95% CI)	82.3 (78.4-85.6)	85.5 (81.9-88.5)
Positive predictive value - % (95% CI)	77.1 (72.4-81.2)	81.1 (76.7-84.8)
Negative predictive value - % (95% CI)	93.8 (91.1-95.7)	95.8 (93.6-97.3)

Table S2:

Category	
Disease, problem, or condition under investigation	Colorectal neoplasm
Special considerations related to	
Sex and gender	Participants in the present study were 63% men and 37% women. Colorectal neoplasms are more prevalent in men than women. (odds ratio: 1.77).[1]
Age	Prevalence increases with age.[1]
Race or ethnic group	Prevalence may be similar between different races.[1]
Geography	Prevalence in the West may be higher than in the East.[2-4]
Other considerations	Prevalence of colorectal neoplasm is uncertain in underdeveloped countries due to the lack of early intervention with colonoscopy and cancer screening programs.
Overall representativeness of this study	The sex ratio well reflect clinical practice as men are likely to have polyps, receive more colonoscopies and need more surveillance colonoscopies after polyp removal. The age distribution (median age of 67) compares well with clinical practice for colonoscopy. Biological sex and age were reported by the participants with reference to their national ID. We did not register ethnicity of participants in the present study. Given the ethnic distribution of the populations in Norway, the UK and Japan, it is likely that the majority of the participants were White and Asian, respectively, while the number of Black patients was limited. Considering there is no significant relationship between the prevalence of colorectal neoplasm and ethnicity, the study results may be representative for any ethnicity.

Table S3: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy, at centre in Norway and sessile serrated polyps considered neoplastic

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	8 9.5 (81.8-94.2)	89.0 (81.4-93.8)
Specificity - % (95% CI)	83.4 (76.7-88.5)	86.7 (80.6-91.0)
Positive predictive value - % (95% CI)	76.8 (68.3-83.6)	80.6 (72.6-86.6)
Negative predictive value - % (95% CI)	93.9 (89.2-96.6)	93.5 (88.8-96.3)
Confidence of optical diagnosis - % high confidence (95%CI)	77.4 (72.2-81.9)	92.9 (89.0-95.4)

Table S4: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy, at centre in Norway and sessile serrated polyps considered non-neoplastic

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	96.3 (90.5-98.6)	96.5 (90.9-98.7)
Specificity - % (95% CI)	81.7 (75.0-87.0)	85.3 (79.1-89.9)
Positive predictive value - % (95% CI)	73.5 (64.8-80.7)	77.9 (69.6-84.4)
Negative predictive value - % (95% CI)	98,0 (94.8-99.3)	98.1 (94.8-99.3)

Table S5: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy, at centre in Japan

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	86.1 (79.9-90.6)	89.3 (83.7-93.2)
Specificity - % (95% CI)	86.5 (81.1-90.5)	89.3 (84.2-92.9)
Positive predictive value - % (95% CI)	82.5 (76.0-87.6)	86.2 (80.0-90.7)
Negative predictive value - % (95% CI)	89.7 (85.0-93.1)	91.8 (87.4-94.8)
Confidence of optical diagnosis - % high confidence (95%CI)	67.9 (62.7-72.7)	93.1 (90.0-95.3)

Table S6: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy, at centre in Japan and sessile serrated polyps considered non-neoplastic

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	87.6 (81.6-91.9)	92.1 (86.9-95.3)
Specificity - % (95% CI)	85.6 (80.3-89.7)	89.6 (84.7-93.1)
Positive predictive value - % (95% CI)	80.8 (74.1-86.1)	86.2 (80.0-90.7)
Negative predictive value - % (95% CI)	91.1 (86.7-94.2)	94.2 (90.2-96.6)

Table S7: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy, at centre in UK

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	93.1 (82.0-97.6)	94.9 (86.0-98.3)
Specificity - % (95% CI)	72.9 (60.5-82.6)	72.6 (60.9-81.9)
Positive predictive value - % (95% CI)	75.1 (63.5-83.9)	74.7 (63.7-83.3)
Negative predictive value - % (95% CI)	91.7 (79.3-97.0)	94.7 (84.3-98.3)
Confidence of optical diagnosis - % high confidence (95%CI)	85.9 (78.3-91.2)	90.8 (85.2-94.4)

Table S8: Performance of Standard and Artificial Intelligence-derived Computer-Aided Diagnosis (CADx) optical diagnosis of small rectosigmoid polyps during colonoscopy, at centre in UK and sessile serrated polyps considered non-neoplastic

	Standard diagnosis	**CADx diagnosis**
Sensitivity - % (95% CI)	93.1 (82.0-97.6)	94.9 (86.0-98.3)
Specificity - % (95% CI)	72.9 (60.5-82.6)	72.6 (60.9-81.9)
Positive predictive value - % (95% CI)	75.1 (63.5-83.9)	74.7 (63.7-83.3)
Negative predictive value - % (95% CI)	91.7 (79.3-97.0)	94.7 (84.3-98.3)

References

1. Corley DA, Jensen CD, Marks AR, et al. Variation of adenoma prevalence by age, sex, race, and colon location in a large population: implications for screening and quality programs. Clin Gastroenterol Hepatol 2013;11:172-80.
2. Almadi MA, Alharbi O, Azzam N, Wadera J, Sadaf N, Aljebreen AM. Prevalence and characteristics of colonic polyps and adenomas in 2654 colonoscopies in Saudi Arabia. Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association 2014;20:154-61.
3. Aswakul P, Prachayakul V, Lohsiriwat V, Bunyaarunnate T, Kachintorn U. Screening colonoscopy from a large single center of Thailand - something needs to be changed? Asian Pacific journal of cancer prevention : APJCP 2012;13:1361-4.
4. Hong W, Dong L, Stock S, Basharat Z, Zippi M, Zhou M. Prevalence and characteristics of colonic adenoma in mainland China. Cancer management and research 2018;10:2743-55.

Article III

Speedometer for withdrawal time monitoring during colonoscopy: A clinical implementation trial

*Ishita Barua[1,2], *Masashi Misawa[3], Jeremy R Glissen Brown[4], Trent Walradt[4], Shin-ei Kudo[3], Sunil G Sheth[5] Judy Nee[5], Johanna Iturrino[5], Rupa Mukherjee[5], Catherine P Cheney[4], Mandeep S Sawhney[4], Douglas K Pleskow[4], Kensaku Mori[6], Magnus Løberg[1,2], Mette Kalager[1,2], Paulina Wieszczy[1,7], Michael Bretthauer[1,2], **Tyler M Berzin[4] and **Yuichi Mori[1,2, 3].

*Contributed equally as co-first authors.

**Contributed equally as co-last authors.

[1]Clinical Effectiveness Research Group, Institute of Health and Society, University of Oslo, Oslo, Norway.

[2]Clinical Effectiveness Research Group, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway.

[3]Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan.

[4]Center for Advanced Endoscopy, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, USA.

[5]Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA.

[6]Graduate School of Informatics, Nagoya University, Nagoya, Japan.

[7]Department of Gastroenterology, Hepatology and Clinical Oncology, Centre of Postgraduate Medical Education, Warsaw, Poland

Word count: Abstract: 236; Running text: 2004

Corresponding author:

Ishita Barua, M.D.

Clinical Effectiveness Research Group

Institute of Health and Society

University of Oslo, Oslo, Norway

ishita.barua@medisin.uio.no

+47 48184943

Disclosures:

I.B. has received a travel grant from Olympus Norway; M.M. has received consultant fees and speaking honoraria from Olympus Corp and has ownership interest with Cybernet Corporation.; M.B. has received consultant fees from Cybernet Systems Corp. Y.M. has received consultant fees and speaking honoraria from Olympus Corp and has ownership interest with Cybernet Corporation.; T.M.B. has received consultant fees from Wision AI, Fujifilm, Magentiq Eye, and Medtronic; All other authors have nothing to disclose.

This trial was investigator-initiated and funded by the Philanthrope Committee at BIDMC. This grant covered fees for developing the study database, data center and technical equipment. Cybernet System Corporation (Tokyo, Japan) loaned the AI software and one high-specification computer at no cost for the trial. The entities did not play any role in the design, conduct, data analysis, or interpretation of the trial, and were not involved in content or writing of the manuscript.

Funding: I.B., Fulbright Fellowship from the US-Norway Fulbright Foundation for Educational Exchange; **T.M.B.,** Philanthrope Committee at Beth Israel Deaconess Medical

Center; **Y.M.**, European Commission (101026196); Japan Society of Promotion for Science (19KK0421).

ClinicalTrials.gov Identifier: NCT04710251

Abstract

Background

Meticulous inspection of the mucosa during colonoscopy, represents a lengthier withdrawal time, but has been shown to increase adenoma detection rate (ADR). We investigated if artificial intelligence-aided speed monitoring can improve suboptimal withdrawal time.

Methods

We evaluated the implementation of a computer-aided speed monitoring device during colonoscopy at a large academic endoscopy center. After informed consent, patients 18 years or older undergoing colonoscopy between March 5th 2021 and April 29th 2021 were examined without use of the speedometer, and with the speedometer between April 29th 2021 and June 30th 2021. All colonoscopies were recorded, and withdrawal time was assessed based on the recordings in a blinded fashion. We compared mean withdrawal time, percentage of withdrawal time ≥ 6 minutes, and ADR with and without the speedometer.

Results

166 patients in each group were eligible for analyses. Mean withdrawal time was 9 minutes and 6.6 seconds (9.11 minutes) (95%CI 8 minutes and 34.8 seconds to 9 minutes and 39 seconds) without use of the speedometer, and 9 minutes and 9 seconds (9.15 minutes) (95%CI 8 minutes and 45 seconds to 9 minutes and 33.6 seconds) with the speedometer; difference 2.3 seconds (95%CI -42.3-37.7 p=0.91). Adenoma detection rates were 45.2% (95%CI 37.6-52.8) without the speedometer as compared to 45.8% (95%CI 38.2-53.4) with the speedometer (p=0.91). The proportion of colonoscopies with withdrawal time ≥ 6 minutes without the speedometer was 85.5% (95%CI 80.2-90.9) versus 86.7% (95%CI 81.6-91.9) with the speedometer (p=0.75).

Conclusions

Use of artificial intelligence speedometer monitoring during withdrawal did not increase withdrawal time or ADR in colonoscopy.

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer death and the third most common cancer worldwide.[1] Colonoscopy is one of the most commonly used screening tests for prevention and early detection of CRC.[2]

High adenoma detection rates (ADR) are associated with low risk of colorectal cancer after colonoscopy. Therefore, ADR is a key performance indicator in colonoscopy.[3] Withdrawal time has been shown to be a surrogate indicator of ADR.[4] A withdrawal time of at least 6 minutes is recommended to maintain high ADR and thus high-quality colonoscopy. However, there is still considerable variation between colonoscopists in withdrawal time and suboptimal withdrawal time below 6 minutes.[5–7]

We investigated the implementation of a novel artificial intelligence (AI)-aided speedometer that monitors the withdrawal speed and warns the colonoscopist whenever a pre-defined "speed limit" is exceeded in real time during colonoscopy.[8–10] To our knowledge, there is no study which have investigated the specific role of a speedometer in achieving high-quality colonoscopy.[8–11]

Methods

Study design and oversight

We conducted a prospective clinical implementation trial at the Center for Advanced Endoscopy at Beth Israel Deaconess Medical Center (BIDMC), Boston, USA. The study was approved by the local institutional review board and registered at ClinicalTrials.gov (NCT04710251). We obtained written informed consent from all participants before enrolment.

Between March 5th 2021 and April 29th 2021, all participants were examined without the use of a speedometer (pre-implementation period). The speedometer was implemented on

April 29th 2021, and until June 30th 2021, participants were examined with the use of the speedometer (post-implementation period). Nine participants had colonoscopy without the use of the speedometer in the post-implementation period because two endoscopists (out of the nine who participated in the study) joined the study later than the others.

Study population

All patients aged 18 years or older scheduled for colonoscopy at the study center were eligible for enrollment. Exclusion criteria were known colorectal cancer present before colonoscopy, hereditary colorectal polyposis, inflammatory bowel disease, or history of colorectal resection.

Data management

Data obtained in the study were registered using REDCap electronic data capture tools hosted at Beth Israel Deaconess Medical Center.[12,13] The registered data included patient age, sex, and ethnicity, indication for colonoscopy, quality of bowel preparation assessed by the Boston bowel preparation scale (BBPS),[14] cecum intubation, insertion time, withdrawal time, polypectomy (yes/no), number of adenomas, and complications.

Colonoscopy procedures

Preparations and performance of colonoscopy, including bowel preparation, pre-procedure assessment and sedation practices followed ordinary clinical routines at the study center. Colonoscopies were performed using Olympus CLV 180-series or CLV 190-series colonoscopes. Patients underwent sedation at the discretion of the colonoscopist, using either a combination of benzodiazepine and opioids or propofol, both under supervision of a trained anesthesiologist.

Patients received bowel preparation with either 2-liter or 4-liter polyethylene glycol-based preparations, or sodium sulfate-based preparations, or a 30-ounce magnesium citrate-based preparation was used. All bowel preparations were administered as split doses, with the first half of the prescribed bowel preparation in the evening before the procedure and the second half in the morning of the procedure, according to current clinical practice at the study center.

The participating colonoscopists performed all colonoscopies according to clinical routines at the study center. One observer (IB) was present during all colonoscopy procedures during the study to register data and activate the speedometer. The same group of colonoscopists enrolled patients in both the pre- and post-implementation periods. Study colonoscopists were all board-certified endoscopists. We did not offer any pre-trial training of the speedometer before implementation. As the study center is a major teaching hospital for endoscopy training, trainees participated in some procedures under direct supervision by the study colonoscopists. The exact involvement of trainees in every colonoscopy procedure was not recorded.

The speedometer

For withdrawal time monitoring, we used a speedometer device developed by Cybernet System Corporation (Tokyo, Japan). This device was developed and validated in Japan (unpublished data). Because morphology of the colonic mucosa has not been shown to differ between ethnic groups, we believe that the device also produces valid measurements for the North-American population studied in this trial.[15]

The algorithm of the speedometer is based on the Lucas-Kanade method,[16] a differential method for optical flow estimation that combines information from several nearby pixels in a picture, and estimates which direction an object moves so that local changes in

intensity can be measured. This allows measurement of the relative speed of withdrawal during colonoscopy (Supplementary Video).

The speedometer was deployed through a high-specification computer, which was connected to the endoscopy processor. In the trial, the observer present during the whole colonoscopy activated the speedometer immediately after cecum intubation. The device was programmed to activate an acoustic alarm whenever the withdrawal speed exceeded a predefined threshold (Figure 1). This threshold can be set anywhere from level 1 to 20, with level 1 being the lowest level for alarm and level 20 being the highest level for alarm. For this study we chose level 12, which was the alarm threshold communication.

Trial endpoints

The primary endpoint of the trial was difference in withdrawal time between colonoscopies performed with and without the speedometer. Secondary outcome measures included ADR of the participating endoscopists on group level and the proportion of colonoscopies with withdrawal time ≥ 6 minutes.

Endpoint assessments

All colonoscopy procedures in the trial (pre- and post-implementation) were video-recorded. Before analysis, all video recordings were edited by removing all sound and recording dates to mask whether the speedometer was used. After editing, the videos were given to an independent research assistant who kept a scrambling key containing the information about which video recording corresponded to which study period (pre- or post-implementation). This information was not available to the research team and blinded the endpoint assessors to whether the speedometer had been used.

Two independent, blinded assessors (one experienced endoscopist (IB) and one physician who received pre-trial training on colonoscopy videos (TW)) assessed all recordings and registered insertion time, withdrawal time, bowel preparation using BBPS, number of polyps detected, and complications.

Withdrawal time was measured as the duration from the identification of the base of the cecum until exit from the rectum.[5] Time spent on biopsy and polypectomy were subtracted in order to obtain correct withdrawal times.[8–10] All polyps were submitted to histopathology. The histopathology diagnosis was used to categorize polyps into adenomas, non-adenomas, or cancer. Any complication during the procedure was registered by the observer present during the colonoscopy procedures.

After completion of the video assessments and registration of the data, the database was locked to prevent modification on November 29th 2021. The endpoint assessors provided the annotated data, and they were merged with the group labels (pre- or post-implementation) for analyses.

Sample size calculation

Our sample size rationale was based on two sources of information: firstly, a randomized trial from China[8] which showed a withdrawal time difference between standard colonoscopies and colonoscopies with a speedometer in a combination with a blind spot detector was 1.6 minutes, with a standard deviation (SD) of 2.5 minutes, and secondly unpublished data collected from 125 colonoscopy cases performed by expert colonoscopists at the study center which revealed a withdrawal time of 5.9 minutes with a standard deviation (SD) of 4.8 minutes. Because the Chinese trial[8] used a non-parametric test to compare withdrawal time between the groups, we assumed that time distribution was not normal, and chose to account for that with +/- 5% to our sample size estimates .

Thus, with a SD of 4.8 minutes in withdrawal time in both groups, statistical power of 80%, an alpha of 5% and no intra-cluster correlation coefficient (ICC) among colonoscopists, the required sample size was 299 patients in total. Expecting a 10% drop out, we planned to enroll 332 patients; 166 patients in the pre-implementation period and 166 patients in the post-implementation period.

Statistical analysis

All analyses were based on modified intention-to-treat analyses, defined as patients with a complete colonoscopy (cecum intubated) and video recording technically assessable. Data were presented as frequencies and proportions for categorical variables and medians and interquartile ranges (IQR) for continuous variables. Histograms were used to assess the distribution of the variables. Normally distributed variables (age) were compared using t-test. Variables with non-normal or ordinal distribution (insertion time, number of adenomas) were compared using non-parametric Wilcoxon's test. Chi-square test or exact Fisher's tests were used to compare the proportions.

For the main endpoint (withdrawal time), mean and 95% confidence intervals were reported. In sensitivity analysis, we used multivariable linear regression models with the withdrawal time as dependent variable to adjust for imbalance in baseline characteristics between the pre-implementation period (without speedometer) and post-implementation period (with speedometer). Only variables for which a significant difference between the groups was observed were included in the model. We calculated adjusted withdrawal time assuming mean value of the confounders. We defined statistical significance if $p<0.05$. All p-values are two-sided. All analyzes were performed using Stata version 16.0 (Texas, USA).

Results

Out of 352 patients who consented to participate in the trial (Figure 1) and underwent colonoscopy during the trial period from March 5th 2021 to June 30th 2021, 342 patients were eligible for analyses, while 10 patients were excluded; 7 due to poor bowel preparation (too poor quality to complete the colonoscopy), 1 because of colonic resection, and 2 due to cecum intubation failure.

Out of the 342 eligible patients, 10 patients had incomplete colonoscopy recordings due to computer hardware failure. Thus, 332 patients were included in the analyses; 166 without the speedometer and 166 with the speedometer. All colonoscopies were performed by 9 colonoscopists.

The median patient age was 61 years, and 53% were women (Table 1). The indication for colonoscopy was polyp surveillance in 48%, screening colonoscopy in 40% and clinical signs or symptoms in 11%. The median colonoscopy insertion time was 6.2 minutes (IQR 4.1-9.2) (Table 1). We did not observe any complications related to the colonoscopy.

Speedometer effects

Mean withdrawal time was 9 minutes and 6.6 seconds (9.11 minutes) (95%CI 8 minutes and 34.8 seconds to 9 minutes and 39 seconds) without use of the speedometer, and 9 minutes and 9 seconds (9.15 minutes) (95%CI 8 minutes and 45 seconds to 9 minutes and 33.6 seconds) with the speedometer, for a difference of 2.3 seconds (95%CI -42.3-37.7 p=0.91).

We found a significant difference in the distribution between men and women enrolled in the pre-implementation period and the post-implementation period (58.4% men and 41.6% women vs. 47% men and 53% women, p =0.04), Thus, we performed sensitivity analyses to adjust for this, but found no differences to the outcomes (see supplementary Table 1-4).

Adenoma detection rates were 45.2% (95%CI 37.6-52.8) without the speedometer as compared to 45.8% (95%CI 38.2-53.4) with the speedometer (p=0.91).

The proportion of colonoscopies with withdrawal time ≥ 6 minutes without the speedometer was 85.5% (95%CI 80.2-90.9) versus 86.7% (95%CI 81.6-91.9) with the speedometer (p=0.75).

The mean number of adenomas per colonoscopy was 0.71 (95%CI 0.553-0.857) without the speedometer vs. 0.73 (95%CI 0.577-0.881) with the speedometer (p=0.83).

Discussion

Missed adenomas are the main contributing factors for interval CRC after colonoscopy screening.[17,18] Standardized minimum withdrawal times have been shown to correlate with improved adenoma detection. A speedometer for withdrawal time monitoring could help in maintaining an ideal withdrawal speed even more consistently throughout the duration of the procedure, and thus reduce both recognition errors of polyps and exposure errors of colorectal surface during colonoscopy. However, our study found no benefit from using the speedometer to increase withdrawal time. Although several clinical trials have investigated the effectiveness of the combined use of speedometer and blind-spot detection AI tools, no previous studies have investigated the isolated impact of a speedometer.[8–11]

Several factors may explain the discordant results between our trial and other studies that have evaluated speedometers. First, our study was performed at a teaching hospital with colonoscopies performed by endoscopists with different experience levels. A similar study from China[8] involving less experienced endoscopists with only 1-3 years of endoscopy training found a significant withdrawal time difference with the use of a combined device (p<0.0001). Thus, by involving only non-expert endoscopists, the difference in withdrawal time might be larger. Additionally, in U.S. centers, where colonoscopy is part of national

screening guidelines, withdrawal time is often already proactively recorded, which may encourage standardization of withdrawal time compared to countries where there is no national colonoscopy screening program.

Another factor that distinguishes our study from the previously RCTs is the sole use of speedometer. The previous RCTs evaluated two or more AI technologies at the same time. Three randomized trials investigated the combined use of speedometer and computer-aided detection. [8–10] Two studies also used a blind spot detector in addition.[8,10,11] Two trials showed a significant increase in withdrawal time,[8,9] and the trial which showed the largest increment in withdrawal time, included a blind spot detector.[8] Sole use of the speedometer in our study may have resulted in no difference in withdrawal time.

The Hawthorne effect may have also contributed to our findings.[19] Since our trial design was non-blinded and an observer was present during procedures in both the pre- and post-intervention period, operational bias might have influenced the result. Withdrawal time in unmonitored endoscopists has been shown to be shorter compared to withdrawal time in endoscopists that are aware of the monitoring, and ADR increases with the awareness of monitoring.[6] Trials that evaluated the efficiency of AI-devices with a blinded design could address this bias.

One of the concerns with a new clinical device is the potential for distraction.[20] An audible alarm from the speedometer, especially if it goes off frequently, may interrupt the endoscopist's focus leading to errors. However, our results show that there was no difference in performance between these two groups and thus there is no evidence to suggest that withdrawal time monitoring causes distraction during colonoscopy.

In conclusion, we found no significant increment in withdrawal time difference, ADR or proportion of colonoscopies with withdrawal time ≥ 6 minutes by comparing standard colonoscopy and colonoscopy performed with a speedometer.

Figure 1: The speedometer deployed during colonoscopy withdrawal.

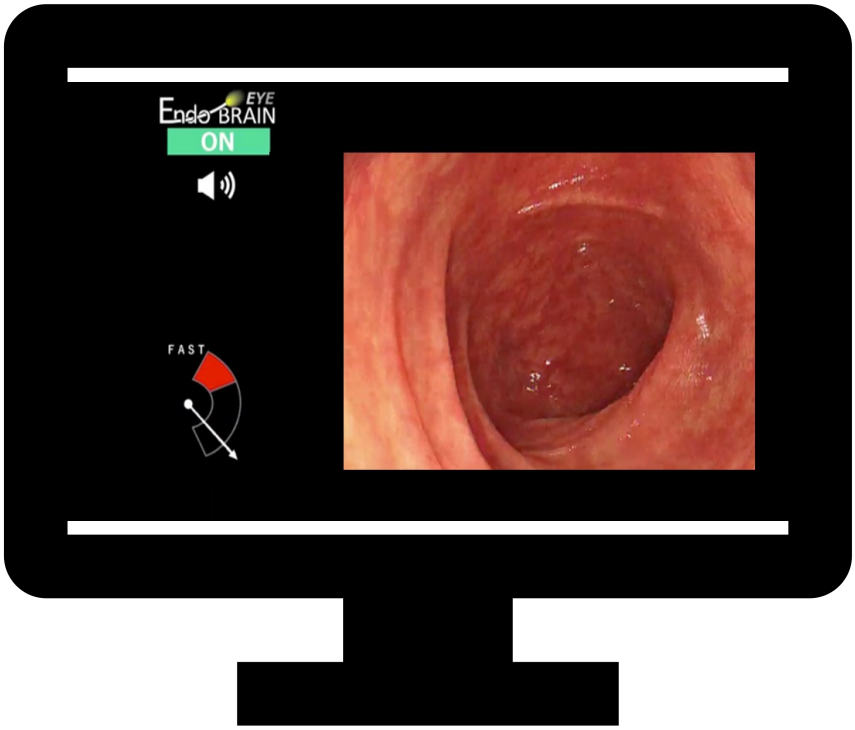


Table 1: Baseline characteristics of the included patients and detected polyps.

Patients	**Total (n=332)**		**Without the speedometer (n=166)**		**With the speedometer (n=166)**		**P-value**
Median age (IQR*) - years	61	(52-69)	61	(55-69)	62	(51-69)	0.802
Sex							0.037
Men - no. (%)	157	(47.3)	69	(41.6)	88	(53.0)	
Women - no. (%)	175	(52.7)	97	(58.4)	78	(47)	
Race							0.874
White - no. (%)	241	(72.6)	119	(71.7)	122	(73.5)	
African American (Black) - no. (%)	65	(19.6)	34	(20.5)	31	(18.7)	
Asian - no. (%)	22	(6.6)	10	(6.0)	12	(7.2)	
American Indian - no. (%)	1	(0.3)	1	(0.6)	0		
Other - no. (%)	3	(0.9)	2	(1.2)	1	(0.6)	
Ethnicity							0.835
Hispanic - no. (%)	25	(7.5)	13	(7.8)	12	(7.2)	
Non-Hispanic - no. (%)	307	(92.5)	153	(92.2)	154	(92.8)	
Colonoscopies	**(n=332)**		**(n=166)**		**(n=166)**		
Indication - no. (%)							0.123
Diagnostic colonoscopy	38	(11.4)	24	(14.5)	14	(8.4)	
Screening colonoscopy	134	(40.4)	60	(36.1)	74	(44.6)	
Surveillance colonoscopy after polypectomy	160	(48.2)	82	(49.4)	78	(47.0)	
Insertion time – median minutes (IQR)	6.2	(4.1-9.2)	6.4	(4.3-9.3)	5.9	(4.0-9.1)	0.507
Preparation quality - BBPS scale** of good or very good quality - no. (%)	310	(93.4)	153	(92.2)	157	(94.6)	0.779
Polypectomy							1.000
No	108	(32.5)	54	(32.5)	54	(32.5)	
Yes	224	(67.5)	112	(67.5)	112	(67.5)	
Number of colonoscopies done by the endoscopists			**(n=166)**		**(n=166)**		0.177
Endoscopist no 1	116	(34.9)	67	(40.4)	49	(29.5)	
Endoscopist no 2	62	(18.7)	34	(20.5)	28	(16.9)	
Endoscopist no 3	52	(15.7)	20	(12.0)	32	(19.3)	
Endoscopist no 4	26	(7.8)	10	(6.0)	16	(9.6)	
Endoscopist no 5	25	(7.5)	12	(7.2)	13	(7.8)	
Endoscopist no 6	22	(6.6)	11	(6.6)	11	(6.6)	
Endoscopist no 7	19	(5.7)	10	(6.0)	9	(5.4)	
Endoscopist no 8	8	(2.4)	2	(1.2)	6	(3.6)	
Endoscopist no 9	2	(0.6)	0		2	(1.2)	
Histopathology	**(n=400)**		**(n=201)**		**(n=199)**		
Adenomas - no. (%)	238	(59.5)	117	(58.2)	121	(60.8)	0.597
Non-adenomas	160	(40.0)	82	(40.8)	78	(39.2)	0.382
Inflammatory polyp	6	(1.5)	1	(1.2)	5	(6.4)	
Non-epithelial polyps	6	(1.5)	2	(2.4)	4	(5.1)	
Lipoma	1	(0.3)	1	(1.2)	0		
SSL	47	(11.8)	27	(32.9)	20	(25.6)	
Hyperplastic polyps	87	(21.8)	44	(53.7)	43	(55.1)	
Others (colonic mucosa)	13	(3.3)	7	(8.5)	6	(7.7)	
Adenocarcinomas	2	(0.5)	2	(1.0)	0		0.499
Lost specimens	2		2	(1.0)	0		0.499

*IQR: Interquartile range **Boston bowel preparation score: Boston Bowel preparation scale: 9-point assessment scale for cleaning quality during colonoscopy. Colon is divided into three segments: proximal, transverse and distal. Each segment is classified from 0 to 3 depending on the degree of soiling. The sum total of the three segments represents the degree of soiling (≤ 5 points: poor bowel preparation; 6–7 good bowel preparation, ≥ 8 very good bowel preparation).[14]

Figure 2: Study flow

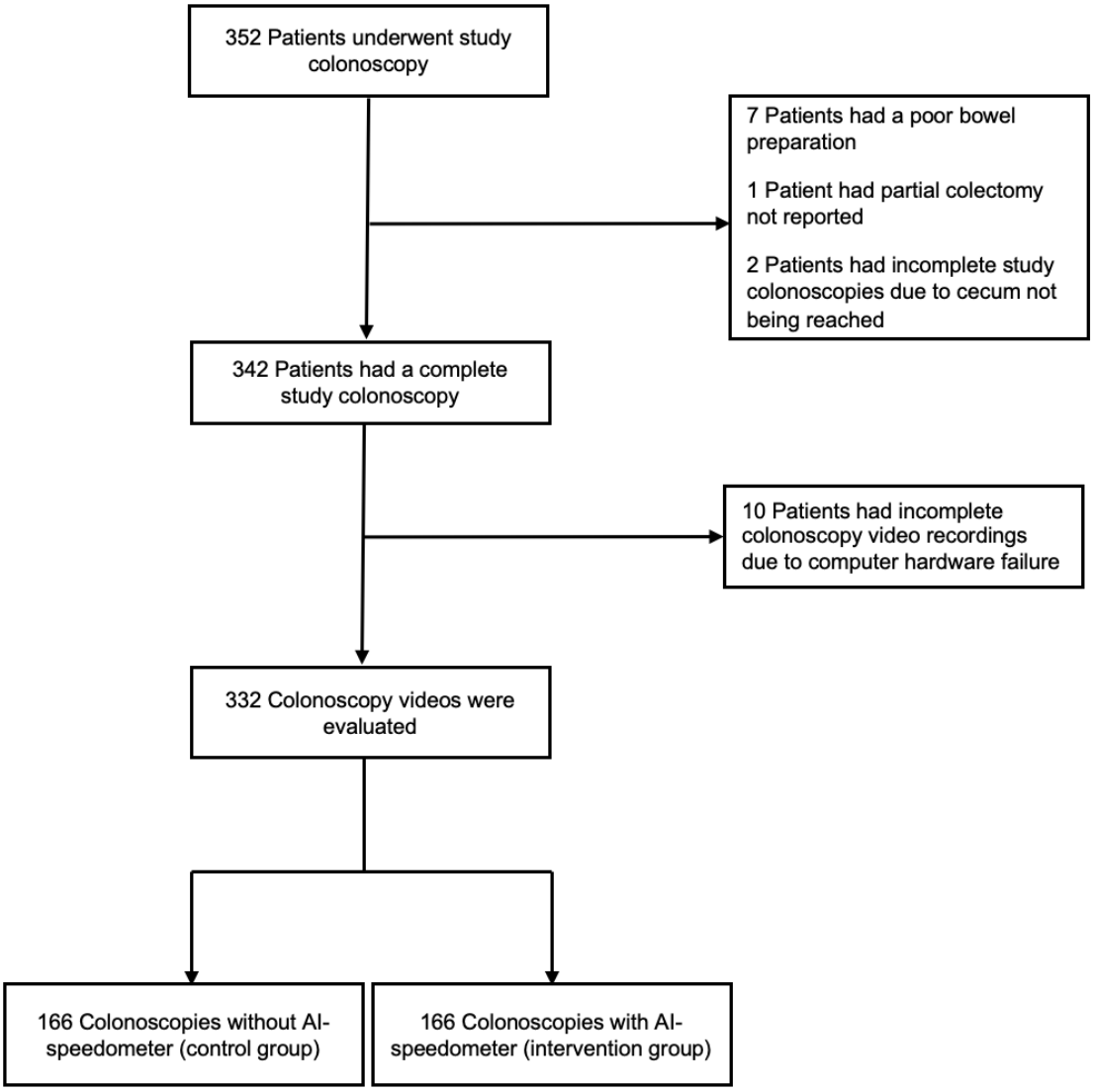


Table 2: Withdrawal time without and with AI-speedometer in minutes.

Variable	Withdrawal time in minutes (95% CI)	P-value	Withdrawal time difference between colonoscopy with and without AI-speedometer (95% CI)*	P-value
Without AI-speedometer	9.11 (8.64-9.59)	0.908	0.04 (-0.63-0.71)	0.908
With AI-speedometer	9.15 (8.68-9.62)			

*Estimates from the linear regression model.

Figure 3: Withdrawal time in seconds comparing coloscopy without and with AI-speedometer.

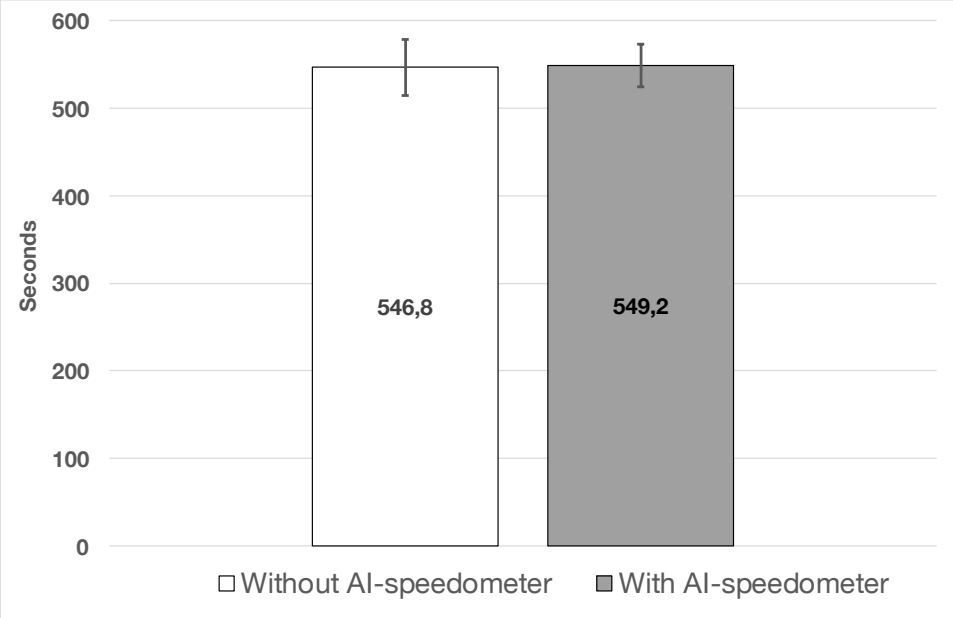


Table 4: ADR and proportion of exams with >6 min withdrawal time (95% CI) without and with AI-speedometer.

Variable	ADR (95% CI)	P-value	Proportion of exams with >6 min withdrawal time (95% CI)	P-value	Mean number of adenomas per colonoscopy (95% CI)	P-value
Without AI-speedometer	45.2% (37.6-52.8)	0.912	85.5% (80.2-90.9)	0.751	0.705 (0.553-0.857)	0.826**
With AI-speedometer	45.8% (38.2-53.4)		86.7% (81.6-91.9)		0.729 (0.577-0.881)	

**P-value from univariable linear regression. In the primary analysis Wilcoxon test was used to compare the distribution of the number of adenomas.

Figure 4: Adenoma detection rate (ADR) (unadjusted) in colonoscopy done without a speedometer and with a speedometer, and the proportion of colonoscopies with withdrawal time ≥ 6 minutes

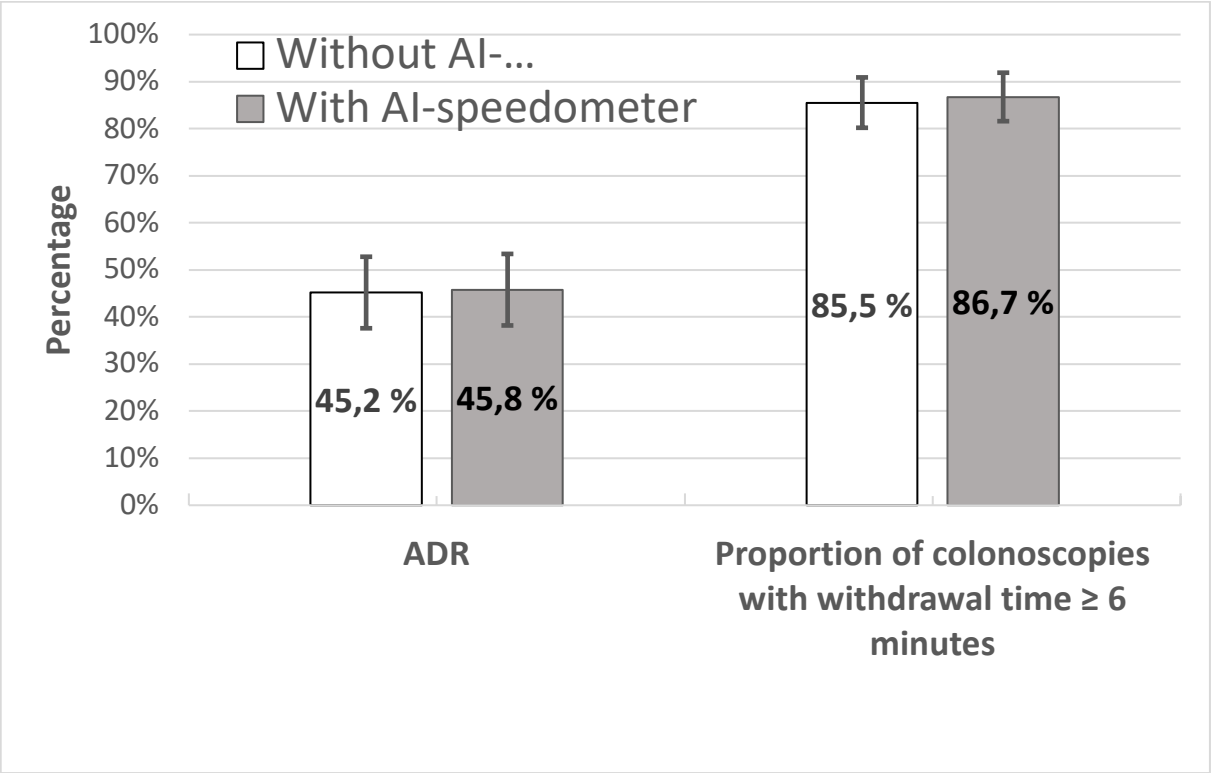
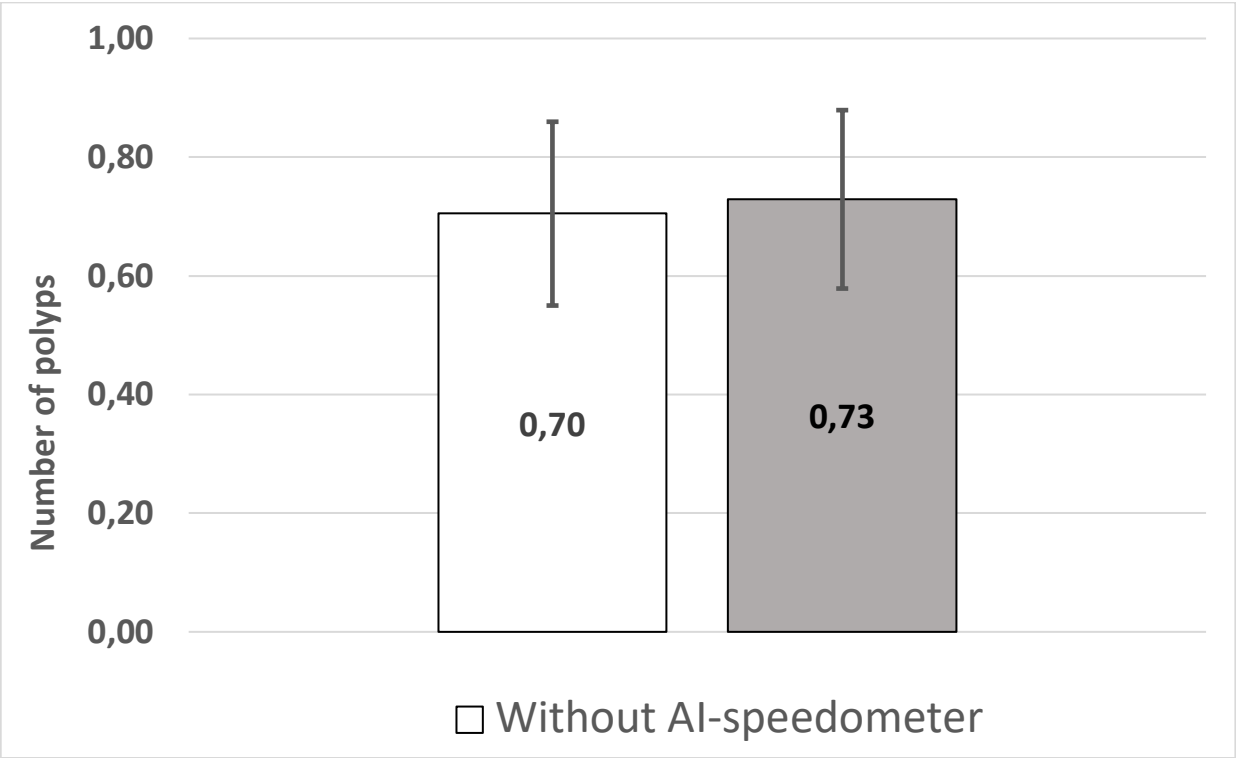


Figure 5: Number of adenomas per colonoscopy (unadjusted) done without a speedometer and with a speedometer.



Supplementary

Supplementary Video: Withdrawal speed monitoring software for colonoscopy.

Supplementary table 1: Sensitivity analysis with sex-adjusted withdrawal time in minutes.

	Withdrawal time difference between colonoscopy with and without AI-speedometer (95% CI)*	**P-value**	**Withdrawal time in minutes (95% CI) for colonoscopy with AI-speedometer (95% CI)***	**Withdrawal time in minutes (95% CI) for colonoscopy without AI-speedometer (95% CI)***
Unadjusted	0.04 (-0.63-0.71)	0.908	9.15 (8.68-9.62)	9.11 (8.64-9.59)
Adjusted**	-0.01 (-0.68-0.66)	0.975	9.13 (8.66-9.60)	9.14 (8.67-9.61)

*Estimates from the linear regression model. **Adjusted for patients' sex.

Supplementary table 2: Sensitivity analysis with sex-adjusted ADR.

Variable	**Unadjusted ADR (95% CI)**	**P-value**	**Adjusted ADR (95% CI)***	**P-value**
Without AI-speedometer	45.2% (37.6-52.8)	0.912	45.9% (38.2-53.5)	0.873
With AI-speedometer	45.8% (38.2-53.4)		45.0% (37.3-52.6)	

*Estimates from the logistic regression model adjusted for patients' sex.

Supplementary table 3: Sensitivity analysis with sex-adjusted proportion of exams with >6 min withdrawal time (95% CI).

Variable	**Unadjusted proportion of exams with >6 min withdrawal time (95% CI)**	**P-value**	**Adjusted proportion of exams with >6 min withdrawal time (95% CI)***	**P-value**
Without AI-speedometer	85.5% (80.2-90.9)	0.751	85.8% (80.5-91.1)	0.827
With AI-speedometer	86.7% (81.6-91.9)		86.6% (81.4-91.8)	

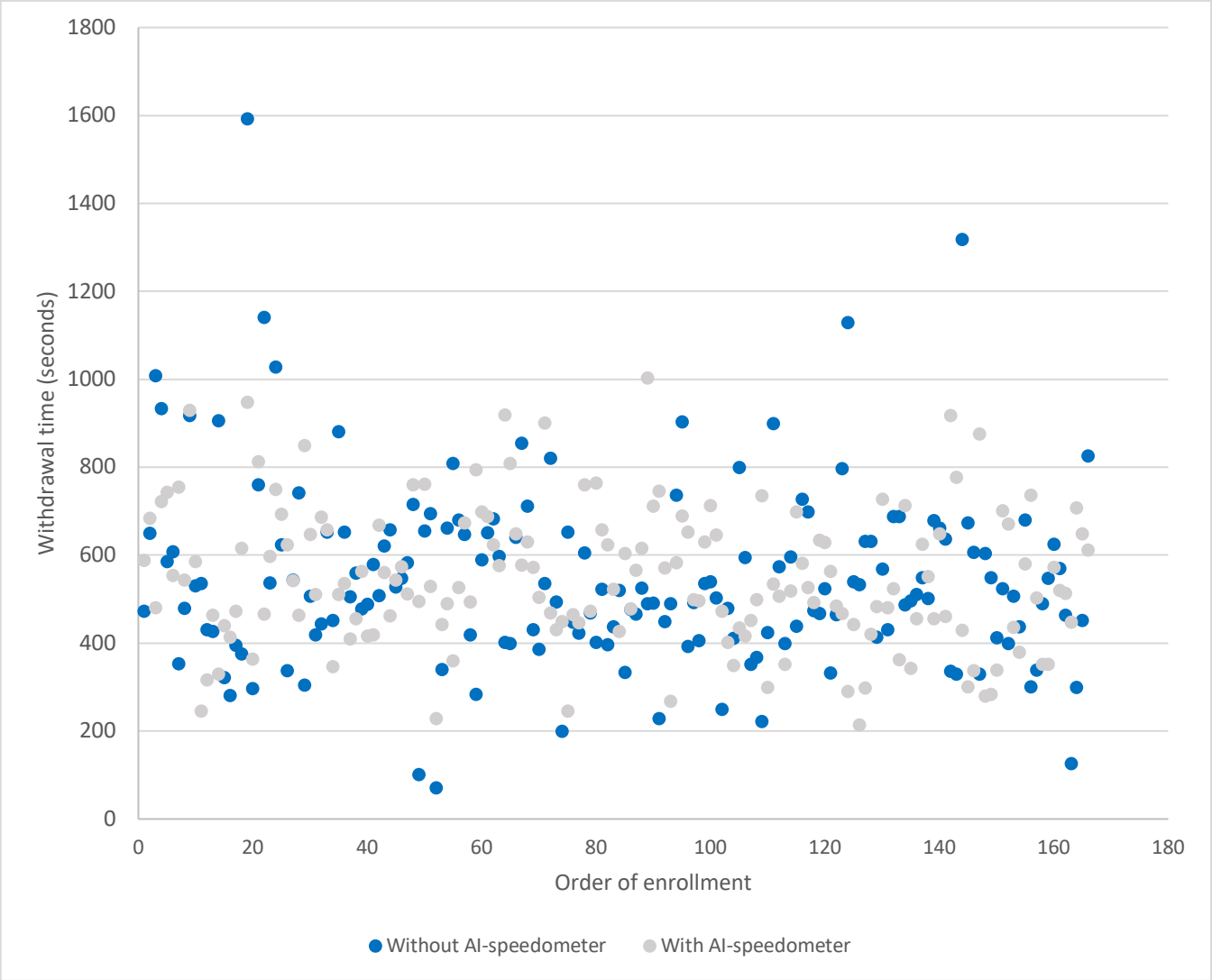
*Estimates from the logistic regression model adjusted for patients' sex.

Supplementary table 4: Sensitivity analysis with sex-adjusted mean number of adenomas per colonoscopy.

Variable	Unadjusted mean number of adenomas per colonoscopy (95% CI)	P-value	Adjusted mean number of adenomas per colonoscopy (95% CI)*	P-value
Without AI-speedometer	0.705 (0.553-0.857)	0.826**	0.720 (0.568-0.871)	0.959
With AI-speedometer	0.729 (0.577-0.881)		0.714 (0.563-0.866)	

*Estimates from the linear regression model adjusted for patients' sex. **P-value from univariable linear regression. In the primary analysis Wilcoxon test was used to compare the distribution of the number of adenomas.

Supplementary figure 1: Withdrawal time (seconds) with and without a speedometer in order of enrollment.



References

1. Sung, H. *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Helsingen, L. M. *et al*. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ* **367**, 12 (2019).
3. Michal Kaminski, A. F. *et al*. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. (2017) doi:10.1055/s-0043-103411.
4. Kaminski, M. F. *et al*. Quality Indicators for Colonoscopy and the Risk of Interval Cancer. *N. Engl. J. Med.* **362**, 1795–1803 (2010).
5. Barclay, R. L., Vicari, J. J., Doughty, A. S., Johanson, J. F. & Greenlaw, R. L. Colonoscopic Withdrawal Times and Adenoma Detection during Screening Colonoscopy. *N. Engl. J. Med.* **355**, 2533–2541 (2006).
6. Vavricka, S. R. *et al*. Monitoring colonoscopy withdrawal time significantly improves the adenoma detection rate and the performance of endoscopists. *Endoscopy* **48**, 256–262 (2016).
7. Shaukat, A. *et al*. Longer Withdrawal Time Is Associated With a Reduced Incidence of Interval Cancer After Screening Colonoscopy. *Gastroenterology* **149**, 952–957 (2015).
8. Gong, D. *et al*. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol. Hepatol.* (2020) doi:10.1016/s2468-1253(19)30413-3.
9. Su, J. R. *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.* **91**, 415-424.e4 (2020).
10. Yao, L. *et al*. An artificial intelligence-based quality improvement system significantly improved the efficacy of computer-aided detection system in colonoscopy: A four group parallel study. *Endoscopy* (2021) doi:10.1055/A-1706-6174.
11. A Single Center Study on Comparing the Different Function of EndoAngel in Improving the Quality of Colonscopy - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/study/NCT04453956?term=NCT04453956&draw=2&rank=1.
12. Harris, P. A. *et al*. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
13. Harris, P. A. *et al*. The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
14. Lai, E. J., Calderwood, A. H., Doros, G., Fix, O. K. & Jacobson, B. C. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest. Endosc.* **69**, 620–625 (2009).
15. Penn, E., Garrow, D. & Romagnuolo, J. Influence of race and sex on prevalence and recurrence of colon polyps. *Arch. Intern. Med.* **170**, 1127–1132 (2010).
16. Lucas, B. D. & Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision.
17. Zhao, S. *et al*. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. *Gastroenterology* **156**, 1661-1674.e11 (2019).
18. Robertson, D. J. *et al*. Colorectal cancers soon after colonoscopy: A pooled multicohort analysis. *Gut* **63**, 949–956 (2014).
19. McCambridge, J., Witton, J. & Elbourne, D. R. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *J. Clin. Epidemiol.* **67**, 267 (2014).
20. Barua, I., Bretthauer, M. & Mori, Y. Computer-aided quality assessment (CAQ): the next step for artificial intelligence in colonoscopy? *Mini-invasive Surg.* (2022) doi:10.20517/2574-1225.2022.5.