

Artificial intelligence for colonoscopic polyp detection: High performance versus human nature

James E. East MD(Res), FRCP; Jens Rittscher PhD

Colorectal cancer is a preventable disease(1). This statement is now widely accepted through screening via endoscopy, CT colonography, or faecal blood and DNA test; however, this statement should be qualified. Most colorectal cancer is preventable, dependant on the performance of the endoscopist performing the colonoscopy and polypectomy, which is the final common pathway in prevention-based programs. We know that endoscopist performance varies widely and is closely linked to post-colonoscopy colorectal cancer (PCCRC) rates though adenoma detection rate (ADR) measurement, with a 1% absolute increase in ADR leading to a 3% relative decrease in PCCRC rates(2). Feedback with quality assurance data has been shown to modify ADR and the associated PCCR rates, but this is challenging to deliver at scale(3). More straightforward educational interventions “evidence bundles” can also improve ADR with durable effects especially for poor performers(4). Devices can also modify ADR, such as wider-angle instruments, advanced endoscopic imaging, or caps and cuffs, though there are no data yet that this links to reduced PCCRC(5).

Artificial intelligence (AI) and deep learning-based techniques have exploded into the wider endoscopic consciousness in the last year or two with multiple studies suggesting that AI can improve adenoma detection, lesion characterization, procedure quality, and so forth. AI, today predominantly manifested by deep learning, provides powerful models that exceed human performance in certain visual interpretation tasks in endoscopy. It is therefore timely to look across the available data, which are often smaller studies as technologies develop, to see whether by combining studies appropriately, we can gain further insights and greater precision around performance estimates and test the generalizability of learnt models. Excitingly, the systematic review and associated meta-analysis of three randomized controlled trials by Aziz et al. in this issue of JGH suggest that AI does significantly enhance polyp and adenoma detection, with narrow confidence intervals(6).

Most striking is the absolute size of the effect, absolute increase in ADR of 12.1% (risk ratio 1.58, 95% confidence interval 1.39–1.80) from a modest baseline of 20.8% exceeds most other available interventions. In the Corley et al. Kaiser Permanente dataset, this would have moved the average endoscopist from ADR quintile 2 who have an unchanged rate of PCCRC compared with quintile 1 (worst performers), to quintile 4 (second best performers) who have a relative reduction in PCCRC rate of 0.51.2 If we compare this with relative effect sizes in a recent network meta-analysis for add-on devices (e.g. cuff of caps; odds ratio 1.18), or advanced imaging (e.g. narrow band imaging; odds ratio 1.21), we can see that the impact on ADR is modest for these technologies compared with the effect of AI for adenoma detection(5).

However, we should apply some caution to the results. Despite careful inclusion of randomized controlled trials only, all of the studies came from China, which may not be fully applicable to international practice, although a more recent AI detection study by Repici and colleagues based in Italy provides very similar results in terms of risk ratio and absolute increase in adenoma detection(7). We might also question whether transposing the results here onto the effects seen in the Kaiser Permanente dataset is truly appropriate. This is in part as a US healthcare dataset variation in PCCRC rate may not map onto the current Chinese derived data, but also for a more fundamental reason.

In the Corley et al. dataset, ADR is a function of both endoscopist mucosal visualization (through cleaning the mucosa, appropriate insufflation, pressing down and looking behind folds, and re-examining flexures) and the ability of the endoscopist to recognize sometimes subtle lesions on the exposed mucosal surface. Therefore, the reductions in PCCRC may reflect increased removal of small

adenomas, but are also likely to reflect more meticulous and comprehensive mucosal inspection where fewer lesions are missed in hard-to-see areas. We do not know the relative contribution of these two factors to PCCRC rates as measured through ADR; however, we do know that the AI algorithms described in the meta-analysis only contribute to the detection of adenomas already in the visual field and that the AI-assisted ADR (AI-ADR) does not reflect extra efforts to enhance mucosal visualization. If it turns out that comprehensiveness of mucosal inspection is the more important aspect in reducing PCCRC rates, then the striking enhancement of ADR by AI may not in fact lead to the reduction in PCCRC that we might expect. Against this interpretation is the sub-analysis presented where intriguingly and slightly unexpectedly, the enhanced detection rate for large lesions was quite similar to that for diminutive ones. This is in contrast to back-to-back colonoscopy studies where miss rates for diminutive lesions are much higher than for larger polyps(8).

Therefore, widespread use of AI for polyp detection (computer-aided detection [CAdE]) may reduce PCCRC rates, but by a difficult to determine amount without very large studies. Nevertheless, CAdE is already commercially available from multiple companies as certified medical devices and is in clinical use (Discovery AI, Pentax; CAdE EYE, Fujinon; GI Genius, Medtronic). Many endoscopists, including low performers, will have their ADRs enhanced and will perhaps meet performance criteria for ADR that they were previously unable to achieve, or only with very slow, meticulous, high effort examinations. Some health maintenance organizations and other payers have previously stipulated minimum levels of ADR, and if endoscopists cannot achieve them, they would not pay the procedure fee. Should then payers now set a different and higher ADR threshold for AI-ADR? Equally where volume of colonoscopy is key to business success, less scrupulous endoscopists might be tempted to perform a very rapid AI-enhanced colonoscopy where with AI support they could still meet the minimum ADR standard, perhaps with quite limited mucosal inspection.

There is a risk therefore that AI might allow “gaming” of the ADR quality assurance standard. How should those assessing and trying to enhance colonoscopic quality in order to reduce PCCRC approach AI, which will rapidly become the de facto standard of care in a few years? Reporting the use of AI in procedures would be a first step, especially if this was captured in administrative databases. If use of AI was associated with differential payment, this would likely happen automatically, but it is not yet clear whether payers are willing to reward the purchase and integrating of AI polyp detection systems (CAdE) into colonoscopy pathways, for reasons above. We may need to set different quality standards for AI-ADR procedures, and this is a matter of urgency given that there may be very few non-AI-supported procedures in a short time frame, driven by patient demand.

One potential solution is to enhance both polyp detection and mucosal visualization with AI systems—specifically, to have an AI solution that measures in real time and feeds back to the colonoscopist the proportion of mucosa visualized with clear levels of confidence and adequate bowel preparation. The field now needs to bridge the gap between studies on selected video images and continuous real-time performance on video. The analogy to building systems that enable the different levels of autonomous driving provides a good indication of the level of difficulty, and levels of automated support should also be clearly defined by the AI community. The technical challenges to this are significantly greater than polyp segmentation in a video image but would mean that both components that we think define a high-quality examination, and that should result in the lowest PCCRC rate, would be met. Google and other groups are already working on this problem(9).

With comprehensive AI colonoscopist support, almost all colorectal cancer might really be preventable. AI-enhanced polyp detection alone might not be enough.

Conflict of interest

JEE has served on clinical advisory boards for Lumendi, Boston Scientific, and Paion; has served on the clinical advisory board and owns share options in Satisfai Health; and reports speaker fees from Falk.

Funding

James East and Jens Rittscher are funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

References

1. Lauby-Secretan B, Vilahur N, Bianchini F et al. The IARC perspective on colorectal cancer screening. *N Engl J Med* 2018 378: 1734– 1740.
2. 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.
3. Kaminski MF, Wieszczyn P, Rupinski M et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. *Gastroenterology* 2017 153: 98– 105.
4. Neilson LJ, East JE, Rajasekhar PT et al. Sustained colonoscopy quality improvement using a simple intervention bundle. *Endoscopy* 2020 52: 285– 292.
5. Facciorusso A, Triantafyllou K, Murad MH et al. Compared abilities of endoscopic techniques to increase colon adenoma detection rates: a network meta-analysis. *Clin Gastroenterol Hepatol* 2019 17: 2439– 2454.
6. Aziz M, Fatima R, Dong C et al. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: a systematic review with meta-analysis. *J Gastroenterol Hepatol* 2020; 35: 1676– 1683
7. 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.
8. Zimmermann-Fraedrich K, Sehner S, Rex DK et al. Right-sided location not associated with missed colorectal adenomas in an individual-level reanalysis of tandem colonoscopy studies. *Gastroenterology* 2019 157: 660– 671.
9. Freedman D, Blau Y, Katzir L et al. Detecting deficient coverage in colonoscopies. *IEEE Trans Med Imaging* 2020 <https://doi.org/10.1109/TMI.2020.2994221>