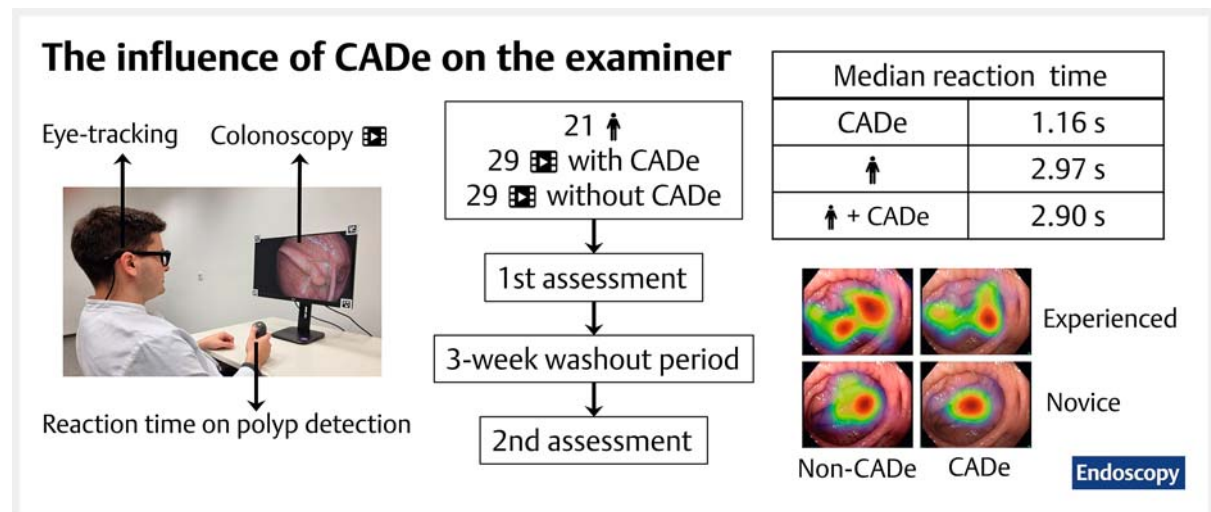# The influence of computer-aided polyp detection systems on reaction time for polyp detection and eye gaze

OPEN ACCESS

INFOGRAPHIC

Authors

Joel Troya[1, *] ⓘ, Daniel Fitting[1, *], Markus Brand[1] ⓘ, Boban Sudarevic[1], Jakob Nikolas Kather[2], Alexander Meining[1], Alexander Hann[1] ⓘ

Institutions

1　Interventional and Experimental Endoscopy (InExEn), Internal Medicine II, University Hospital Würzburg, Würzburg, Germany
2　Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany

🌐 Tables 1 s, 2 s; Figs. 1 s–4 s
Supplementary material is available under
https://doi.org/10.1055/a-1770-7353

🌐 Scan this QR-Code for the author commentary.

Corresponding author

Alexander Hann, Universitätsklinikum Würzburg, Medizinische Klinik und Poliklinik II, Oberdürrbacher Str. 6, 97080 Würzburg, Deutschland
hann_a@ukw.de

ABSTRACT

Background Multiple computer-aided systems for polyp detection (CADe) have been introduced into clinical practice, with an unclear effect on examiner behavior. This study aimed to measure the influence of a CADe system on

* These authors contributed equally to this work.

reaction time, mucosa misinterpretation, and changes in visual gaze pattern.

Methods Participants with variable levels of colonoscopy experience viewed video sequences (n = 29) while eye movement was tracked. Using a crossover design, videos were presented in two assessments, with and without CADe support. Reaction time for polyp detection and eye-tracking metrics were evaluated.

Results 21 participants performed 1218 experiments. CADe was significantly faster in detecting polyps compared with participants (median 1.16 seconds [99%CI 0.40–3.43] vs. 2.97 seconds [99%CI 2.53–3.77], respectively). However, the reaction time of participants when using CADe (median 2.90 seconds [99%CI 2.55–3.38]) was similar to that without CADe. CADe increased misinterpretation of normal mucosa and reduced the eye travel distance.

Conclusions Results confirm that CADe systems detect polyps faster than humans. However, use of CADe did not improve human reaction times. It increased misinterpretation of normal mucosa and decreased the eye travel distance. Possible consequences of these findings might be prolonged examination time and deskilling.

Introduction

Computer-aided systems for polyp detection (CADe) are expected to improve the quality of care in screening colonoscopy for colorectal cancer (CRC). The main indicator for the quality of screening colonoscopy is the adenoma detection rate (ADR) [1,2]. A meta-analysis of randomized studies evaluating the effect of CADe systems on ADR showed a significant increase in ADR with the use of artificial intelligence (AI) [3]. Even when compared with existing advanced imaging techniques and distal attachment devices, AI systems presented superior benefits, as outlined in a recent systematic review and network meta-analysis [4].

Several CADe systems have already been introduced to the market and are used with increasing frequency [5–7]. Thus, scientific investigation of this new technology is needed to understand the effect of CADe systems on the examiner. Hassan et al. showed that the first detection time (FDT) of the proposed CADe system was shorter than the reaction time of endoscopists when a polyp appeared on the screen during colonoscopy [8]. A measurement of the reaction time of the examiner when using the CADe system was not performed in that study. A short FDT could allow the detection of additional polyps that appear for only a short period of time. However, an important limitation of CADe systems is false-positive detections. Identification and classification of false positives are important and are the subject of current research [9]. However, few data are available on the effect of false positives on the examiner and on the ability to react to CADe findings. Most importantly, potential overreliance on CADe and deskilling need to be analyzed. Less-experienced endoscopists might be particularly susceptible to these effects and should be included as a dedicated subgroup in CADe research [10].

Eye-tracking assessment of colonoscopists enables analysis of visual gaze pattern (VGP) during colonoscopy [11–13]. The VGP of expert colonoscopists differs from that of novices [12]. Therefore, the assessment of VGP has the potential to reveal behavioral changes induced by CADe systems.

This study aimed to assess the influence of a CADe system on the reaction time of examiners. In addition, the influence of false-positive detections on the misinterpretation of mucosal surfaces and changes in VGP of examiners were analyzed.

Examiners were further stratified into novices and experienced endoscopy staff in order to account for this confounding factor.

Methods

Study design

The crossover design of the study included two different arms (see **Fig. 1s** in the online-only Supplementary Material). Each arm contained two assessments with a washout period of 3 weeks in between. Participants reviewed 29 colonoscopy video sequences per assessment while wearing eye-tracking glasses (**Fig. 2s**). The same video sequences were presented with and without the overlay of bounding boxes generated by a commercially available CADe system (GI Genius, version March 2020; Medtronic GmbH, Meerbusch, Germany), without audio cues. A total of 17 videos were presented containing a single polyp each; the remaining videos contained only false-positive detections of different duration. The selection of these latter videos was stratified according to the duration of the false-positive activations as proposed by Hassan et al. [9]: five videos presented mild (<1 second), three moderate (1–3 seconds), and four severe (>3 seconds) false-positive activations. In each video sequence, participants had to discern whether the bounding boxes presented by the CADe were actually polyps or not. In the first assessment, participants viewed the videos with or without CADe; in the second assessment, the alternative format was viewed, so that each participant viewed the videos once with CADe and once without CADe. Video order was kept the same. Participants knew beforehand whether the assessment was with or without the use of CADe. Further information about participants, materials, and statistical analysis are described in the Supplementary Material. Video sequence characteristics can be found in **Table 1s**.

Study end points

The measurement of the reaction time per polyp of the examiner with CADe support, the examiner without CADe support, and the FDT of the CADe system itself were set as primary end points. The study objective was the comparison of the different reaction times. Reaction time was defined as the time needed to press the button after the first appearance of the polyp in a video sequence.

Misinterpretation of the mucosa that led to false-positive detections of the participant and eye-tracking metrics were secondary end points. Regarding misinterpretations, we performed a comparative analysis between misinterpretations with and without CADe support. In addition, eye-tracking metrics including the distance of eye travel and the percentage of time an examiner inspected CADe bounding boxes were assessed. In addition, a heatmap containing gaze data of an experienced examiner and a novice, with and without CADe support, using a video with mild false-positive activations was generated to exemplify the different eye travel patterns. A *P* value of <0.01 indicated statistical significance.

Ethical considerations

The study was approved by the local ethical committee (No. 2021033001). All procedures were in accordance with the Helsinki Declaration of 1964 and later versions. Written informed consent for publication from the person shown in **Fig. 2 s** was obtained prior to submission.
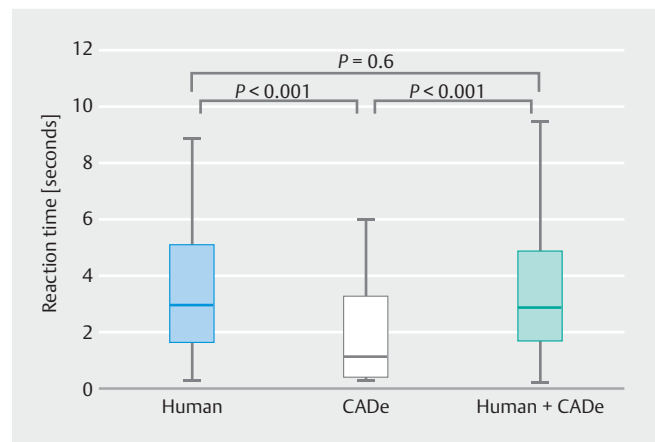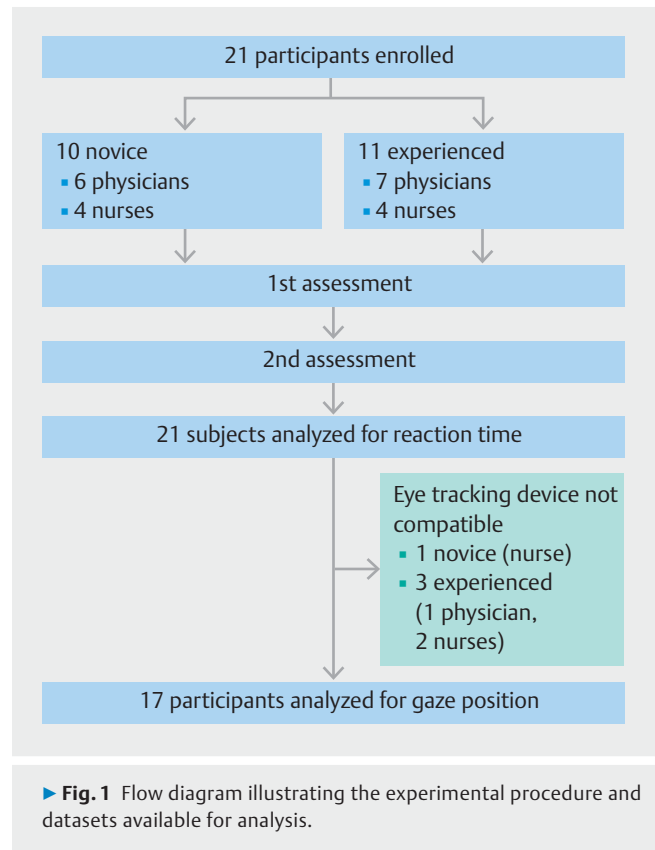
Results

Participants

A total of 21 participants were enrolled in the study (▶ **Fig. 1**) and assessed 29 video sequences with and without the CADe system, leading to a total of 1218 experiments. The ratio of professional experience was balanced, with 10 novice and 11 experienced participants. The reaction time was analyzed in all experiments. A comparison of the median reaction time paired distribution at the first assessment with the one at the second assessment confirmed the effectiveness of the 3-week washout period. Owing to optical interference of prescription glasses with the eye-tracking device, the eye tracking data of one novice and three experienced participants could not be analyzed.

Reaction time for polyp detection

Out of 17 polyps visible in the videos, experienced and novice examiners both detected a median of 16 polyps (99 %CI 16–16 for both). The use of the CADe system did not change the number of polyps detected.

The median FDT of the CADe system itself was 1.16 seconds (99 %CI 0.40–3.43). This was significantly faster than the median reaction time of participants (2.97 seconds [99 %CI 2.53–3.77]). This finding held true when the six participants with the fastest reaction times were compared with the CADe FDT. Surprisingly, the median reaction time of participants using the CADe system (2.90 seconds [99 %CI 2.55–3.38]) was not statistically different from that without CADe support (▶ **Fig. 2**).

We performed a subgroup analysis to estimate the impact of experience on the reaction time and how the CADe influenced polyp detection by novices. The median reaction time of experienced examiners was 2.54 seconds, which was faster than that of novices (3.38 seconds); however, this difference was not statistically significant. CADe support did not significantly decrease the reaction time of either of the two groups (**Fig. 3 s**). There was no significant difference in reaction time between the subgroups of professional roles (**Table 2 s**).

▶ **Fig. 1** Flow diagram illustrating the experimental procedure and datasets available for analysis.

▶ **Fig. 2** Distribution of the reaction time of 21 participants for the detection of polyps with and without the support of a computer-aided detection (CADe) device (green and blue, respectively). Additionally displayed is the distribution of the CADe first detection time (white). Boxes extend from the lower to the upper quartile values of the data with a line at the median. Whiskers denote 1.5 × interquartile range.

▶ **Table 1** Eye travel distance with and without computer-aided detection.

	Eye travel distance, median (99%CI), cm		P value
	Without CADe	CADe	
All videos			
▪ All participants	248.86 (221.39–282.88)	232.68 (210.43–262.33)	<0.001
▪ Experienced	247.89 (209.68–299.26)	227.09 (192.39–278.32)	
▪ Novice	255.66 (221.40–297.99)	247.89 (216.31–266.20)	
False-positive videos only			
▪ All participants	368.95 (331.75–395.10)	329.76 (293.81–387.00)	

CADe, computer-aided detection.

Analyzing the five longest FDTs of the CADe system (range 3.3–6.03 seconds) revealed that four of those were also included in the five longest reaction times of participants without CADe help (range of medians 5.83–19.2 seconds). Thus, polyps that were recognized with difficulty (i. e. longer FDTs) by the CADe system were also recognized with difficulty by participants.

Misinterpretation of polyps

A total of 29 videos were used to analyze how often the examiner misinterpreted the displayed mucosa for a polyp. Participants falsely identified a polyp in a median of 4 cases (99%CI 2–5) without CADe and a median of 6 cases (99%CI 4–7) with CADe support ($P = 0.001$, n = 21). Regardless of the experience level, both groups significantly misidentified more normal mucosa for a polyp when using CADe support (**Fig. 4 s**).

Considering the classification of false-positive detections published previously [9], participants falsely identified a polyp significantly more often in videos containing moderate and severe activations than in videos containing mild activations (99% CI 0–1, 1–1, 0–0, respectively).

Eye gaze

To further analyze the influence of the CADe system on the gaze pattern of participants, the eye travel distance was measured. The eyes of participants watching videos without a CADe system traveled significantly longer distances than when participants viewed the same videos with the support of a CADe system (▶ **Table 1**). This difference was also significant for experts, novices, and when videos containing only false positives were analyzed. ▶ **Fig. 3** is an example of the area covered by the gaze of a novice and an experienced examiner on a video containing no polyp.

To assess whether and how long participants visualized false-positive detections generated by the CADe system, we analyzed the percentage of time an examiner inspected false-positive bounding boxes (▶ **Table 2**). Participants inspected the bounding boxes for < 50% of the time that the boxes were displayed on the screen. This percentage was lowest for mild false-positive activations in both novice and experienced groups. In addition, novices spent significantly more time than experienced examiners inspecting the CADe false-positive detections.
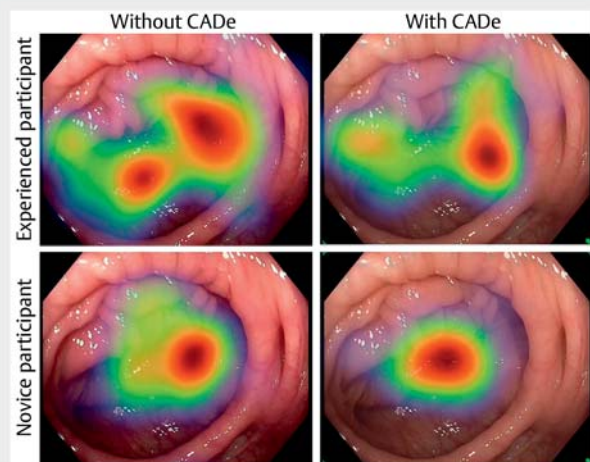
▶ **Fig. 3** Heatmap representing the eye movement of an experienced and a novice participant watching a video with and without the use of computer-aided detection (CADe). For improved visualization, the heatmap was overlaid on a still image of this video. The warmer the color of the overlay, the more the participant visualized the specific area.

Discussion

In this study, we analyzed the influence of a commercially available CADe system on endoscopy professionals regarding time to polyp detection, misinterpretation of colonic mucosa, and VGP changes. Longer withdrawal times lead to improved ADR rates [14], but examination time is limited in clinical practice. Evaluation of the difference in FDT between the CADe system and the reaction time of the participants revealed a significant difference of 1.81 seconds. This is in agreement with the study by Hassan et al., who reported a difference of 1.27 seconds between colonoscopists and the CADe system [8]. Thus, we can confirm that a CADe system detects faster than humans. A fast CADe system could analyze more suspicious mucosal areas in the same time span, thereby potentially leading to higher ADRs.

In our study, experienced participants presented a nonsignificant trend toward faster reaction times compared with novi-

▶ **Table 2** Percentage of time an examiner inspected false-positive bounding boxes relative to duration of the bounding box on screen.

	Inspection time, % (95 %CI)		
	Experienced	Novices	All
False-positive activation length*			
▪ Mild	22.86 (18.10–34.29)	28.10 (15.24–33.81)	24.76 (18.10–32.38)
▪ Moderate	47.64 (30.19–63.68)	51.89 (36.08–69.34)	50.94 (35.85–63.68)
▪ Severe	42.14 (23.67–55.51)	56.53 (38.01–77.35)	47.75 (39.18–63.78)
▪ All	42.68 (30.14–50.44)	49.50 (41.31–63.28)	43.79 (40.71–54.89)

* Mild = <1 second; moderate = 1–3 seconds; severe = >3 seconds.

ces. However, we also observed that there was no shortening of reaction time when CADe system support was used. An influencing factor might have been the false-positive detections of the CADe system. Reaction time may have been prolonged because the participant had to critically verify all bounding boxes. However, in general, the shorter FDT of the CADe system offers the potential to point the examiner in the right direction for polyp detection, thereby remodeling the VGP of endoscopists.

To further evaluate the effect of false-positive CADe detections on the examiner, we analyzed videos without polyps. In these cases, the use of the CADe system led to misinterpretation of normal mucosa. This was true for both novices and experienced participants. Based on these data, we hypothesize that overreliance is a potential drawback of CADe systems, and that a thorough evaluation of bounding boxes is mandatory.

Evaluation of participants by eye-tracking revealed VGP remodeling influenced by the CADe system. Videos examined with AI support significantly reduced eye travel distance. This was pronounced in the sequences without polyps where examiners were expected to show their visual polyp search pattern. Effort, as expressed by eye travel distance, decreased, and gaze remained more focused, presumably waiting for a bounding box to appear. This might be efficient but risks missing a polyp that is not captured by the system. To further analyze the changes in VGP, we generated a heatmap of the principal search pattern for polyps using data from one experienced and one novice participant. This is consistent with a previous report by Lami et al. describing the VGPs of colonoscopists with different polyp detection rates, where the distribution of fixations to the so-called "bottom U" of the screen was positively correlated with polyp detection rate [12]. Extrapolating this finding raises the concern that CADe systems may have an impact on the learning curve of colonoscopy trainees by preventing them from developing the visual "bottom U" pattern of high-performing endoscopists [10].

Regarding false-positive detections, participants inspected false-positive bounding boxes for <50 % of the time that the bounding boxes were displayed on the screen. The percentage decreased even more significantly for short activations. Hassan et al. previously questioned the relevance of these short activations; thus, developers of CADe systems might consider suppressing them [9].

Limitations of the study include the experimental design using only short video sequences that were chosen in order to increase the number of voluntary participants.

In conclusion, we confirmed the superiority of CADe systems compared with human examiners regarding the time to polyp detection. However, use of CADe did not result in a shorter time to polyp detection by human examiners. The use of the CADe system led to greater misinterpretation of the mucosa and reduction in eye travel distance during mucosal inspection. Analysis of false-positive detections and eye tracking metrics revealed marked changes, influenced by CADe, and suggests a potential risk of overreliance and deskilling when these systems are used.

Competing interests

The authors declare that they have no conflict of interest.

Funding

References

[1] Corley DA, Jensen CD, Marks AR et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014; 370: 1298–1306

[2] Kaminski MF, Robertson DJ, Senore C et al. Optimizing the quality of colorectal cancer screening worldwide. Gastroenterology 2020; 158: 404–417

[3] Hassan C, Spadaccini M, Iannone A et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021; 93: 77–85

[4] Spadaccini M, Iannone A, Maselli R et al. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. Lancet Gastroenterol Hepatol 2021; 6: 793–802

[5] Pfeifer L, Neufert C, Leppkes M et al. Computer-aided detection of colorectal polyps using a newly generated deep convolutional neural network: from development to first clinical experience. Eur J Gastroenterol Hepatol 2021; 33: e662–e669

[6] Repici A, Badalamenti M, Maselli R et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020; 159: 512–520

[7] Weigt J, Repici A, Antonelli G et al. Performance of a new integrated computer-assisted system (CADe/CADx) for detection and characterization of colorectal neoplasia. Endoscopy 2022; 54: 180–184

[8] Hassan C, Wallace MB, Sharma P et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. Gut 2020; 69: 799–800

[9] Hassan C, Badalamenti M, Maselli R et al. Computer-aided detection-assisted colonoscopy: classification and relevance of false positives. Gastrointest Endosc 2020; 92: 900–904

[10] Bisschops R, East JE, Hassan C et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – update 2019. Endoscopy 2019; 51: 1155–1179

[11] Almansa C, Shahid MW, Heckman MG et al. Association between visual gaze patterns and adenoma detection rate during colonoscopy: a preliminary investigation. Am J Gastroenterol 2011; 106: 1070–1074

[12] Lami M, Singh H, Dilley JH et al. Gaze patterns hold key to unlocking successful search strategies and increasing polyp detection rate in colonoscopy. Endoscopy 2018; 50: 701–707

[13] Meining A, Atasoy S, Chung A et al. "Eye-tracking" for assessment of image perception in gastrointestinal endoscopy with narrow-band imaging compared with white-light endoscopy. Endoscopy 2010; 42: 652–655

[14] Barclay RL, Vicari JJ, Doughty AS et al. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. N Engl J Med 2006; 355: 2533–2541