

HealthTech draft guidance

Artificial intelligence (AI) technologies to help detect or characterise colorectal polyps

November 2025

Guidance development process

NICE diagnostics guidance evaluates diagnostic tests and technologies. It provides evidence-based recommendations about how accurate and effective the tests are, as well as their costs. The guidance supports healthcare professionals and commissioners to choose the best diagnostic options to improve patient care. NICE aims to enhance the quality of diagnostic practices in the UK, leading to better health outcomes for patients.

Find out more on the [NICE webpage on the diagnostics assessment programme](#).

NICE is producing this guidance on the use of AI technologies to help detect or characterise colorectal polyps in the NHS in England. The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts.

This document has been prepared for consultation with the stakeholders. It summarises the evidence and views that have been considered and sets out the recommendations made by the committee. NICE invites comments from the stakeholders for this evaluation and the public. This document should be read along with the [evidence](#).

The committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?

- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the recommendations sound and a suitable basis for guidance to the NHS?
- Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of age, disability, gender reassignment, pregnancy and maternity, race, religion or belief, sex or sexual orientation?

After consultation:

- Based on the consultation comments received, the committee may meet again.
- If committee meets again, it will consider the evidence, this evaluation consultation document and comments from stakeholders.
- The committee will then prepare the final draft guidance, which will go through a resolution process before the final guidance is agreed.

Note that this document is not NICE's final guidance on AI technologies to help detect or characterise colorectal polyps. The recommendations in section 1 may change after consultation.

More details are available in [NICE's health technology evaluations: the manual](#).

Key dates:

Closing date for comments: 11th December 2025

Second committee meeting: 7th January 2025

1 Recommendations

For people who do not have diagnosed IBD or Lynch syndrome

Can be used during the evidence generation period to help detect colorectal polyps

1.1 Five artificial intelligence (AI) technologies can be used in the NHS during the evidence generation period as options to help detect colorectal polyps during colonoscopy, for people who do not have diagnosed inflammatory bowel disease (IBD) or Lynch syndrome. The technologies are:

- CAD EYE
- ENDO-AID
- EndoScreener
- GI Genius
- MAGENTIQ-COLO.

These technologies can only be used:

- if the evidence outlined in the [evidence generation plan for AI technologies to help detect or characterise colorectal polyps](#) is being generated
- as long as they have appropriate regulatory approval including NHS England's Digital Technology Assessment Criteria (DTAC) approval.

1.2 The companies must confirm that agreements are in place to generate the evidence. NICE will contact the companies annually to confirm that evidence is being generated and analysed as planned. NICE may revise or withdraw the guidance if these conditions are not met.

- 1.3 At the end of the evidence generation period (4 years), the companies should submit the evidence to NICE in a format that can be used for decision making. NICE will review the evidence and assess if the technology can be routinely adopted in the NHS.

More research is needed to help detect colorectal polyps

- 1.4 More research is needed on 5 AI technologies used to help detect colorectal polyps during colonoscopy. The technologies are:

- Argus
- CADDIE
- Discovery
- ENDOANGEL
- Endoscopic Multimedia Information System (EMIS).

More research is needed to help characterise colorectal polyps

- 1.5 More research is needed on 4 AI technologies used to help characterise colorectal polyps. The technologies are:

- CADDIE
- CAD EYE
- GI Genius
- MAGENTIQ-COLO.

For people with diagnosed IBD or Lynch syndrome

- 1.6 More research is needed on 10 AI technologies to help detect or characterise colorectal polyps for people with diagnosed IBD or Lynch syndrome. The technologies are:

- Argus (for detecting)
- CAD EYE (for detecting and characterising)
- CADDIE (for detecting and characterising)
- Discovery (for detecting)
- EMIS (for detecting)

- ENDO-AID (for detecting)
- ENDOANGEL (for detecting)
- EndoScreener (for detecting)
- GI Genius (for detecting and characterising)
- MAGENTIQ-COLO (for detecting and characterising).

What this means in practice

For AI technologies that can be used during the evidence generation period

The 5 AI technologies listed in recommendation 1.1 can be used in the NHS as options to help detect colorectal polyps during the evidence generation period (4 years) and paid for using core NHS funding. During this time, more evidence will be collected to address any uncertainties. Companies are responsible for organising funding for evidence generation activities.

After this, NICE will review this guidance and the recommendations may change. Take this into account when negotiating the length of contracts and licence costs.

Potential benefits of use in the NHS during the evidence generation period

- **Clinical benefit:** Clinical evidence shows that AI technologies can help endoscopists detect adenomatous polyps (which can develop into colorectal cancer) during colonoscopy which might have otherwise been missed. Early detection means earlier access to appropriate treatment or surveillance pathways as needed. This could reduce the number of future colorectal cancer cases and improve clinical outcomes.

- **Resources:** AI technologies could reduce the need for more intensive healthcare resources later, such as for emergency presentations or more advanced stages of cancer.

Managing the risk of use in the NHS during the evidence generation period

- **Costs:** Results from economic analyses suggest that the AI technologies could be cost effective. But these results are uncertain. By collecting more evidence on how effective the AI technologies are, and the numbers of colorectal cancers missed and prevented in the future, this uncertainty could be reduced. Trusts should consider the costs of the AI technologies, particularly upfront costs, when implementing them.
- **Clinical risk:** The AI technologies should only be used as an aid to detect colorectal polyps. The technologies cannot replace endoscopist review. AI technologies for characterising colorectal polyps should only be used in research because evidence for characterisation is more limited.
- **Equality:** Evidence generation could help address equality by collecting data on differences in how effective the AI technologies are for some groups of people compared with others.

For AI technologies that need more research

There is not enough evidence to support NHS funding of the:

- 5 AI technologies to help detect colorectal polyps (see recommendation 1.4)
- 4 AI technologies to help characterise colorectal polyps (see recommendation 1.5)
- 10 AI technologies to help detect or characterise colorectal polyps for people with IBD or Lynch syndrome (see recommendation 1.6).

When more research is needed, access to the AI technologies should be through company, research or non-core NHS funding, and clinical or financial risks should be managed appropriately.

What evidence generation is needed

More evidence needs to be generated on the effect of the AI technologies on:

- polyp detection rate by type and size
- post-colonoscopy colorectal cancer rates
- clinical management.

The [evidence generation plan](#) gives further information on the prioritised evidence gaps and outcomes, ongoing studies and potential real-world data sources. It includes how the evidence gaps could be resolved through real-world evidence studies.

What research is needed

More research is needed:

- on the effect of the AI technologies on polyp detection rate:
 - by type and size
 - for people with IBD or Lynch syndrome.
- for AI technologies that can characterise colorectal polyps, on the:
 - sensitivity and specificity of using the AI technologies alongside endoscopist judgement to characterise different types of polyps
 - ability to differentiate between potentially cancerous polyps (including sessile serrated lesions; SSLs) and polyps that are not.

Why the committee made these recommendations

Clinical evidence shows that 5 AI technologies (CAD EYE, ENDO-AID, EndoScreener, GI Genius and MAGENTIQ-COLO) increase the percentage of colonoscopies in which one or more of a type of potentially cancerous polyp, called an adenoma, is detected. But evidence suggests that the

technologies mostly increase the number of smaller adenomas found, which are less likely than large polyps to develop into bowel cancer. It is uncertain whether the technologies increase the number of large adenomas found. There is also not enough evidence to show how effective the technologies are at helping to detect SSLs, which are another type of potentially cancerous polyp. So, it is uncertain in practice whether using AI technologies increases the number of cancer cases detected during colonoscopy. This means that their cost effectiveness is also uncertain. These 5 AI technologies can be used while evidence is generated.

There is limited evidence on the effectiveness of Argus, CADDIE, Discovery, ENDOANGEL and EMIS in detecting colorectal polyps. So, it is uncertain whether they increase the percentage of colonoscopies in which one or more adenoma is found. Also, it was not possible to estimate the cost-effectiveness for CADDIE or ENDOANGEL. So, more research is needed on these 5 AI technologies.

People with IBD or Lynch syndrome were excluded from a lot of the clinical trials and their bowels may look different, or they may have different types of polyps. So, it's uncertain whether the AI technologies will work as well in people with these conditions. So, more research is needed for these people.

Some AI technologies include a function to help the endoscopist characterise the polyp and decide if it is potentially cancerous without removing it or examining it in a laboratory. There is less evidence for this function, and the research is of lower quality. So, more research is needed on AI technologies to help characterise colorectal polyps.

2 Information about the technologies

The technologies

- 2.1 Polyps are small growths in the bowel. It is important to detect and identify these because some may develop into bowel (colorectal) cancer if left untreated. Colonoscopy is a procedure that allows

healthcare professionals to look inside the bowel using a thin, flexible tube with a camera to check for problems such as polyps or early signs of cancer. Colonoscopy is usually done as an outpatient procedure, with the option of sedation or pain relief.

- 2.2 Artificial intelligence (AI) software supports polyp detection during colonoscopy by highlighting areas of concern for the endoscopist in a process known as computer-aided detection (CAdE). Some AI software also offers computer-aided diagnosis (CAdx), which helps assess polyp features to guide decisions about removal. Ten AI technologies to assist colonoscopy in the detection of polyps were identified as relevant to the NHS. All 10 technologies featured CAdE and 4 also featured CAdx. The technologies are listed in table 1. The cost of the technologies is presented in table 2 of the external assessment report.

Table 1 Features of each technology

Technology name	Company	Regulatory approval status	Does the technology have CADx functionality?
Argus	Endosoft	In process	No
CADDIE	Odin Vision	CE marked	Yes
CAD EYE	Fujifilm Healthcare UK	CE class IIa	Yes
Discovery	Pentax Medical UK	CE class IIa	No
Endoscopic Multimedia Information System (EMIS)	EndoPrev LLC	In process	No
ENDO-AID	Olympus Medical Systems	CE class IIa	No
ENDOANGEL	Wuhan ENDOANGEL Medical Technology Company	CE marked	No
EndoScreener	Wision AI	CE class IIa	No
GI Genius	Medtronic	CE class IIa	Yes
MAGENTIQ-COLO	Magentiq Eye	CE class IIa	Yes

3 Committee discussion

The diagnostics advisory committee considered evidence on artificial intelligence (AI) technologies to help detect or characterise colorectal polyps from several sources. This included evidence submitted by the companies, a review of clinical and cost evidence by the external assessment group (EAG), and responses from stakeholders. Full details are available in the [project documents for this guidance](#).

The condition

- 3.1 Colorectal polyps may cause no symptoms (asymptomatic) but some may bleed, and some may cause abdominal pain or changes in bowel habits. Risk factors for colorectal polyps include older age, family medical history, lifestyle factors and conditions such as

inflammatory bowel disease (IBD). Most colorectal cancers develop from polyps, so finding and removing them early is vital.

- 3.2 Colonoscopy allows healthcare professionals to detect and remove polyps before they become cancerous. This can reduce the risk of bowel cancer by up to 90%. Two types of polyps are especially important: adenomatous polyps, which are precancerous, and sessile serrated lesions (SSLs), which are harder to spot but can also develop into cancer. Polyps that are less than 5mm are less likely to be cancerous and are more likely to be missed in standard care. Early and accurate detection helps ensure people get the right treatment at the right time.

Diagnostic pathways

- 3.3 In England, colonoscopy is used:
- to screen for colorectal cancer in adults aged 50 and over after a positive faecal immunochemical test (FIT)
 - for surveillance in people with a higher risk of developing colorectal cancer, such as people with IBD, a family medical history of colorectal cancer, or previous polyps or cancer
 - to help investigate symptoms consistent with colorectal cancer such as bleeding, pain or anaemia.
- 3.4 Therapeutic colonoscopy allows the endoscopist to remove polyps and take tissue samples during the procedure to confirm a diagnosis through a microscope (histology). But this has a cost, so the bowel cancer screening programme recommends that in certain cases, where the polyp is small and appears benign, the polyp can be removed but not analysed. This is known as the 'resect-and-discard' strategy. Similarly, the European Society of Gastrointestinal Endoscopy recommends that diminutive (less than 5 mm) rectosigmoid polyps that are predicted to be non-

adenomatous with high confidence can be left in place. This is known as the 'diagnose-and-leave strategy'.

Unmet need

- 3.5 The demand for colonoscopy in the NHS is high, and missed polyps can lead to a delayed diagnosis of colorectal cancer. Clinical experts explained that smaller polyps and particular types of polyps such as SSLs can be more difficult to detect. AI software using computer-aided detection (CAdE) may help by acting as a second observer, improving polyp detection, and potentially leading to better clinical outcomes. Some software also uses computer-aided diagnosis (CAdx) to support diagnosis and characterisation, potentially helping guide decisions about polyp removal and reducing unnecessary procedures. Using AI software could also help to reduce variation in polyp detection between endoscopists with different levels of skill and experience.

Clinical effectiveness

Overview of evidence base

- 3.6 The EAG searched for, identified and reviewed studies that used AI software during live colonoscopy, which reflects real-world use. It used studies in which AI software supported, not replaced, endoscopists. A wide range of different outcomes was reported across the studies. For CAdE, diagnostic accuracy data was limited and often not relevant to clinical practice. Adenoma detection rate (ADR) was the most widely reported outcome and was used as the main outcome by the EAG in its economic modelling. ADR is the percentage of screening colonoscopies in which at least one adenomatous (precancerous) polyp is detected. For CAdx, the EAG used diagnostic accuracy studies that supported endoscopists to characterise polyps and compared this to histology as the reference standard.

- 3.7 The EAG identified and reviewed 70 studies on AI-assisted colonoscopy, which focused mainly on polyp detection (CAdE). Most of the studies were randomised controlled trials comparing AI-assisted colonoscopy with standard colonoscopy. The quality and quantity of evidence varied across AI technologies, with some supported by more robust data than others.
- 3.8 The EAG did meta-analyses for each AI software individually, focusing mainly on ADR as the key outcome.

Suitability of ADR as an effectiveness measure

- 3.9 ADR was the only outcome that was widely reported in RCTs for all the technologies. The EAG explained that there is evidence to show that an improvement in ADR leads to a reduction in the number of colorectal cancer cases detected post colonoscopy. The EAG added that diagnostic accuracy data was rarely reported for using the software to aid detection. It said that this is because the software aims to improve on the gold standard which is colonoscopy without AI software.
- 3.10 Clinical experts noted that ADR only tells us about the proportion of colonoscopies in which at least one adenomatous polyp was detected and not whether the software is accurate in detecting all polyps. So, some potentially cancerous polyps could still be left behind even when ADR is improved. The clinical experts explained that the size and type of polyp detected is very important because smaller polyps (less than 5mm) are much less likely to become cancerous. Whereas advanced or larger adenomas are associated with a higher risk. They also noted that particular types of polyps called SSLs are important because they are harder to detect and so could be missed but also have the potential to become cancerous. The clinical experts expressed concern that AI software may primarily pick up smaller, less advanced polyps and said that

an endoscopist is less likely to miss more advanced or larger polyps without AI. Additionally, the clinical experts were concerned that AI may not help detect SSLs, which can also cause cancer.

- 3.11 The committee concluded that ADR is a useful outcome to provide evidence on whether the software improves detection overall. But it said that reporting ADR separately by polyp type and size is important to assess whether the software may be able to reduce colorectal cancer rates. They said that this is because only increasing the detection of smaller polyps, which are less likely to become cancerous, may have a very small effect on colorectal cancer rates.

Effectiveness for detection

- 3.12 The EAG reported pooled ADR data for each technology, which showed a statistically significant improvement for most technologies. It also acknowledged the nationwide study of AI in adenoma detection during colonoscopy (NAIAD) as an important piece of UK-based real-world evidence supporting one of the technologies, noting its large sample size and practical relevance. Where possible, the EAG reported ADR separately for advanced and non-advanced adenomas, SSLs and by polyp size. It also reported other outcomes when available, such as the number of adenomas found per colonoscopy. For advanced adenomas and SSLs, the AI software improved the ADR, but this was not statistically significant for any technology. For non-advanced adenomas, the impact of AI was greater, suggesting that the software is better at detecting smaller polyps. AI technologies may improve detection of smaller polyps (5 mm or less, and 6 mm to 9 mm) compared to polyps 10 mm or more, but the evidence is inconsistent.

- 3.13 The committee concluded that using the AI technologies improves ADR overall. But it said that it is uncertain whether the technologies improve the detection of more clinically significant polyps such as advanced or larger adenomas or SSLs. So, there is also uncertainty about whether the increase in ADR overall would translate into a reduction in colorectal cancer rates.

Quality of evidence for different technologies

- 3.14 The EAG said that all the studies reported that the addition of AI led to improved ADRs as measured by risk ratios. But it said that this was not statistically significant for 3 of the 10 technologies (Argus, Discovery and Endoscopic Multimedia Information System [EMIS]). These 3 technologies generally had fewer trials with fewer people enrolled.
- 3.15 The committee concluded that these 3 technologies could not be recommended for use with evidence generation because it was too uncertain whether they could improve ADR. They said that further research is needed.

Evidence in screening, symptomatic and surveillance populations

- 3.16 Most of the studies were done in mixed populations. The EAG explored subgroup analyses of CADe by colonoscopy indication (screening, symptomatic and surveillance). No strong differences in AI software performance were found across patient subgroups, although the EAG considered that subtle differences could not be ruled out.
- 3.17 Clinical experts discussed differences between populations and noted that the risk of them having polyps or cancer would differ but that for most populations the software should work the same. They noted that in the bowel cancer screening programme endoscopists must have special accreditation and could be expected to be more

skilled. So, the scope to benefit from the addition of AI software for polyp detection may be lower in this population.

- 3.18 The EAG noted they had explored endoscopist experience and skill in another subgroup analysis. It said that the evidence appeared mixed, with some studies suggesting greater benefit for less experienced endoscopists and others suggesting the opposite. But the EAG said that inconsistent reporting, differing definitions of experience and limited data made interpretation difficult, which limited confidence in the conclusions. Clinical experts added that it is difficult to define experience and skill because baseline ADR can be affected by the population and case mix. They also said that years of experience does not necessarily correlate with increased skill.
- 3.19 The committee concluded that there is no clear evidence to suggest that the AI technologies are less effective in some populations.

Populations excluded from studies

- 3.20 The committee noted that some high-risk groups for colorectal cancer, such as people with Lynch syndrome, familial adenomatous polyposis (FAP) or IBD were excluded from most studies. Clinical experts discussed whether the software might work differently in these groups. They noted that people with IBD or Lynch syndrome may have differences in their bowel appearance and have different polyp types. So, they thought that the AI technologies may not work as well for these people, particularly if people with these conditions were not represented well in datasets used to train the technologies. Clinical experts said that people with FAP would usually present with lots of smaller polyps but that these would look the same as a polyp in someone without this condition. So, they said that it's more likely that the AI software would perform

in a similar way for people with this condition as for people without it.

- 3.21 The committee concluded that more research is needed to understand how well the AI technologies work for people with IBD or Lynch syndrome before they can be used in the NHS for people with these conditions.

Overreliance

- 3.22 The committee discussed the potential for deskilling among endoscopists because of overreliance on AI software. The EAG noted that the large real-world NAIAD study evaluated one of the technologies in an NHS setting across 3 phases. Phase 1 was before the introduction of the software and phase 3 was withdrawal of the software. But the committee judged that it would be difficult to draw conclusions on whether any changes seen in ADR between phases 1 and 3 were caused by overreliance or for other reasons such as changes in case mix. Clinical experts also noted studies done in other countries which showed a reduction in ADR from baseline after AI is removed. The committee concluded that it is difficult to control for other factors that may affect detection rates. So, they said that it is difficult to draw any strong conclusions from studies not specifically designed to investigate the potential of AI software to deskill endoscopists.

False positives

- 3.23 The committee raised the lack of specificity data for computer-aided detection. It noted that there was not a lot of information available on how many times the technologies incorrectly flagged something as a polyp, known as a false positive. The EAG confirmed that there was limited data on this for detection of polyps. But it noted that what was reported did not suggest that this was a problem. It added that definitions varied but it was usually

defined as disagreement between the AI software and the endoscopist. The committee concluded that false positives should be identified by the endoscopist. But it said that an increase in detection and therefore removal of polyps that are benign may increase laboratory costs and could affect surveillance intervals and lead to an increase in colonoscopies. The committee concluded that more evidence is needed on the impact of introducing the technologies on the management of a polyp. It said that more evidence is particularly needed on changes in decisions on patient follow up, surveillance intervals, and additional excision and testing of polyps.

Effectiveness for characterisation

3.24 The EAG found that evidence on the CADx functionality of the technologies was limited and inconsistent, with variable diagnostic accuracy and several methodological concerns. Key concerns included:

- the autonomous use of the software (without endoscopist review), which does not reflect how the technologies would be used in the NHS
- only reporting the results for polyps diagnosed with ‘high confidence’, which could limit the clinical relevance and bias the results.

As a result, the EAG considered the evidence insufficient to support strong conclusions about the effectiveness of the technologies used in this context.

3.25 Clinical experts also expressed concern with how some of the technologies classify SSLs as non-significant polyps. They said that this is not an accurate characterisation because they have the potential to cause cancer.

- 3.26 The committee concluded that further research is needed to address some of the methodological concerns identified by the EAG. It stated that the diagnostic accuracy (sensitivity and specificity) of the AI technologies when used alongside endoscopist judgement to characterise different types of polyps would be an important outcome. It also noted that the AI technologies should be able to accurately differentiate between polyps that may and may not potentially cause cancer, including SSLs.

Procedure time

- 3.27 Patient experts noted that procedure length is a key consideration for patients and questioned whether AI technologies would influence this. The EAG explained that there was some evidence that procedure length increases with use of the technologies but that this was only by 1 to 2 minutes. The committee concluded that the AI technologies do not appear to have a substantive impact on procedure length.

Cost effectiveness

Clinical-effectiveness inputs

- 3.28 The EAG developed a decision-tree model to simulate the impact of the AI technologies on colonoscopy. A hypothetical blended cohort of people (to reflect screening, surveillance and symptomatic groups) was assigned to 1 of 5 disease states. These were no significant pathology, IBD, low-risk adenoma present, advanced adenoma present, and colorectal cancer. It was assumed that the colonoscopy would either detect all adenomas present or would miss some. Sensitivity for standard colonoscopy without AI was estimated using published adenoma miss rates. The AI effectiveness was modelled using ADR as a proxy for sensitivity by applying the risk ratio from the meta-analyses to the colonoscopy sensitivity. Where data allowed, the EAG used ADR

for advanced and non-advanced adenomas to reflect the improvements in detection for high and low-risk adenomas. But data limitations meant that a single ADR estimate was applied for some technologies, which would likely overstate the effect of the software on the detection rate for higher-risk adenomas.

- 3.29 The committee considered the use of ADR as a surrogate outcome for sensitivity and noted it may not fully capture clinical benefit. It said that this was particularly in relation to long-term outcomes like post-colonoscopy colorectal cancer. The committee highlighted its concerns that the clinical evidence suggested that ADR improvements may be driven by increased detection of small, low-risk polyps. It recalled that these may not translate into meaningful health gains if the improvement in ADR does not result in fewer cases of colorectal cancer. This view was confirmed by a clinical expert, who stated that less than 0.1% of small (less than 5 mm) polyps develop into colorectal cancer. The committee noted that this would not be captured in the model for technologies that did not report ADR for advanced and non-advanced adenomas separately. The committee also recalled that even when ADR was reported for advanced adenomas separately, the results were not statistically significant (see [section 3.12](#)). So, the committee said that there was still uncertainty in the results for these technologies.

Model structure

- 3.30 The decision-tree model developed by the EAG only modelled the initial colonoscopy procedure. Lifetime-cost and quality-adjusted life year (QALY) outcomes were applied at the end of each branch to reflect the lifetime consequences of missing polyps. These were adjusted to reflect a delayed diagnosis for people with missed or misdiagnosed adenomas or colorectal cancer. The values for lifetime-cost and QALY outcomes were ultimately derived from the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel)

individual patient simulation model developed by the Sheffield Centre for Health and Related Research (SCHARR).

- 3.31 This approach of using lifetime 'pay offs' meant that additional or future colonoscopies were not modelled to include AI. The committee questioned the impact of the assumption in the model that all colonoscopies after the initial index colonoscopy would be done without AI. The committee judged that the assumption undermined the ability to assess the full impact of AI technologies on long-term outcomes and did not reflect reality. The EAG explained that this was a limitation of the model that could not be easily resolved. It said that this was because it would need access to the MiMiC-Bowel model, which it did not have. The EAG added that addressing this concern would likely improve estimated QALYs in the intervention arm because AI would continue to pick up polyps that would otherwise be missed. But it said that similarly, costs may also increase to account for ongoing costs of the technologies. The EAG estimated that the impact would be small.
- 3.32 Clinical experts also noted some further limitations of using the MiMiC-Bowel model in that it does not take SSLs into account. So, they said that it may not reflect the outcomes associated with detection and non-detection of different types of polyps.
- 3.33 The committee also noted that the inputs used for long-term QALYs were key drivers of the model results. The committee agreed that within these inputs a reduction in the rates of post-colonoscopy colorectal cancer and earlier detection would be driving the model results.

Uncaptured costs

- 3.34 The committee also raised concerns that the economic model did not fully account for the downstream impact of diagnostic decisions. For instance, increased detection of small polyps, driven

by AI technologies, could lead to more people being placed onto surveillance pathways. This could result in additional colonoscopies and follow up and increased histology costs with no clear corresponding clinical benefit. The need for surveillance triggered by guideline thresholds (for example, 5 or more polyps) was also not modelled, because ADR does not capture the incidence of multiple polyps.

- 3.35 The committee concluded that there are likely some costs related to the use of AI technologies that have not been captured in the model.

Uncertain costs for some technologies

- 3.36 Eight of the 10 technologies were included in the analysis, with 2 technologies being excluded because of a lack of technology cost data. So, no health-economic results were available for CADDIE or ENDOANGEL. The committee concluded that the cost effectiveness of these 2 technologies was too uncertain to be able to make a recommendation for use in the NHS outside of research.

Cost-effectiveness conclusions

- 3.37 All the modelled technologies were estimated to be cost effective compared with standard colonoscopy without the use of AI. All but 1 were estimated to be less costly and more effective compared with standard colonoscopy. But, in all cases the differences in costs and QALYs were small, and the EAG advised caution in interpreting the results. They noted that the results were very uncertain, with the probability of each intervention being cost effective estimated as approximately 50% at NICE's usual willingness-to-pay range (£20,000 to £30,000).
- 3.38 The committee noted that the cost-effectiveness estimates were highly sensitive to inputs and assumptions in the model, with some members noting that even minor changes could reverse the

conclusions. The committee recalled the uncertainty in the clinical evidence about whether AI can improve the detection of clinically significant polyps. It recalled that this may not be well represented in the model because of a lack of clinical data to inform the differences in improvement between different polyp types and sizes. The committee questioned whether the technologies would remain cost effective if this was truly reflected in the model, along with other costs that may not be fully captured (see [section 3.13](#) and [section 3.29](#)). The committee concluded that the economic case for adopting the technologies remains uncertain.

3.39 The committee explored how this uncertainty could be reduced. It agreed that the key clinical benefit of detecting more polyps, and the key driver in the model, is a reduction in rates of post-colonoscopy colorectal cancer. They noted that to collect data on this outcome would need studies to have a long follow-up period. But clinical experts advised that a study on one of the technologies was already collecting this data. They said that this could be informative about the link between increased polyp detection with AI and any reduction in cancer rates.

3.40 The committee concluded that the structure and inputs of the economic model developed by the EAG were broadly appropriate. It said that the results suggested that the technologies could be cost effective but that they were still very uncertain. The committee agreed that data on post-colonoscopy colorectal cancer rates would be beneficial and could reduce the uncertainty about the clinical and economic impact of the AI technologies before routine adoption.

Equality considerations

3.41 Many studies excluded key groups such as people with IBD, Lynch syndrome, FAP or prior colorectal cancer. The EAG noted that AI

technologies may not be adequately validated for these populations because of poor reporting of training data and limited subgroup representation, making reliable conclusions difficult. The committee recommended further research before the technologies can be used in some of these groups (see [section 3.21](#)).

- 3.42 A clinical expert highlighted that people from ethnic minority backgrounds often present for colonoscopy later and experience delays in diagnosis. A clinical expert also noted that there was a lower uptake of screening among these groups and that this could potentially lead to underrepresentation in training datasets and biased performance. Representatives from 2 of the companies confirmed that their technologies had been trained on datasets derived from a broad range of providers and countries to ensure equal representation.

4 Committee members and NICE project team

This topic was considered by [specialist committee members appointed for this topic](#) and [NICE's diagnostics advisory committee](#), which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the technology to be evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The [minutes of each committee meeting](#), which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Chair

Professor Neil Hawkins

Vice chair, diagnostics advisory committee

NICE project team

Each evaluation is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the evaluation), a technical adviser, a project manager and an associate director.

Iain Willits

Technical lead

Judith Shore

Technical adviser

Toni Gasse

Project manager

Lizzy Latimer

Associate director

ISBN: [to be added at publication]