NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Evidence generation plan

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

November 2025

1 Purpose of this document

NICE's assessment of AI technologies to help detect or characterise colorectal polyps recommends that more evidence is generated while 5 of the technologies are being used in the NHS. The other technologies that were assessed can only be used in research and are not covered in this plan.

This plan outlines the evidence gaps and what data needs to be collected for a NICE review of the technologies again in the future. It is not a study protocol but suggests an approach to generating the information needed to address the evidence gaps. Evidence generated through other study approaches will also be considered. For assessing comparative treatment effects, well-conducted randomised controlled trials are the preferred source of evidence.

The companies are responsible for ensuring that data collection and analysis takes place.

Guidance on commissioning and procurement of the technologies will be provided by NHS England, who are developing a digital health technology policy framework to further outline commissioning pathways.

NICE will withdraw the guidance if the companies do not meet the conditions in section 4 on monitoring.

After 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 all the evidence and assess whether the technologies can be routinely adopted in the NHS.

2 Evidence gaps

This section describes the evidence gaps, why they need to be addressed and their relative importance for future committee decision making.

The committee will not be able to make a positive recommendation without the essential evidence gaps (see section 2.1) being addressed. The company can strengthen the evidence base by also addressing as many other evidence gaps (see section 2.2) as possible. This will help the committee to make a recommendation by ensuring it has a better understanding of the patient or healthcare system benefits of the technology.

2.1 Essential evidence for future committee decision making

Improvement in adenoma detection rate (ADR) by polyp type and size

The committee said that evidence showed that the 5 AI software technologies significantly increase adenoma detection rate (ADR). But it concluded that there was not enough evidence to determine whether the software increases detection of advanced adenomas or sessile serrated lesions (SSLs). This is important because these polyps are more likely to develop into cancer. The committee needs more evidence, categorised by polyp type and size, on whether using the software leads to an improvement in ADR for advanced adenomas and SSLs.

Change in post-colonoscopy colorectal cancer rates

While a significant increase in ADR was seen when AI software was used, it was not clear if this translated into a change in the number of cases of colorectal cancer detected post colonoscopy. The committee said that there was not enough evidence on the type and size of adenomas that the software helped to detect. This means that the improved ADR may be caused by increased numbers of small adenomas being detected. Small adenomas are less likely to develop into colorectal cancer.

The committee would like more evidence on whether using these AI software technologies leads to changes in post-colonoscopy colorectal cancer rates.

Impact on clinical management

The committee noted that there was a lack of data about how the increased identification and removal of polyps may impact on costs and surveillance intervals. It was concerned that it could lead to an increase in the overall number of colonoscopies with no clear corresponding clinical benefit. The committee concluded that more evidence is needed on the impact of introducing the AI software on clinical management following polyp identification. More evidence is particularly needed on the effect of the software on decisions made about follow up, surveillance intervals and additional excision and testing of polyps.

3 Approach to evidence generation

3.1 Evidence gaps and ongoing studies

Table 1 summarises the evidence gaps and ongoing studies that might address them. Information about evidence status is derived from the external assessment group's report; evidence not meeting the scope and inclusion criteria is not included. The table shows the evidence available to the committee when the guidance was published.

Table 1 Evidence gaps and ongoing studies

Evidence gap	CAD EYE	ENDO-AID	EndoScreener	GI Genius	MAGENTIQ-COLO
Improvement in ADR by polyp type and size	Good evidence Ongoing study	Good evidence Ongoing study	No evidence Ongoing study	Good evidence Ongoing study	No evidence Ongoing study
Change in post-coloscopy colorectal cancer rates	No evidence	No evidence	No evidence	No evidence	No evidence
Impact on clinical management	No evidence	No evidence	No evidence	No evidence	No evidence

Abbreviations: ADR, adenoma detection rate.

3.2 Data sources

The Future of real time endoscopy AI (FORE AI) trial is collecting data which may address some of the evidence gaps (see section 3.3).

There are ongoing studies looking at diagnostic accuracy for the following AI software technologies:

- CAD EYE - for people with Lynch syndrome (2 studies)
- GI Genius (2 studies)
- MAGENTIQ-COLO (1 study).

There are ongoing studies looking at colorectal cancer rates after AI-supported polyp detection with ENDO-AID (1study).

There are several real-world data collections with different strengths and weaknesses that could potentially support evidence generation. NICE's real-world evidence framework provides detailed guidance on assessing the suitability of a real-world data source to answer a specific research question. There are existing real-world data registries for colonoscopy outcomes, including the National Endoscopy Database.

The quality and coverage of real-world data collections is of key importance when used to generate evidence. Active monitoring and follow up through a central coordinating point is an effective and viable approach of ensuring good-quality data with broad coverage.

3.3 Evidence collection plan

FORE AI study

The FORE AI study may provide evidence on ADR by polyp type and size. It may also help advise on which polyp features detected by AI are most likely to be associated with the development of colorectal cancer. The FORE AI study is a prospective observational study which will collect video and histopathology data from a subset of people who have consented to the CONSCOP2 study. In CONSCOP2, people are randomised to high-definition white-light colonoscopy with or without indigo carmine dye spray. In the FORE AI study, a recording of the colonoscopy video stream will be made, and the CADDIE software will subsequently run on this footage. The accuracy in detection and diagnosis of polyps from the video will be compared with colonoscopists using CADDIE. Everyone in the CONSCOP2 study will be followed up for 3 years. If someone enters a surveillance pathway then their clinical data will be collected. If they return to a routine screening pathway then cancer registries will be analysed to determine if colorectal cancer develops. This means that data on colorectal cancer rates can be compared to the AI polyp detection reports. This data can be used to determine whether specific clinical outcomes, such as ADR, correlate with colorectal cancer development.

The FORE AI study uses the CADDIE AI system for polyp detection, but the committee concluded that evidence about the correlation between increase in ADR using AI and colorectal cancer would likely apply to all 5 technologies.

Diagnostic accuracy study

Because the FORE AI trial is investigating a specific AI-supported polyp detection technology (CADDIE), companies may want to consider doing their own diagnostic accuracy study. A diagnostic accuracy study is used to assess the agreement

between 2 or more methods. The study would assess the agreement between the diagnosis decision reached for each case of suspected cancer by:

- AI-supported polyp detection (intervention)
- endoscopist polyp detection alone (comparator)
- a reference standard.

Video colonoscopy footage would be prospectively assessed by AI. A comparison between the AI-supported polyp detection, endoscopist polyp detection and a reference standard would allow an assessment of the diagnostic accuracy of the AI software technology compared with standard colonoscopy.

Observational cohort study with a historical control

To understand the impact of AI-supported polyp detection technologies on post-colonoscopy colorectal cancer rates, an observational study with a historical control should be done. An observational cohort study will allow assessment of whether the AI software technology impacts on clinical management following polyp detection, and on rates of colorectal cancer following colonoscopy. This information could also be gathered with any other scientifically appropriate approach.

For both cohorts within the study, data should be collected on the:

- total number, type and size of adenomas or other lesions detected during a colonoscopy
- number of resections completed within a colonoscopy, and the histopathological results for resections
- proportion of people put on a surveillance pathway, and frequency of follow-up colonoscopies done
- proportion of people diagnosed with post-colonoscopy colorectal cancer.

It is anticipated that national endoscopy databases will provide the relevant clinical information and should be contacted to assist with data collection.

3.4 Data to be collected

The following information has been identified for collection:

Diagnostic accuracy study

- ADR, classified by adenoma type and size, AI software technology and highly skilled endoscopist
- proportion of people with resectable polyps
- histopathological results for resected polyps, by AI software technology and by highly skilled endoscopist
- whether or not the AI was able to process colonoscopy video footage correctly.

Observational cohort study

- patient information, for example age, sex and ethnicity
- ADR, classified by adenoma type and size
- number of resections, and number of unresectable lesions
- proportion of people referred onto a surveillance pathway
- proportion of people who develop post-colonoscopy colorectal cancer.

Data collection should follow a predefined protocol, and quality assurance processes should be put in place to ensure the integrity and consistency of data collection. See NICE's real-world evidence framework, which provides guidance on the planning, conduct and reporting of real-world evidence studies.

3.5 Evidence generation period

The evidence generation period will be 4 years to allow for setting up the study, implementing the AI software technology, data collection, time for follow up to detect colorectal cancer rates, data analysis and reporting.

3.6 Following best practice in study methodology

Following best practice in conducting studies is paramount to ensuring the reliability and validity of the research findings. Adherence to rigorous guidelines and established standards is crucial for generating credible evidence that can ultimately improve patient care. The NICE real-world evidence framework details some key considerations.

Within the context of a conditional recommendation a key factor to consider as part of the informed consent process is to ensure that patients (and their carers, as appropriate) understand that data will be collected to address the evidence gaps identified in section 2. Where applicable this should take account of NICE guidance about shared decision making.

4 Monitoring

NICE will contact the companies:

- within 6 months of publication of this plan to confirm agreements are in place to generate the evidence
- annually to confirm that the data is being collected and analysed as planned.

The companies should tell NICE as soon as possible about anything that may affect ongoing evidence generation, including any:

- substantial risk that the evidence will not be collected as planned
- new safety concerns
- significant changes to the AI software technology that could affect the evidence generation process.

If data collection is expected to end later than planned, the companies should contact NICE to arrange an extension. NICE reserves the right to withdraw the guidance if data collection is delayed, or if it is unlikely to resolve the evidence gaps.

5 Minimum evidence standards

All 5 AI software technologies have some clinical evidence suggesting that they increase ADR as measured by risk ratios. However the committee concluded that ADR does not capture the full clinical benefit, particularly in relation to long-term outcomes like post-colonoscopy colorectal cancer rates. Evidence relating to this was not seen for any technologies. None of the technologies reported any safety concerns.

Further to this, the committee has indicated that it may be able to recommend AI software technologies in this topic area, in the future, that have evidence for the impact of the technology on the:

- improvement of ADR, by adenoma type and size
- change in post-colonoscopy colorectal cancer rates
- clinical management following polyp identification.

6 Implementation considerations

The following considerations around implementing the evidence generation process have been identified through working with system partners:

Evidence generation period

- To collect useful data on post-colonoscopy colorectal cancer rates, the length of data collection will need to be sufficient. Clinical advice suggests that 3 years of follow-up data is appropriate, so companies will need to be efficient in starting data collection.
- Companies will need to be proactive about contacting national data registries, to ensure that data sharing agreements are in place.

Equalities

- Data should be collected on ethnicity and analysed to assess any differences in diagnostic accuracy when the AI software is used for people from different ethnic backgrounds.

System considerations

- When introducing AI-supported polyp detection, care should be taken to ensure that endoscopist skill and experience is not lost or fails to develop.
- Companies should use dedicated and incentivised research staff, to reduce the burden on non-research staff.
- Real-world data should be used when possible.

ISBN: [to be added at publication]