Evaluation of artificial intelligence system for adenoma detection in Lynch Syndrome. A randomized, parallel, multicenter, controlled trial. TIMELY study.

Brief title: Evaluation of ArTificial Intelligence System for adenoMa dEtection in Lynch Syndrome.

Authors:

Oswaldo Ortiz[1], Maria Daca-Alvarez[1], Liseth Rivero-Sanchez[1], Antonio Z Gimeno-Garcia[2], Marta Carrillo-Palau[2], Victoria Alvarez[3], Alejandro Ledo-Rodriguez[3], Luigi Ricciardiello[4-5], Chiera Pierantoni[6], Robert Hüneburg[7], Jacob Nattermann[7], Raf Bisschops[8], Sabine Tejpar[8], Alain Huerta[9], Faust Riu Pons[10], Cristina Alvarez-Urturi[10], Jorge López-Vicente[11], Alessandro Repici[12], Cessare Hassan[12-13], Lucia Cid[14], Giulia Martina Cavestro[15], Cristina Romero-Mascarell[16], Jordi Gordillo[16], Ignasi Puig[17], Maite Herraiz[18], Maite Betes[18], Jesús Herrero[19], Rodrigo Jover[20], Francesc Balaguer[1]* and Maria Pellisé[1]* on behalf of the Timely Study group.

*Francesc Balaguer and Maria Pellisé share senior co-authorship

Corresponding Author:

Maria Pellisé

Carrer Villarroel 170

Barcelona, Spain

Email: mpellise@clinic.cat

Collaborators:

1. Hospital Clinic Barcelona, Gastroenterology department. Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), IDIBAPS (Institut d´Investigacions Biomédiques August Pi i Sunyer), University of Barcelona, Barcelona, Spain.
2. Hospital Universitario de Canarias, Digestive System Service, Santa Cruz de Tenerife, Spain.
3. Complejo Hospitalario de Pontevedra, Department of Gastroenterology, Pontevedra, Spain.
4. IRCCS Azienda Ospedaliero Universitaria di Bologna, Bologna, Italy; Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy.
5. Gastroenterology, Hepatology and Nutrition, University of Texas at MD Anderson Cancer Center, Houston, TX, USA
6. IRCCS Azienda Ospedaliero Universitaria di Bologna, Bologna, Italy
7. Department of Internal Medicine I, University Hospital Bonn, Germany. National Center for Hereditary Tumor Syndromes, University Hospital Bonn, Germany.European Reference Network for Genetic Tumor Risk Syndromes (ERN Genturis)
8. Gastroenterology Department, University Hospital Leuven.
9. Hospital Galdakao-Usansolo, Department of Gastroenterology, Galdakao, Spain.
10. Gastroenterology department. Hospital del Mar Research Institute, Barcelona, Spain.
11. Hospital Universitario de Móstoles, Digestive System Service, Móstoles, Spain.
12. Gastroenterology Department, IRCCS Humanitas Research Hospital, Milan, Italy.
13. Department of Biomedical Sciences, Humanitas University, Milan, Italy.
14. Hospital Alvaro Cunqueiro, Galicia, Spain. Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur). SERGAS-UVIGO
15. Gastroenterology and Gastrointestinal Endoscopy Unit, Vita-Salute San Raffaele University, IRCCS San Raffaele Scientific Institute, Milan, Italy.
16. Hospital de la Santa Creu i Sant Pau, Gastroenterology Unit, Barcelona, Spain.
17. Gastrointestinal Oncology Endoscopy and Surgery (GOES) research group, Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), 08500 Vic, Spain. Digestive Diseases Department, Althaia Xarxa Assistencial Universitària de Manresa, 08243 Manresa, Spain; Facultat de Medicina, Universitat de Vic-Central de Cataluña (UVIC-UCC), 08500 Vic, Spain.

18. University of Navarra Clinic-IdiSNA, Gastroenterology Department, Pamplona, Spain.
19. Complexo Hospitalario Universitario de Ourense, Instituto de Investigación Biomédica Galicia Sur, CIBERehd, Ourense. Spain.
20. Hospital Universitario de Alicante, Pais Valencia, Spain.

Author contributions

Abbrevations:

ADR adenoma detection rate

AI Artificial intelligence

APC adenomas per colonoscopy

CADe Computer aided diagnosis for detection

LS Lynch Syndrome

RR Adjusted Rate ratio.

RCT Randomized controlled trial.

WLE High-definition white light endoscopy

WORD COUNT ABSTRACT: 467 words

WORD COUNT MANUSCRIPT: 3575 words inclusive of main text.

TABLE/FIGURE COUNT: Tables 2, figures 2.

SUPPLEMENTARY MATERIAL: Supplementary tables 8.

RESEARCH IN CONTEXT PANEL

Evidence before this study

We searched on Pubmed for publications between January 1, 2006, and May 19th, 2024, with the keywords “Artificial intelligence” AND “Lynch syndrome” with no restrictions of language. Our research retrieves one case report and one pilot RCT evaluating a CADe system (CAD-EYE) versus high-definition white light for the detection of adenomas during colonoscopy(1). The mentioned study analyzed 96 patients after exclusions and showed a non-statistically significant improvement of adenoma detection rate 26.1%[95%CI14.3–41.1] vs.36.0% [22.9–50.8]; p=0.379 that authors attributed to the low sample size.

Added value of this study

Our study is the first adequately powered multicenter randomized controlled trial evaluating a CADe system during colonoscopy in Lynch syndrome carriers. In the intention-to-treat analysis, the mean number of adenomas per colonoscopy was 0.64 (SD 1.57) for the intervention group and 0.64 (SD 1.17) for the WLE group, respectively [adjusted rate ratio (RR) = 1.03 (95% CI 0.72-1.47); p = 0.87]. Our results differ from the previous pilot trial(1) and from previous trials and meta-analyses where the same CADe system showed improved ADR in different settings and populations (2–5). However, they are consistent with recent pragmatic trials where CADe did not show superiority over WLE (6,7)

Implications of all the available evidence

There is no additional benefit from integrating the AI detection system for detection of adenomas in Lynch syndrome carriers. Better understanding of endoscopist and AI interactions and factors and an adequately exposure of the mucosa could affect the AI detection systems performance during colonoscopy. In this context, proper training of endoscopists and adherence to high-quality standard procedures remain the cornerstone of colonoscopy procedures for Lynch syndrome.

ABSTRACT

Background: Computer-aided artificial intelligence-based polyp detection systems (CADe) have been demonstrated to increase small polyp detection during colonoscopy in the average-risk population. Lynch syndrome (LS) represents an ideal target population for CADe since adenomas, the primary cancer precursor lesions, are characterized by their small size and higher likelihood of displaying advanced histology. The potential clinical value of CADe in LS remains unknown.

Methods: A prospective, multicenter (17 centers from Spain, Italy, Germany, Belgium) parallel, randomized controlled study was conducted to compare the mean number of adenomas per colonoscopy (APC) between the CADe (Gi Genius Medtronic©) intervention arm and the high-definition white-light endoscopy (WLE) control arm. We enrolled consecutive individuals with 18 years or more harboring (likely) pathogenic *MLH1, MSH2, MSH6*, or *EpCam* variants. Patients were consecutively randomized (1:1) to either the WLE or CAD arm. A center-stratified randomization sequence was generated through a computer-generated system which consists in a separate randomization list for each center according to block-permuted randomization. Allocation was automatically provided by the online AEG-Redcap database. Patients were unaware of the random assignment but endoscopists were unmasked. The main outcome was to compare the mean number of adenomas per colonoscopy between both groups calculated by dividing the total number of adenomas detected by the total number of colonoscopies. The procedures, management and resection of lesions was made according to clinical practice. Histopathology was the gold standard. An intention to treat and per protocol analysis was conducted. The trial is registered in clinical trials *ClinicalTrials.gov* (NCT04909671).

Findings: 414 patients were randomized to receive CADe (n=214) or WLE (n=216). Baseline characteristics of patients/procedures were well distributed between groups. Out of the 430 individuals, 256 (59.5%) were female: 127 out of 214 (59.3%) in the CADe arm and 129 out of 216 (59.7%) in the WLE arm. In the intention-to-treat analysis, the mean number of APC was 0.64 (SD 1.57) for the CADe group and 0.64 (SD 1.17) for the WLE group, respectively [adjusted rate ratio (RR) = 1.03 (95% CI 0.72-1.47); p = 0.87]. The results for the main outcome in the per-protocol analysis were in concordance with the intention-to- treat analysis: CADe 0.64 (SD 1.59) and 0.64 for WLE (SD 1.17) [RR=0.99 (95% CI 0.69-1.42); p=0.95]. No adverse events were reported during the trial.

Interpretation: In this multicenter international trial CADe did not improve the detection of adenomas in Lynch syndrome individuals. High-quality procedures and thorough inspection and exposure of the colonic mucosa remain the cornerstone in surveillance of Lynch syndrome.

Funding: Spanish Gastroenterology Association, Spanish Society of Digestive Endoscopy, European Society of Gastrointestinal Endoscopy, Societat Catalana de Digestologia and Instituto Carlos III (PI19/01050; PI19/01867; PI22/00470; ICI22/00063) and Beca de la Marato´ de TV3 2020 (Beca la Marato - 201932-30). Co-funded by the European Union. CIBERehd is funded by the Instituto de Salud Carlos III.
.

INTRODUCTION

Artificial intelligence (AI) has been integrated into the endoscopy clinical setting in recent years delivering real-time outputs for the detection and characterization of colorectal lesions. There is mounting evidence suggesting that computer-aided detection systems for colonoscopy (CADe) enhance the detection of colorectal polyps, as demonstrated in various randomized control trials and meta-analyses. However, this evidence primarily pertains to small and non-histologically advanced lesions (2–4,8,9). Additionally, CADe could be a cost-effective strategy as demonstrated in a recent pooled analysis of nine randomized controlled trials. This analysis revealed that CADe increased the proportion of patients requiring intensive surveillance according to US and ESGE guidelines(10). Although the higher adenoma detection rate resulting from CADe could enhance cancer prevention, it may also lead to reduced costs associated with advanced cancer management (11).

Lynch Syndrome (LS) an autosomal dominant disorder caused by a germline pathogenic variant on DNA mismatch repair genes (*MLH1, MSH2, MSH6, PMS2 or EPCAM*) is the most common inherited colorectal cancer (CRC) syndrome (12). In this syndrome, adenoma represents the main precursor lesion, with an accelerated adenoma-carcinoma sequence (13,14). Also, adenomas in LS are usually small with a flat or non-polypoid morphology, which makes the detection of these lesions challenging, especially since they more often display advanced histology despite of their small size (13,15,16). In this regard, back-to back studies in LS have consistently described an elevated adenoma miss rate in LS colonoscopies (17), and colonoscopy quality indicators (i.e. adequate bowel preparation, cecal intubation) have shown to likely impact the potential of colonoscopy in CRC prevention (18).

In this scenario, each adenoma discovered and resected in LS gains increased clinical significance. In a recent pilot study involving 101 LS patients, CADe-assisted colonoscopy demonstrated a trend toward improving adenoma detection rate (ADR) compared to high-definition white light (WLE) (1). Thereby, our study aims to evaluate the performance of CADe in detecting adenomas in LS carriers. (14)

METHODS

Study design and participants

We designed an international, multicenter, parallel, randomized controlled trial comparing CADe (GI genius^{TM}, Medtronic) with WLE in individuals with LS. The study was approved by the local ethical committee (HCB/2021/0322), following the CONSORT (Consolidated Standards of Reporting Trials) and CONSORT AI guidelines. The protocol version 3.0 14/09/2021 is available online. The study coordination was held at Hospital Clinic of Barcelona. A total of 17 participating European centers and 30 endoscopists were involved (see appendix page 6).

Inclusion criteria were age ≥18 years with a (likely) pathogenic germline variant in *MLH1, MSH2, MSH6* or *Epcam* who underwent surveillance colonoscopy. Exclusion criteria were previous history of total colectomy, concomitant inflammatory bowel disease, inability, or refusal to sign the informed consent previous colonoscopy < 12 months, inadequate bowel preparation, incomplete procedure, PMS2 mutation carriers.

Randomization and masking

After signing written informed consent and before starting the procedures, eligible individuals were consecutively randomized (1:1) to either the WLE or CAD arm. Randomization stratified by center, was based on a list of random numbers generated by computer in the coordinating center. This involved a separate randomization list for each center according to block-permuted randomization. The allocation sequences were incorporated into the REDCap electronic database. Once the inclusion criteria and center name were registered by data collectors at each center on the online REDCap database, the button for the randomization was activated . Group allocations were concealed for analysts and pathologists, but not for data collectors and endoscopists.

Personal and clinical data was codified, anonymously registered, and managed using REDCap electronic data capture tools hosted at Asociación Española de Gastroenterología (AEG; www.aegastro.es).

Procedures

All the patients received split dose bowel preparation according to each center protocol. Per procedural antithrombotic and anticoagulant medication were managed

according to applicable guidelines. The procedures were performed under conscious sedation as standard of care for each center. Bowel cleansing was assessed using Boston Bowel Preparation score (BBPS) and was considered as adequate if the score was ≥ 6 (≥ 2 in each segment). All colonoscopies were performed using high-definition white light endoscopes (WLE) leaving to the discretion of the endoscopist to choose between a specific brand or type. The use of add on devices was allowed and registered in the database. The CADe system was switched off during insertion. Endoscopists were informed of the allocation of patients after the informed consent was signed and before the start of the colonoscopy. Cecal intubation was recorded through usual landmarks (identification of appendicular orifice and ileocecal valve).

When appropriate, CADe was activated with white light immediately after the cecum was reached and before starting withdrawal. Procedure and withdrawal times were collected and a minimum of 6 minutes for withdrawal was highly recommended. All lesions were characterized according to the Paris classification (19). Polyp size was estimated before resection by placing an instrument with known size (opened biopsy forceps or polypectomy snare) next to the lesion as a reference. Virtual chromoendoscopy (narrow-band imaging [NBI] Olympus, linked colour imaging, blue light imaging: Fujifilm; I-SCAN: Pentax) and/or magnification were only used for optical diagnosis after identification of the lesion with WLE +/- CADe. Endoscopist's provided the optical diagnosis of each lesion based on NICE, JNET and WASP classification into five categories: hyperplastic polyp, adenoma, sessile serrated lesion, invasive neoplasia and normal/inflammatory mucosa and with their level of confidence for optical diagnosis (high or low).

Lesions were removed following polypectomy techniques according to current guidelines(20). All detected lesions were resected and sent for histopathology in separate containers with the exemption of diminutive (≤ 5 mm) recto-sigmoid polyps with a high confidence optical diagnosis of hyperplastic that were left *in situ* (21).

Immediately after the procedure a short questionnaire addressing subjective measures of the colonoscopy was presented to the endoscopist to assess the following: cecal intubation difficulties, straightness of the scope and perception of complete inspection of the mucosa on the withdrawal. The questionnaire also included two questions to subjectively ascertain the reliance on CADe system of each endoscopist.

At the start of the study, all participating endoscopists were required to have performed at least 10 colonoscopies with the Gi genius CADe System. The endoscopists were not blinded for the arm of the study due to the inherent functionality of CADe.

GI genius^{TM} (Medtronic version 1.0) is an artificial intelligence device based on a deep learning system using convolutional neural network and validated on a dataset of white-light endoscopy videos from high-quality randomized controlled trials for the detection of colorectal lesions during colonoscopy. It receives an input on digital image during the procedure and the output consists of encircling the suspicious lesion area within a green box, allowing real time assessment of colonic lesions.

The histopathology of the polyps was evaluated by local expert pathologists specialized in gastrointestinal pathology. All lesions were classified according to the Vienna classification (22).

Outcomes

The primary outcome was the mean number of adenomas per colonoscopy (APC) calculated by dividing the total number of adenomas detected in all the colonoscopies by the total number of colonoscopies.

Secondary outcomes included a comparison between both groups of: withdrawal time; mean number of polyps, flat lesions/adenomas according to Paris classification (19) proximal to splenic flexure, diminutive lesions/adenomas; advanced lesions/adenomas per colonoscopy; adenoma, polyp and proximal serrated detection rate; optical diagnosis between five categories: adenoma, hyperplastic, sessile serrated lesion, invasive neoplasia and inflammatory/normal and degree of confidence of lesions with virtual chromoendoscopy (VCE) and concordance with histopathology; false positives in CAD arm; mean number of non-clinically significant removed lesions; post-colonoscopy colorectal cancer incidence, defined as colorectal cancer diagnosed after a colonoscopy in which no cancer was found (23) that will be calculated 10 years after index procedures and subjective endoscopist measures. (22)

Detection rates were calculated as proportion of individuals with at least one histologically proven lesion of interest detected divided by the number of individuals. The proximal colon was defined as cecum, ascending colon, transverse colon, and splenic flexure. Advanced adenomas were defined as adenomas with 20% or more villous component and/or size of 10 mm or greater and/or high-grade dysplasia. Serrated lesions include hyperplastic polyps, sessile serrated lesions and traditional serrated adenomas.

Advanced serrated lesions were defined as lesions of 10 mm or larger and/or dysplastic component and traditional serrated adenoma. Proximal serrated lesions were defined as serrated lesions proximal to splenic flexure. Advanced lesions were defined as any advanced adenoma and/or advanced serrated lesion. Non-clinically significant lesions were defined as lesions with a histology of normal, inflammatory mucosa and diminutive (<5 mm) recto-sigmoid polyps with a hyperplastic histopathology. Flat lesions included flat-elevated (0-IIa) lesions, flat lesions (0-IIb) and flat slightly depressed (0-IIC) lesions according to Paris classification(19).

During the procedure, false positive alerts were defined as areas signaled as lesion by CADe during two or more seconds in an adequate condition for inspection (i.e. well centered image, without significant bubbles and/or debris on the area) and that were considered as no polyp by the endoscopist with a high confidence of optical diagnosis (24). On the other hand, false positives based on histopathology were defined as lesions, resected by the endoscopist that turned out to have a clinically non-significant histopathology (i.e. normal mucosa, subtle hyperplastic changes, inflammatory changes).

Statistical analysis

We hypothesized an absolute increased in APC of 0.3 in favor of CADe assisted colonoscopy based on previous evidence (4,8). The sample size was calculated based on previous data from a recent meta-analysis in LS showed an APC of 0.49 and a pooled standard deviation of 1.02 in the WLE arm (25). Comparing two negative binomial rates and accepting an alpha risk of 0.05 and a beta risk of 0.2, in a wide range of 'shape parameter' (between 0.5 and 1) which indicates the variability of the expected number of events per patient, the size of the sample per group needed was between 168 and 203 with an anticipated dropout rate of 10% (26).

Categorical variables were presented as frequency and percentages. Quantitative variables are presented as means with standard deviation (SD) when normally distributed and medians and interquartile range (IQR) when non-normally distributed. We compared baseline clinical and demographic characteristics between both arms using x2 test for categorical and two sample t Test for continuous normal distributed and U Mann Whitney for non-normal distribution variables. Two-side p-value of 0.05 were used as a threshold for statistical significance for all outcomes. The primary intention to treat analysis included all randomized individuals and per-protocol analysis was based on all randomized patients except post-randomization exclusions.

Generalized lineal model using negative log-binomial adjusted by center were performed to consider overdispersion for comparisons of number of lesions detected between both arms and for main and secondary outcomes. Results are expressed as risk ratio and its two-sided 95% confidence interval for primary and secondary outcomes. We calculated the ADR for each endoscopist and classified the ones with more than 10 procedures into low and high detectors based on the mean ADR. Sensitivity analysis at endoscopist level between these categories was calculated using the negative log-binomial adjusted by center.

All statistical analysis were performed using SPSS version 25.0 (IBM Corp., Armonk, New York, USA). The study was registered in *ClinicalTrials.gov* (NCT04909671).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, interpretation or writing of the report. The AI devices were loaned by Medtronic in the context of the ESGE research grant agreement.

RESULTS

A total of 456 individuals with LS were eligible for enrollment between - September 13th 2021 to April 6th, 2023. After exclusions, 430 patients were finally randomized (214 with CADe and 216 with WLE). Sixteen participants were excluded during the procedure and after randomization, resulting on 414 patients for the per protocol analysis (204 with CADe and 210 with WLE) (Figure 1). Baseline characteristics of individuals and procedures were well distributed between both groups (Table 1). No adverse events were reported. Mean age was 48.9 (Standard deviation [SD] 14.26). The median of withdrawal time between both groups were 13.1 min [IQR 10-16 min] for CADe vs 12.6 min [IQR 9-15 min] for WLE p=0.32). Bowel preparation was adequate on similar proportions between groups 206/214 (96.3%) for CADe versus 210/216 (97.2%) for WLE; p=0.81. Add on devices such as Endocuff was only used in one case. The histopathology of the 575 resected lesions is shown in Appendix p1.

Primary endpoint: mean number of adenomas per colonoscopy.

The mean number of APC between both arms of the study were 0.64 (SD 1.57) for CADe *vs* 0.64 (SD 1.17) for WLE (adjusted rate ratio (RR)= 1.03 [95% CI, 0.72-1.47]; p=0.87) (Table 2; Figure 2a). The findings were similar in the per-protocol analysis (See Appendix p2)

Secondary endpoints

Adenoma and proximal serrated lesion detection rates

The adenoma detection rate was 70/214 (32.7%) for CADe and 79/216 (36.1%) WLE, (RR=0.92 [95% CI, 0.75-1.13]; p=0.42) and for proximal serrated polyp detection rates CADe 47/210 (23.0%) vs. WLE 38/216 (18.1%), (RR=1.15 [95% CI 0.89-1.49]; p=0.28) (Table 2).

Adenoma characteristics

Subgroup analysis based on adenoma characteristics revealed no statistically significant results when comparison was made between the two groups in terms of size (categorized as diminutive (1-4 mm), small (5-9 mm), or ≥10 mm), morphology (flat versus polypoid), or location (proximal versus distal to the splenic flexure), as detailed in (Table 2 and Figure 2b-d). Mean number of advanced adenomas was 0.04 (SD 0.19) for

CADe vs. 0.07 (SD 0.27) for WLE (p=0.24). Five adenocarcinomas were detected (two in CADe arm and three in WLE arm) (Table 2).

Serrated lesions

The mean number of serrated lesions per colonoscopy were: 0.57 (SD 0.94) for CADe *vs*. 0.47 (SD 0.96) for WLE (RR=1.17; [95% CI, 0.83-1.65]; p=0.39) (**Table 2**). No statistically significant results were found on subgroups analysis according to location (proximal or distal to splenic flexure), and advanced serrated lesions (serrated lesions ≥10 mm or with dysplasia). The number of serrated lesions sized 5-9mm was superior in the study arm: 0.14 (SD 0.45) for CADe *vs*. 0.06 (SD 0.24) for WLE (RR=2.31 [95% CI, 1.15-4.63]; p=0.02). (Table 2)

Subgroup analysis by endoscopists

Performance of every endoscopist (percentage of procedures, adenoma detection rate and APC) are shown in Appendix p3**.** After excluding the endoscopist with less than 10 procedures (N= 12), we grouped the endoscopists based on their adenoma detection rates: high (ADR ≥35%) and low (ADR<35%) detectors. Eight endoscopist were considered low detectors (ADR 27.6% [range 15.4%-30.8%]) APC=0.54; (SD=1.54) and 10 were considered high detectors (ADR 41.8% [range 35.7%-50.0%]) APC=0.73; (SD=1.24). Among low detectors, there was a trend of higher APC in the CADe arm (CADe 0.71 vs. WLE 0.45; p=0.51). Conversely, for high detectors the trend for better performance on APC was observed in the WLE arm when compared to CADe arm (CADe 0.70 vs WLE 0.75; p=0.74)**.**

Optical diagnosis assessment of lesions

The concordance of all resected lesions and high confidence optical diagnosis cases in relation to histopathology are shown in Appendix p4.

False positives rate in the CADe group

A total of 192 CADe alerts were considered as false positive by the endoscopists (median of false positive per procedure 0.98). The false positive rate (defined as the percentage of colonoscopies with at least 1 false positive) was 77/204 (37.5%).

Based on histopathology the total number of false positive lesions for both groups were 58 (median of 0.14 per procedure). There was a higher histopathological

false positive rate in the CADe arm 0.21 (SD 0.62) versus the WLE arm 0.07 (SD 0.29) [RR=2.93 (95% CI, 1.48-5.79); p=0.002].

Subjective measures

The observed results in the perceived procedure difficulties between both arms are shown in Appendix p5. In 109/195 (55.9%) of CADe-assisted procedures, endoscopists acknowledged the system's aid in detecting at least one lesion**.**

Analysis by endoscopist subcategories revealed a higher perceived benefit among higher detectors compared to low detectors (62.1% vs 56.3%; p=0.02) (See Appendix p5).

DISCUSSION

To our knowledge this study marks the first sufficiently powered randomized controlled trial to explore the effectiveness of CADe in individuals with Lynch syndrome. Our findings indicate no significant enhancement in adenoma detection using CADe compared to WLE. Furthermore, analyses of adenomas by size, morphology, or location within the colon did not reveal any differences, nor in the overall adenoma detection rate. However, CADe was more effective in identifying serrated lesions measuring 5-9 mm. Notably, among lower performing detectors, there was a trend towards increased APC rates, whereas higher false positives based on histopathological analysis were observed in the CADe group.

Our findings diverge from those of a recent randomized pilot study with 91 LS carriers that hinted at a potentially higher ADR (36% vs. 26.1%; p=0.379) and mean APC (0.6 vs 0.43; p=1) with CADe, although these differences did not reach statistical significance (1). In the referenced study, which examined a population similar to ours, a distinct AI system (CAD-EYE by Fujifilm Japan) was utilized. Interestingly, secondary outcome analysis indicated a favorable signal for the use of CADe in identifying flat adenomas (Paris IIb). However, our study, which was adequately powered, did not confirm this hypothesis. Our results also differ from previous trials and meta-analyses where the same CADe system showed improved ADR in different settings and populations (2–5). However, they are consistent with recent pragmatic trials where CADe did not show superiority over WLE (6,7). Several factors could explain these differences. Firstly, our study involved a larger number of endoscopists from multiple centers, potentially enhancing external validity compared to previous trials conducted at expert centers with fewer endoscopists. Additionally, the longer inclusion period in our study may have minimized the Hawthorne effect, which could have favored CADe in previous trials with more concentrated recruitment efforts. Furthermore, differences in the study population may have influenced the outcomes. Interestingly, in our study, almost one third of the centers were non-academic, and approximately half of the endoscopists were younger than 40 years old. Despite not specifically selecting endoscopists based on their experience, they were all familiar with LS patient surveillance. This might potentially have led to higher quality standards among them, resulting in a higher overall lesion APC and detection rate, and a potential diminished benefit of CADe, particularly among high detector endoscopists. As a matter of fact, the ADR (34.7%) and APC (0.64) reported in

our study in the WLE arm is the highest ADR reported so far in this population. Nevertheless, the results have considerable external validity for previously explained reasons: a relatively large number of endoscopists and centers included with participation of non-academic centers and young endoscopists. On the other hand, the Hawthorne effect induced by the trial setting, the lack of concealment and the competition between human and an AI tool might also have triggered this high performance in both arms of the study. A similar situation occurred in the largest multicenter trial on CADe including individuals from a FIT based population screening program, where the high ADR in the control group was also used as an argument to justify the non-superiority of CADe towards WLE (27).

Recent studies have suggested that CADe may offer benefits when utilized by endoscopists with lower detection rates(28). Our trial revealed consistent results with a non-significant trend towards CADe being more beneficial among low detector endoscopists. Based on this, one could argue that CADe could still be a useful tool in non-expert hands. However, in the context of LS where the carcinogenic pathway is accelerated and the risk of post-colonoscopy CRC is heightened (29), it remains imperative to uphold high-quality standards during colonoscopy procedures. Emphasizing the importance of thorough mucosal inspection is paramount, as it may outweigh the utility of ancillary techniques like CADe. It is essential to note that CADe can only detect lesions that are adequately exposed by the endoscopist. Furthermore, our findings suggest a potential learning effect among investigators over time. Interestingly, the detection rate in the control arm of our study was even higher than in previous recent trials (30), indicating that there is room for improvement and emphasizing the critical importance of the quality of colonoscopy procedures (18).

Additionally, our trial identified several potential drawbacks associated with the use of CADe. Firstly, the false positive rate based on histopathology was 2.8-fold higher in the CADe arm, which raises concerns about the possibility of overtreatment. However, it is essential to interpret this data cautiously, as we did not conduct recuts or central revision of histology. Interestingly, a recent observational study comparing optical diagnoses based on artificial intelligence and pathology for diminutive polyps revealed higher concordance for adenoma between the artificial intelligence and the expert endoscopists optical diagnosis in lesions reported as "non-adenoma" by pathology(31).

Secondly, regarding endoscopist based subgroup analysis, our study is not statistically powered to address this endpoint properly and this might explain the lack of statistically significant differences. Furthermore, our study observed an increase in the detection of lesions with serrated histology, particularly those sized between 5-9 mm, which were predominantly hyperplastic. Identifying these lesions, which pose minimal risks of developing cancer in the Lynch syndrome population, could potentially lead to shorter surveillance colonoscopies. We hypothesize that this finding may be attributed to a change in endoscopists' behavior secondary to a Hawthorne effect: the presence of a green box signaling a lesion with an optical diagnosis of a hyperplastic polyp larger than 5 mm might prompt its resection rather than disregarding it.

In conclusion, our multicenter international RCT did not demonstrate a benefit of CADe over standard WLE for adenoma detection in LS carriers. These findings underscore the importance of maintaining high-quality colonoscopy practices, while also highlighting potential limitations and challenges associated with CADe implementation in this population.

Contributors

Conception and design: F.B., M.P., O.O., M.D.A., L. R.S.

Analysis and interpretation of the data: O.O., M.P., F.B.

Drafting of the article: O. O., F. B., M. P.

Critical revision for important intellectual content and final accepts of the manuscript: All authors.

Obtaining of funding: M.P., O. O.

Collection and assembly of data: O. O., M. D.A., L. R.S., A. G.G., M. C., V. A., A. L., L. R., C. P., R. H., J.N., R. B., S. T., A. H., F. R. P., C. A.U., J. L. V., A. R., C. H., L. C., G. M.C., C. R., J. G., I. P., M. H., M. B., S. T., J. H., R. J..

Declaration of interests

AR received payment or honoraria for lectures, presentations, speakers bureaus or educational events from Boston Scientific, Fujifilm, NEC and Olympus.

A.Z.G received grants from Fundación de la Sociedad Española de Endoscopia Digestiva and University Hospital of the Canary Islands.

C.H. received a grant from AIRC under IG 2022 - ID. 27843 project and consulting fees from Boston Scientific, Fujifilm, Olympus and Odin.

I.P. received consulting fees from Fujifilm and supporting for attending travels and meetings from Fujifilm and Casen.

J.N. received grants or contracts Deutsche Forschungsgemeinschaft (DFG), Deutsches Zentrum für Infektionsforschung (DZIF), Deutsche Krebshilfe, Hector-Stiftung, Bundesministerium für Bildung und Forschung – BMBF. Also, payment honoraria for lectures and presentation and attending travels and meetings from Falk foundation.

RB received research funding, speaker, and advisory fees from Medtronic, grants and contracts form Pentax, Digital endoscopy, Fujifilm, payment or honaria for lectures and presentations from Medtronic, Pentax and Fujifilm, receipt of equipment from Fujifilm, Pentax and payment for expert testimony from Pentax. R. B. is also supported by a grant of Research Foundation Flanders (FWO).

R.H. received grants or contracts from Deutsche Forschungsgemeinschaft (DFG), Deutsches Zentrum für Infektionsforschung (DZIF), Deutsche Krebshilfe, Hector-Stiftung, Bundesministerium für Bildung und Forschung – BMBF. Also received payment for lectures, presentations from Falk Foundation and MSD Sharp&Dohme and receipment of equipment from Fujifilm.

MP received speaker fee from Medtronic and advisory and speaker fees form Fujifilm and Olympus.

FB received consultant fees from Olympus, Nouscom, Sysmex, and Norgine, and editorial fees from Elsevier.

R.J. received contracts from Medtronic through a research grant institution.

The rest of authors declare no competing interests.

Data sharing

OO can be contact at the corresponding mail oaortiz@clinic.cat regarding data sharing requests. Individual participant data reported in this article will be stored at RedCap aegastro (https://www.aegastro.es) and will be Shared after de-identified participant data if required. This data will be available immediately after the publication and ending after 60 months following article publication.

Acknowledgments

Timely study group (see appendix page 7)

Beca Grupo de Trabajo EndoCAR group from Spanish Gastroenterology Association (AEG) and Spanish Society of Digestive Endoscopy (SEED)

ESGE Medtronic Artificial Intelligence Research Award grant 2021.

Ajut per la Iniciaciò a la Recerca. Societat catalana de Digestologia 2021.

Instituto Carlos III | PI19/01050. Rio Hortega contract 2021.

Funded by Instituto de Salud Carlos III (PI19/01050; PI19/01867; PI22/00470; ICI22/00063) and Beca de la Marato´ de TV3 2020 (Beca la Marato - 201932-30). Co-funded by the European Union. CIBERehd is funded by the Instituto de Salud Carlos III.

REFERENCES

1. Hüneburg R, Bucksch K, Schmeißer F, Heling D, Marwitz T, Aretz S, et al. Real-time use of artificial intelligence (CADEYE) in colorectal cancer surveillance of patients with Lynch syndrome—A randomized controlled pilot trial (CADLY). United European Gastroenterol J. 2023 Feb 1;11(1):60–8.
2. Repici A, Spadaccini M, Antonelli G, Correale L, Maselli R, Galtieri PA, et al. Artificial intelligence and colonoscopy experience: Lessons from two randomised trials. Gut. 2022 Apr 1;71(4):757–65.
3. Barua I, Vinsard DG, Jodal HC, Loberg M, Kalager M, Holme O, et al. Artificial intelligence for polyp detection during colonoscopy: A systematic review and meta-analysis. Endoscopy. 2021;53(3):277–84.
4. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc [Internet]. 2021;93(1):77-85.e6. Available from: https://doi.org/10.1016/j.gie.2020.06.059
5. Hassan C, Spadaccini M, Mori Y, Foroutan F, Facciorusso A, Gkolfakis P, et al. Real-Time Computer-Aided Detection of Colorectal Neoplasia During Colonoscopy. Ann Intern Med. 2023 Sep;176(9):1209–20.
6. Levy I, Bruckmayer L, Klang E, Ben-Horin S, Kopylov U. Artificial Intelligence-Aided Colonoscopy Does Not Increase Adenoma Detection Rate in Routine Clinical Practice. American Journal of Gastroenterology. 2022 Nov 1;117(11):1871–3.
7. Ladabaum U, Shepard J, Weng Y, Desai M, Singer SJ, Mannalithara A. Computer-aided Detection of Polyps Does Not Improve Colonoscopist Performance in a Pragmatic Implementation Trial. Vol. 164, Gastroenterology. W.B. Saunders; 2023. p. 481-483.e6.
8. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. Gastroenterology. 2020 Aug;159(2).
9. Huang D, Shen J, Hong J, Zhang Y, Dai S, Du N, et al. Effect of artificial intelligence-aided colonoscopy for adenoma and polyp detection: a meta-analysis of randomized clinical trials. Vol. 37, International Journal of Colorectal Disease. Springer Science and Business Media Deutschland GmbH; 2022. p. 495–506.
10. Areia M, Mori Y, Correale L, Repici A, Bretthauer M, Sharma P, et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. Lancet Digit Health. 2022 Jun 1;4(6):e436–44.
11. Mori Y, Wang P, Løberg M, Misawa M, Repici A, Spadaccini M, et al. Impact of Artificial Intelligence on Colonoscopy Surveillance After Polyp Removal: A Pooled Analysis of Randomized Trials. Clinical Gastroenterology and Hepatology. 2023 Apr 1;21(4):949-959.e2.
12. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895–2015. Nat Rev Cancer. 2015 Mar 12;15(3).
13. De Jong AE, Morreau H, Van Puijenbroek M, Eilers PHC, Wijnen J, Nagengast FM, et al. The role of mismatch repair gene defects in the development of adenomas in patients with HNPCC. Gastroenterology [Internet]. 2004 Jan;126(1):42–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0016508503017116

14. Edelstein DL, Axilbund J, Baxter M, Hylind LM, Romans K, Griffin CA, et al. Rapid Development of Colorectal Neoplasia in Patients With Lynch Syndrome. Clinical Gastroenterology and Hepatology. 2011 Apr;9(4):340–3.
15. Rijcken FEM. Proximal adenomas in hereditary non-polyposis colorectal cancer are prone to rapid malignant transformation. Gut [Internet]. 2002 Mar 1;50(3):382–6. Available from: http://gut.bmj.com/cgi/doi/10.1136/gut.50.3.382
16. Rondagh E, Gulikers S, Gómez-García E, Vanlingen Y, Detisch Y, Winkens B, et al. Nonpolypoid colorectal neoplasms: a challenge in endoscopic surveillance of patients with Lynch syndrome. Endoscopy [Internet]. 2013 Feb 25;45(04):257–64. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0032-1326195
17. Stoffel EM, Kim Turgeon D, Stockwell DH, Zhao L, Normolle DP, Tuck MK, et al. Missed adenomas during colonoscopic surveillance in individuals with Lynch syndrome (hereditary nonpolyposis colorectal cancer). Cancer Prevention Research. 2008 Nov;1(6):470–5.
18. Sánchez A, Roos VH, Navarro M, Pineda M, Caballol B, Moreno L, et al. Quality of colonoscopy is associated with adenoma detection and post-colonoscopy colorectal cancer prevention in Lynch syndrome. Clinical Gastroenterology and Hepatology. 2020 Nov;
19. Endoscopic Classification Review Group. Update on the Paris Classification of Superficial Neoplastic Lesions in the Digestive Tract. Endoscopy [Internet]. 2005 Jun;37(6):570–8. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-2005-861352
20. Ferlitsch M, Hassan C, Bisschops R, Bhandari P, Dinis-Ribeiro M, Risio M, et al. Colorectal polypectomy and endoscopic mucosal resection: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2024. Endoscopy [Internet]. 2024 Apr 26; Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/a-2304-3219
21. Rivero-Sánchez L, Gavric A, Herrero J, Remedios D, Alvarez V, Albéniz E, et al. The diagnose and leave in strategy for diminutive rectosigmoid polyps in Lynch syndrome: A post hoc analysis from a randomized controlled trial. Endoscopy. 2022 Jan 1;54(1):27–34.
22. Dixon MF. Gastrointestinal epithelial neoplasia: Vienna revisited. Gut. 2002 Jul 1;51(1).
23. Rutter MD, Beintaris I, Valori R, Chiu HM, Corley DA, Cuatrecasas M, et al. World Endoscopy Organization Consensus Statements on Post-Colonoscopy and Post-Imaging Colorectal Cancer. Gastroenterology. 2018 Sep 1;155(3):909-925.e3.
24. Holzwanger EA, Bilal M, Glissen Brown JR, Singh S, Becq A, Ernest-Suarez K, et al. Benchmarking definitions of false-positive alerts during computer-aided polyp detection in colonoscopy. Endoscopy. 2021 Sep 1;53(9):937–40.
25. Houwen BBSL, Mostafavi N, Vleugels JLA, Hüneburg R, Lamberti C, Rivero-Sánchez L, et al. Dye-Based Chromoendoscopy in Patients With Lynch Syndrome. American Journal of Gastroenterology. 2021;Publish Ah:825–8.
26. Zhu H, Lakkis H. Sample size calculation for comparing two negative binomial rates. Stat Med. 2014 Feb 10;33(3):376–87.

27. Mangas-Sanjuan C, de-Castro L, Cubiella J, Díez-Redondo P, Suárez A, Pellisé M, et al. Role of Artificial Intelligence in Colonoscopy Detection of Advanced Neoplasias. Ann Intern Med. 2023 Sep;176(9):1145–52.
28. Lau LHS, Ho JCL, Lai JCT, Ho AHY, Wu CWK, Lo VWH, et al. Effect of Real-Time Computer-Aided Polyp Detection System (ENDO-AID) on Adenoma Detection in Endoscopists-in-Training: A Randomized Trial. Clinical Gastroenterology and Hepatology. 2023;
29. Morris EJA, Rutter MD, Finan PJ, Thomas JD, Valori R. Post-colonoscopy colorectal cancer (PCCRC) rates vary considerably depending on the method used to calculate them: a retrospective observational population-based study of PCCRC in the English National Health Service. Gut [Internet]. 2015 Aug;64(8):1248–56. Available from: http://gut.bmj.com/lookup/doi/10.1136/gutjnl-2014-308362
30. Houwen BBSL, Hazewinkel Y, Pellisé M, Rivero-Sánchez L, Balaguer F, Bisschops R, et al. Linked Colour imaging for the detection of polyps in patients with Lynch syndrome: A multicentre, parallel randomised controlled trial. Gut [Internet]. 2021;1–8. Available from: https://gut.bmj.com/content/gutjnl/early/2021/03/18/gutjnl-2020-323132.full.pdf
31. Shahidi N, Rex DK, Kaltenbach T, Rastogi A, Ghalehjegh SH, Byrne MF. Use of Endoscopic Impression, Artificial Intelligence, and Pathologist Interpretation to Resolve Discrepancies Between Endoscopy and Pathology Analyses of Diminutive Colorectal Polyps. Gastroenterology. 2020 Feb 1;158(3):783-785.e1.