Annals of Gastroenterology (2026) **39,** 1-8

Artificial intelligence-assisted colonoscopy: adenoma detection, health equity, and implementation barriers

Rowan Bandaranaike[a], Beth Rutland[b], Travis Rutland[b]

Southeast Health, Dothan, Alabama, USA

Abstract

Computer-aided detection (CADe) has been proposed to raise the adenoma detection rate (ADR) and to blunt the operator-dependent variability that undermines colonoscopy quality. We examined how far that promise holds. Drawing on major CADe meta-analyses (each pooling 21-48 largely overlapping randomized trials, the largest with 36,201 patients), the 2025 American Gastroenterological Association (AGA) Living Guideline and 2025 European Society of Gastrointestinal Endoscopy (ESGE) Position Statement, real-world implementation series, cost-effectiveness models, and the health-equity literature through January 2026, we ask not only whether CADe works, but where, for whom, and at what cost. The meta-analytic signal is consistent: ADR rises by roughly one fifth (rate ratio 1.20-1.21; about 8 absolute percentage points, e.g., 44.7% vs. 36.7%) and the adenoma miss rate falls by about half (rate ratio 0.45-0.47). Almost all of that additional yield represents diminutive adenomas. Three findings temper the enthusiasm: real-world series show a smaller benefit than trials, and such benefit is often non-significant; ADR gains appear to erode across the procedural day, consistent with alert fatigue; and no published trial we identified reports algorithm performance by race or ethnicity, despite documented differences in polyp morphology and distribution. The 2 guidelines from 2025 disagreed, with the AGA making no recommendation and the ESGE weakly favoring CADe; that split itself reflects how finely the benefits and harms are balanced. We argue that CADe makes most sense for lower-detecting endoscopists, but should not be widely adopted until it is proven to work equally well for all patients.

Keywords Computer-aided detection, colonoscopy, adenoma detection rate, adenoma miss rate, colorectal cancer

Ann Gastroenterol 2026; 39 (XX): 1-8

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer death in the United States (US), accounting for an estimated 153,020 new diagnoses and 52,550 deaths each year [1]. Colonoscopy anchors screening and prevention, but its protective effect is only as good as the examiner. The adenoma detection rate (ADR), the share of screening examinations in which at least 1 adenoma is found, ranges from under 10% to over 50% among endoscopists [2], and that spread has consequences: each 1-point rise in ADR corresponds to roughly a 3% fall in post-colonoscopy CRC (PCCRC) risk [3], a relationship recently reaffirmed in a large cohort by Pilonis *et al*, whose higher-ADR physicians had fewer subsequent cancers among their patients [4].

Computer-aided detection (CADe), deep-learning systems that flag polyps in real time, was designed to close that gap by catching what the human eye misses. Several platforms now hold regulatory clearance: these include GI Genius (Medtronic; Dublin, Ireland), EndoScreener (Wuhan EndoAngel; Wuhan, China), CAD EYE (Fujifilm; Tokyo, Japan) and DISCOVERY (Pentax; Tokyo, Japan) [5]. Their rapid uptake has produced an unusually large trial literature, but also unresolved questions about how well the technology performs outside trials, whether it is cost-effective, and whether it works equally well for every patient.

Two developments make those questions pressing. Early-onset CRC (diagnosed before age 50) is climbing by about

[a]Department of Internal Medicine, Southeast Health, Dothan, Alabama, USA (Rowan Bandaranaike); [b]Digestive Health Specialists, Dothan, Alabama, USA (Beth Rutland, Travis Rutland)

Conflict of Interest: None

Correspondence to: Rowan Bandaranaike, MD, Department of Internal Medicine, Southeast Health, 1108 Ross Clark Circle, Dothan, AL 36301, USA, e-mail: rowan16b@gmail.com

Received 7 April 2026; accepted 12 July 2026;
published online 25 August 2026

DOI: https://doi.org/10.20524/aog.2026.1107

2% annually and disproportionately affects racial and ethnic minorities [1]. And in 2025 the 2 bodies clinicians look to for guidance disagreed: the American Gastroenterological Association (AGA) declined to recommend for or against routine CADe [6], while the European Society of Gastrointestinal Endoscopy (ESGE) weakly endorsed it [7], a divergence that captures how unsettled the evidence remains.

Prior systematic reviews and meta-analyses have quantified the pooled effect of CADe on surrogate endpoints, such as ADR and adenoma miss rate (AMR), with considerable rigor [8-11]. What they have largely ignored are the problems that surface only in practice: the shortfall between trial efficacy and real-world effectiveness, the newer signals of alert fatigue and deskilling, and the near-total absence of equity-focused validation. This review draws those strands together. Rather than generating new pooled estimates, we read the existing literature (clinical outcomes, implementation experience and health equity) as a single problem, with the aim of giving clinicians usable guidance on when, and for whom, CADe has earned its place.

Materials and methods

This is an interpretive narrative review, not a systematic review. We chose that form deliberately: our question spans domains (trial efficacy, implementation, alert fatigue, deskilling, cost-effectiveness and health equity) that no single PRISMA-style synthesis captures well. We performed no original meta-analysis, and every statistic reported here (rate ratios, odds ratios, confidence intervals, P-values) is drawn directly from the cited primary sources; none was recalculated or independently verified by us.

We searched PubMed/MEDLINE and Google Scholar from January 2018 through January 2026, combining terms including "artificial intelligence," "computer-aided detection," "CADe," "colonoscopy,", "adenoma detection rate," "adenoma miss rate," "advanced adenoma," "sessile serrated lesion," "health equity," "algorithmic bias," "racial disparities," "colorectal cancer disparities," "cost-effectiveness," "alert fatigue," "deskilling," and "real-world implementation." We limited results to English-language publications, hand-searched the reference lists of included articles and relevant reviews, and last updated the search on January 15, 2026.

We gave priority to randomized controlled trials (RCTs) of CADe-assisted versus standard colonoscopy; published systematic reviews and meta-analyses; large observational cohorts and real-world implementation studies; cost-effectiveness and health-economic analyses; work on algorithmic bias, training-dataset composition, or equity in artificial intelligence (AI)-assisted endoscopy; and clinical practice guidelines from major societies. We excluded case reports, editorials without original data, conference abstracts lacking full-text publication, and studies of computer-aided diagnosis (CADx) with no CADe component.

Guidelines reviewed included the AGA Living Clinical Practice Guideline (2025) [6], the ESGE Position Statement informed by the British Medical Journal (BMJ) Rapid Recommendation initiative (2025) [7], the American College of Gastroenterology (ACG) Quality Indicators for Colonoscopy (2024) [12], and the AGA Clinical Practice Update on AI in colon polyp diagnosis and management (2023) [13]. Primary outcomes of interest were ADR, AMR, adenomas per colonoscopy (APC), and advanced colorectal neoplasia (ACN) detection; secondary topics included sessile serrated lesion (SSL) detection, non-neoplastic resection burden, procedural metrics, real-world effectiveness, alert fatigue, deskilling, cost-effectiveness, algorithmic bias, and equity. Quantitative findings are summarized in Tables 1-4.

Results

Adenoma detection: meta-analytic evidence

Across RCTs and their pooled analyses, the direction of effect is consistent: CADe raises adenoma detection (Table 1). The largest synthesis, by Soleymanjahi *et al* (44 RCTs, 36,201 patients), reported a higher ADR with CADe (44.7% vs. 36.7%; rate ratio [RR] 1.21, 95% confidence interval [CI] 1.15-1.28), higher APC (incidence rate difference 0.22, 95%CI 0.16-0.28), and an AMR that was roughly halved (RR 0.47, 95%CI 0.36-0.60) [9]. Makar *et al* (28 RCTs, 23,861 participants) found much the same, a 20% increase in ADR (RR 1.20, 95%CI 1.14-1.27) and a 55% drop in AMR (RR 0.45, 95%CI 0.37-0.54) [10], as did the earlier analysis by Hassan *et al* (21 RCTs, 18,232 patients: 44.0% vs. 35.9%; relative risk 1.24, 95%CI 1.16-1.33; ~55% miss-rate reduction) [11]. In a network meta-analysis of 48 RCTs (38,986 patients), Shinozaki *et al* compared individual platforms and found ENDO-AID (RR 1.26, 95% credible interval [CrI] 1.14-1.40), CAD-EYE (RR 1.18, 95%CrI 1.10-1.26), and GI Genius (RR 1.15, 95%CrI 1.08-1.22) supported by moderate-confidence evidence [8].

That concordance is reassuring but easy to over-read. The analyses were conducted independently, yet they draw on nearly the same underlying trials: Soleymanjahi's 44, Makar's 28, Hassan's 21, and Shinozaki's 48 overlap heavily (summarized in the Table 1 footnote). Their agreement therefore reflects repeated interrogation of 1 shared evidence base, not 4 independent replications. What it does establish is that the ADR benefit holds up across different search dates, inclusion criteria and analytic methods—a genuine, if surrogate-level, signal. Head-to-head comparisons between systems are still missing, so the apparent ranking among platforms may owe as much to differences in populations, endoscopist skill and study design as to any real superiority.

The individual trials behind these pools are instructive precisely because they disagree (Table 2). Repici *et al*, at a single Italian center, found that CADe (GI Genius) raised ADR from 40.4% to 54.8% (P<0.001), with a number needed to treat of 7 [14]; Shaukat *et al* confirmed an APC gain across screening and surveillance in a US multicenter trial [15]. However, Wei *et al*, in the community-based AI-SEE trial (769 patients), found nothing: ADR 35.9% vs. 37.2% (P=0.77), APC 0.73 vs.

Table 1 Pivotal meta-analyses of CADe-assisted versus standard colonoscopy

First author (year) [ref.]	Design	RCTs (n)	Patients (N)	ADR effect (95% CI)	AMR effect (95%CI)	Other key outcomes	Key caveats
Soleymanjahi (2024) [9]	SR & meta-analysis	44	36,201	44.7% vs. 36.7%; RR 1.21 (1.15-1.28)	RR 0.47 (0.36-0.60); ~53% reduction	APC IRD 0.22 (0.16-0.28)	Largest dataset; trials overlap substantially with other meta-analyses
Makar (2025) [10]	SR & meta-analysis	28	23,861	RR 1.20 (1.14-1.27)	RR 0.45 (0.37-0.54); ~55% reduction	APC WMD 0.21 (0.14-0.29); diminutive-only IRR 1.46 (1.19-1.80); non-neoplastic RR 1.39 (1.23-1.57); SSL NS	Benefit concentrated in diminutive (≤5 mm) lesions
Hassan (2023) [11]	SR & meta-analysis	21	18,232	44.0% vs. 35.9%; RR 1.24 (1.16-1.33)	~55% relative reduction	Non-neoplastic 0.52 vs. 0.34 per colonoscopy; no advanced-adenoma benefit	Earlier dataset; a subset of trials in later meta-analyses
Shinozaki (2026) [8]	Network meta-analysis	48	38,986 (network)	Several systems improved ADR vs. standard; ENDO-AID (RR 1.26), CAD-EYE (RR 1.18), and GI Genius (RR 1.15) supported by moderate-confidence evidence	NR	SSL detection (SSLDR) was a secondary outcome and varied across systems	Indirect comparisons; no head-to-head RCTs

These 4 meta-analyses were conducted independently but draw on a largely overlapping set of the same underlying CADe randomized trials; their concordance reflects repeated analysis of a shared evidence base rather than independent replication

ADR, adenoma detection rate; AMR, adenoma miss rate; APC, adenomas per colonoscopy; CADe, computer-aided detection; CI, confidence interval; IRD, incidence rate difference; IRR, incidence rate ratio; NR, not reported; NS, not significant; RCT, randomized controlled trial; RR, rate or risk ratio; SR, systematic review; SSL, sessile serrated lesion; WMD, weighted mean difference

Table 2 Landmark randomized controlled trials of CADe-assisted colonoscopy

Author (year) [ref.]	Setting/country	CADe system	N	Control ADR	CADe ADR	Key finding
Repici (2020) [14]	Single-center, Italy	GI Genius	685	40.4%	54.8%	ADR significantly improved (P<0.001); number needed to treat=7
Shaukat (2022) [15]	Multicenter, USA	SKOUT	NR	NR	NR	Improved adenomas per colonoscopy in both screening and surveillance populations
Wei – AI-SEE (2023) [16]	Multicenter, community, USA	EndoVigilant	769	37.2%	35.9%	No significant ADR difference (P=0.77) and no difference in APC (0.73 vs. 0.67); a negative real-world community RCT in which CADe did not improve adenoma detection

Control and CADe ADR values are shown only where reported in the primary trials summarized in this review

ADR, adenoma detection rate; CADe, computer-aided detection; NR, not reported in the cited source

0.67 (P=0.50) [16]—a reminder that efficacy in referral centers need not survive the move to community practice. Spadaccini *et al* offer the clearest explanation for that scatter: in a meta-regression of roughly 20,000 patients, both a lower control-arm ADR and a longer withdrawal time were independent predictors of how much CADe helped, and control-arm ADR emerged as the principal driver of between-trial heterogeneity [18]. The investigators treated baseline ADR as a continuous variable and identified no discrete cut-point above which the benefit disappears, so the exact inflection remains unknown. Limiting the analysis to expert endoscopists attenuated but did not erase the effect (RR 1.19, 95%CI 1.11-1.27) [10]. Read together, these data cast CADe less as a universal upgrade than as a leveling tool, most useful where baseline performance is weakest.

Advanced neoplasia and sessile serrated lesion detection

The lesions that actually drive cancer prevention, ACN, benefit only marginally. Pooling 22 studies (N=19,645), the AGA systematic review found CADe edged out standard colonoscopy on ACN detection rate (12.7% vs. 11.5%; RR 1.16, 95%CI 1.02-1.32) but showed essentially no difference in ACN per colonoscopy (incidence rate difference 0.01, 95%CI -0.01

Table 3 Real-world implementation and human–AI interaction studies

Author (year) [ref.]	Design	N	Key outcome	Caveat/interpretation
Patel (2024) [19]	SR & meta-analysis of 8 nonrandomized studies	9782	ADR 44% vs. 38%, RR 1.11 (0.97-1.28), NS; APC 0.93 vs. 0.79, NS	Point estimate suggests~11% relative gain, but CI crosses 1; real-world effect imprecise and non-significant
Nehme (2023) [20]	Pragmatic implementation, US tertiary center	NR	CADe activated in only 52.1% of cases; no APC improvement (1.08 vs. 1.04, P=0.65)	Non-activation attributed to operator factors (false positives 82.4%, distraction 58.8%); unavailable vs. available-but-not-activated not separately reported
Nayar (2026) [21]	Real-world, single tertiary center	4028	ADR 38.6% vs. 34.2%, RR 1.129 (1.013-1.257), P<0.05	Improvement driven predominantly by 5 mm lesions
Huang (2026) [22]	Secondary analysis (alert fatigue)	1700	CADe ADR 49.9% (first 3 cases/day) vs. 39.9% (later); APC benefit early+0.27 (P<0.001), late+0.11 (P=0.51)	Reported tests were CADe vs. unassisted within each window; a direct early- vs. late-day significance test was not reported
Takasu (2025) [23]	Retrospective, single center, Japan	NR	Non-CADe-room ADR 13.5% (after) vs. 31.6% (before), P<0.001	Retrospective, non-randomized, cannot establish causation; magnitude implausibly large for true deskilling and likely confounded; hypothesis-generating only

ADR, adenoma detection rate; APC, adenomas per colonoscopy; CADe, computer-aided detection; CI, confidence interval; NR, not reported; NS, not significant; RR, rate or risk ratio; SR, systematic review

Table 4 Cost-effectiveness analyses of CADe

Author (year) [ref.]	Model/approach	Key assumption(s)	Finding
Areia (2022) [24]	US population microsimulation	Assumed ADR increase from~25% to 37% (a 12-percentage-point absolute gain)	Cost-saving: 7194 fewer CRC cases and 2089 fewer deaths per year, $290 million saved per year. The assumed effect exceeds the pooled RCT estimate (a 20% relative gain on a 25% baseline is only ~5 percentage points), so the conclusion is sensitive to an optimistic input.
Thiruvengadam (2023) [25]	Microsimulation with threshold/sensitivity analysis	Threshold inputs: minimum ADR gain or maximum per-procedure cost	Cost-effective only if ADR improves ≥4 percentage points (26% to ≥30%) or CADe costs <$579 per procedure, thresholds closer to pooled and real-world estimates.

The 2 analyses differ in aim: Areia *et al*. modeled a base-case projection under an optimistic assumed effect size, whereas Thiruvengadam *et al*. performed threshold analysis to identify the minimum effect or price required for cost-effectiveness. This difference in analytic purpose, not a genuine contradiction, explains the apparent discrepancy between a 12-percentage-point assumption and a 4-percentage-point threshold
ADR, adenoma detection rate; CADe, computer-aided detection; CRC, colorectal cancer; RCT, randomized controlled trial

to 0.02) [6]. The 2 endpoints answer different questions, and here they come apart. Detection rate is a per-patient proportion; ACN per colonoscopy is a lesion count. CADe can nudge a few patients from 0 advanced lesions to 1, moving the proportion without adding to the total lesion burden, because advanced lesions are uncommon, and the extra pickups are mostly diminutive. The message is sobering: more patients cross the "at least 1 advanced lesion" threshold, but the total pool of advanced lesions barely grows, so the incremental effect on the endpoint that matters most for cancer prevention is small.

SSL detection remains inconsistent. Shinozaki *et al* treated it as a secondary outcome and saw it vary by system [8], while Makar *et al* found no significant improvement in either SSL detection (RR 1.10, 95%CI 0.93-1.30; P=0.27) or SSL miss rate (RR 0.44, 95%CI 0.16-1.19; P=0.11) [10]. Given how central the serrated pathway has become to our understanding of interval cancers, that unreliability is a real shortcoming.

The extra lesions CADe finds are overwhelmingly diminutive (≤5 mm) tubular adenomas of low malignant potential [10]. In a size-stratified analysis by Makar *et al*, the incremental yield reached significance only for diminutive lesions (incidence rate ratio [IRR] 1.46, 95%CI 1.19-1.80), with non-significant trends for small (6-9 mm; IRR 1.11, 95%CI 0.94-1.31) and large (≥10 mm; IRR 1.24, 95%CI 0.94-1.62) lesions [10]. Whether detecting more of these tiny adenomas helps patients is genuinely uncertain, since their risk of harboring advanced histology or progressing within a surveillance interval is very low [6,17]. CADe also drives up removal of non-neoplastic polyps (RR 1.39, 95%CI 1.23-1.57) [10], about 0.18 additional non-neoplastic polypectomies per colonoscopy in 1 analysis (0.52 vs. 0.34) [11], adding procedure time, pathology cost and patient anxiety for no clinical return. Downstream, the AGA panel's microsimulation projected that CADe would generate an estimated 635 additional surveillance colonoscopies per 10,000 people over 10 years, a modeled rather than observed figure, largely because more detected adenomas, especially diminutive ones, shorten recommended intervals [6]. That surveillance tail has real consequences for endoscopy capacity and for patients obliged to return more often.

Real-world implementation and alert fatigue

Outside trials, the benefit shrinks (Table 3). Patel *et al* pooled 8 nonrandomized studies (9782 patients) and found no significant difference in ADR (44% vs. 38%; risk ratio 1.11, 95%CI 0.97-1.28) or APC (0.93 vs. 0.79; mean difference 0.14, 95%CI -0.04 to 0.32) [19]. The 11% relative ADR gain in the point estimate looks encouraging, until the confidence interval, which crosses 1, is taken seriously: in the real world, the effect is imprecise and did not reach significance. Nehme *et al* watched what actually happens at the bedside in a US tertiary center, where CADe was switched on in just 52.1% of cases and produced no APC gain (1.08 vs. 1.04; P=0.65); endoscopists blamed false-positive signals (82.4%) and distraction (58.8%) [20]. That study could not separate cases in which CADe was unavailable from those in which it was available but deliberately left off. Nayar *et al*, in a more recent series of 4028 colonoscopies, did find a modest but significant ADR gain (38.6% vs. 34.2%; RR 1.129, 95%CI 1.013-1.257; P<0.05), though again the lift came mostly from ≤5 mm lesions [21]. The efficacy-effectiveness gap has several plausible sources: no Hawthorne effect in routine work, uneven engagement with alerts, inconsistent activation, and differences in case mix and bowel preparation [17].

More unsettling is the possibility that CADe's benefit decays within a single working day. Huang *et al*, analyzing 1700 colonoscopies at the University of Montreal, found CADe-assisted ADR far higher in the day's first 3 procedures (49.9%, 95%CI 45.5-54.3) than in later ones (39.9%, 95%CI 33.8-46.3), while unassisted ADR held steady (39.4% vs. 40.9%) [22]. The APC advantage told the same story: significant early (+0.27; P<0.001), gone late (+0.11; P=0.51) [22]. One caveat matters: the reported tests compared CADe with unassisted colonoscopy within each time window, not early-day against late-day CADe directly, so the apparent within-CADe decline is descriptive rather than formally tested. Even so, the pattern fits alert fatigue (the same erosion of responsiveness documented for monitor alarms in the intensive care unit) and, if real, implies that CADe helps most when the endoscopist is freshest.

A related and more speculative worry is deskilling. Takasu *et al*, in a retrospective Japanese study, reported that once CADe was introduced in 3 of 4 rooms, colonoscopies still performed without it detected fewer lesions than before its arrival (ADR 13.5% vs. 31.6%; polyp detection rate 18.2% vs. 39.8%; both P<0.001) [23]. This should be read very cautiously. A more-than-halving of ADR is far too large to represent genuine skill loss over the study window, and is much more likely due to confounding: secular trends, case-mix and selection differences, and systematic differences between the non-CADe room and the CADe rooms. Because rooms were not randomized, the design cannot support causal claims. We therefore treat it as hypothesis-generating: it raises, without demonstrating, the concern that reliance on CADe could dull independent detection, a concern that bears most directly on trainees learning colonoscopy in an assisted environment [17]. Adequately controlled prospective work is needed before anything stronger can be said.

Cost-effectiveness

Whether CADe saves money depends almost entirely on what one assumes about its effect size (Table 4). Areia *et al*, modeling the US population, concluded it was cost-saving (7194 fewer CRC cases, 2089 fewer deaths, and $290 million saved annually), but only by assuming CADe lifts ADR from about 25% to 37% [24]. That 12-point absolute jump is the problem: a 20% relative gain on a 25% baseline yields roughly 5 points (to about 30%), not 12, so the cost-saving verdict rests on an optimistic input. Thiruvengadam *et al* approached the question from the other end, asking what CADe would have to deliver to pay for itself, and found the bar modest: an ADR gain of at least 4 points (26% to ≥30%) or a per-procedure cost below $579 [25]. The 2 studies are often read as contradictory, but they are not: one is a base-case projection under a generous assumption, the other a threshold analysis—and the Thiruvengadam *et al* requirements sit much closer to pooled and real-world estimates. Both rely on RCT-derived ADR gains that probably overstate the real-world benefit. The AGA panel had similar reservations, noting that its modeling suggested CADe could substantially increase follow-up procedures and cost [6]. Reimbursement is the practical sticking point: with no dedicated Current Procedural Terminology code for AI-assisted colonoscopy in the US, practices currently absorb the cost of these systems unreimbursed.

Health equity and algorithmic bias

The equity questions are the most consequential, and the least answered. CRC burden in the US is far from uniform: incidence is highest among Alaska Native (88.5 per 100,000), American Indian (46.0), and Black (41.7) individuals, compared to 35.7 in White individuals, with a parallel gradient in mortality (35.9, 17.5 and 17.6 vs. 13.1 per 100,000) [1]. CRC death rates in Black Americans run about a third higher than in non-Hispanic White Americans (17.6 vs. 13.1 per 100,000; a ratio of 1.34) [26]. These gaps are multifactorial (differences in risk-factor prevalence, screening access and care quality [27]) and are compounded by tumor biology: Black patients are roughly 30% more likely to present with proximal, right-sided tumors, which carry worse outcomes and may be less amenable to screening [28]. May *et al* showed that these Black/White disparities span the entire care continuum, yet vanish in health systems that equalize access [27].

Against that backdrop, we found no CADe trial reporting performance by race or ethnicity. Because we did not design a search specifically to yield such a study, we frame this as absence of evidence within the literature we reviewed, rather than proof that none exists, a gap the commentary literature has flagged but not itself resolved by systematic search [29]. The gap still matters, because polyp morphology, size, location and mucosal appearance differ across populations, and an algorithm is only as general as its training data. Systems built largely on East Asian or European images may not transfer cleanly to Black, Hispanic or other underrepresented patients [29].

Amirmohammadi *et al* reviewed publicly available colonoscopy imaging databases and found demographic information on polyps and patients underreported in nearly all of them [30], which means the field cannot presently confirm that its algorithms have been validated across the populations they will serve. The AGA panel raised the same concern directly, warning that questions about how models were derived could sow mistrust among patients of different backgrounds, and that because CADe is costly and unreimbursed, it may reach only well-resourced settings and widen access gaps [6]. The point generalizes beyond the US: recommending CADe could deepen inequity between well- and under-resourced health systems worldwide [6]. Broader frameworks for auditing bias in clinical AI, such as that of Adegunle *et al*, argue for exactly the equity audits and transparent reporting standards that CADe has so far escaped [31].

Current guideline recommendations

The guidelines remain split, and the split is informative. Using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework, the 2025 AGA Living Guideline concluded that no recommendation could be made for or against CADe [6], citing very low certainty for the critical outcomes (11 fewer CRC cases and 2 fewer CRC deaths per 10,000 people over 10 years), the projected surveillance burden (a modeled 635 additional colonoscopies per 10,000), and cost and resource concerns [6]. The 2025 ESGE Position Statement, drawing on the BMJ Rapid Recommendation process, arrived at a weak recommendation in favor [7]: the panel granted that the evidence is limited, the absolute benefits small, and the patient burden real, but judged that most well-informed patients already committed to colonoscopy would still want the assistance [7]. The ACG's 2024 Quality Indicators declined to recommend for or against routine use, insisting instead that ADR remains the governing quality metric, with or without AI [12]. These are not so much conflicting readings of the data as different vantage points on the same close call: population-level versus individual-patient framing, and the unavoidable difficulty of recommending anything on the strength of a surrogate endpoint when the cancer outcomes that matter remain unmeasured.

Discussion

Taken together, the CADe literature tells a coherent but cautionary story. Its central tension is a mismatch between a robust, reproducible gain in a surrogate (ADR) and the absence of any demonstrated gain in what that surrogate stands for: fewer cancers. The ADR-PCCRC link is well grounded in observational data [3,4], but the AGA technical review and its source literature suggest the relationship is nonlinear, and probably flattens at the higher baseline ADRs now routinely achieved [6,17]. Where exactly it flattens is unclear; the plateau sometimes placed near 45% is best treated as the panel's reading of observational data, not a validated cutoff. The practical implication is uncomfortable: because CADe adds mostly diminutive adenomas, in populations that are already detecting effectively the marginal cancer-prevention payoff may be close to nil. No prospective trial has shown that a CADe-driven ADR gain reduces cancer incidence or mortality, and microsimulation implies modest benefit at best over a decade [6,25]. This is the deepest weakness in the case for CADe: the whole rationale assumes that the ADR-PCCRC relationship built on all adenomas applies equally to the incremental, mostly ≤5 mm lesions CADe supplies, the very lesions the size-stratified data single out as the only category with a significant yield [10].

The real-world attenuation compounds the problem and reshapes both practice and economic modeling [19-21]. Huang *et al*'s alert-fatigue signal suggests that human-AI collaboration itself degrades across the day, which argues for concrete countermeasures (periodic recalibration, scheduled breaks, or alternating CADe-on and CADe-off blocks) and for concentrating reliance where it works best: i.e. early [22]. Takasu *et al*'s deskilling signal, weak as it is, aims the same warning at training programs, where independent polyp recognition could atrophy if it is never practiced unassisted; that single confounded retrospective study cannot prove the point, but the AGA panel likewise flagged CADe's educational role as an open question [6,23].

Of everything unresolved, the equity gap is the one we would place first. Whether CADe narrows or widens CRC disparities depends entirely on 2 things the field has not secured: that the algorithms perform equitably across populations, and that they reach the patients with the most to gain [29,30]. The lesson from May *et al* is that equitable access, not the technology itself, is what erased disparities where they were erased [27], a caution against assuming that simply deploying CADe will help those who most need help.

We read our own contribution as one of emphasis rather than new computation. Several recent syntheses have quantified CADe efficacy rigorously [8-11], and reviews by Hassan *et al* and Putera *et al* have mapped its controversies and its expanding role across polyp management [17,32]. Relative to those, we lean harder on 3 underweighted areas: the alert-fatigue and deskilling signals that bear on real deployment [22,23]; the equity gap, including absent demographic reporting in training data and the potential for bias to widen disparities [29,30]; and a reading of the divergent 2025 AGA and ESGE positions that yields concrete, practice-level guidance rather than another pooled estimate [6,7]. We claim no methodological novelty, only a different weighting of the same evidence.

The limits of this review follow from its form. As a narrative rather than systematic synthesis, it used no PRISMA protocol or independent analysis, and is therefore open to selection bias in what we found and how we read it; we tried to blunt that by searching multiple databases, hand-searching references, and privileging the strongest evidence for outcome data. Publication bias and incomplete capture remain possible. As emphasized throughout, the pivotal meta-analyses share most of their

trials, so their agreement is not independent confirmation. Most implementation and equity studies are observational and thus confounded. PCCRC is rare enough that even pooled data may be underpowered for the outcomes that matter. And our equity argument, though well-anchored in the wider AI-fairness literature, is largely hypothesis-generating for CADe specifically; no study we identified has shown differential performance across racial groups: the problem is that the question has not been asked.

Several priorities follow. First, prospective studies with PCCRC reduction as the primary endpoint are needed to establish whether CADe delivers real clinical value; given how rare PCCRC is, this probably means large, long-follow-up international registries [6,7]. Second, algorithm validation across diverse populations, with mandatory transparent reporting of performance by race, ethnicity, age and sex, is essential, and we would require regulators to provide demographic performance data at premarket approval [29,30]. Third, cost-effectiveness analyses should be rebuilt on real-world rather than RCT-derived ADR gains, and should carry the surveillance tail and downstream utilization [24,25]. Fourth, implementation research on alert fatigue, deskilling, workflow and reimbursement will decide whether the technology's promise survives contact with practice [22,23]. Fifth, pairing CADe with CADx to offset the non-neoplastic resection burden deserves study—a gap the AGA panel also identified [6].

From the evidence, a few practical positions follow. CADe is on firmest ground for lower-detecting endoscopists, who show the largest relative and absolute gains [18]; we use an ADR of 25% as a pragmatic deployment threshold only because it is the established minimum benchmark for screening colonoscopy in mixed-sex populations [2], not because any trial has defined a CADe-specific cutoff—the meta-regression treats baseline ADR as continuous, so 25% is a heuristic, not a boundary [18]. For endoscopists already at or above that mark, CADe offers diminishing returns, and any decision should weigh its cost against a shrinking marginal benefit. Training programs should let residents build independent detection skills before layering CADe on top, and should teach the technology's limits explicitly [17,23]. Units adopting CADe should watch for alert fatigue, especially late in the day, and consider breaks, alternating activation or rotation to keep engagement u 2]. And before any institution buys in, it should demand from manufacturers the training demographics and subgroup performance data that equity requires, with professional societies pressing for regulatory standards to match [29,30].

CADe reliably improves adenoma detection at little procedural cost and could work as a scalable equalizer, most of all for lower-detecting endoscopists. But whether finding more diminutive adenomas helps patients is still unproven, the major guidelines disagree, and the largest gaps (long-term outcomes, cost-effectiveness, and above all equity-focused validation) remain open. No trial we identified reports CADe performance by race or ethnicity, and the datasets behind these algorithms are demographically opaque. In the absence of deliberate, equity-first development, AI-assisted colonoscopy could just as easily widen as narrow the CRC disparities that already fall hardest on Black and other minority patients. Getting this right will take not only better trials but a decision to build the technology so that it works for everyone.

Acknowledgments

Use of large language model (LLM) tools. In preparing this manuscript, the authors used a large language model (LLM) tool (Open Evidence) to assist with literature identification, drafting, and language editing. All references were verified by the authors against their primary sources. The authors reviewed and edited all content and take full responsibility for the accuracy, originality and integrity of the final manuscript.

References

1. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin* 2023;**73**:233-254.
2. Shaukat A, Kahi CJ, Burke CA, Rabeneck L, Sauer BG, Rex DK. ACG clinical guidelines: colorectal cancer screening 2021. *Am J Gastroenterol* 2021;**116**:458-479.
3. 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.
4. Pilonis ND, Spychalski P, Kalager M, et al. Adenoma detection rates by physicians and subsequent colorectal cancer risk. *JAMA* 2025;**333**:400-407.
5. Khalaf K, Rizkala T, Repici A. The use of artificial intelligence in colonoscopic evaluations. *Curr Opin Gastroenterol* 2025;**41**:3-8.
6. Sultan S, Shung DL, Kolb JM, et al. AGA living clinical practice guideline on computer-aided detection-assisted colonoscopy. *Gastroenterology* 2025;**168**:691-700.
7. Bretthauer M, Ahmed J, Antonelli G, et al. Use of computer-assisted detection (CADe) colonoscopy in colorectal cancer screening and surveillance: European Society of Gastrointestinal Endoscopy (ESGE) position statement. *Endoscopy* 2025;**57**:667-673.
8. Shinozaki S, Watanabe J, Kanno T, Nakazawa K, Yano T. Comparative performance of artificial intelligence-based computer-aided detection systems for colorectal polyps: a systematic review and network meta-analysis. *Dig Endosc* 2026;**38**:e70138.
9. Soleymanjahi S, Huebner J, Elmansy L, et al. Artificial intelligence-assisted colonoscopy for polyp detection: a systematic review and meta-analysis. *Ann Intern Med* 2024;**177**:1652-1663.
10. Makar J, Abdelmalak J, Con D, Hafeez B, Garg M. Use of artificial intelligence improves colonoscopy performance in adenoma detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2025;**101**:68-81.
11. Hassan C, Spadaccini M, Mori Y, et al. Real-time computer-aided detection of colorectal neoplasia during colonoscopy: a systematic review and meta-analysis. *Ann Intern Med* 2023;**176**:1209-1220.
12. Rex DK, Anderson JC, Butterly LF, et al. Quality indicators for colonoscopy. *Am J Gastroenterol* 2024;**119**:1754-1780.
13. Samarasena J, Yang D, Berzin TM. AGA clinical practice update on the role of artificial intelligence in colon polyp diagnosis and management: commentary. *Gastroenterology* 2023;**165**:1568-1573.
14. 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.

15. Shaukat A, Lichtenstein DR, Somers SC, et al; SKOUT™ Registration Study Team. Computer-aided detection improves adenomas per colonoscopy for screening and surveillance colonoscopy: a randomized trial. *Gastroenterology* 2022;**163**:732-741.
16. Wei MT, Shankar U, Parvin R, et al. Evaluation of computer-aided detection during colonoscopy in the community (AI-SEE): a multicenter randomized clinical trial. *Am J Gastroenterol* 2023;**118**:1841-1847.
17. Putera M, Koh FHX, Li JW. Controversies in computer-assisted detection in colonoscopy. *Digestion* 2026 Feb 5 [Online ahead of print]. doi:10.1159/000550268
18. Spadaccini M, Hassan C, Mori Y, et al; CADe Study group. Variability in computer-aided detection effect on adenoma detection rate in randomized controlled trials: a meta-regression analysis. *Dig Liver Dis* 2025;**57**:1141-1148.
19. Patel HK, Mori Y, Hassan C, et al. Lack of effectiveness of computer aided detection for colorectal neoplasia: a systematic review and meta-analysis of nonrandomized studies. *Clin Gastroenterol Hepatol* 2024;**22**:971-980.
20. Nehme F, Coronel E, Barringer DA, et al. Performance and attitudes toward real-time computer-aided polyp detection during colonoscopy in a large tertiary referral center in the United States. *Gastrointest Endosc* 2023;**98**:100-109.
21. Nayar KD, Yakout A, Bakheet N, et al. Outcomes of artificial intelligence-enhanced colonoscopy in a tertiary clinical setting. *Am J Gastroenterol* 2026;**121**:974-981.
22. Huang F, Michal V, Djinbachian R, et al. Computer-aided colonoscopy alert fatigue and its effect on adenoma detection. *Am J Gastroenterol* 2026;**121**:899-906.
23. Takasu A, Kogure H, Dai Z, et al. Impact of introducing artificial intelligence on colonoscopy: a retrospective study on potential benefits and drawbacks. *J Gastroenterol Hepatol* 2025;**40**:2258-2266.
24. Areia M, Mori Y, Correale L, et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. *Lancet Digit Health* 2022;**4**:e436-e444.
25. Thiruvengadam NR, Coté GA, Gupta S, et al. An evaluation of critical factors for the cost-effectiveness of real-time computer-aided detection: sensitivity and threshold analyses using a microsimulation model. *Gastroenterology* 2023;**164**:906-920.
26. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin* 2020;**70**:145-164.
27. May FP, Mehtsun WT, Jemal A, Gupta S. Black-White disparities across the colorectal cancer care continuum in the USA. *Nat Rev Gastroenterol Hepatol* 2025;**22**:603-618.
28. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer statistics for African American/Black People 2022. *CA Cancer J Clin* 2022;**72**:202-229.
29. Uche-Anya E, Anyane-Yeboa A, Berzin TM, Ghassemi M, May FP. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. *Gut* 2022;**71**:1909-1915.
30. Amirmohammadi E, Shalbaf A, Esteki A, et al. Application of artificial intelligence in colonoscopy imaging for polyp analysis-A systematic review. *Comput Methods Programs Biomed* 2026;**277**:109239.
31. Adegunle F, Chhatwal K, Arab S, et al. Bias and oversight in clinical AI: a review of decision support tools and equity frameworks. *J Gen Intern Med* 2026;**41**:1957-1968.
32. Hassan C, Bisschops R, Sharma P, Mori Y. Colon cancer screening, surveillance, and treatment: novel artificial intelligence driving strategies in the management of colon lesions. *Gastroenterology* 2025;**169**:444-455.