NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Quantitative faecal immunochemical tests to guide colorectal cancer pathway referral in primary care. A systematic review, meta-analysis and cost-effectiveness analysis

Sue Harnan,[1*] Aline Navega Biz,[1] Jean Hamilton,[1] Sophie Whyte,[1] Emma Simpson,[1] Shijie Ren,[1] Katy Cooper,[1] Mark Clowes,[1] Muti Abulafi,[2] Alex Ball,[3] Sally C Benton,[4] Richard Booth,[2] Rachel Carten,[5] Stephanie Edgar,[6] Willie Hamilton,[7] Matt Kurien,[1,3] Louise Merriman,[6] Kevin Monahan,[8,9] Laura Heathcote[1] and Matt Stevenson[1]

[1]Sheffield Centre for Health and Related Research (SCHARR), School of Medicine and Population Health, University of Sheffield, Sheffield, UK
[2]Croydon Health Services NHS Trust, London, UK
[3]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[4]Royal Surrey NHS Foundation Trust, Guildford, UK
[5]Buckinghamshire Healthcare NHS Trust, Aylesbury, UK
[6]South Yorkshire CCG, Sheffield, UK
[7]University of Exeter, Exeter, UK
[8]St Mark's The National Bowel Hospital, London, UK
[9]Imperial College London, London, UK

[*]**Corresponding author** s.harnan@sheffield.ac.uk

Published September 2025
DOI: 10.3310/AHPE4211

Scientific summary

Quantitative faecal immunochemical tests to guide colorectal cancer pathway referral in primary care. A systematic review, meta-analysis and cost-effectiveness analysis

Health Technology Assessment 2025; Vol. 29: No. 46
DOI: 10.3310/AHPE4211

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Early diagnosis and treatment of colorectal cancer (CRC) in people presenting to primary care with symptoms can improve survival and cure rates. The introduction of the National Institute for Health and Care Excellence (NICE) National Guideline 12 (NG12) in 2015 to expand symptoms-based criteria for referral to secondary care led to an increase in the number of urgent 2-week wait (2WW) suspected cancer pathway referrals, but no corresponding increase in the proportion of patients investigated through 2WW who had cancer. This has led to pressure on colonoscopy capacity and to long waiting times in some areas of the UK, especially in the non-urgent [18-week wait (18WW)] referral pathway.

Quantitative faecal immunochemical tests (FITs) are designed to detect occult (small amounts) of blood in stool samples (faecal haemoglobin) using antibodies specific to human haemoglobin. They are currently used in patients with low-risk symptoms in primary care [as described in NICE Diagnostics Guidance 30 (DG30)], but not in patients with high-/medium-risk symptoms as defined in NG12, who are instead referred directly to secondary care. There is evidence that FITs are a better predictor of CRC risk in patients than symptoms alone and could result in fewer referrals of people without CRC to the 2WW pathway. Therefore, triage with FITs for all patients could avoid unnecessary referrals, patient anxiety, time off work and loss of economic productivity, and rare adverse events associated with colonoscopy such as bleeding, perforation and death. Those who are likely to have CRC could be prioritised more effectively, potentially reducing time to diagnosis. The released colonoscopy capacity could allow non-urgent referrals to be seen more quickly. The extent to which colonoscopy capacity is released and time to diagnosis is affected will depend in part on the threshold used to define a positive FIT result, with a higher threshold resulting in fewer referrals but a greater chance of missing disease.

Objectives

The decision problem in the NICE scope was 'What is the most clinically and cost-effective way to use quantitative FITs to reduce the number of people without significant bowel pathology who are referred to the suspected cancer pathway for CRC, taking into consideration potential colonoscopy capacity constraints for urgent and non-urgent referrals?' Eight FITs were within the scope of the assessment, namely HM-JACKarc, FOB Gold, OC-Sensor, NS-Prime, IDK TurbiFIT, IDK Hemoglobin ELISA (IDK Hb), IDK Hb/Hp complex ELISA (IDK Hb/Hp) and QuikRead go.

The decision problem was addressed through a systematic review of evidence relating to the accuracy of the tests, a statistical synthesis to pool data across studies, and an economic model that aimed to estimate the cost-effectiveness of FIT strategies based on diagnostic accuracy, the number of colonoscopies undertaken and the impact on time to diagnosis.

Methods

Clinical evidence review methods

Searches were conducted across four databases and six registries in December 2022. The titles and abstracts of records retrieved were screened by one reviewer, with the first 20% checked by a second reviewer before the remainder were screened. Records for which the full text was obtained were checked for inclusion by two reviewers. Data extraction and quality assessment were conducted by one reviewer and checked by a second. Study quality was assessed using quality assessment of diagnostic test accuracy studies version 2.

As no randomised controlled trials were identified, studies were included if they reported the diagnostic test accuracy of FIT in patients presenting to primary care, or referred from primary care, with signs or symptoms of CRC. Studies

reporting data for 'dual FIT', whereby patients are asked to provide two samples from different bowel movements, were also included. All thresholds for defining a FIT were eligible for inclusion. The reference standard was not restricted but expected to comprise colonoscopy, computed tomography colonography (CTC), other imaging tests or records follow-up. Studies were also subgrouped according to several patient characteristics (anaemia, age, sex, ethnicity and medication or other blood disorders that might affect FIT). Test failure rates, uptake of FITs, time to colonoscopy, time to diagnosis and patient-reported outcomes such as health-related quality of life, preference and anxiety were also sought.

The statistical synthesis pooled estimates of sensitivity and specificity at all reported diagnostic thresholds and provided summary estimates at all possible thresholds within the observed range. Studies were synthesised for each test separately. Sensitivity analyses investigated the effects of population type and reference standard, where data allowed.

Cost-effectiveness methods

A mathematical model was developed to simulate the experiences of patients presenting to primary care with symptoms of CRC. Three interventions were evaluated: intervention 1, the use of a single FIT threshold to determine whether a person would be referred to the 2WW pathway or follow the safety-netting pathway; intervention 2, the use of two thresholds to determine if a patient would be referred to the 2WW pathway, be referred to a intermediate pathway or follow the safety-netting pathway; and intervention 3, which represented current practice, whereby all patients at NG12 high/medium risk were referred to the 2WW pathway and the remainder received a FIT and were subsequently assigned to the 2WW pathway or the safety-netting pathway. For the purpose of the economic model, patients receiving safety netting took one of four possible pathways: referral to the 2WW pathway due to ongoing clinical concerns, referral to non-urgent referral pathway (18WW), watchful waiting, or being offered a second FIT (repeat FIT).

The model was populated by published literature (synthesised where appropriate), Grey Literature, estimates provided by clinical experts and costs of FIT reported by the relevant companies. An initial decision tree model was used to categorise patients in terms of their true underlying disease status, whether NG12 high/medium risk or DG30 low risk for intervention 3, and whether a FIT result was true positive, false positive, true negative or false negative. Following this, state-transition models were used to model patient survival, costs incurred and quality-adjusted life-years (QALYs) gained. The model assumed that the proportional reduction in the total number of patients referred to the 2WW and 18WW pathways would translate directly into an equivalent proportional reduction in time before diagnosis for patients in these pathways.

Outputs from this model included the life-years gained, QALYs and costs associated with each FIT strategy, the number of 2WW and 18WW referrals, the numbers receiving repeat FIT and allocated to the watch and wait pathway, the number of colonoscopies undertaken, and the mean time to a diagnosis of CRC, advanced adenomas (AAs) and inflammatory bowel disease (IBD). To explore cost-effectiveness, incremental net monetary benefit (iNMB) was used as it allowed an easy comparison between FIT strategies that varied both by the specific FIT brand and by the threshold(s) used to denote a positive, intermediate (in intervention 2) or negative FIT result.

Thirteen scenario analyses were performed, explored the impact of (1) decreasing the underlying times before diagnosis associated with current care; (2) increasing the underlying times before diagnosis associated with current care; assuming the loss of a full day's health (3) for people receiving a colonoscopy or (4) for every month of delay associated with a definitive diagnosis for those in the 2WW or 18WW pathways and also for those with underlying bowel disease not in these pathways to account for patient anxiety while undiagnosed but with symptoms; (5) assuming the use of dual FITs; (6) setting the prevalence of AAs and IBD to zero; (7) using a lower return rate for FITs; (8) assuming an alternative diagnostic accuracy of current FITs in low-risk patients in intervention 3; (9) an increase in general practitioner resource required for patients in the watch and wait and repeat FIT pathways without underlying disease from 1.9 appointments to 2.9 appointments; (10) assuming a lower price associated with FITs used in intervention 3; and (11) assuming FITs to have perfect accuracy (sensitivity and specificity = 1.0) and return rate of 100%, to test an extreme scenario in which no patients are missed by test or wrongly sent to 2WW; and (12) reduction in and (13) increase in the prevalence of CRCs, AAs and IBD by 50%.

Results

Clinical evidence results

Forty-nine studies were included in the review, across all tests and all subgroups and outcomes.

There were risk of bias and/or applicability concerns with all the studies. Studies recruiting in secondary care generally scored as being at high risk of bias for patient selection, as some primary care patients were not recruited, and studies recruiting in primary care generally scored as being at high risk of bias for the reference standard, as not all patients received colonoscopy or CTC. Various other sources of bias were also noted.

As there was only a small number of head-to-head comparative studies, comparative test accuracy was not formally quantified. Considering a threshold of 10 µg/g, the results for sensitivity and specificity, respectively, were as follows: HM-JACKarc (*n* = 16 studies), 89.5% [95% credible interval (CrI) 84.6% to 93.4%] and 82.8% (95% CrI 75.2% to 89.6%); OC-Sensor (*n* = 11 studies), 89.8% (95% CrI 85.9% to 93.3%) and 77.6% (95% CrI 64.3% to 88.6%); and FOB Gold (*n* = 3 studies), 87% (95% CrI 67.3% to 98.3%) and 88.4% (95% CrI 81.7% to 94.2%). No synthesis was conducted for QuikRead go, NS-Prime or IDK tests, as there was only one study for each. The estimates of sensitivity and specificity at 10 µg/g, respectively, were as follows: QuikRead go, 92.90% [95% confidence interval (CI) 68.5% to 98.7%] and 70.10% (95% CI 66.1% to 73.8%); and NS-Prime, 71.40% (95% CI 35.9% to 91.8%) and 83.60% (95% CI 78.2% to 87.9%). The study of IDK Hb and IDK Hb/Hp only reported data at 2 µg/g, and the sensitivity and specificity were calculated as follows: IDK Hb 87% (95% CI 84.4% to 89.6%) and 88.1% (95% CI 85.6% to 90.6%); and IDK Hb/Hp, 82.6% (95% CI 79.6% to 85.6%) and 80.8% (95% CI 77.7% to 83.9%). No diagnostic test accuracy data that met the inclusion criteria for the review were found for the combined use of IDK Hb + Hb/Hp or for IDK TurbiFIT tests.

The sensitivity analyses showed that the results were similar when studies were subgrouped according to population type (all patients presenting to primary care, high-risk patients, low-risk patients) and when subgrouped according to the reference standard used, as the credible intervals overlapped.

Four studies reported data using a dual FIT strategy. Where both were reported, the sensitivity was higher but specificity lower using dual FIT (either FIT positive) than when using only the first FIT result.

The three included comparative diagnostic test accuracy studies concluded that there were some differences between tests, but none concluded whether (and what) different FIT cut-off values would be required for each test.

Across patient characteristic subgroups (anaemia, *n* = 11 studies; age, *n* = 3 studies; sex, *n* = 3 studies; and people taking medications that may affect FIT results, *n* = 3 studies), evidence was limited and sometimes inconsistent. It was not possible to conclude what or whether different FIT thresholds may be required. No studies were identified according to ethnicity or for people with other blood disorders that may affect FIT results.

Eight studies reported data on the accuracy of FITs for AAs and IBD. Uncertainty was high, with a large amount of heterogeneity between studies.

Eleven studies reported test failure rates largely between 2% and 5%. The non-return rate in the study most closely matching the decision problem was 9.4%. For dual FIT, non-return rates appeared generally higher.

Two studies reported patient perspectives. The authors concluded that most patients found FIT acceptable, but strategies are needed to engage patients with more negative views of FIT, and shared decision-making should be considered for patients dissatisfied with relying on a negative FIT result. Generalisability may have been affected by the fact that all patients included had been referred to secondary care.

One study reported on the impact of sociodemographic factors on FIT return rates and found higher return rates for female patients than for male patients, for patients aged ≥ 65 years than for those aged < 65 years, for White patients than for patients in Asian, Black and mixed/other ethnicity groups, and for the least socioeconomically deprived quintile than for all other quintiles. Suggested strategies for addressing demographic differences in FIT return rate included

following up after FIT non-return, using multiple languages, using shared decision-making and providing patient counselling to address concerns.

Cost-effectiveness results

For the vast majority of FIT strategies, the iNMB was positive compared with current care regardless of the cost-effectiveness threshold used, or whether one or two thresholds were used. These conclusions were robust to the sensitivity analyses undertaken. The iNMBs were typically in the range of £200–350 per patient, driven by the reduction in the costs of colonoscopy, although there was a slight decrease in patient health predominantly attributable to patients who had a false-negative FIT result and who would have received a colonoscopy under current practice. A robust estimation of with which FIT brand and at which threshold(s) the iNMB was highest was not achievable given the uncertainty in model parameters and in the inherently simplified modelling structure.

Discussion

The systematic review identified diagnostic test accuracy data for seven of the nine tests. Only one relatively small (*n* analysed < 700, CRC events < 25) study was identified for each of QuikRead go, NS-Prime, IDK Hb and IDK Hb/Hp. The statistical synthesis produced summary estimates of sensitivity and specificity across all possible thresholds where data allowed. There were insufficient data to statistically synthesise the comparative diagnostic test accuracy between tests. Dual FIT studies were identified for only three tests (HM-JACKarc, OC-Sensor and QuikRead go). There were insufficient and inconsistent data relating to patient characteristics (anaemia, age, sex, ethnicity, medication that might affect FITs, other blood disorders that might affect FITs), and no conclusions could be drawn on whether different thresholds should be used. FIT was found to be generally acceptable, but return rates may differ according to sociodemographic factors, and interventions may be needed to improve uptake. There were limitations to both the evidence base and the systematic review that should be taken into consideration when interpreting the evidence.

For all FIT brands there are strategies with which the iNMB is positive compared with current care, although all are associated with a slight decrease in patient health. The exact brand of FIT and threshold(s) that generate the greatest iNMB (at a selected threshold) could not be robustly determined due to the similarity of iNMB values, parameter uncertainty and the possibility of omissions from the model structure.

Suggested research priorities

Research priorities include investigating the comparative diagnostic test accuracy between tests, and whether different thresholds are required for patients with characteristics that may affect FIT accuracy. Consideration should be given when designing studies to the patient population recruited and the reference standard used as analyses were not conclusive regarding the impact of these factors.

Study registration

This study is registered as PROSPERO CRD42022383580.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR135637) and is published in full in Health Technology Assessment; Vol. 29, No. 46. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

NICE TAR and DAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR135637. The protocol was agreed in December 2022. The draft manuscript began editorial review in September 2023 and was accepted for publication in October 2024. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).