



HHS Public Access

Author manuscript

Gastroenterology. Author manuscript; available in PMC 2024 October 29.

Published in final edited form as:

Gastroenterology. 2023 March ; 164(3): 481–483.e6. doi:10.1053/j.gastro.2022.12.004.

Computer-aided Detection of Polyps Does Not Improve Colonoscopist Performance in a Pragmatic Implementation Trial

URI LADABAUM,

Division of Gastroenterology and Hepatology, Department of Medicine, Stanford University School of Medicine, Stanford, California

JOHN SHEPARD,

Critical Care Quality and Strategic Initiatives, Stanford Health Care, Stanford, California

YINGJIE WENG,

Quantitative Sciences Unit, Department of Medicine, Stanford University School of Medicine, Stanford, California

MANISHA DESAI,

Quantitative Sciences Unit, Department of Medicine, Stanford University School of Medicine, Stanford, California

SARA J. SINGER,

Department of Medicine, Stanford University School of Medicine, Stanford University Graduate School of Business, Stanford, California

AJITHA MANNALITHARA

Division of Gastroenterology and Hepatology, Department of Medicine, Stanford University School of Medicine, Stanford, California

Correspondence Address correspondence to: Uri Ladabaum, MD, MS, Division of Gastroenterology and Hepatology, Stanford University School of Medicine, 430 Broadway Street, Pavilion C, 3rd Floor C-326, Redwood City, California 94063-6341. uri.ladabaum@stanford.edu.

Conflicts of interest

This author discloses the following: Uri Ladabaum is on the advisory board for UniversalDx and Lean Medical, and is a consultant for Medtronic, Clinical Genomics, Guardant Health, Freenome, and Geneoscopy. The remaining authors disclose no conflicts.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org and at <https://doi.org/10.1053/j.gastro.2022.12.004>.

CRedit Authorship Contributions

Uri Ladabaum, MD, MS (Conceptualization: Lead; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Project administration: Lead; Supervision: Lead; Visualization: Equal; Writing – original draft: Lead). John Shepard, MBA, MHA (Data curation: Supporting; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Equal; Writing – review & editing: Supporting). Yingjie Weng, MHS (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Equal; Investigation: Supporting; Methodology: Supporting; Validation: Equal; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Desai Manisha, PhD (Conceptualization: Supporting; Formal analysis: Equal; Investigation: Supporting; Methodology: Equal; Supervision: Equal; Validation: Equal; Visualization: Supporting; Writing – review & editing: Supporting). Sara Singer, MBA, PhD (Methodology: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Ajitha Mannalithara, PhD (Conceptualization: Supporting; Data curation: Lead; Formal analysis: Lead; Investigation: Equal; Methodology: Equal; Project administration: Supporting; Resources: Supporting; Software: Lead; Validation: Equal; Visualization: Lead; Writing – original draft: Supporting; Writing – review & editing: Supporting).

Artificial intelligence (AI), including computer-aided detection (CADe), could revolutionize endoscopy. The adenoma detection rate (ADR) is inversely associated with the risk of postcolonoscopy colorectal cancer.¹ The first CADe device approved in the United States (GI Genius; Medtronic, Minneapolis, MN) significantly increased the ADR and adenomas per colonoscopy (APC)^{2,3} and decreased the adenoma miss rate⁴ in randomized trials.

We assessed the CADe device in a 3-month trial that leveraged our Stanford Colonoscopy Quality Assurance Program⁵ infrastructure to address a research priority identified by a Delphi process with international experts: studies of real-world endoscopist–AI interaction in intended clinical pathways, reporting relevant patient outcomes.⁶ We performed a pragmatic implementation study in routine practice of the impact of CADe on a comprehensive set of colonoscopy quality metrics. By design, we used a minimalist deployment strategy, including standard startup training, but no additional measures that could affect endoscopist behavior. We hypothesized that lesion detection rates would be higher (particularly for endoscopists with lower baseline detection rates), procedure times would be longer, and non-neoplastic resection rates would be higher with vs without CADe.

The Supplementary Material details our methods. We conducted a retrospective pragmatic trial with historical and concurrent control subjects. CADe devices were installed in our health system's largest outpatient endoscopy unit ("CADe site") for a 3-month evaluation (February to May 2022, the implementation period). Our system's 5 other units served as control sites. After the CADe devices were returned, we first assessed CADe use; then, using a difference-in-difference approach,⁷ we analyzed whether quality metrics changed as hypothesized in the CADe site, compared with control sites, during the implementation vs preimplementation periods, matching each endoscopist's number of colonoscopies. This approach accounts for a possible period effect independent of CADe use and for differences between study sites and is preferred over a simple comparison of metrics between sites with or without CADe. Endoscopists at control sites were not made aware of the CADe trial, but we made no effort to limit casual communication. Endoscopists were not aware of any hypotheses. The Stanford Institutional Review Board approved the study.

During the implementation period, CADe was used in 1008 of 1037 (97.2%) eligible colonoscopies. Of these, 619 were performed for screening/surveillance by 24 endoscopists who participate in our quality assurance program. The implementation and preimplementation period study cohorts in the CADe and control sites were comparable across demographics and colonoscopy indications (Supplementary Table 1).

During the implementation period in the CADe site, ADR was 40.1% (95% confidence interval [CI], 36.2%–44.0%) and mean APC was 0.78 (95% CI, 0.68–0.90) with CADe vs 41.8% (95% CI, 37.9%–45.8%; $P = .44$) and 0.89 (95% CI, 0.77–1.02; $P = .23$), respectively, during the preimplementation period without CADe (Figure 1, Supplementary Table 1). The detection rates for sessile serrated lesions, advanced adenomas or sessile serrated lesions, and lesion multiplicity were also comparable across periods (Supplementary Table 1, Supplementary Figure 1). In the control sites, all detection metric results without CADe use were comparable between the implementation and preimplementation periods (Supplementary Table 1, Supplementary Figure 1).

No statistically significant effect of CADe on ADR (odds ratio, 1.14; 95% CI, 0.83–1.56; $P = .41$), APC (OR, 1.08; 95% CI, 0.80–1.45; $P = .63$) or any other detection metric was detected by difference-in-difference analyses accounting for within-endoscopist correlation and adjusting for patient age and sex and procedure indication (Supplementary Table 1, Supplementary Figure 1). No effects of CADe on procedure times and non-neoplastic lesion resection rates were seen (Supplementary Table 1, Supplementary Figure 1). CADe use did not substantially mitigate differences in performance for ADR or APC (Figure 1) or for any other metric between lower vs higher tertiles of metric-specific baseline performance (Supplementary Table 1, Supplementary Figure 1).

Our results contrast sharply with those of randomized trials.^{2–4} Despite very high enthusiasm for trialing the technology, CADe use was not associated with improved detection rates. Although a ceiling effect might apply to high performers, it would not apply to lower performers. Given that CADe clearly identifies polyps,⁸ we must consider whether chance or subtle aspects of endoscopist behavior might explain our results. We caution against dismissing our study as an outlier, given a recent report of lower detection rates with vs without CADe.⁹

We were interested in the impact of a real-world, open-label implementation of CADe. We simply made CADe available, without any interventions beyond encouragement and basic startup training. We made no attempt to influence performance and had no discussions about hypotheses.

Perhaps there truly was a higher detection rate attributable to CADe *in exposed mucosa* in our study, but counterbalancing factors emerged. Some endoscopists may have dismissed suspected adenomas or sessile serrated lesions that were not highlighted by CADe, may have made errors in diagnosis and decisions about resection, or may have dismissed true-positive CADe prompts. Most concerning would be if, inadvertently, CADe use was accompanied by a simultaneous unconscious degradation in the *quality of mucosal exposure*, possibly due to a false sense of comfort that CADe would ensure a high-quality examination.

In contrast, the selected endoscopists in the randomized trials knew the study design and hypotheses, must have been cognizant that they could influence results on a nascent technology, and could not be blinded. It is possible that CADe in these trials encouraged better mucosal exposure or more careful lesion appraisal.

Substantial research from organizational and implementation sciences¹⁰ suggests that how new technologies are deployed influences outcomes. Ensuring clinicians' trust in and acceptance of a technology could result in more effective application. Attention to an implementation process (eg, intentional planning for deployment, discussion about achieving the technology's potential, reflection after deployment) could improve results.

We remain optimistic about CADe, which clearly identifies polyps.⁸ However, a minimalist deployment strategy may not ensure success. It may take a suite of AI features to maximize impact, including real-time assessment of mucosal exposure, CADe, lesion sizing, computer-aided diagnosis, assessment of resection adequacy, and support in generating endoscopy reports. Future challenges include ensuring that CADe detects subtle and

high-risk lesions, on which current modules were not trained. Whether AI will reduce postcolonoscopy colorectal cancer and mortality is a critical question.

In summary, our results contrast sharply with those of randomized trials. In real-world practice, CADe implementation without attention to endoscopist inclination and behavior may not achieve the intended results. Better understanding of subtle factors at the interface of technology and endoscopist performance, including mucosal exposure, could inform the development of multidimensional AI suites to promote uniformly high quality in endoscopy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Data Availability

Data and study materials cannot be made available to other researchers. Analytic methods are described in the Supplementary Material.

Abbreviations used in this paper:

ADR	adenoma detection rate
AI	artificial intelligence
APC	adenoma per colonoscopy
CADe	computer-aided detection

References

1. Schottinger JE, et al. *JAMA* 2022;327:2114–2122. [PubMed: 35670788]
2. Repici A, et al. *Gastroenterology* 2020;159:512–520. [PubMed: 32371116]
3. Repici A, et al. *Gut* 2022;71:757–765. [PubMed: 34187845]
4. Wallace MB, et al. *Gastroenterology* 2022;163:295–304. [PubMed: 35304117]
5. Ladabaum U, et al. *Am J Gastroenterol* 2021;116:1365–1370. [PubMed: 34183571]
6. Ahmad OF, et al. *Endoscopy* 2021;53:893–901. [PubMed: 33167043]
7. Wing C, et al. *Annu Rev Public Health* 2018;39:453–469. [PubMed: 29328877]
8. Hassan C, et al. *Gut* 2020;69:799–800. [PubMed: 31615835]
9. Levy I, et al. *Am J Gastroenterol* 2022;117:1871–1873. [PubMed: 36001408]
10. Nilsen P *Implement Sci* 2015;10:53. [PubMed: 25895742]

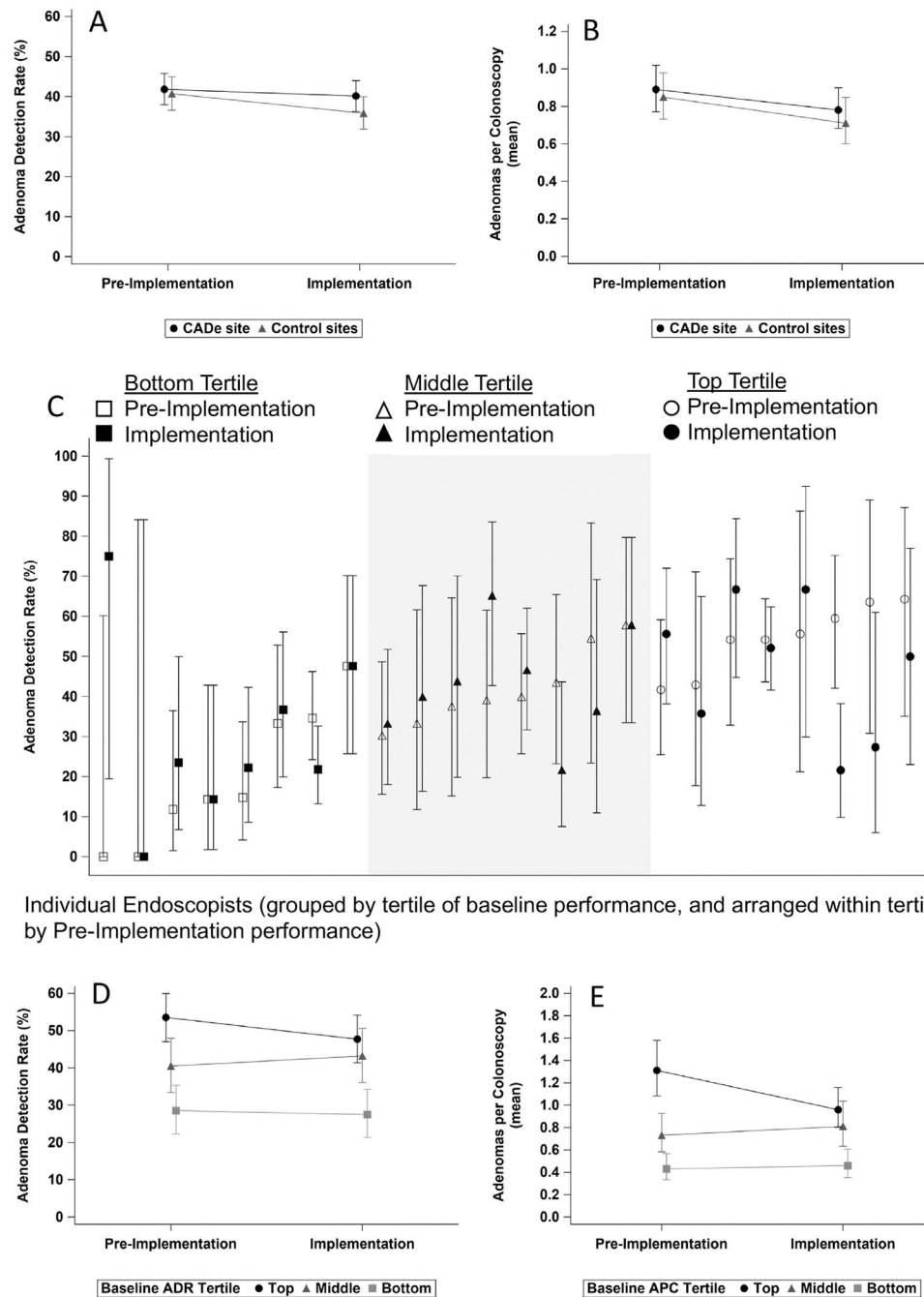


Figure 1. (A) Adenoma detection rate (ADR) and (B) adenomas per colonoscopy (APC) during the preimplementation and implementation periods in the computer-aided detection (CADe) and control sites. (C) Individual endoscopist ADR during the preimplementation and implementation periods in the CADe site, grouped by tertiles of endoscopist 12-month baseline ADR. (D) ADR and (E) APC during the preimplementation and implementation

periods in the CADe site, aggregated by tertiles of endoscopist 12-month baseline metric-specific performance.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript