NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 29 • Issue 46 • September 2025

ISSN 2046-4924

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

Sue Harnan, Aline Navega Biz, Jean Hamilton, Sophie Whyte, Emma Simpson, Shijie Ren, Katy Cooper, Mark Clowes, Muti Abulafi, Alex Ball, Sally C Benton, Richard Booth, Rachel Carten, Stephanie Edgar, Willie Hamilton, Matt Kurien, Louise Merriman, Kevin Monahan, Laura Heathcote and Matt Stevenson

DOI 10.3310/AHPE4211

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

This report should be referenced as follows:

Harnan S, Navega Biz A, Hamilton J, Whyte S, Simpson E, Ren S, *et al.* Quantitative faecal immunochemical tests to guide colorectal cancer pathway referral in primary care. A systematic review, meta-analysis and cost-effectiveness analysis. *Health Technol Assess* 2025;**29**(46). https://doi.org/10.3310/AHPE4211

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).

Abstract

Background: Faecal immunochemical tests may be better than symptoms alone at identifying which patients who present to primary care with symptoms are at high risk of colorectal cancer and should have a colonoscopy. This could reduce waiting lists and patient anxiety/discomfort and enable earlier treatment of colorectal cancer. The threshold used will affect how well faecal immunochemical tests work, with a higher threshold resulting in fewer referrals but a greater chance of missing disease.

Objective: What is the most clinically effective and cost-effective way to use faecal immunochemical tests to reduce the number of people without significant bowel pathology who are referred to the suspected cancer pathway for colorectal cancer, taking into consideration potential colonoscopy capacity constraints for urgent and non-urgent referrals? Tests were HM-JACKarc, OC-Sensor, FOB Gold, NS-Prime, QuikRead go, IDK TurbiFIT, IDK Hb, IDK Hb/Hp complex and IDKHb+Hb/Hp ELISAs.

Design: Systematic review, meta-analysis and cost-effectiveness analyses were conducted.

Review methods: Searches across four databases and six registries were conducted (December 2022). Diagnostic accuracy studies conducted in patients presenting to or referred from primary care with symptoms suggestive of colorectal cancer using any reference standard were included. Risk of bias was assessed with quality assessment of diagnostic test accuracy studies version 2. For each test, sensitivity and specificity were pooled at all reported thresholds and summary estimates were provided at all possible thresholds within the observed range. Comparative accuracy between tests was considered. Other outcomes, for example test uptake, failure and patient acceptability, were also extracted.

Cost-effectiveness analysis methods: A mathematical model was developed to compare three different diagnostic strategies that used quantitative faecal immunochemical tests in primary care patients with symptoms of colorectal cancer to determine subsequent management pathways. The model assessed the health outcomes and costs associated with each strategy over a lifetime horizon from the perspective of the United Kingdom National Health Service and Personal Social Services, using evidence from published literature and other sources.

Results: Syntheses of sensitivity and specificity were conducted for HM-JACKarc (n = 16 studies), OC-Sensor (n = 11 studies) and FOB Gold (n = 3 studies). No synthesis was conducted for QuikRead go, NS-Prime IDK Hb or IDK Hb/Hp as there was only one study for each. No eligible studies were found for IDK Hb+Hb/Hp or for IDK TurbiFIT. Other outcomes (e.g. patient acceptability) were also synthesised. Model results suggest that faecal immunochemical tests generate a positive incremental net monetary benefit compared with current care, typically in the range of £200–350 per patient, regardless of the threshold used, for the majority of faecal immunochemical tests strategies assessed. These conclusions were robust to the sensitivity analyses undertaken.

Conclusions: For all faecal immunochemical test brands, there are strategies at which the incremental net monetary benefit is positive compared with current care. The exact brand and threshold(s) that generate the greatest incremental net monetary benefit could not be robustly determined due to the similarity of incremental net monetary benefit values, parameter uncertainty and the possibility of omissions from the model structure.

Future work: More data are needed on comparative diagnostic test accuracy and whether different thresholds should be used in some patients (e.g. anaemic, male/female, younger/older).

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.

Contents

List of tables vii

List of figures xi

List of supplementary material xiii

List of abbreviations xiv

Plain language summary xv

Scientific summary xvi

Chapter 1 Background and definition of the decision problem **1**
Condition and aetiology 1
Aetiology, pathology and prognosis 1
Epidemiology and incidence 1
Burden of disease 1
Current service provision 2
National Guideline 12 high-/medium-risk and Diagnostic Guideline 30 low-risk patients 2
Speciality guidance during the height of the COVID-19 pandemic 4
Association of Coloproctology of Great Britain and Ireland/British Society of Gastroenterology guideline and National Health Service England letter 4
Description of the decision problem 5
Purpose of the decision to be made 5
Place of the intervention in the treatment pathway 6
Definition of the intervention 6
Populations and relevant subgroups 10
Relevant comparators 10
Healthcare setting 11
Outcomes 11
Other considerations 12
Areas that are outside the scope of the appraisal and therefore do not require any detailed assessment (e.g. key factors for which evidence is already accepted) 14

Chapter 2 Clinical evidence **15**
Methods 15
Summary of the approach to the review 15
Population 15
Interventions 16
Comparators 16
Outcomes 17
Study design 17
Search strategy 18
Study selection 18
Data extraction strategy 19
Quality assessment strategy 19
Synthesis strategy 19

The analysis plan and rationale 20
Rationale for the analysis plan 20
The analysis plan 21
Main analysis: diagnostic test accuracy for colorectal cancer for each test individually 22
Additional analysis 1: diagnostic test accuracy for colorectal cancer for all tests together 22
Additional analysis 2: impact of reference standard on diagnostic test accuracy estimates 23
Results 23
Main analysis: HM-JACKarc 24
Main analysis: OC-sensor 36
Main analysis: FOB Gold 41
Main analysis: QuikRead go 44
Main analysis: NS-Prime 46
Main analysis: IDK tests 48
Main analysis: dual faecal immunochemical test 50
Additional analysis 1: synthesis of all tests together in a single analysis 51
Additional analysis 2: reference standard sensitivity analysis 51
Main analysis: summary 53
Comparative diagnostic test accuracy studies 53
Subgroup analyses by patient characteristics 53
Advanced adenomas and inflammatory bowel disease outcomes 55
Other outcomes 55
Data selected to enter the cost-effectiveness model 55

Chapter 3 Cost-effectiveness **57**
Review of existing published health economic analyses 57
Review and critical appraisal of economic analyses provided by test manufacturers 57
Independent economic evaluation 57
Scope of the Evidence Assessment Group economic analysis 57
Conceptualisation of the model 58
Model structure 59
Evidence sources used to inform the model parameters 67
Methods of model evaluation 78
Model verification and validation 79
Cost-effectiveness results 79
Conclusions from the cost-effectiveness analyses undertaken in the Evidence Assessment Group's base case 90
Deterministic scenario analyses 90

Chapter 4 Discussion and conclusions **93**
Statement of principal findings 93
Clinical effectiveness: principal findings 93
Cost-effectiveness: principal findings 95
Strengths and limitations of the assessment 96
Strengths and limitations of the clinical evidence base 96
Strengths and limitations relating to the health economic analysis 98
Uncertainties 98
Generalisability 98
Implications for service provision 99
Suggested research priorities 99
Use of patient and public involvement 100
Equality, diversity and inclusion 100

Additional information 101

References 104

Appendix 1 Literature search strategies 118

Appendix 2 Statistical methods and additional results 129

Appendix 3 Quality assessment of diagnostic test accuracy studies version 2 scoring scheme and scores with reasons for all studies 132

Appendix 4 Clinical review: table of excluded studies with rationale 156

Appendix 5 Additional analyses 157

Appendix 6 Subgroup analyses and other outcomes 161

Appendix 7 Review of existing published health economic analyses 182

Appendix 8 Model estimation of the impact of additional time to diagnosis on colorectal cancer outcomes 188

Appendix 9 Methods for model evaluation: table with distributions used for each group of parameters in Evidence Assessment Group probabilistic analyses 195

Appendix 10 Results for the deterministic scenario analyses ran by the Evidence Assessment Group 197

List of tables

TABLE 1 Summary of interventions 8

TABLE 2 Summary of studies entering the analysis, by test and study population 23

TABLE 3 Study and patient characteristics of HM-JACKarc studies 28

TABLE 4 Summary sensitivity and specificity at specific thresholds for HM-JACKarc 35

TABLE 5 Study and patient characteristics of OC-Sensor studies 37

TABLE 6 Summary sensitivity and specificity at specific thresholds for OC-Sensor 42

TABLE 7 Study and patient characteristics of FOB Gold studies 44

TABLE 8 Study and patient characteristics and diagnostic test accuracy of QuikRead go studies 47

TABLE 9 Study and patient characteristics and diagnostic test accuracy of the NS-Prime study 48

TABLE 10 Study and patient characteristics of IDK Haemoglobin (human) and haemoglobin/haptoglobin complex ELISA tests 50

TABLE 11 Study and patient characteristics of studies reporting data for dual FIT 52

TABLE 12 Summary sensitivity and specificity at selected thresholds 54

TABLE 13 Sensitivity and specificity reported in studies comparing different tests within the same patients 56

TABLE 14 Scope of the EAG economic analysis 58

TABLE 15 Evidence sources used in the model 68

TABLE 16 Parameters related to disease prevalence and disease stage or severity used in the base-case analysis 69

TABLE 17 Proportion of patients receiving each of the management pathways following FIT results,[a] based on clinical advice provided to the EAG 71

TABLE 18 Proportion of patients receiving each pathway at their lower-GI referral 72

TABLE 19 Estimates of accuracy for imaging tests used in patients in 2WW and 18WW 72

TABLE 20 Complications, QALY losses and costs associated with colonoscopy included in the short-term model 73

TABLE 21 Estimated diagnostic delays by each type of pathway and diagnostic result 74

TABLE 22 Proportion of patients by disease severity and type and health utilities applied in the EAG model for IBD patients 75

TABLE 23 Applied in the EAG model for IBD patients **75**

TABLE 24 Test costs assumed in EAG analysis **76**

TABLE 25 Tabulated results for FOB Gold using one threshold **82**

TABLE 26 Tabulated results for FOB Gold using two thresholds **83**

TABLE 27 Tabulated results for HM JACKarc using one threshold **84**

TABLE 28 Tabulated results for HM JACKarc using two thresholds **85**

TABLE 29 Tabulated results for IDK Haemoglobin using one threshold **87**

TABLE 30 Tabulated results for IDK Haemoglobin/haptoglobin using one threshold **87**

TABLE 31 Tabulated results for NS-Prime **88**

TABLE 32 Tabulated results for OC-Sensor using one threshold **89**

TABLE 33 Tabulated results for OC-Sensor using two thresholds **91**

TABLE 34 Tabulated results for QuikRead go **92**

TABLE 35 Parameters used to inform priors for syntheses with < 5 studies **129**

TABLE 36 Meta-analysis model fit statistics **130**

TABLE 37 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability **134**

TABLE 38 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability **135**

TABLE 39 FOB-Gold studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **135**

TABLE 40 QuikRead go studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **136**

TABLE 41 NS-Prime study: SCHARR's assessment of risk of bias and applicability **136**

TABLE 42 IDK studies: SCHARR's assessment of risk of bias and applicability **136**

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **137**

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **144**

TABLE 45 FOB Gold studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **152**

TABLE 46 QuikRead go studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **153**

TABLE 47 NS-Prime study: SCHARR's assessment of risk of bias and applicability, with reasons for scores **154**

TABLE 48 FOB Gold studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores **155**

TABLE 49 Summary sensitivity and specificity at specific thresholds for all tests together **159**

TABLE 50 Sensitivity and specificity of studies reporting data for patients with anaemia **162**

TABLE 51 Sensitivity and specificity by age **165**

TABLE 52 Sensitivity and specificity by sex **166**

TABLE 53 Sensitivity and specificity for patients taking medications that may affect the risk of GI bleeding **168**

TABLE 54 Studies reporting data on AAs and IBD **170**

TABLE 55 Summary sensitivity and specificity at selected thresholds for AA outcome **171**

TABLE 56 Summary sensitivity and specificity at selected thresholds for IBD outcomes **172**

TABLE 57 Studies issuing FIT in primary care or issuing dual FIT, and reporting test failure rates, test uptake and number of repeat tests **174**

TABLE 58 Studies reporting other outcomes listed in the NICE scope **176**

TABLE 59 Study characteristics of patient acceptability studies **177**

TABLE 60 Patient characteristics of patient acceptability studies **178**

TABLE 61 Patient acceptability results, Georgiou Delisle *et al.*[32] **179**

TABLE 62 Patient acceptability results, MacLean *et al.*[33] **179**

TABLE 63 Study characteristics of equity study[31] **180**

TABLE 64 Results of equity study by Bailey *et al.*[31] **181**

TABLE 65 Existing economic evaluations: analytic scope **185**

TABLE 66 Existing economic evaluations: modelling approach, main assumptions, definition of health states and summary of HRQoL included **186**

TABLE 67 Age distributions **190**

TABLE 68 Transition probabilities used in the model[93] **191**

TABLE 69 Expected discounted life-years, QALYs, and inflated treatment costs by age and stage at diagnosis **193**

TABLE 70 Estimated outcomes by additional time to diagnosis for CRC **193**

TABLE 71 Impact of additional time to diagnosis of HRA **194**

TABLE 72 Distributions used in EAG probabilistic analyses **195**

TABLE 73 Deterministic sensitivity analyses results for HM JACKarc using one threshold **197**

TABLE 74 Tabulated results for HM JACKarc using one threshold (Scenario 1) **198**

TABLE 75 Tabulated results for HM JACKarc using one threshold (Scenario 2) **199**

TABLE 76 Tabulated results for HM JACKarc using one threshold (Scenario 3) **200**

TABLE 77 Tabulated results for HM JACKarc using one threshold (Scenario 4) **201**

TABLE 78 Tabulated results for HM JACKarc using one threshold (Scenario 5) **202**

TABLE 79 Tabulated results for HM JACKarc using one threshold (Scenario 6) **203**

TABLE 80 Tabulated results for HM JACKarc using one threshold (Scenario 7) **204**

TABLE 81 Tabulated results for HM JACKarc using one threshold (Scenario 8) **205**

TABLE 82 Tabulated results for HM JACKarc using one threshold (Scenario 9) **206**

TABLE 83 Tabulated results for HM JACKarc using one threshold (Scenario 10) **207**

TABLE 84 Tabulated results for HM JACKarc using one threshold (Scenario 11) **208**

TABLE 85 Tabulated results for HM JACKarc using one threshold (Scenario 12) **209**

TABLE 86 Tabulated results for HM JACKarc using one threshold (Scenario 13) **210**

List of figures

FIGURE 1 The diagnostic pathway for patients presenting to primary care with symptoms of CRC **2**

FIGURE 2 The diagnostic pathway for patients presenting to primary care with symptoms of CRC as recommended in the ACPGBI/BSG guideline **4**

FIGURE 3 Proposed new pathway incorporating FITs for all patients in primary care: (a) using a single FIT threshold; and (b) using two FIT thresholds to create an intermediate risk group who would follow a different diagnostic pathway **7**

FIGURE 4 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of study selection **25**

FIGURE 5 Observed data and summary sensitivity and specificity for HM JACKarc **34**

FIGURE 6 Observed data and summary sensitivity and specificity for OC-Sensor **42**

FIGURE 7 Observed data and summary sensitivity and specificity for the FOB Gold primary analysis **45**

FIGURE 8 Observed data and summary sensitivity and specificity for the FOB Gold Primary analysis, with study information **45**

FIGURE 9 Sensitivity and specificity for all tests **53**

FIGURE 10 Evidence Assessment Group model: decision tree structure, intervention 1 **60**

FIGURE 11 Evidence Assessment Group model: decision tree structure, intervention 2 **61**

FIGURE 12 Evidence Assessment Group model: decision tree structure, intervention 3 **62**

FIGURE 13 Evidence Assessment Group model: state transition model **63**

FIGURE 14 Net monetary benefit for intervention 1 assuming a threshold of £20,000 per QALY gained and low-intensity safety netting **80**

FIGURE 15 Net monetary benefit for intervention 2 assuming a threshold of £20,000 per QALY gained and low-intensity safety netting **81**

FIGURE 16 The proportion of patients diagnosed by category **81**

FIGURE 17 Observed data and summary sensitivity and specificity for all tests together **158**

FIGURE 18 Summary sensitivity and specificity, reference standard sensitivity analysis **160**

FIGURE 19 Observed data and summary sensitivity and specificity for AA outcomes **171**

FIGURE 20 Observed data and summary sensitivity and specificity for all tests **172**

FIGURE 21 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram, review for economic evaluations and HRQoL studies **184**

FIGURE 22 Model diagram **189**

FIGURE 23 Stage distribution of CRC in the 2WW population in England **190**

FIGURE 24 Change in CRC stage distribution by additional time to diagnosis **192**

List of supplementary material

Report Supplementary Material 1	Worked example relating to potential bias of excluding FIT negative patients
Report Supplementary Material 2	Diagnostic test accuracy data entering the statistical synthesis
Report Supplementary Material 3	PPV and NPV for selected thresholds
Report Supplementary Material 4	Test uptake and repeat tests data from secondary care
Report Supplementary Material 5	EAG survey collected from EAG's clinical advisors
Report Supplementary Material 6	Additional health economic results

Supplementary material can be found on the NIHR Journals Library report page (https://doi.org/10.3310/AHPE4211).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

µg/g or µg Hb/g	micrograms of haemoglobin per gram of faeces
18WW	18-week wait
2WW	2-week wait
AA	advanced adenoma
ACPGBI	Association of Coloproctology of Great Britain and Ireland
AE	adverse event
A&E	accident and emergency
BSG	British Society of Gastroenterology
CD	Crohn's disease
CRC	colorectal cancer
CT	computed tomography
CTC	computed tomography colonography
DG30	Diagnostics Guidance 30
DIC	deviance information criterion
EAG	Evidence Assessment Group
ELISA	enzyme-linked immunosorbent assay
FIT	faecal immunochemical test(s)
fOBT	faecal occult blood test
GI	gastrointestinal
GP	general practitioner
HRQoL	health-related quality of life
IBD	inflammatory bowel disease
ICER	incremental cost-effectiveness ratio
IDA	iron-deficiency anaemia
iFOBT	immunochemical faecal occult blood test
iNMB	incremental net monetary benefit
LYG	life-years gained
NCRAS	National Cancer Registration and Analysis Service
NG12	National Guideline 12
NICE	National Institute for Health and Care Excellence
NMB	net monetary benefit
NSBP	no significant bowel pathology
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSS	Personal Social Services
QALY	quality-adjusted life-year
QUADAS-2	quality assessment of diagnostic test accuracy studies version 2
RCT	randomised controlled trial
ROC	receiver operating characteristic
TA	Technology Appraisal
UC	ulcerative colitis

Plain language summary

When a person visits their doctor with symptoms suggestive of colorectal cancer, people with high-risk symptoms are sent to have a test called a colonoscopy, and people with low-risk symptoms get a faecal immunochemical test. Colonoscopies, where a special camera is inserted into the anus, are very good at spotting colorectal cancer, but they are unpleasant and expensive. Long colonoscopy waiting lists mean that people are diagnosed later, when their cancer is harder to treat.

Faecal immunocheical tests use a poo sample and are quick and easy to do at home but are less good at spotting colorectal cancer. If a person has a positive result from a faecal immunochemical test, they are sent for a colonoscopy. If the faecal immunochemical test result is negative, the person is given advice by the doctor, such as to come back if their symptoms continue or worsen. The doctor can still send people for a colonoscopy if they are worried about the symptoms.

This project aimed to see if faecal immunochemical tests could be used instead of a colonoscopy in patients with high-risk symptoms, to reduce waiting lists and improve chances of survival through quicker treatment. We did a systematic review to find all relevant research studies about faecal immunochemical tests. We built a mathematical model to estimate the impact of faecal immunochemical tests on the health of patients and on National Health Service costs. The model used evidence from the systematic review, from other sources such as scientific studies and clinical opinion, and assumptions.

The model showed that using faecal immunochemical tests would shorten waiting lists and lower costs. However, the health of patients overall was slightly lower because even though some people were diagnosed more quickly, faecal immunochemical tests missed cancer in a small number of people, who had a lower chance of survival because their cancer was diagnosed later. These conclusions remained true using different assumptions in the model.

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.

Chapter 1 Background and definition of the decision problem

This chapter reproduces some content previously published by National Institute for Health and Care Excellence (NICE) as part of its scope, and as part of the study protocol produced by the Evidence Assessment Group (EAG).

Condition and aetiology

Aetiology, pathology and prognosis

The aetiology of colorectal cancer (CRC), like that of many cancers, is multifactorial and involves an interplay of hereditary, environmental and lifestyle factors. Around two-thirds of cases occur in people with no hereditary predisposition and are caused by a wide range of genetic mutations and epigenetic aberrations that may occur as a result of potentially modifiable risk factors.[1] A history of CRC in the family is evident in around 25% of cases, with around 5% attributable to hereditary cancer syndromes.[1,2]

Colonic polyps are abnormal growths in the lining of the bowel. These are usually asymptomatic and are a common finding during colonoscopy. Although most polyps do not become cancerous, most CRCs arise from colonic polyps, and their removal significantly reduces the risk of CRC. There is a greater risk of progression to CRC in people with large and/or multiple polyps, but such progression usually takes many years.[3] Therefore, an incidental but important consequence of colonoscopic investigations for CRC may be the opportunity to identify and remove polyps.

The prognosis of CRC depends on disease stage. Most people with early CRC can be cured, but late-stage disease is associated with a low 5-year survival rate. Therefore, early identification is desirable.

Epidemiology and incidence

Colorectal cancer is the fourth most common form of cancer in the UK. Approximately 42,000 new cases of CRC are diagnosed each year, resulting in around 16,800 CRC-related deaths annually.[4] The Global Burden of Disease study[5] estimates that there were 1.8 million [95% confidence interval (CI) 1.8 to 1.9 million] incident cases globally of CRC in 2017, with an age-standardised rate of 23.2 (95% CI 22.7 to 23.7) per 100,000 person-years, an increase of 9.5% since 1990. The regions with the highest incidence were Australasia, high-income Asia Pacific, high-income North America and Europe. Incidence was higher in men than in women in all regions.

Colorectal cancer is predominantly a disease of older adults, although in recent years its incidence has increased sharply in younger adults aged 20–39 years.[6] It is historically a disease of affluence in the UK, but the influence of socioeconomic factors has also changed in recent years, as surveillance data showed an increased risk among adults from areas with higher deprivation between 1996 and 2010 for men[7,8] and in the 2010s for women.[9]

Burden of disease

The Global Burden of Disease study estimates that, in the UK in 2019, 0.33 disability-adjusted life-years were lost per 100,000 person-years, a number that has fallen since the 1990s, when it was estimated to be 0.48 per 100,000 person-years.[5]

Current service provision

National Guideline 12 high-/medium-risk and Diagnostic Guideline 30 low-risk patients

National Guideline 12 (NG12) describes the diagnostic pathway for patients presenting to primary care with symptoms suggestive of CRC[10] (*Figure 1*). In accordance with this guideline, patients with the following symptoms (referred to as NG12 high-/medium-risk patients in this assessment) should be referred to secondary care with an urgent 2-week wait (2WW) suspected CRC referral. NG12[10] states:

'Refer adults using a suspected cancer pathway referral (for an appointment within 2 weeks) for CRC if:

- they are aged 40 and over with unexplained weight loss and abdominal pain or
- they are aged 50 and over with unexplained rectal bleeding or
- they are aged 60 and over with:
  - iron-deficiency anaemia or
  - changes in their bowel habit, or
- tests show occult blood in their faeces

Consider a suspected cancer pathway referral (for an appointment within 2 weeks) for CRC in adults with a rectal or abdominal mass.

Consider a suspected cancer pathway referral (for an appointment within 2 weeks) for CRC in adults aged under 50 with rectal bleeding and any of the following unexplained symptoms or findings:

- abdominal pain
- change in bowel habit
- weight loss
- iron deficiency anaemia'.

In July 2017, NG12[10] was partially updated by Diagnostics Guidance 30 (DG30).[11] In this update, the guaiac faecal occult blood test (gFOBT), which had been recommended for use in low-risk patients, was replaced with faecal immunochemical test (FIT). NG12 states:

'Offer testing with quantitative faecal immunochemical tests (see the NICE diagnostics guidance on quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care) to assess for colorectal cancer in adults without rectal bleeding who:

- are aged 50 and over with unexplained:
  - abdominal pain or
  - weight loss, or
- are aged under 60 with:
  - changes in their bowel habit or
  - iron-deficiency anaemia, or
  - are aged 60 and over and have anaemia even in the absence of iron deficiency'.

National Institute for Health and Care Excellence guidance is prepared for the NHS in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

The review undertaken for DG30 showed that the specificity of FIT was high (> 90%), and hence it was recommended for use as a rule in test at a threshold of 10 micrograms of haemoglobin per gram of faeces (µg Hb/g, hereafter referred to as µg/g). Patients testing positive using a FIT should be referred on to the 2WW suspected CRC pathway.

What happens in secondary care following referral is thought to vary across England: it may be to a specialist who will order further tests [colonoscopy, computed tomography colonography (CTC), or other tests as they see fit] or may be a direct referral by a general practitioner (GP) to an imaging test such as colonoscopy or CTC. The choice of imaging test may depend on local practice guidelines or age and comorbidities that contraindicate colonoscopy. CTC may be necessary where colonoscopy fails or is inappropriate. Colon capsule endoscopy is a relatively new imaging modality, whereby a small capsule containing a camera is swallowed in order to image the digestive tract, which is used in some areas of the UK. During colonoscopy, a biopsy may be taken for histological confirmation, unless this is contraindicated (e.g. blood-clotting disorders).

It is recommended that patients testing negative using FIT are followed up in primary care. This should include 'safety netting' as described for all cancer pathways in NG12, to avoid missing disease (cancer or otherwise) in people with negative FIT results (see *Safety netting*).[10] Safety netting in NG12 includes an awareness of the possibility of false negatives, and re-testing either after a period of time or on the emergence of new symptoms, or the recurrence, persistence or worsening of existing symptoms.[10] Safety netting may also include strategies for diagnosing other gastrointestinal (GI) conditions such as inflammatory bowel disease [IBD, a term used to describe Crohn's disease (CD) and ulcerative colitis (UC)], and further monitoring for colorectal or other types of cancer.

Speciality guidance during the height of the COVID-19 pandemic

In November 2020, NICE issued a specialty guide for patient management during the COVID-19 pandemic on triaging patients with lower-GI symptoms, which was supported by the British Society of Gastroenterology (BSG). The advice was to continue to refer in accordance with NG12, but that FIT could be used to help clinicians prioritise referrals. People with > 100 µg/g and no colonoscopy within the last 3 years, or who had symptoms considered by a specialist GI surgeon/gastroenterologist to warrant urgent investigation, would be referred for urgent colonoscopy or computerised tomography (CT), which could be CTC or plain CT. People with between 10 and 100 µg/g, or people with > 100 µg/g who have had a colonoscopy requiring no further investigation in the last 3 years, would be referred for prioritised colonoscopy or colonic imaging (CTC, plain CT, or colon capsule endoscopy). People with < 10 µg/g would be managed using safety-netting processes (see *Safety netting*).

Association of Coloproctology of Great Britain and Ireland/British Society of Gastroenterology guideline and National Health Service England letter

In 2022, the Association of Coloproctology of Great Britain and Ireland (ACPGBI) and the BSG published guidance on FIT in patients with signs or symptoms of suspected CRC (ACPGBI/BSG guidance).[12] This guidance was based on a systematic review of the available evidence and expert opinion and was agreed by consensus. An economic evaluation was not conducted. In October 2022, NHS England published letters[13,14] endorsing the use of the ACPGBI and BSG guideline on FITs in primary care, stating that it should be implemented in full.

The ACPGBI/BSG guideline recommends that FITs should be used in primary care to identify people with clinical features of CRC for referral for urgent investigation, using a threshold of 10 µg Hb/g. Those with a FIT result indicating faecal Hb ≥ 10 µg Hb/g should be referred on the urgent 2WW suspected CRC pathway in secondary care. Those not meeting these criteria and with no ongoing clinical concerns can be managed in primary care or referred on to an alternative pathway. The pathway is represented diagrammatically in *Figure 2*.

The ACPGBI/BSG guideline notes that FIT should not be the sole determinant of referral. Patients without symptoms were not considered in the guideline and should not be referred on the basis of a positive FIT result, except within the context of the national screening programme. Patients with negative FIT results should not be excluded from referral;

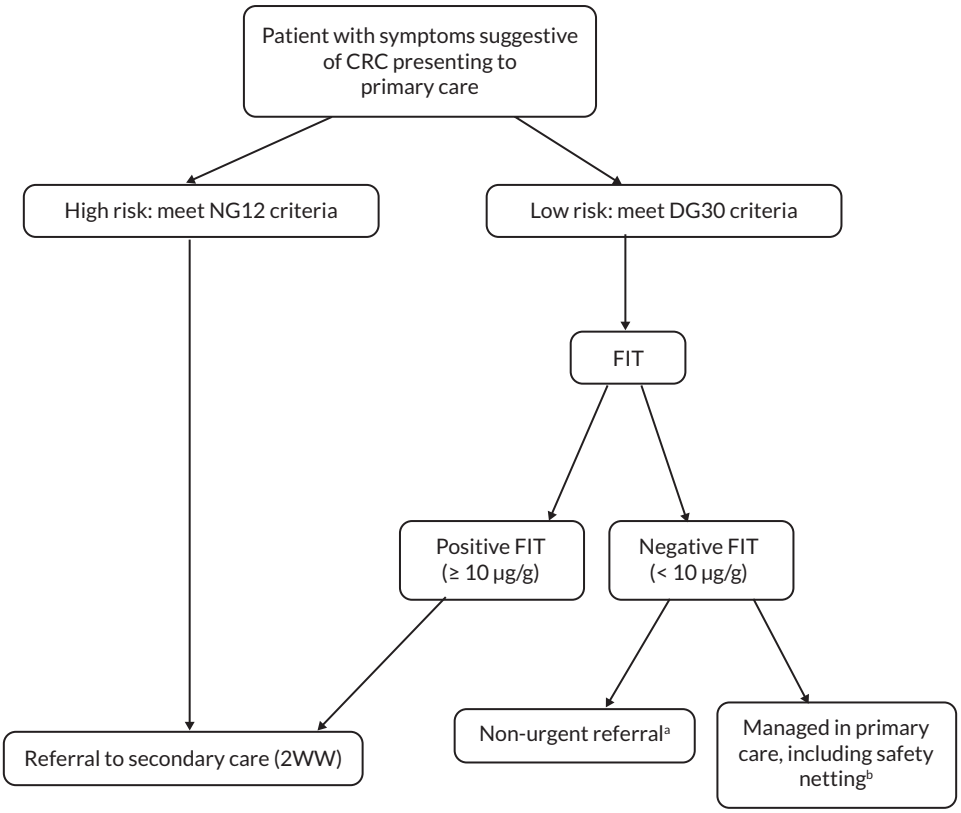

FIGURE 1 The diagnostic pathway for patients presenting to primary care with symptoms of CRC. Based on NG12[10] and DG30.[11] 2WW, two week wait; DG30, diagnostic guideline 30; FIT, faecal immunochemical test; NG12, national guideline 12; *t*, threshold. a, Non-urgent referral is part of current care for some FIT-negative patients, based on clinical judgement; b, Safety netting is discussed in *Safety netting*.

where FIT is < 10 µg Hb/g but there are persistent and unexplained symptoms that concern the GP, the patient should be referred to secondary care for evaluation. This referral may be to routine or urgent pathways, but not necessarily to the CRC pathway. Those with abdominal mass should be referred and a FIT ordered at the same time for use in secondary care. Those with anal/rectal mass or anal ulceration should be referred on to the urgent 2WW suspected CRC pathway without a FIT.

The NHS England letter also contains recommendations on safety netting for people with negative FIT results. This is discussed in more detail in *Safety netting*.

The ACPGBI/BSG guideline also includes recommendations for patients who fail to complete their FIT. These include informing the patient that their clinical assessment is incomplete and encouraging them to return the test. If the patient still does not return the FIT, existing national and local guidelines should be used to assess their risk of CRC. A limited evidence base suggested that people from ethnic minorities may be less likely to return the test, possibly due to hygiene concerns. Clinical advisors to the EAG noted the use in primary care of software (e.g. AccuRx) to send text message reminders and list non-completers for follow-up, although this may not be implemented consistently across regions.

Description of the decision problem

Purpose of the decision to be made

Early diagnosis and treatment of CRC in people presenting to primary care with symptoms can improve survival and cure rates. NICE NG12,[10] introduced in 2015, expanded the symptoms-based referral criteria recommended in NICE Clinical Guideline 27 (2005, now unavailable) to a wider set of symptoms. This resulted in an increase in the number of 2WW referrals, but there was no corresponding increase in the proportion of patients investigated who have cancer.[15] Indeed, in 2018, of 392,588 referrals made with suspected cancer on the 2WW pathway in England, only 13,168 (3.3%)

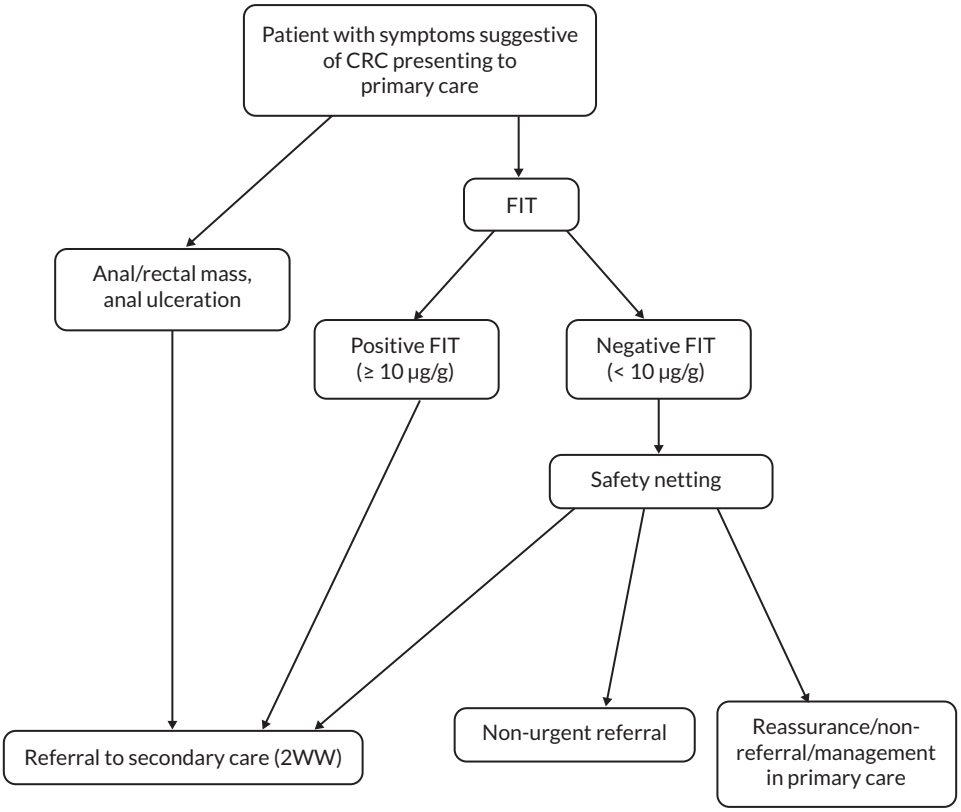

FIGURE 2 The diagnostic pathway for patients presenting to primary care with symptoms of CRC as recommended in the ACPGBI/BSG guideline.[12]

had a cancer. In addition, in August 2022, 28% of people seen by a specialist for suspected CRC were not seen within 2 weeks of urgent referral, and 53% did not have a diagnosis within 28 days (NHS cancer waiting times, August 2022, www.england.nhs.uk/statistics/statistical-work-areas/cancer-waiting-times/monthly-prov-cwt/2022-23-monthly-provider-cancer-waiting-times-statistics/provider-based-cancer-waiting-times-for-august-2022-23-provisional/). Of 15,053 people treated for lower-GI cancer in 2020–1 under a suspected cancer pathway referral, only 50.6% received treatment within 62 days following an urgent GP referral (compared with an operational standard of 85%).

National Institute for Health and Care Excellence also heard that waiting times for the non-urgent referrals are extremely long in some geographical areas. Among patients who present in primary care with symptoms of CRC, non-urgent referrals, usually with an 18-week wait (18WW) target, may be made for patients who do not meet the criteria for a 2WW referral, but for whom there is clinical concern. This may be because the GP suspects that another bowel pathology could be present, such as IBD. A delay in diagnosis for these patients could result in worse quality of life and other patient outcomes.

The reasons for the increased waiting list times for colonoscopy are unclear and may be attributable to a backlog that accumulated during the COVID-19 pandemic and/or referrals exceeding capacity.

National Institute for Health and Care Excellence heard from consultation with stakeholders and the NHS that the current symptom-based referral pathway, using the NG12 and DG30 criteria, is difficult for GPs to implement. The ACPGBI/BSG guideline[12] and the meta-analysis that informed the guidelines[16] also found that there is no clinically significant difference in sensitivity when FIT is used in patients presenting with DG30 and NG12 symptoms as well as those presenting with certain individual symptoms (rectal bleeding, iron deficiency anaemia and abdominal pain), although this guideline did not consider the impact on cost-effectiveness.

There is evidence that FIT is a better predictor of CRC risk in patients than symptoms alone[17] and could result in fewer referrals of people without CRC to the urgent 2WW suspected CRC pathway. Therefore, triage with FIT could mean

that people who are unlikely to have CRC may avoid unnecessary referral for investigations. They would also avoid the associated disadvantages of referral such as patient anxiety, time off work and loss of economic productivity, as well as the rare adverse events (AEs) associated with colonoscopy, for example bleeding, perforation and death. Furthermore, those likely to have CRC can be prioritised more effectively,[18] leading to a reduction in time to diagnosis. This may also release colonoscopy capacity to allow people on non-urgent referral pathways to be seen more quickly. The extent to which colonoscopy capacity is released will depend in part on the threshold used to define a positive test for the symptomatic patients.

The medical technologies topic oversight group identified FIT as an adjunct to clinical assessment in guiding referral for people with high-risk symptoms in primary care as suitable for guidance development by the Diagnostics Assessment Programme on the basis of a briefing note. The topic scoping completed in April 2020 but was paused due to changes in clinical pathways due to the COVID-19 pandemic. Following exceptional surveillance of NICE NG12 and NICE DG30, it was decided to resume the topic but rescope to take into account the changes to clinical practice.

As a result of the rescoping exercise, and of the scoping workshop on 11 October 2022 and the assessment subgroup meeting on 2 November 2022, the need to identify the optimal way to use FIT to reduce the number of people without significant bowel pathology who are referred to the suspected CRC pathway, taking into consideration the threshold used to define a positive test, and the potential colonoscopy capacity constraints for urgent and non-urgent referrals, was identified as an objective of this assessment.

Place of the intervention in the treatment pathway

This assessment considered the use of FITs in people presenting to primary care with GI symptoms indicating a risk of CRC. The treatment pathway and proposed position for FIT are shown in *Figure 3*.

Faecal immunochemical tests were to be evaluated as an adjunct to clinical assessment to guide referral of a symptomatic population to the suspected CRC pathway. Consistent with the ACPGBI/BSG guideline, this population included both those meeting NG12 criteria for an urgent 2WW suspected CRC referral and those meeting DG30 criteria for a FIT, and excluded those with rectal or anal mass, or anal ulceration (who should go straight to an urgent 2WW suspected CRC referral, termed ‘bypass symptoms’ in this assessment). Patients would receive the test in primary care, and the result of the test would be used to determine who would proceed to secondary care and who would be followed up in primary care with safety netting.

Definition of the intervention

Quantitative FITs are designed to detect occult (small amounts) of blood in stool samples (faecal haemoglobin) using antibodies specific to human haemoglobin.

Faecal immunochemical tests are available as quantitative tests [using immunoturbidimetric or enzyme-linked immunosorbent assay (ELISA) methods to measure haemoglobin concentration] or qualitative tests (using immunochromatographic test devices to detect haemoglobin). In line with DG30, this evaluation will focus on quantitative FIT.

Immunoturbidimetric FIT contains particles that are coated in antibodies specific to human haemoglobin. Six of the tests within the scope of this assessment use this methodology (see *Table 1*). The antibodies bind to haemoglobin present in the faecal sample, creating complexes that are detected using turbidimetry (how much light is absorbed when passed through a solution).

Enzyme-linked immunosorbent assay FIT uses antibodies specific to human haemoglobin to bind haemoglobin in the faecal sample to the surface of microtiter wells. Two of the tests within the scope for this assessment use this methodology (see *Table 1*). This is then treated with chemicals to produce a colour change. The intensity of the colour is proportional to the amount of haemoglobin in the sample. Some assays may also include antibodies for human haptoglobin. Haptoglobin is a protein produced by the liver that binds to haemoglobin, making it less likely to break down during transit through the GI tract. The detection of haptoglobin is claimed to increase the likelihood of detecting lesions higher in the colon.

Different FIT may report outcomes using either concentration of haemoglobin in the sampling device buffer (nanograms Hb/ml buffer) or concentration of haemoglobin by mass of faeces (μg/g). As the amount and type of buffer used vary between manufacturers, the World Endoscopy Organization's expert working group on FIT for CRC screening recommended that μg/g should be used as a standard measure that can be compared easily between tests.[19]

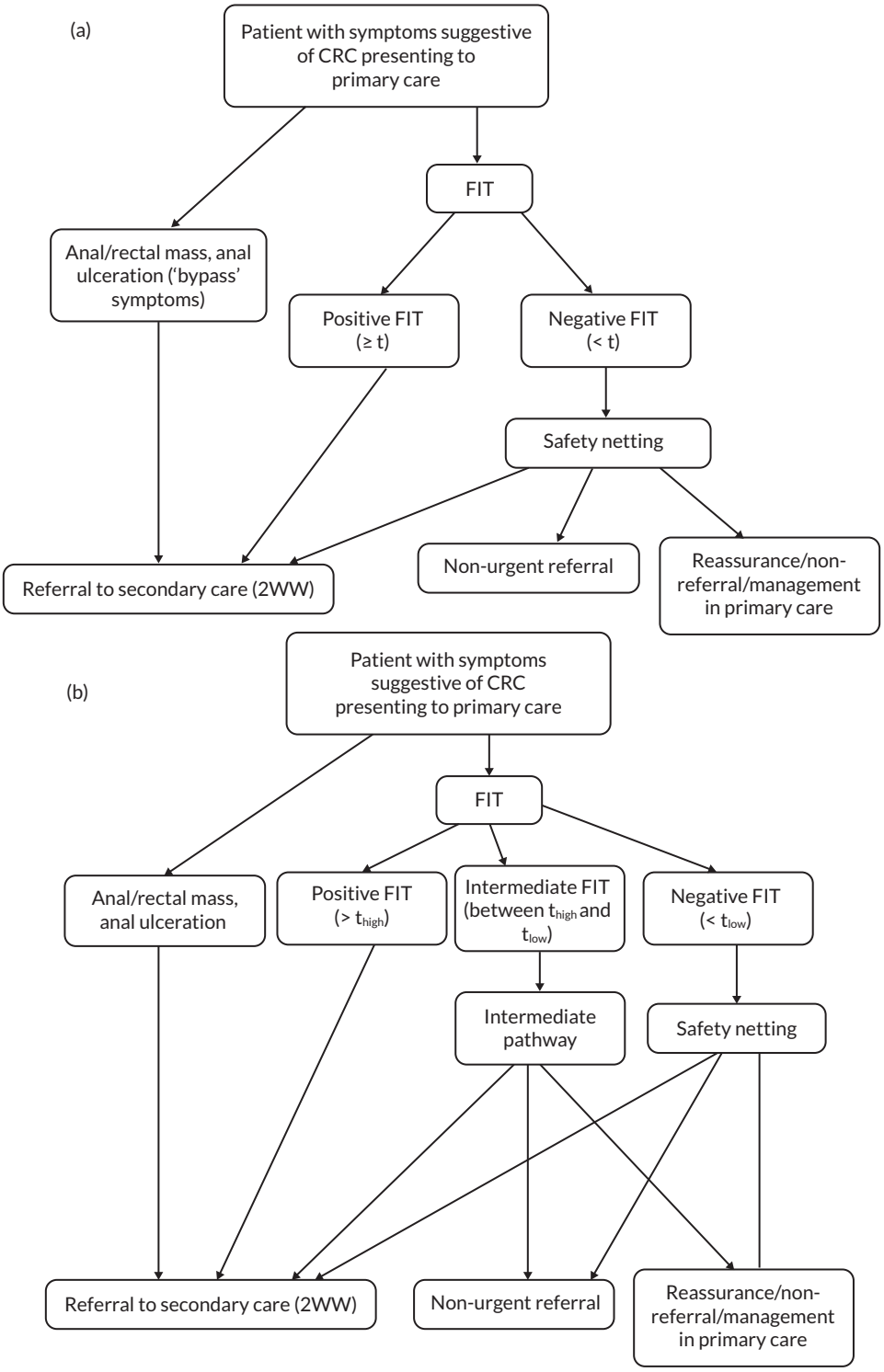

FIGURE 3 Proposed new pathway incorporating FITs for all patients in primary care: (a) using a single FIT threshold; and (b) using two FIT thresholds to create an intermediate risk group who would follow a different diagnostic pathway. *t*, threshold; t_{high}, higher threshold; t_{low}, lower threshold.

TABLE 1 Summary of interventions

Test	Test principle	Analyser compatibility	Sample size required (mg)	Measuring range (µg/g)	Limit of detection (µg/g)	Limit of quantitation (µg/g)	Throughput
HM-JACKarc	Immunoturbidimetry	HM JACKarc analyser	2	7–400	2	7	200 samples per hour
FOB Gold	Immunoturbidimetry	Various	10	Varies according to the analyser used	Varies according to the analyser used	Varies according to the analyser used	Dependent on the analyser used
OC-Sensor	Immunoturbidimetry	OC-Sensor PLEDIA	10	2–50,000	2	2	320 samples per hour
	Immunoturbidimetry	OC-Sensor iO	10	2–200	2	4	88 samples per hour
NS-Prime	Immunoturbidimetry	NS-Prime analyser	10	4–240	4	10	300 tests per hour
IDK TurbiFIT	Immunoturbidimetry	Various	15	Varies according to the analyser used	Varies according to the analyser used	Varies according to the analyser used	Depends on the analyser used
IDK Hemoglobin ELISA	ELISA	Various [ELISA plate reader with a photometer (Dynex DS2 and DSX systems)]	15	0.18–50	0.15	0.18	Depends on the analyser used
IDK Hb/Hp complex ELISA	ELISA		15	0.25–50 µg Hb/Hp/g	0.16 µg Hb/Hp/g	0.25 µg Hb/Hp/g	Depends on the analyser used
QuikRead go iFOBT (point-of-care test)	Immunoturbidimetry	QuikRead Go analyser	10	10,200	2.5	9.5	< 2 minutes per test

Hb, haemoglobin; Hp, haptoglobin.

Note
Accuracy should be estimated according to analyser used if data are available. Information was provided by companies to NICE or taken from the test's instructions for use document or website.
Source: adapted from table 1 in the NICE scope,[21] © NICE 2022. Quantitative faecal immunochemical testing to guide colorectal cancer pathway referral in primary care. Available from www.nice.org.uk/guidance/dg56. All rights reserved. Subject to Notice of rights www.nice.org.uk/terms-and-conditions#notice-of-rights.
NICE guidance is prepared for the National Health Service in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

Strategies and thresholds for using faecal immunochemical tests as a triage tool

As the test is quantitative, thresholds may be varied to achieve optimal clinical effectiveness and cost-effectiveness outcomes with respect to colonoscopy capacity, quality-adjusted life-years (QALYs) or net monetary benefit (NMB).

Strategies using one FIT threshold were to be investigated, where a FIT result above a threshold resulted in referral to the urgent 2WW suspected CRC pathway, while a FIT result below the threshold would result in safety netting (see *Safety netting* and *Short-term decision-tree component of the model*). A strategy using two FIT thresholds (t_1 and t_2) was also to be considered (see *Figure 3b*) and is described in *Use of two FIT thresholds to guide referral, and the intermediate group pathway* and *Short-term decision-tree component of the model*. A strategy using two FITs (dual FIT) was also of interest (see *Dual testing* and *Scenario analyses*).

Several FITs are within the scope of this assessment. These are described in *HM-JACKarc system* to *QuikRead go iFOBT* and are summarised in *Table 1*.

HM-JACKarc system

The HM-JACKarc system (Minaris Medical Co., Ltd, Tokyo, Japan) is a fully automated quantitative immunoturbidimetric FIT system. The system comprises a sample collection device (designed to measure 2 mg of faeces) that contains 2 ml of stabilising buffer, latex agglutination reagent and buffer solution. The assay is compatible with the HM JACKarc analyser, which can process up to 200 samples per hour, with a maximum capacity of 80 samples per run.

FOB Gold

FOB Gold (Sentinel/Sysmex, Milan, Italy) is an automated quantitative immunoturbidimetric FIT system. It comprises faecal sample collection tubes (the SENTiFIT pierceTube faecal collection device) that collect 10 mg of faeces in 1.7 ml of buffer, and latex agglutination reagent. The FOB Gold kit is compatible with Sentinel's own SENTiFIT analyser as well as those manufactured by five other companies. The performance characteristics of the assay vary depending on which analyser is used. The throughput of the test depends on the clinical chemistry analyser used to process the samples, but 270 samples can be run per hour on the SENTiFIT 270.

OC-Sensor

The OC-Sensor (Eiken Chemical, Japan/MAST Diagnostics, Bootle, UK) is a quantitative immunoturbidimetric FIT. It comprises faecal sample collection tubes, latex reagent and buffer. The OCAuto sampling bottles can hold 10 mg of faeces.

The test can be run on either the OC-Sensor PLEDIA or the OC-Sensor iO analyser, which differ in the number of samples they are able to process. The OC-Sensor PLEDIA can process up to 320 samples per hour, with a capacity of 200 samples per run. The OC-Sensor iO can process up to 88 samples per hour with a maximum capacity of 20 samples per run.

MAST Diagnostics states that the OC-Sensor iO will soon be replaced by the OC-Sensor Ceres.

NS-Prime

The NS-Prime (Alfresa, Tokyo, Japan/Abbott, Illinois, USA) is an automated quantitative immunoturbidimetric FIT system. The NS-Prime comprises a specimen collection container which collects 10 mg of faeces in 1.9 ml of buffer solution.[20] The test is run on the NS-Prime clinical chemistry analyser.

The NS-Prime haemoglobin reagent is specific to the NS-Prime analyser and cannot be used on other platforms. The NS-Prime analyser can run up to 220 samples at the same time, processing 300 tests per hour.

IDK TurbiFIT

The IDK TurbiFIT assay (Immundiagnostik, Bensheim, Germany) is an immunoturbidimetric FIT compatible with a range of automated clinical chemistry analysers from 16 manufacturers. The TurbiFIT kit comprises reagents, control samples and calibration samples. IDK TurbiTUBE sample collection devices are available separately, which collect 15 mg of faeces in 1.5 ml of buffer. The performance characteristics and throughput of the assay vary depending on which analyser is used.

IDK Haemoglobin (human) and haemoglobin/haptoglobin complex enzyme-linked immunosorbent assay tests

The IDK haemoglobin (human) ELISA (Immundiagnostik) is an immunoassay for the quantitative determination of human haemoglobin in faeces. It consists of:

- a microtiter plate, pre-coated in antibodies
- buffers for washing, extraction and sample dilution
- conjugate peroxidase-labelled antibodies
- standards and controls
- tetramethylbenzidine substrate (to induce the colour change).

The test requires an ELISA plate reader with a photometer (Dynex DS2 and DSX systems; Dynex, Chantilly, VA, USA) to determine the result. The throughput of the test depends on the clinical chemistry analyser used to process the samples.

The company also produces the IDK haemoglobin/haptoglobin complex ELISA, which is similar but uses anti-haptoglobin antibodies in the coated microtiter plate. The company recommends using this test in addition to a haemoglobin test to improve sensitivity for detecting bleeding adenomas or cancers of the upper intestine.

QuikRead go immunochemical faecal occult blood test

The QuikRead go (Aidian, Espoo, Finland) is a point-of-care analyser that can be used for a number of different diagnostic tests, including the immunochemical faecal occult blood test (iFOBT) which is an immunoturbidimetric test. The kits contain reagent capsules and buffer in prefilled cuvettes. Faecal sampling sets and control materials are supplied separately. A single sample can be run at a time, and it takes < 2 minutes for the test result to be displayed.

Populations and relevant subgroups

The population of interest was people presenting to primary care with symptoms or signs indicating a risk of CRC, as defined in NG12 and DG30.

Certain symptoms may indicate that patients should be referred directly to the urgent 2WW suspected CRC pathway (people with palpable rectal or anal mass or anal ulceration, termed 'bypass symptoms' in this assessment), and these patients were excluded from the scope. In contrast to DG30 (see *NS-Prime*), rectal bleeding was not considered a symptom that would preclude the use of FIT, as clinicians indicated during the scoping process for this assessment that FITs can be used in those with rectal bleeding.

Some reports suggest that faecal haemoglobin levels may differ according to certain patient characteristics. If confirmed, different cut-off values may be needed according to the following characteristics:

- age
- sex
- ethnicity
- people taking medications or with conditions that increase the risk of GI bleeding
- people with blood disorders (e.g. beta thalassemia) that could affect the performance of the test
- people with anaemia [including iron-deficiency anaemia (IDA)].

This assessment aimed to identify diagnostic test accuracy data within these subgroups to help inform whether alternative thresholds may be required to achieve accuracy equivalent to that for patients without these characteristics. Economic modelling was not planned for these subgroups.

Relevant comparators

Current practice corresponds to standard care according to NG12 and DG30 (see *National Guideline 12 high-/medium-risk and Diagnostic Guideline 30 low-risk patients*). This includes:

- clinical assessment and referral for further investigation in secondary care
- use of FIT (threshold of 10 µg/g) to guide referral only for those with 'low-risk' symptoms without rectal bleeding (in line with NICE DG30).

Feedback from clinical experts and stakeholders during the scoping stage of this assessment was that stratification by symptoms is a poor predictor of risk of CRC. Any resulting guidance that differentiates between the risk groups currently defined in NICE guidance would not address this problem. Therefore, despite the possibility of differential cost-effectiveness by subgroup, NICE's scope[22] stated that the intervention arm should not subgroup according to NG12 high-risk and DG30 low-risk categories and should not exclude those with active rectal bleeding, to prevent recommendations being made according to symptom-based criteria. Consequently, the comparator was a blended group of people who would currently be considered under the guidance of NG12 and DG30.

The NICE scope noted that the comparators for the modelling may differ.

Healthcare setting

The assessment related to the use of FITs in primary care.

Outcomes

The NICE scope[22] states that intermediate outcomes of interest may include:

- diagnostic accuracy at different FIT thresholds for CRC, AA and IBD
- risk of CRC (and IBD and AAs) in relevant subgroups according to FIT threshold
- test failure rates
- prognostic implications of false-negative results
- uptake (completion) of FITs in primary care
- number/proportion of people referred to secondary care
- number/proportion of people followed up in primary care
- duration of validity of negative test (implications for follow-up)
- number/proportion of urgent (2WW suspected cancer) specialist appointments
- number/proportion of urgent (2WW suspected cancer) colonoscopy/CTCs
- number/proportion of non-urgent colonoscopy/CTCs
- time to colonoscopy/CTC
- time to diagnosis of CRC or other conditions
- number/proportion of colonoscopy/CTCs that do not detect CRC
- number/proportion of colonoscopy/CTCs that do not detect significant bowel pathology
- number/proportion of people presenting to emergency departments with symptoms of CRC.

The NICE scope[22] states that clinical outcomes for consideration may include:

- number of CRC diagnoses
- number/proportion of CRC diagnoses from urgent referrals
- stage of detected cancers
- number/proportion of people identified with other bowel pathologies
- number/proportion of people with AAs detected, or detected and treated
- morbidity including AEs associated with colonoscopy
- mortality.

The NICE scope[22] states that patient-reported outcomes for consideration may include:

- health-related quality of life (HRQoL)
- anxiety associated with waiting for referral or test results due to diagnostic delays, and further diagnostic workup
- preference for FIT versus colonoscopy.

The NICE scope[22] states that costs were to be considered from an NHS and Personal Social Services (PSS) perspective. Costs for consideration included:

- cost of equipment, reagents and consumables for FITs
- cost of staff and associated training
- medical costs arising from testing and care, including further follow-up and safety netting
- medical costs of AEs that arise from testing or further diagnostic workup, including those associated with false test results and inappropriate treatment.

A lifetime horizon was to be used. The cost-effectiveness of FIT versus usual practice was to be expressed in terms of the incremental cost per QALY gained [incremental cost-effectiveness ratio (ICER)]. Net health benefit was to be used

when comparing multiple interventions, but the EAG has presented NMB to aid the committee's interpretations of the results of the economic analyses.

Other considerations

There is known to be heterogeneity within care pathways across the country and this was to be investigated in the assessment.

Faecal immunochemical test threshold for referral

The FIT cut-off value recommended in DG30 was 10 µg/g, as the committee concluded that this gave the test enough sensitivity to reliably rule out CRC in the low-risk population. FIT thresholds may be varied for two reasons:

- To optimise the treatment pathway for clinical effectiveness (QALYs) or cost-effectiveness (in terms of net health benefit) and to investigate the impact on numbers/proportions of referrals
- Because faecal haemoglobin levels are thought to differ according to certain patient characteristics (see *Populations and relevant subgroups*), different cut-off values may be needed for these subgroups to avoid potential equity issues.

Both reasons for threshold alteration were to be considered in the assessment.

Use of two faecal immunochemical test thresholds to guide referral, and the intermediate group pathway

Two FIT thresholds could be used to define low- (FIT lower than t_{low}), intermediate- (FIT between t_{low} and t_{high}) and high-risk populations (FIT > t_{high}). In this strategy, people in the intermediate-risk group (with FIT between t_{low} and t_{high}) may have more intensive monitoring of their condition than in the low-risk group or be referred to a specialist safety-netting pathway (see *Safety netting*). The management pathway for the intermediate group was unclear and was addressed in the modelling.

Measurements and diagnostic test accuracy of different tests and analysers

Different tests, different analysers and different combinations of tests and analysers (see *Table 1*) may have different measuring ranges, may give different absolute measurements and may have different test accuracy. The NICE scope[22] notes that accuracy should be analysed according to the test–analyser combination. This was considered in the clinical review of evidence.

Use of faecal immunochemical tests alongside bypass referral

As already noted, clinical experts advised NICE that rectal bleeding would no longer be considered a reason to bypass FIT. Both the NHS England letter and the ACPGBI/BSG guideline stated that the presence of a palpable rectal or anal mass, or anal ulceration, were symptoms that indicated that patients should move straight to a 2WW referral, thereby bypassing FIT. Some clinical experts said that FIT could still be useful alongside referral to help choose the method of further investigation, and may be required by some secondary care centres. As the bypass symptoms are not part of the decision problem population, this assessment did not include those symptoms in the modelling.

Dual testing

Two FITs can be used to guide referral. There are two main ways in which two tests can be used, and in this assessment, these are termed 'dual FIT' and 'repeat FIT'.

'Dual FIT' was defined in the NICE scope[22] as using two samples from different bowel movements rather than a single sample from one bowel movement. The scope notes that it is different from using FIT as part of a safety-netting programme, which we are calling in this assessment 'repeat FIT'. Repeat FIT has also been defined elsewhere as referring to the use of a second FIT after a decision to refer or not refer has been made.[23] Repeat FIT usually takes place weeks or months later (see *Safety netting*) as a result of continuing or worsening symptoms, whereas dual FIT is given to all patients on the basis of their initial consultation.

This assessment considered dual FIT as a testing strategy. Based on clinical expert opinion, people would be referred to the suspected cancer pathway if either FIT sample was positive. Dual FIT may result in fewer false-negative results, more false-positive results and higher costs of FIT testing.

Repeat FIT is considered as part of safety netting within the economic modelling for this assessment (see *Probability of following each of the pathways following faecal immunochemical test result*). Studies reporting data on repeat FIT are reported in *Test uptake and repeat tests*.

Safety netting

Clinically, safety netting refers to various strategies and processes used in the diagnostic pathway to avoid missing disease (cancer or otherwise). In the context of the CRC pathway, this is most usually for those who are not initially referred to secondary care. This section outlines some of the available recommendations on safety netting, which clinical advisors to the EAG indicated are implemented to differing extents across the country.

Diagnostics Guidance 30 modelling assumed the following for safety netting (persons with 'negative FIT'): (1) if they had cancer, they would have a delay in diagnosis of < 12 months as they would re-present with continuing or worsening symptoms; and (2) for those without cancer a proportion would also have persistent symptoms, some of whom would receive colonoscopy, and some would receive a repeat FIT. For (2), proportions were estimated based on clinical opinion (two clinicians who provided quite different estimates; table 26 in DG30[24]); the DG30 EAG assumed that 32.5% of patients who tested negative with FIT/gFOBT would have persistent symptoms and would receive colonoscopy and 20% had repeat FIT.

National Guideline 12 recommends safety netting for people with symptoms associated with an increased risk of cancer who do not meet the criteria for referral or other investigative action across all cancer pathways. This may be planned within a timeframe agreed with the person, or initiated by the person if their symptoms recur, persist or worsen. The guideline also states the possibility of false-negative results from FIT. The ACPGBI/BSG guideline recommends that safety-netting protocols should include advice and strategies for the diagnosis of colorectal and extracolonic cancers, as well as other serious GI conditions.

The recent NHS England letter stated that the ACPGBI/BSG guideline[12] should be implemented in full and provided recommendations for safety netting. These stated that clinical teams should consider:

- 'Providing the patient with clear information about who to contact if they develop new symptoms or if their existing symptoms worsen.
- Using advice and guidance via eRS (electronic referral system) to guide management of patients with persistent or troublesome symptoms.
- Offering a second FIT if ongoing clinical concerns remain. (NB, this is called "repeat FIT" in this assessment.)
- Referral to a non-specific-symptoms urgent cancer pathway, if appropriate and there are ongoing concerns about possible cancer.
- Management of FIT negative patients in an outpatient setting following referral on a non-urgent pathway. For example, the North Central London Cancer Alliance has developed a FIT negative, non-urgent referral pathway, as has Oxford University Hospitals NHS Foundation Trust'.

The electronic referral system is used in some areas as a means of communication between primary and secondary care for advice and guidance, while in others it may be used only to make and track referrals. Other methods of communication may be used between primary and secondary care for advice and guidance.

Safety netting was to be included as part of the diagnostic pathway of patients with negative FIT results in this assessment, exploring different assumptions about its composition (see *Short-term decision-tree component of the model* and *Probability of following each of the pathways following faecal immunochemical test result*).

Other conditions with gastrointestinal symptoms

Patients presenting with symptoms of CRC may have other GI pathologies such as IBD (CD or UC), diverticular disease or AAs. Colonoscopy is required to diagnose IBD and to identify and treat AAs.

The COLOFIT project conceptual modelling has opted to explicitly include IBD (both CD and UC) in the model because of the known impact of a delayed diagnosis on prognosis, costs and quality of life. Other bowel diseases

were not modelled explicitly because of a lack of clarity around whether a diagnostic delay is likely to cause harm. In this assessment, a similar approach was taken, and IBD and AAs were included within the scope of the modelling; the pathways for these patients were also considered to have an impact in outcomes due to a delay in diagnosis. A delay in diagnosis for IBD may worsen quality of life and patient outcomes, while AAs are largely asymptomatic and colonoscopic findings in AAs are largely incidental, but some may eventually develop into CRC if not treated, which may have an impact on patients' lifetime survival, HRQoL and costs (see *Model structure*).

Urgent 2-week wait suspected colorectal cancer pathway and secondary care management

Clinical advisors indicated during the scoping process for this assessment that there was heterogeneity in current practice regarding what happens in secondary care when a patient is referred to the urgent 2WW suspected CRC pathway. This was to be appropriately represented in the project.

Non-urgent referral pathway

Clinical advisors indicated during the scoping process for this assessment that there was heterogeneity in current practice regarding what the non-urgent referral pathway entails. This was to be appropriately captured in the project.

Non-completers of faecal immunochemical tests

A proportion of patients do not return their FIT. Based on the systematic review conducted for DG30, FIT were returned by 41% (in a study where patients were sent an invitation to participate along with their referral letter) to 98% (in a study where patients were given the specimen collection device at their initial consultation with a gastroenterologist) for patients using OC Sensor, and 56–66% patients using HM-JACKarc. This was to be taken into account in the project.

Areas that are outside the scope of the appraisal and therefore do not require any detailed assessment (e.g. key factors for which evidence is already accepted)

Evidence on the equivalence of tests and test-analyser combinations (e.g. Bland–Altman plots) was not sought or statistically synthesised by the EAG. Evidence submitted by companies relating to equivalence was to be considered by the EAG to inform modelling scenarios.

The development of a risk prediction model using FIT and clinical characteristics was not within the scope of the assessment. This type of work was being conducted by other groups (e.g. the NICE FIT group, COLOFIT). A review of risk prediction models is also not within the scope of this assessment, as this work is being conducted by the COLOFIT group.

Chapter 2 Clinical evidence

This chapter reproduces some content previously published by NICE as part of the study protocol produced by the EAG. © NICE 2022. Quantitative faecal immunochemical testing to guide colorectal cancer pathway referral in primary care. Available from www.nice.org.uk/guidance/dg56. All rights reserved. Subject to Notice of rights www.nice.org.uk/terms-and-conditions#notice-of-rights.

National Institute for Health and Care Excellence guidance is prepared for the NHS in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

Methods

A systematic review was conducted to identify clinical efficacy and diagnostic test accuracy studies relevant to the decision problem. Clinical efficacy studies refer to 'end-to-end' studies that compare two different testing strategies using a randomised controlled trial (RCT) design, whereas diagnostic test accuracy studies refer to studies that report intermediate outcomes such as sensitivity and specificity using a cohort or cross-sectional design. The main review question was:

- What is the clinical efficacy and diagnostic test accuracy of FITs for patients presenting to primary care with symptoms and signs suggestive of CRC?

The review also aims to identify, in these studies, other outcomes of relevance to the decision-making process such as test uptake, test failure rates, patient perspectives and the impact of sociodemographic factors on the outcomes of interest. The full list of outcomes can be found in *Outcomes*.

Summary of the approach to the review

The ACPGBI/BSG guideline[12] was based on a recent systematic review of the literature relating to clinical efficacy and diagnostic test accuracy.[16] Some of the authors of that review were clinical advisors to the EAG (Mr Muti Abulafi, Mr Kevin Monahan, Dr Richard Booth, Dr Rachel Carten) and they shared their review work as a basis for the review for this assessment. There were some notable differences in scope between the review for this assessment and the ACPGBI/BSG, namely that a limited number of thresholds were eligible for inclusion in ACPGBI/BSG, different subgroup analyses were planned, and the focus was not on the recruitment of patients in primary care only. To ensure that all threshold and relevant subgroup data were identified, the list of ACPGBI/BSG excluded studies was scrutinised to identify studies relevant to this assessment, and where data were not extracted for all thresholds and subgroups reported in a study, the original study was revisited to perform de novo data extraction as detailed in *Data extraction strategy*. Studies not relevant to this assessment that were included in the ACPGBI/BSG review were excluded.

The protocol for this review was registered in the International prospective register of systematic reviews (PROSPERO, registration number CRD42022383580).

Population

Studies were included if they recruited people presenting to primary care with signs or symptoms indicating a risk of CRC. Signs and symptoms of CRC were defined as those described in NG12 and DG30 (see *National Guideline 12 high-/medium-risk and Diagnostic Guideline 30 low-risk patients*), although studies were not excluded if the recruitment criteria were wider than those listed in NG12 and DG30, or were narrower. Studies reporting data relating to the subgroups specified in the population section (see *Populations and relevant subgroups*) of the decision problem (e.g. age, sex) were also included (hereafter these are called 'patient characteristics studies'), but studies reporting on very narrow populations that did not relate to a subgroup of interest, such as those with rectal bleeding only, were excluded. Studies that did not recruit only patients presenting to or referred from primary care (e.g. those that included people referred from secondary care) or that did not recruit only symptomatic patients (e.g. those that included people undergoing

population-level screening or referred as a result of screening, polyp surveillance, or with a family history of CRC) were excluded.

A tiered approach to inclusion was taken. Where no or few data for a given test or subgroup were identified, studies that recruited somewhat different populations (e.g. that recruited patients referred from secondary care as well as from primary care) were considered for inclusion if generalisability was thought to be reasonable. Where criteria have been widened, this is noted in the report. Decisions around generalisability were made on the basis of the proportion of out-of-scope participants, and on the likely impact of a given patient spectrum. In particular, studies exclusively of screening or surveillance populations were not considered generalisable.

Interventions

Studies were included if they reported data using any of the test–analyser combinations listed in *Definition of the intervention* and in *Table 1*. Data relating to all thresholds were included. Studies reporting dual testing (see *Dual testing*, hereafter referred to as 'dual FIT') were included. Each test was considered individually, but an analysis by test-analyser was not conducted as an assumption of equivalence between devices was considered reasonable by the EAG's clinical advisors for OC-Sensor devices, and there were too few studies to conduct such an analysis for FOB Gold, and the same assumption of equivalence has been made.

Comparators

For the review of clinical efficacy

End-to-end RCT studies that compared one diagnostic strategy with current standard of care (under NG12/DG30; see *Figure 1*) in England were eligible for inclusion.

For the review of diagnostic test accuracy studies and comparative diagnostic test accuracy studies

Studies were included if the reference standard was full colonic imaging via colonoscopy or CTC, or if some patients received other reference standards such as index-test-dependent differential references standards comprising imaging for FIT-positive patients and records follow-up for FIT-negative patients. This was a change to the published protocol, in which a tiered approach was planned, which prioritised studies with 100% colonoscopy or CTC reference standards in the first instance.

It is the EAG's view that all the reference standards available were subject to limitations. Full colonic imaging using colonoscopy or CTC is not 100% accurate,[25,26] and consequently it may be preferable that studies using this as a reference standard should also perform additional follow-up via medical records to identify missed cases. This was rarely if ever done in the studies identified by this review. This reference standard is also not suitable for some patients, such as those who are elderly/infirm and those with rectal bleeding. Studies using this reference standard may therefore exclude some patients from their analysis, which may reduce the generalisability of the findings. Reliance only on records follow-up for some patients may also result in missed diagnoses, for example through incomplete record-keeping, patients moving away, or patients dying from another cause before a diagnosis has been reached. Records follow-up may also incorrectly classify some patients as false negatives when follow-up is long (e.g. in the order of years rather than months), allowing time for cancers that were not present at the time of the index test to have developed. It may also be less sensitive to non-cancer diagnoses, as record-keeping for such conditions may be less complete. The recent ACPGBI/BSG review found some numerical differences in diagnostic test accuracy between studies when comparing studies with a full colonic imaging reference standard with those with a differential reference standard comprising a mix of imaging and records follow-up. However, the difference was not statistically significant.

There is some evidence[27] that most patients with a missed CRC will re-present to primary or emergency care within 6 months of their initial consultation, mitigating some of the concerns with follow-up reference standards. Furthermore, the exclusion of studies that did not give all patients full colonic imaging would have largely excluded studies that recruited all patients in primary care (see *Rationale for the analysis plan*), skewing the patient spectrum away from the population of most interest. As a result, all reference standards were eligible for inclusion in the review, and a sensitivity analysis was performed to include only studies with > 90% colonoscopy or CTC, as was done in line with the ACPGBI/BSG review.[16]

No adjustment for imperfect reference standards was attempted in the statistical synthesis for this assessment.

Comparative diagnostic test accuracy studies that compared two or more of the tests or test–analyser combinations listed in *Definition of the intervention* with each other were included, as long as they included a valid reference standard.

Outcomes

For the review of end-to-end clinical efficacy studies, the following outcomes were eligible for inclusion:

- number of CRC diagnoses
- number/proportion of CRC diagnoses from urgent referrals
- stage of detected cancers
- number/proportion of people identified with other bowel pathologies
- number/proportion of people with advanced adenomas (AAs) detected, or detected and treated
- morbidity including AEs associated with colonoscopy
- mortality
- HRQoL
- anxiety associated with waiting for referral or test results due to diagnostic delays, and further diagnostic workup
- preference for FIT versus colonoscopy
- risk of CRC (and IBD and AAs) in relevant subgroups according to FIT threshold
- test failure rates
- prognostic implications of false-negative results
- uptake (completion) of FIT in primary care, to include with respect to cultural, demographic or socioeconomic factors
- number/proportion of people referred to secondary care
- number/proportion of people followed up in primary care
- duration of validity of negative test (implications for follow-up)
- number/proportion of urgent (2WW suspected CRC) specialist appointments
- number/proportion of urgent (2WW suspected CRC) colonoscopy/CTCs
- number/proportion of non-urgent colonoscopy/CTCs
- time to colonoscopy/CTC
- time to diagnosis of CRC or other conditions
- number/proportion of colonoscopy/CTCs that do not detect CRC
- number/proportion of colonoscopy/CTCs that do not detect significant bowel pathology
- number/proportion of people presenting to emergency departments with symptoms of CRC.

For the review of diagnostic test accuracy studies, the following outcomes were eligible for inclusion and extraction:

- Number of true positives, true negatives, false positives and false negatives, only where all four statistics were reported or could be calculated for CRC, or for IBD or AAs. IBD and AA data were extracted only from studies that also reported CRC diagnostic test accuracy data.
- Other outcomes as listed for the clinical efficacy studies.

Where an outcome was not identified by the review, and was required by the model, these were subsequently reviewed in the searches for modelling parameters (see *Evidence sources used to inform the model parameters*).

Study design

For the review of end-to-end clinical efficacy studies, RCTs or non-RCTs were eligible for inclusion.

For the review of diagnostic test accuracy and comparative diagnostic test accuracy, only cohort or cross-sectional studies that recruited patients regardless of eventual diagnosis were eligible for inclusion (i.e. studies that avoided a case–control design).

Studies not published in the English language were eligible for inclusion if sufficient data could be extracted from non-English-language full texts, or from an existing English-language abstract. Conference abstracts and non-peer-reviewed

reports were eligible if the data were presented in a succinct and accessible manner (e.g. a manuscript prepared for submission to a journal), if sufficient methodological details were reported to allow critical appraisal of the study quality, and if results were reported in sufficient detail. Where there were gaps in the available literature, exclusion criteria for conference abstracts and non-English-language papers could be relaxed.

Search strategy

A systematic literature review was undertaken to identify evidence on the intervention (FIT assays) and target condition (CRC), following the guidelines developed by the Centre for Reviews and Dissemination[28] for reviews in health care and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.[29]

Searches were run in December 2022 based on those conducted for the ACPGBI/BSG review (March 2022), which was in turn based on the searches for DG30 (March 2016). Facets of the searches were limited to either 2022 onwards or 2016 onwards, depending on whether the ACPGBI/BSG (for which searches were done in 2022) or DG30 review (for which searches were done in 2016) had searched that facet. Search strategies used subject headings and free-text terms including both generic and product names for the interventions and were optimised for each database. No language restrictions were applied. The search strings are reproduced in full in *Appendix 1*.

Databases searched included:

- MEDLINE-ALL (via Ovid), including Epub Ahead of Print, In-Process Citations and Daily Update
- EMBASE (via Ovid)
- Cochrane Database of Systematic Reviews (CDSR) (via Wiley)
- Cochrane Central Register of Controlled Trials (CENTRAL) (via Wiley).

The following additional sources were searched to identify relevant HTA reports, ongoing reviews and clinical trials (respectively):

- INAHTA (searched 13 December 2022)
- NIHR HTA programme website (searched 13 December 2022)
- PROSPERO (searched 13 December 2022)
- ClinicalTrials.gov (searched 13 December 2022)
- EU Trials Register (searched 13 December 2022)
- WHO ICTRP (searched 13 December 2022).

Retrieved records from all searches were downloaded into EndNote for deduplication and eligibility screening. Reference lists in included articles and relevant systematic reviews were checked for additional studies. Clinical experts were consulted to ensure that no relevant studies had been missed.

Study selection

Studies were selected for inclusion in the review if they met the inclusion criteria detailed in *Population* to *Study design*. Titles and abstracts were considered for inclusion against the criteria by one reviewer, with a minimum 10% sample checked by a second reviewer. This was conducted in increments of 100 until 100% sensitivity was achieved, and before the remainder were screened, to train both reviewers in implementing the criteria. Sensitivity of 100% was achieved (all relevant studies were identified by both reviewers) during the second batch of 100 records, although specificity was somewhat lower for both reviewers (both included some additional irrelevant titles), which was dealt with during the full-text sift. Full texts were obtained and considered for inclusion by one reviewer, with decisions checked by a second reviewer. Any discrepancies were resolved through discussion.

Where multiple publications relating to the same study were identified, only those with relevant outcome data not published in the others were included. Where there was a crossover in the locations and dates of recruitment between two or more studies, the largest was included, unless the other publication(s) reported more thresholds or was a better match for the patient populations of interest (see *Study categorisation*), in which case a decision on which to include was based on a consideration of all these factors.

Data extraction strategy

The data extraction form used by ACPGBI/BSG guideline group was used as a basis for a de novo data extraction form, which was piloted on three studies and adapted as necessary. Several fields were added including fields relating to study population type (see *Study categorisation*) and study and patient characteristics (see *Populations and relevant subgroups*). Study recruitment dates and locations were extracted to aid an assessment of 'crossover' with other studies to avoid double counting of patients. Data relating to diagnostic test accuracy were extracted as the absolute numbers of true positives, true negatives, false positives and false negatives, where available, or as sensitivity and specificity that was later transformed into true positives, true negatives, and so on, as described in *Appendix 2*. The final list of fields extracted included first author and date; year of recruitment; location; study name; inclusion and exclusion criteria; population characteristics (age, sex, medications that increase GI bleeding, blood disorders); test–analyser combination; index test methods; reference standard; *N* recruited; *N* missing from analysis; *N* analysed; outcome (CRC, AA or IBD); *N* with outcome; threshold; diagnostic accuracy metrics; and any additional outcomes as described in *Outcomes*.

Data included in the ACPGBI/BSG data extraction form were checked by an EAG reviewer against the original publication and checked for completeness against the inclusion criteria for the review for this assessment (e.g. additional thresholds or subgroups). Additional data were extracted where necessary and checked by a second reviewer. Disagreements were resolved through discussion. Authors were contacted to provide missing data or resolve data ambiguities where of key importance to the review.

Data extractions for a small number of studies were not checked by a second reviewer due to time constraints. These include an update to the Nottingham study[30] (see *Main analysis: OC-Sensor*), which was received shortly before the report deadline, and studies included in the reviews of patient preferences and the impact of socioeconomic factors (see *Patient perspectives* and *Sociodemographic factors*).[30–33]

Quality assessment strategy

Quality assessment of diagnostic test accuracy studies version 2 (QUADAS-2)[34] was used to assess the quality of the included studies. The scoring scheme is provided in *Appendix 3*. Scores were assigned by one reviewer and checked by a second, with disagreements resolved through discussion. For the review of comparative diagnostic test accuracy (n = 3 included studies), quality assessment using QUADAS-C[35] was planned but was not completed due to time constraints.

Synthesis strategy

Narrative synthesis methods

Study and patient characteristics were summarised narratively for all the main analyses. Where there were insufficient data for a statistical synthesis, outcomes were synthesised narratively. Where a statistical synthesis was performed, a narrative synthesis of outcomes was not provided in the interest of brevity.

Methods for the meta-analysis of diagnostic test accuracy

Diagnostic accuracy was considered separately for each FIT assay type. For tests where data were available from more than one study, pooled estimates of diagnostic parameters were estimated using the modelling approach described in Jones *et al.*[36] The model accommodates estimates of sensitivity and specificity at more than one explicit diagnostic threshold per study. Pooled estimates are produced at all possible thresholds, even where data for a given threshold have not been reported by an empirical study included in the review. Selected thresholds, based on clinical opinion about the most clinically relevant, are presented in this report. The model is summarised in *Appendix 2* and full details are provided in the original publication.

A random-effects meta-analysis was used to account for the heterogeneity between studies that is generally expected in diagnostic accuracy studies. Reasons for the heterogeneity in sensitivity and specificity between studies according to study population type (described in *Population types among included studies*) and reference standard received (see *Reference standards*) were explored using subgroup analyses.

Summary sensitivity and specificity for each test/fitted model were evaluated based on the mean values of the four sets of study-level random effects (mu_1, mu_2, $m\sigma_1$, $m\sigma_2$). As described in Jones *et al.*,[36] the summary sensitivity and specificity at any threshold value, Ct, can be calculated as:

$$\text{logit}(1 - \text{Specif icity}(C_t)) = \frac{(m_{u1} - \log_e(C_t))}{\exp(m_{\sigma 1})},$$
$$\text{logit}(\text{Sensitivity}(C_t)) = \frac{(m_{u2} - \log_e(C_t))}{\exp(m_{\sigma 2})}. \quad (1)$$

Summary sensitivity and specificity were evaluated for thresholds ranging from 2 (the smallest threshold evaluated in the included studies) to 401 (the largest reported threshold).

Results are displayed as receiver operating characteristic (ROC) plots with summary ROC curves of sensitivity versus 1 – specificity. Sensitivity and specificity are also plotted individually against threshold with 95% credible intervals (CrIs) for the summary estimates illustrating the range of likely values for average diagnostic accuracy of the synthesised studies. 95% prediction intervals (PrIs) are also shown, illustrating the between-study heterogeneity and providing a range of values that might be expected in a future study.

Summary sensitivity and specificity are plotted for the full range thresholds (2–401) for all FIT types. Numerical results and 95% CrIs are presented in tables for selected thresholds, only where the selected thresholds are within the range of values evaluated in the reported studies, to avoid extrapolating beyond the observed data.

Analyses were conducted in R (The R Foundation for Statistical Computing, Vienna, Austria)[37] using the JAGS Markov chain Monte Carlo sampler and the RJAGS interface package.[38] Convergence with the target posterior distributions was assessed using the Gelman–Rubin statistic[39] for three chains with different initial values. For all analyses, a burn-in of 50,000 iterations of the Markov chain was used, with a further 30,000 iterations retained to estimate parameters after thinning by retaining every 10th sample. Model fit penalising for complexity was compared using the deviance information criterion (DIC).[40] Models with lower values of the DIC are preferred. Model fit for all presented analyses is provided in *Appendix 2*.

The analysis plan and rationale

The analysis plan was formulated in response to the available data, following the principles set out in the EAG's protocol and taking into consideration the issues outlined in *Impact of specific symptoms on faecal immunochemical test sensitivity and specificity* to *Reference standards*.

Rationale for the analysis plan

Impact of specific symptoms on faecal immunochemical test sensitivity and specificity

The EAG heard from clinical advisors that FIT is a better predictor of CRC risk than symptoms alone, and that the sensitivity and specificity of FIT may not differ according to the symptoms reported at presentation. The ACPGBI/BSG review[16] showed that the sensitivity was similar in studies recruiting NG12 high-/medium-risk patients to that in studies recruiting DG30 low-risk patients (88.7%, 95% CI 84.4% to 92.0%, and 88.7%, 95% CI 78.1% to 95.3%, respectively), but that the specificity was numerically different (78.5%, 95% CI 73.0% to 83.2%, and 88.5%, 95% CI 87.1% to 89.9%, respectively). As specificity affects estimates of cost-effectiveness, the EAG decided to subgroup studies according to population type to allow exploration of any potential difference.

Population types among included studies

Missing patients

The population for this appraisal was all patients presenting to primary care with signs and symptoms suggestive of CRC, as listed in NG12 and DG30. A number of studies were encountered that included both NG12 high-/medium-risk and DG30 low-risk patients, but only those who reached secondary care (e.g. recruited all on the 2WW). Such studies

will be likely to include nearly all NG12 high-/medium-risk patients, as all of these patients should be referred to secondary care as per the pathways outlined in *Current service provision* and *Figure 1*, but will likely exclude a proportion of DG30 low-risk patients who are not referred and stay in primary care. If the assumption that symptoms do not have an impact on FIT sensitivity and specificity is incorrect (see *Impact of specific symptoms on faecal immunochemical test sensitivity and specificity*), it would be important to avoid excluding patients who did not make it to secondary care as this would alter the patient spectrum and may bias the estimates of diagnostic test accuracy.

Enrichment with faecal immunochemical test positives

In addition, studies that recruited patients who had reached the 2WW may well include a proportion of DG30 low-risk patients who were referred on the basis of a positive FIT result received in primary care before referral (if the region's GPs were using FIT to guide referral in accordance with DG30). As DG30 low-risk FIT-positive patients (both true positive and false positives) are usually referred, and DG30 low-risk FIT-negative patients (both false negatives and true negative) are usually not, the patient spectrum will be enriched with DG30 low-risk FIT-positive patients while excluding most DG30 low-risk FIT-negative patients. The exclusion of DG30 low-risk patients whose first FIT was negative is likely to impact on both sensitivity and specificity and likely to result in an overestimation of sensitivity (because disproportionately fewer false negatives are included) and an underestimation of specificity (because disproportionately fewer true negatives are included). A worked example is provided in *Report Supplementary Material 1* to demonstrate this issue. The extent of this bias will depend on the numbers affected by the referral practice and is not known.

Economic model requirements

It was also useful to the model if diagnostic test accuracy data were available for NG12 high-/medium-risk and DG30 low-risk patients separately for the following reasons:

- If test accuracy differs according to population, estimates for diagnostic accuracy in the comparator arm would need to come from studies that recruited DG30 patients.
- Estimates of the prevalence of CRC in DG30 patients would also be required by the model, as would prevalence for the whole population presenting to primary care (i.e. DG30 + NG12 patients).

Reference standards

Comparators discusses the relative merits of the different reference standards encountered in this review. In summary, all reference standards have limitations, and the restriction to only studies using > 90% colonoscopy or CTC would result in the exclusion of most studies that recruited a spectrum of patients closest to being representative of the target population (all patients in primary care) and a greater dependence on studies that may be enriched with FIT-positive patients and excludes some of the primary care patients. The worked example in *Report Supplementary Material 1* considers the impact of an imperfect refence standard on estimates of diagnostic test accuracy.

The analysis plan

Study categorisation

For the reasons given in *Population types among included studies*, the studies have been broadly categorised as follows:

- **Population type 1:** Studies closest to being a representative spectrum of all patients presenting to primary care with symptoms of CRC who meet NG12 or DG30 criteria (minus bypass symptoms). This was for studies that recruited the full spectrum of patients, or those with some minor differences in recruitment criteria (wider or narrower than NG12 and DG30), and where a prior FIT result did not influence recruitment.
- **Population type 2:** Studies closest to being a representative spectrum of NG12 high-/medium-risk patients. This was for studies that recruited NG12 high-/medium-risk patients (minus bypass symptoms). These were often studies that had recruited patients in secondary care who were referred to the 2WW (i.e. population type 4 studies; see below) and had reported a subgroup specifically of NG12 high-/medium-risk patients. Because all or nearly all NG12 high/medium-risk patients should be referred to secondary care, studies recruiting in secondary care should recruit most NG12 high-/medium-risk patients.

- **Population type 3:** Studies closest to being a representative spectrum of DG30 low-risk patients. This was for studies that recruited a representative spectrum of DG30 low-risk patients. These were likely to have been recruited in primary care, for example in areas using FIT in accordance with DG30.
- **Population type 4:** Unclear/likely unrepresentative spectrum. This was for studies that were not population type 1, 2 or 3 studies, or where it was not clear what criteria were used to select patients either for FIT testing or for referral or both. In particular, this included studies that recruited patients in secondary care who were referred to the 2WW, which is likely to be a mix of NG12 high-/medium-risk, DG30 low-risk FIT positives (if implemented in primary care at the time of recruitment), and others that GPs have concerns about. It also included studies from countries that did not use NG12 or DG30 and did not state what their criteria were as it would be unclear how representative such a spectrum would be.
  - We note that it could be assumed that studies recruiting patients in the 2WW are likely to be predominantly NG12 high/medium risk, but we also expect these studies to be enriched with FIT-positive patients, as described above. Equally, studies in other countries recruiting patients who have been referred to secondary care are likely to be similar to NG12 high-/medium-risk patients, but again the similarity is unknown.

Categorising studies according to population type was done by the systematic review team, and was at times difficult to judge. Clinical advisors to the EAG were contacted in cases of doubt, and authors were contacted for more detail where there was uncertainty. This did not always lead to complete clarity, partly because it was difficult for the authors to tell how well GPs adhered to guidelines about who to give FIT to in primary care and/or who to refer to secondary care. In cases where uncertainty in referral criteria was unresolved, these studies were placed in population type 4.

For each analysis, careful consideration was given to the location and dates of recruitment as described in *Study selection*.

Main, subgroup and sensitivity analyses

As a conservative approach, the EAG considered restricting the analysis to population type 1 studies as the main analysis, but, as can be seen in *Table 2*, very few studies recruited a population wide enough to be considered 'all patients', and even among these the population was often wider or narrower in some way, especially with respect to bypass symptoms (rectal/anal mass or anal ulceration). The EAG therefore included all study types and explored the impact of each through a series of sensitivity analyses.

The following analyses were conducted where > 1 study was available for analysis.

Main analysis: diagnostic test accuracy for colorectal cancer for each test individually

- Each test analysed separately, including all study population types 1–4 together.
  - Sensitivity analysis with type 4 studies removed as these may be enriched with FIT positives, and under the assumption that specific symptoms do not alter FIT sensitivity and specificity.
  - Subgroup analysis according to study population type, under the assumption that FIT sensitivity and specificity are affected by specific symptoms:
    - study population type 1 (all presenting to primary care)
    - study population type 2 (NG12 high/medium risk)
    - study population type 3 (DG30 low risk).

Additional analysis 1: diagnostic test accuracy for colorectal cancer for all tests together

This analysis was run to allow the investigation of the impact of study population type and reference standards on a larger sample of studies and because these factors were unlikely to interact with test type. It was also used to inform the priors used when < 5 studies were being synthesised (see *Appendix 2*, *Statistical methods for the evidence synthesis*).

- All tests analysed together, including all study population types 1–4 together.
  - Sensitivity analysis removing type 4 studies.
  - Subgroup analysis according to:
    - study population type 1 (all presenting to primary care)
    - study population type 2 (NG12 high/medium risk)
    - study population type 3 (DG30 low risk).

This set of studies would also provide estimates of prevalence for each of population types 1–3 for the economic model, as prevalence should not be affected by test type. Similar analyses were planned for diagnostic test accuracy for AA and IBD separately, and undertaken where data allowed.

Additional analysis 2: impact of reference standard on diagnostic test accuracy estimates

A sensitivity analysis was undertaken which restricted to studies in which > 90% of patients received colonoscopy or CTC as the reference standard, in order to investigate the effect of the reference standard on estimates of diagnostic test accuracy. This was carried out for all tests together and for tests separately where data allowed.

Results

The report is structured as follows. The discussion in *Discussion and conclusions* provides an overview of the evidence base along with a discussion of limitations and a comparison with other recent reviews. This, along with *Main analysis: summary*, which summarises test accuracy for the tests at selected thresholds, serves as a good summary of the evidence base.

The main analyses for each test are then provided separately in *Main analysis: HM-JACKarc* to *Main analysis: IDK tests*. Dual faecal immunochemical test studies are reported in *Main analysis: dual FIT*. Additional analysis 1 (all tests together, and subgrouped by population type) is described in *Additional analysis 1: synthesis of all tests together in a single analysis*, and additional analysis 2 (sensitivity analysis for the reference standard) in *Additional analysis 2: reference standard sensitivity analysis*. A summary of the main analyses is provided in *Main analysis: summary*. A summary of comparative diagnostic test accuracy studies is provided in *Comparative diagnostic test accuracy studies*, and the data entering these analyses in *Report Supplementary Material 2*. Sections, each with an associated appendix, are provided relating to subgroup analyses according to patient characteristics (see *Subgroup analyses by patient characteristics*) and for studies

TABLE 2 Summary of studies entering the analysis, by test and study population

Test	Main analysis	Population type				Patient characteristics subgroups	Dual FIT	Any analysis
		1: all patients	2: NG12 high/ medium risk	3: DG30 low risk	4: unclear/ unrepresentative			
HM-JACKarc	16	5	4	2	8	Anaemia, sex, age, medications	2	18
OC-Sensor	11	3	1	1	7	Anaemia, sex, medications	1	17
FOB-Gold	3	0	1	0	2	0	0	3
QuikRead go	1	0	1	0	0	0	1	2
NS-Prime	1	0	1	0	0	0	0	1
IDK TurbiFIT	0	0	0	0	0	0	0	0
IDK Hb, Hb/ Hp complex	1	0	0	0	1	0	0	1

reporting the AA and IBD studies (see *Advanced adenomas and inflammatory bowel disease outcomes*). Appendices are also dedicated to studies reporting non-diagnostic test accuracy data including test failures, uptake and repeat tests (see *Test uptake and repeat tests*) time to diagnosis and other outcomes (see *'Time to' and other outcomes*), patient acceptability (see *Patient perspectives*) and sociodemographic factors (see *Sociodemographic factors*).

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for the selection of studies is provided in *Figure 4*. A total of 1874 records were retrieved by the database and registry searches, of which 1774 were excluded on the basis of their title and/or abstract. The full text of the remaining 100 records were retrieved and assessed for eligibility against the study selection criteria. Thirty records[18,27,32,33,41–66] were included in the review. A further 184 records were identified through other sources including nominations by experts or stakeholders (n = 5), screening of studies included in other reviews (n = 137), company submissions (n = 40 not already identified by SCHARR searches) and the checking of references in other included studies (n = 2). Two of the 40 studies submitted by companies were received after the first committee meeting, in response to the appraisal consultation document produced by NICE.[67] From these other sources, 19 publications[17,30,31,68–83] were included. In total, 49 publications were included in the review. The records excluded on the basis of their full text are listed in *Appendix 4*, along with reasons for their exclusion.

No end-to-end studies were identified. Among the 49 included publications, 16 studies reported across 21 publications[17,18,27,46–48,50,52,56,60,61,63,65,66,71–75,81,82] reported diagnostic test accuracy data for HM-JACKarc, 17 studies reported across 18 publications[30,41–46,49,51,54,55,57,62,64,68,70,76–78] reported diagnostic test accuracy data for OC-Sensor, three studies reported across three publications[44,59,83] reported diagnostic test accuracy data for FOB-Gold, two studies[58,80] reported diagnostic test accuracy data on QuikRead go, one study[44] reported data for NS-Prime, no studies reported data for IDK TurbiFIT or IDH Hb+Hb/Hp and one study[79] reported data for IDK Hb and Hb/Hp complex separately. Diagnostic test accuracy data relating to patients subgrouped according to the patient characteristics listed in *Populations and relevant subgroups* were identified in 17 publications,[27,43,52,54,65,66,68,70–73,75–78,82,84] one of which was outside the inclusion criteria for the review but was included in a subgroup due to the small number of data.[84] Data on dual FIT were identified in four publications[51,75,80,81] for HM-JACKarc, OC-Sensor and QuikRead go only. Other outcomes such as test uptake, failed and repeat tests, time to diagnosis, patient preference and sociodemographic factors were reported across 30 publications.[18,42–46,48–62,65,68,69,72,75,80–83]

It should be noted that across the evidence base, it was often unclear whether patients with bypass symptoms (rectal or anal mass or anal ulceration) were excluded. Equally, a number of studies excluded patients with rectal bleeding, which may affect the patient spectrum. These issues are not dealt with in detail because of the poor level of reporting on these factors but should be noted as a potential limitation of the evidence base.

Main analysis: HM-JACKarc

No end-to-end studies were identified. Seventeen studies reported across 21 publications[17,18,27,46–48,50,52,56,60,61,63,65,66,71–75,81,82] reported diagnostic test accuracy data for HM-JACKarc (*Table 3*). Studies with multiple publications include the NICE FIT study[17,72,73] and a study from Tayside, UK, with two publications.[60,61] Three publications from Oxford[27,63,66] comprise a series of different data cuts from a single registry analysis (CSS-BIO-3 4730). These have been counted as two separate studies: one study that recruited January to March 2016,[63] and one study reported over two publications with different but overlapping recruitment dates (recruitment dates March 2017 to March 2020[27] and March 2017 to December 2020[66]), but because both also report unique analyses, both publications were included in the review.

Sixteen studies (17 publications)[18,27,46–48,50,52,56,60,61,63,65,72,74,75,81,82] contributed to the main analysis. Patient characteristic subgroup data (see *Populations and relevant subgroups*) were reported by eight studies,[27,52,65,66,71,73,75,82] one of which was a study not included in the main analysis because it did not report recruitment dates (so double counting of patients could not be ascertained)[71] and two of which[66,73] were from studies included in the main analysis (NICE FIT and the Oxford cohort), but the subgroup analysis was reported in a separate publication. Subgroup data are discussed in *Subgroup analyses by patient characteristics* and its associated appendix. Two studies[75,81] reported data on dual FIT, one[75] reported data for both single and dual FIT, and one[81] reported data for dual FIT only (see *Main analysis: dual faecal immunochemical test*). One further study reported data for repeat FIT in Scotland (see *Patient perspectives* and its

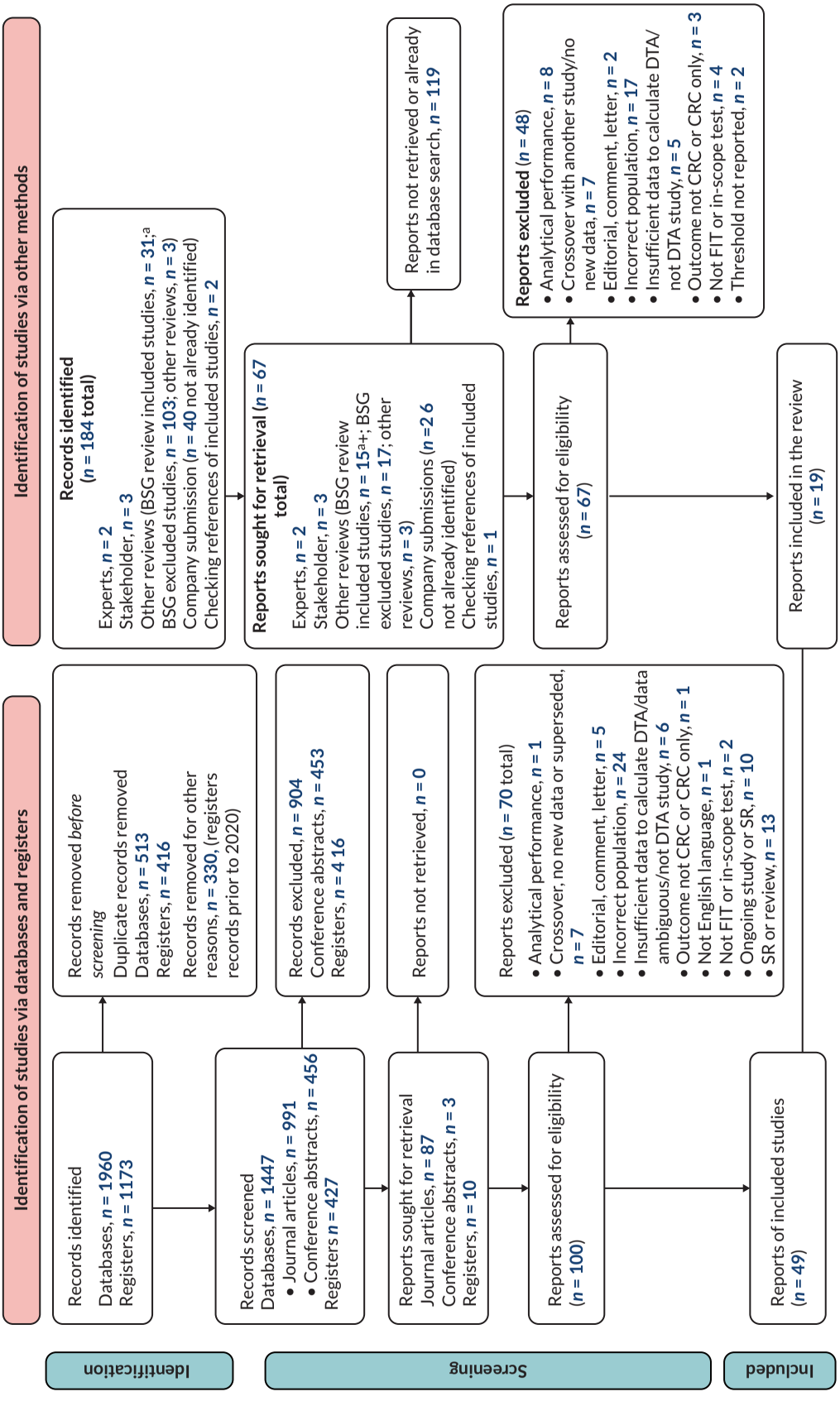

FIGURE 4 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of study selection.[85] a, *N* = 16/31 records included in the BSG review were captured by database searches and were therefore not sought for retrieval. Reproduced and adapted in accordance with Creative Commons Attribution (CC BY 4.0) license.

associated appendix),[53] and another reported a comparison of several different FIT[44] (see *Comparative diagnostic test accuracy studies*) and was conducted in a subset of the NICE FIT study.[17,72,73] One of the NICE FIT publications[17] has been included as it reports AA and IBD data, but it has not been used in the analyses relating to CRC as these data are reported in the other NICE FIT publication.[72]

Main analysis

Across the 16 studies (17 publications)[18,27,46–48,50,52,56,60,61,63,65,72,74,75,81,82] included in the main analysis, thresholds ranged from 2[17,18,60,61,72,82] to 401.[47] Among these, an NG12 high-/medium-risk subgroup was included from two population type 4 studies[72,74] as these were likely to be a representative spectrum of NG12 high/medium patients, and to ensure that the sample was not enriched with FIT-positive patients who had received FIT in primary care (note that the NICE FIT study did not include patients who were FIT positive in primary care) and/or because the additional patients were not a full and exclusive spectrum of DG30 low-risk patients. The same was not done to the one type 4 study[27] that also reported a population type 3 subgroup analysis[66] because the study was not enriched with FIT positives. All studies were in the UK; five (six publications)[50,52,56,60,61,75] were in Scotland and one was in Wales,[65] with the remainder in England. Sample sizes ranged from 175[48] to 9896[27] and the prevalence of CRC ranged from 1.06%[27] to 6.36%.[74] Patient characteristics (sex, ethnicity, blood disorders, medications, anaemia) were rarely or never reported. Age was usually reported as a median, which ranged from 58[63] to 72[47] years among studies of types 1–4. The proportion who were male ranged from 41.4%[53] to 50%.[81] The reference standard was records follow-up in five studies (six publications),[27,52,56,60,61,63] secondary care follow-up comprising various imaging tests in four[47,48,81,82] and colonoscopy or CTC in the remaining seven.[18,50,56,65,72,74,75]

Population type 1 studies

Five studies (six publications)[18,52,56,60,61,75] were included in this category. Four studies (five publications)[52,56,60,61,75] were from Scotland, where the use of FIT has been encouraged in a wider group of patients than in England, encompassing both NG12 high-/medium-risk and DG30 low-risk patients with some differences (see footnotes to *Table 3* for Gerrard 2023). CRC prevalence among these ranged from 1.29%[56] to 3.05%,[75] possibly indicating heterogeneity in the criteria used to select patients for FIT across these studies, or in how well GPs adhered to guidelines. One further study[18] was conducted in London at a time when all NG12 and DG30 patients were referred to secondary care by GPs without the use of FIT. The prevalence of CRC in this study is higher than in the Scottish studies, at 4.03%. The EAG notes that it is likely that not all DG30 patients were referred as GPs would use judgement when making referrals. It reports patients subgrouped by NG12 high/medium risk and DG30 low risk and therefore contributed to three population type subgroup analyses (type 1, type 2, type 3).

Population type 2 studies

Four studies[18,73,74,81] were considered by the EAG to be population type 2 studies because they recruited or reported patients referred to secondary care who met the NG12 high-/medium-risk referral criteria. Two of these are subgroups of studies that recruited all patients who were referred to the 2WW (a study from Croydon and the NICE FIT study).[18,72] All four studies are likely to recruit a fairly representative spectrum of patients meeting NG12 criteria who present to primary care, although where additional criteria, such as a requirement to have undergone a colonoscopy,[18,72] were also used to select patients (see *Table 3*, column 4), some patients may have been systematically excluded (e.g. elderly patients). The studies reported thresholds ranging from 2[18,72] to 150,[72] had sample sizes ranging from 160[18] to 7194[72] and had CRC prevalence from 3.57%[72] to 6.36%.[74]

All four had a reference standard that comprised full colonic imaging in secondary care. Two studies[18,72] did not report whether patients presenting with rectal/anal masses or anal ulceration were included, while two others[74,81] reported small proportions with these symptoms. No data relating to subgroups were reported.

Population type 3 studies

Two studies[18,66] were considered by the EAG to be population type 3 studies because they reported a subgroup of DG30 patients. Both were subgroups of larger studies.[18,27] The studies report thresholds of 2 and 10 µg/g, had sample sizes of 138[18] and 166,[66] and had CRC prevalence of 1.45%[18] and 0.84%.[66] The reference standard was colonoscopy in one and records follow-up in another. Neither study reported whether patients presenting with rectal/anal masses or anal ulceration were included, although as these symptoms are not DG30 criteria, it could be assumed that such

patients were not recruited. One study[66] reports multiple patient characteristic subgroups (see *Table 3*, final column), but for the wider population (type 4).

Population type 4 studies

Eight studies[27,46–48,50,63,65,82] were categorised as population type 4 studies. Four studies[46–48,82] included patients referred to secondary care on the 2WW pathway, which may mean that the sample is enriched with patients who received FIT in primary care and had a positive result compared with samples not selected on the basis of a positive FIT, for example NG12 high risk. Two studies, one from Scotland in 2013[50] and one from Wales in 2020,[65] included patients referred to secondary care using unclear criteria. The two Oxford studies[27,63] recruited patients given FIT in primary care and are likely to have recruited populations closer to DG30 low-risk type 2 studies, but both included some patients outside these criteria. In at least one of these, some of the additional patients had symptoms likely to indicate lower risk of CRC, including inflammation, thrombocytosis and being tired all the time.

The reference standard was imaging including colonoscopy in all six studies that recruited patients in secondary care,[46–48,50,65] but it was not always clear how many had colonoscopy or CTC. The two studies of patients receiving FIT in primary care[27,63] used records follow-up as a reference standard. Two studies[65,82] reported some patients presenting with rectal/anal masses or anal ulceration and one reported 0% of such patients,[27] while the remaining studies were unclear for some or all of these criteria. Three studies[27,65,82] reported patient characteristic subgroups (see *Table 3*, final column).

One study[53] reported data relating to repeat FIT. Data were collected from three regions in Scotland (Tayside, Greater Glasgow and Clyde, and Highlands). Patients who returned two FIT more than 1 week apart, but within 1 year apart, were analysed, but it was unclear what criteria were used to select patients for FIT. The threshold was 10 µg/g, and records follow-up was the reference standard. The prevalence of CRC was low in this group (0.73%).

Other studies

The other studies reporting patient characteristics subgroup data and dual FIT data are reported in *Main analysis: dual faecal immunochemical test* and *Subgroup analyses by patient characteristics* and their associated appendices.

Quality assessment

Quality assessment of diagnostic test accuracy studies version 2[34] was applied to studies reporting diagnostic test accuracy data only. QUADAS-2 asks questions relating to internal validity (risk of bias) and external validity (applicability). Each is discussed separately in the following two paragraphs.

Risk of bias

Appendix 3, *Table 37*, summarises the risk of bias and applicability scores as assessed by the authors of this review. Full risk-of-bias scores with reasons for the scores are provided in *Appendix 3*, *Reasons for scores*, including *Tables 43–48*. For risk of bias, no study scored as low risk for all items, and no item scored as low risk for all studies. The index test scored low risk most often, with only two studies[81,82] scoring high risk because some or all of the reported thresholds were derived to optimise accuracy. Where patient selection was at risk of bias it was because a consecutive sample was not recruited and/or because inappropriate exclusions were made, such as excluding people on the basis of not having had a colonoscopy or not having all blood test results available. Owing to the inclusion criteria for the review, all studies avoided a case–control design. The reference standard was rated as being at unclear or high risk of bias for all studies. This was usually due to not all patients receiving a colonoscopy or CTC, or due to it being unclear if the reference standard had been interpreted blind to the index test. Patient flow scored as high risk or unclear in nearly all studies. This was due to a mix of factors, including a lack of clarity about the interval between the index test and the reference standard in nearly all studies, patients receiving different reference standards depending on their FIT result or other factors, and patients being missing from the study.

Applicability

There were concerns about the representativeness of the patients recruited to the studies compared with ‘all those presenting to primary care’ in nearly all studies due to either the exclusion of some patients (study population types 2, 3 and 4) or a lack of clarity about who was included in comparison with the target population. The index test was rated as being at low risk of having poor applicability, except in two cases[27,63] where a few patients had two index tests and if either scored positive this was counted as a positive test, and two studies[81,82] that were rated as being at high

TABLE 3 Study and patient characteristics of HM-JACKarc studies

Study number	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with NICE scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds µg/g	Subgroup data?
Population type 1 studies (all patient presenting to primary care with symptoms meeting NG12 high/medium or DG30 low risk)									
1	D'Souza 2020[18] Croydon, UK November 2016 to October 2017	HM JACKarc analytical system Colonoscopy	All NG12 and DG30 – all symptomatic patients were referred to colonoscopy in this period in this area of London	NR	Mean 60.6 (range 20–90)	• 48.6% • Ethnicity reported[a] • NR	12/298 (4.03%)	2, 10	None
2	Gerrard 2023[75] Lothian, Scotland, UK January 2019 to February 2020	HM-JACKarc Endoscopy or CT with colorectal protocol	Urgent suspected of cancer referrals, criteria for referral[a] are both wider and narrower than NG12 high/medium and DG30 low risk	Wider and narrower than target population[a] Abdominal mass: 3.0% Rectal mass: 2.4%	Median 65 (IQR 56–74)	• 44.3% • NR • 17.8%	135/3426 (3.05%)	10	Anaemia, no anaemia
3	Johnstone 2022[52] Greater Glasgow and Clyde, UK August 2018 to January 2019	HM-JACKarc (personal communication) Records follow-up	All with NG12 high/ medium or DG30 low risk would get FIT (confirmed by author via personal communication)	May be wider Abdominal mass 2.5% Rectal mass 0.9%	Median 59 (range 16–97), *n* = 4968	• 42.3% • NR • IDA 5.4%;[a] anaemia 20.0%	61/4737 (1.29%)	10, 150, 400	Anaemia, no anaemia
4	MacDonald 2022[56] NHS Lanarkshire, UK October 2016 to February 2019	HM-JACKarc Records follow-up	Symptomatic colorectal referrals from primary care, under SIGN 126 and Scottish Referral Guidelines that encompass both NG high risk and DG30 low risk for referral	Includes anorectal or abdominal mass; also includes referrals based on imaging, but from GP care	Median 62 years (range 16–96 years)	• 45.7% • NR • NR	151/5250 (2.88%)	10	None
5	Mowat 2021[61] and 2019[60] NHS Tayside, UK December 2015 to December 2016	HM JACKarc Records follow-up	GPs encouraged to use FIT in patients regardless of the specific lower GI symptoms and perceived risk	NR	Median 65 (range 2–99, IQR 51–75)[60]	• 43.6%[60] • NR • NR	105/5381 (1.95%)	2, 7, 10, 20, 50, 100, 150, 200, 250, 300, 350, 400	None
Population type 2 studies (NG12 high risk)									
1	D'Souza 2020[18] Croydon, UK November 2016 to October 2017	HM JACKarc analytical system Colonoscopy	NG12 high/medium-risk (subgroup of main Croydon study) who underwent colonoscopy	NR	NR for subgroup		8/160 (5.00%)	2, 10	None

TABLE 3 Study and patient characteristics of HM-JACKarc studies (*continued*)

Study number	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with NICE scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds µg/g	Subgroup data?
6	D'Souza 2021[72] D'Souza 2021[17b] NICE FIT October 2017 to December 2019	HM JACKarc analytical system Colonoscopy	Subgroup: NG12 high/ medium risk Full study: 2WW patients (including NG12, DG30, others) who underwent colonoscopy	NR	NG12 high/ medium-risk: mean 65.9 (SD 11.1) Full study: 64.0 (SD 11.9)	NG12 high/medium risk: • 45.7% • Ethnicity reported[a] • IDA 4.2% Full study: • 45.1 • Ethnicity reported[a] • NR	NG12 high risk: 257/7194 (3.57%) Full study: 421/9822 (4.29%)	2, 10, 150	None
7	Farrugia 2020[74] University Hospitals Coventry and Warwickshire NHS Trust, UK January 2015 to March 2017	HM JACKarc automated system Colonoscopy or CT colonography and histology results	NG12 high/med risk[a]	Abdominal/rectal mass *n* = 10	68.6 (error/ range NR)	• 48.9% • NR • Anaemia, including iron deficiency 18.1%	10/519 (6.36%)	10	None
8	Turvill 2018[81] York Hospital, UK February 2016 to March 2017	HM-JACKarc Full colonoscopy or CT colonography or a lesser investigation (such as CT abdomen/pelvis with contrast plus flexible sigmoidoscopy) limited by the identification of pathology	NG12 high/medium risk	4% abdominal mass and 1% rectal mass	Median 69 (IQR 61–76)	• 50% • NR • 18% IDA[a]	27/505 (5.35%)	12	None
Population type 3 (DG30 low risk)									
1	D'Souza 2020[18] Croydon, UK November 2016 to October 2017	HM JACKarc analytical system Colonoscopy	DG30 low risk (subgroup of main Croydon study) who underwent colonoscopy	NR for subgroup	NR for subgroup		2/138 (1.45%)	2, 10	

continued

TABLE 3 Study and patient characteristics of HM-JACKarc studies (*continued*)

Study number	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with NICE scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds µg/g	Subgroup data?
9	Withrow 2022[66] (same study as Nicholson 2020)[27a] Oxfordshire, UK March 2017 to 21 December 2020 CSS-BIO-3 4730	HM JACKarc Records follow-up	Type 3 subgroup from type 4 study – FIT given in primary care for any reason, wider than DG30 low risk alone	NR	Median 61 (IQR 51 to 75)[a]	• 42% • NR • Any anaemia: 26%; IDA: 11%[a]	139/16604 (0.84%)	2, 10	DG30 only subgroup; various anaemia thresholds (men/ women separately); men; women; age </> 40, > 50, > 60, > 70, > 80
Population type 4 (unclear/unrepresentative of all presenting to primary care)									
10	Chapman 2021[46] Nottingham University Hospitals Trust, UK September 2016 to September 2017	HM JACKarc + HM JACKarc analyser 2WW investigations	2WW patients who returned two types of FIT	NR	Median 71.1 (IQR 62.5–78.7)		38/732 (5.19%)	4, 10, 22.6, 150	None
11	Elbeltagi 2022[47] North Yorkshire, UK March to October 2020	HM-JACKarc (personal communication) Colonoscopy or cross-sectional imaging	2WW patients	NR	Median 72 (IQR 63–78)	• NR • NR • NR	52/992 (5.24%)	29 thresholds between 6 and 401 at varying intervals	None
12	Faux 2022[48] Cornwall, UK March to July 2020	HM-JACKarc Colonoscopy or CT abdomen/pelvis, or CT thorax/ abdomen/pelvis	2WW patients	Palpable mass 0%; anal ulceration NR	NR	• NR • NR • NR	6/175 (3.43%)	10	None
13	Godber 2016[50] NHS Lanarkshire, UK June 2013 to December 2013	HM JACKarc analyser Colonoscopy	Referred to colonoscopy in Scotland, 2013, referral criteria unclear	NR	Median 59 (range 16–89), *n* = 507	• 216/507 (42.6%) • NR • 23/484 (4.8%)	11/484 (2.27%)	10	None
9	Nicholson *et al*. 2018[63] Oxfordshire, UK January to March 2016 CSS-BIO-3 4730	HM JACKarc (note that some patients had two test results, any positive was a positive) Records follow-up	Same criteria as DG30 low risk, but unknown proportion outside the criteria	NR	Median 58, range 19–93 years	• 43% • NR • *n* = 62 (denominator unclear)	7/238 (2.94%)	7, 10, 20, 50	None

TABLE 3 Study and patient characteristics of HM-JACKarc studies (*continued*)

Study number	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with NICE scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds µg/g	Subgroup data?
9	Nicholson *et al.* 2020[27] (overlaps with Withrow *et al.* 2022)[66] Oxfordshire, UK March 2017 and March 2020 CSS-BIO-3 4730	HM JACKarc Records follow-up	Same criteria as DG30 low risk, plus some outside the criteria (e.g. inflammation; thrombocytosis; tired all the time)	Palpable rectal or anal mass, or anal ulceration *n* = 0	Median 60 (range 18–101, IQR 51–74)	• 41.4% • NR • Anaemia *n* = 2791/12,509 = 22.3%; Iron deficiency *n* = 1158/12,509 = 9.3%	105/9896 (1.06%)	7, 10, 20, 50, 100, 120, 150	Males; females
14	Tang *et al.* 2022[65] Wales, UK March to June 2020	HM-JACKarc system Colonoscopy or CTC, or MPCT (minimal preparation CT)	All consecutive patients referred from primary care on the USC pathway[a]	Abdominal mass 2.5% Anal lump/mass 2.2% Rectal mass 1.5%	Median 68 (range 21–97) (*n* = 1050)	• 47.4% • NR • New anaemia 11.1%	20/603 (3.32%)	10	IDA
15	Turvill *et al.* 2021[82] Yorkshire and Humber, UK April 2018 to December 2019 Fast track FIT	HM JACKarc Full colonoscopy or CT colonography, or a lesser investigation (such as CT abdomen/ pelvis with contrast or flexible sigmoidoscopy)	2WW patients	Abdominal mass 1.7% Rectal mass 1.6%	Mean 67.3 (SD 11.7) Median 69 (IQR 60–76)	• 44.5% • NR • IDA or other anaemia, 21.9%[a]	151/5040 (3.00%)	2	Anaemia, no anaemia, Males, females, medication (antiplatelets, anticoagulants NSAIDs), age </> 60
Subgroup data only[b]									
16	Cunin 2020[71a] East Sussex, UK NR (must be between 2013 and 2019)	HM JACKarc Various imaging[a]	Type 2: NG12 high/ medium-risk patients with/without IDA	NR	With IDA: median 74 (IQR 65–82) Without IDA: median 72 (IQR 63–79)	With IDA: • 37% • NR • 100% Without IDA: • 41.3% • NR • 0%	With IDA: 20/189 (10.58%) Without IDA: 28/739 (3.79%)	10	IDA, no IDA

continued

TABLE 3 Study and patient characteristics of HM-JACKarc studies (*continued*)

Study number	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with NICE scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds µg/g	Subgroup data?
6	D'Souza 2021[73] NICE FIT October 2017 to December 2019	HM JACKarc analytical system Colonoscopy	Type 4: NG12 high/ medium risk (subgroup of main NICE FIT study) who underwent colonoscopy	NR	Age < 50: 42 (SD 6.5) Age ≥ 50: 66.7 (SD 9.16)	Age < 50: • 40.3% • Ethnicity reported[a] • IDA 5.9%, non-IDA anaemia 1.9% Age 50 + • 45.7% • Ethnicity reported[a] • IDA 4.8%, non-IDA anaemia 5.5%	Age < 50: 16/1103 (1.45%) Age ≥ 50: 313/8719 (3.59%)	2, 10, 150	Age </> 50
Dual FIT									
2	Gerrard 2023[75] Lothian, Scotland, UK March 2020 to July 2021	HM-JACKarc Endoscopy or CT with colorectal protocol.	Type 1: Urgent suspected of cancer referrals, criteria for referral[a] are both wider and narrower than NG12 high/medium and DG30 low risk	Wider and narrower than target population[a] Abdominal mass: 3.2% Rectal mass: 2.4%	Median 65 (IQR 56–74)	• 44.3% • NR • 18.2%	88/2637 (3.34%)	10	0
17	Turvill 2018[81] York Hospital, UK February 2016 to March 2017	HM-JACKarc Full colonoscopy or CT colonography or a lesser investigation (such as CT abdomen/pelvis with contrast plus flexible sigmoidoscopy) limited by the identification of pathology	Type 3: NG12 high/ medium risk	4% abdominal mass and 1% rectal mass	Median 69 (IQR 61–76)	• 50% • NR • 18% IDA • Other characteristics[a]	27/476 (5.67%)	43 (either FIT positive) 2 (both FIT positive)	0

TABLE 3 Study and patient characteristics of HM-JACKarc studies (*continued*)

Study number	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with NICE scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds µg/g	Subgroup data?
Repeat FIT									
18	Johnstone 2022[53] 3 NHS Boards (Tayside, GG&C, Highland), Scotland, UK December 2015 to October 2021[a]	HM-JACKarc Records follow-up	Type 4: Since unclear what criteria used to select patients for FIT. Symptomatic patients who had two FIT between 1 week and 1 year apart	NR	Median (IQR) GG&C 63 (52–74) Tayside 69 (56–78) Highland 69 (57–77)	• 41.4% • NR • NR	42/5761 (0.73%)	10	0

2WW, two week wait; CT, computed tomography; DG30, diagnostic guidance 30; FIT, Faecal immunochemical test; GG&C, Greater Glasgow and Clyde; IDA, iron deficiency anaemia; IQR, interquartile range; NG12, National Guideline 12; NICE, National Institute for Health and Care Excellence; NSAID, non-steroidal anti-inflammatory drug ; NR, not reported; SIGN, Scottish Intercollegiate Guidelines Network.

a Farrugia 2020: study recruited all 2WW patients and reported DG30 low-risk and NG12 high-/medium-risk subgroups. Only the NG12 subgroup has been included in the analysis as the DG30 subgroup is likely to be highly selected and/or enriched with FIT or Guaiac positive patients; Gerrard 2023: inclusion criteria were urgent suspected of cancer or urgent priority referrals with 'high-risk' symptoms: repeated rectal bleeding without obvious rectal cause or blood mixed in stool, persistent change in bowel habit, palpable abdominal or rectal mass, weight loss and/or abdominal pain with or without unexplained IDA; Johnstone 2022:[52] IDA defined as ferritin < 15 µg/l; Johnstone 2022:[53] recruitment dates for each area were Tayside: December 2015 to December 2020, Highland: December 2018 to October 2021, Greater Glasgow and Clyde: September 2018 and December 2020; MacDonald 2022: personal communication with the author indicated that SIGN 126 guidelines and the Scottish Referral Guidelines for Suspected Cancer were used to guide referrals, and that these indicate that both high- and low-risk patients as defined by NG12 should be referred. The paper itself lists 'rectal bleeding, diarrhoea, anaemia, anorectal or abdominal mass, abdominal pain weight loss, change in bowel habit (including faecal incontinence), anorectal symptoms (tenesmus, per rectal pain or mucous) or colorectal abnormalities on imaging, per NHSL pre-existing criteria'; Tang 2022; referral criteria in Wales unclear; Turvill 2021: additional patient characteristic reported, 27% using antiplatelet therapy, anticoagulants or NSAIDs; Turvill 2018: additional patient characteristic reported, 30% were taking NSAID, antiplatelet therapy or anticoagulants; Withrow 2022: Nicholson 2020 was selected for inclusion in the main analysis despite having fewer patients than Withrow 2022 as it reported more thresholds; Cunin 2021: excluded from the main analysis as recruitment dates not reported meaning crossover could not be ascertained, but used in the anaemia subgroup analysis. reference standard was colonoscopy, oesophago-gastroduodenoscopy, computed tomography (CT) scanning or virtual CT colonography; D'Souza 2020:[18] White (62%), Asian (14%), 'other' ethnicity (12%), Black (9%) D'Souza 2021:[72] White 5693 (80.0%), Asian 355 (5.0), Black 253 (3.6), mixed 42 (0.6), Chinese 27 (0.4), not specified 746 (10.5); D'Souza 2021:[73] age < 50 years, White 68.4%, Asian 9.8%, Black 5.4%, mixed 1.1%, Chinese 1.2%, other 18.1%, missing 4.4%.

b The full study population was included in the AA and IBD data analysis as this was not reported for the NG12 subgroup.

risk because some or all of the reported thresholds were derived to optimise accuracy. The reference standard target condition was CRC in all cases and therefore all studies were rated as low risk for this.

Statistical synthesis HM JACKarc

Sixteen studies contributed to the meta-analysis for HM JACKarc. Seven studies provided diagnostic accuracy at a single threshold and the maximum number of thresholds considered in a single study was 103. The final data set provided a total of 151 pairs of sensitivity and specificity, at thresholds between 2 and 401.

Figure 5a displays the results on the ROC plane. Observations from the same study are joined. *Figure 5b* displays the sensitivity and specificity as a function of threshold. Pooled sensitivity and specificity are shown for subgroups based on population type in *Figures 5c* and *5d*, respectively. Sensitivity and specificity for specific thresholds are summarised for all population groups in *Table 4*.

For the analysis of all studies (populations 1–4), sensitivity ranges from 95.9 (95% CrI 92.7 to 97.9; 95% PrI 81.4 to 99.8) at a threshold of 2, to 46.3 (95% CrI 37.4 to 54.9; 95% PrI 21.9 to 70.2) at a threshold of 400. Specificity ranges from 65.1 (95% CrI 55.6 to 74.8; 95% PrI 30.3 to 96.7) at a threshold of 2, to 97.7 (95% CrI 94.7 to 99.2; 95% PrI 78.1 to 100). For the analyses of subgroups by population type, the summary estimates were similar and not statistically significant based on overlap of the 95% CrIs. The summary sensitivity and specificity for population 3 are higher than for the other considered subgroups; however, this analysis was based on only two studies that contributed data at two thresholds (2 and 10). There is, therefore, considerable uncertainty in the pooled estimates, and these should be interpreted with caution.

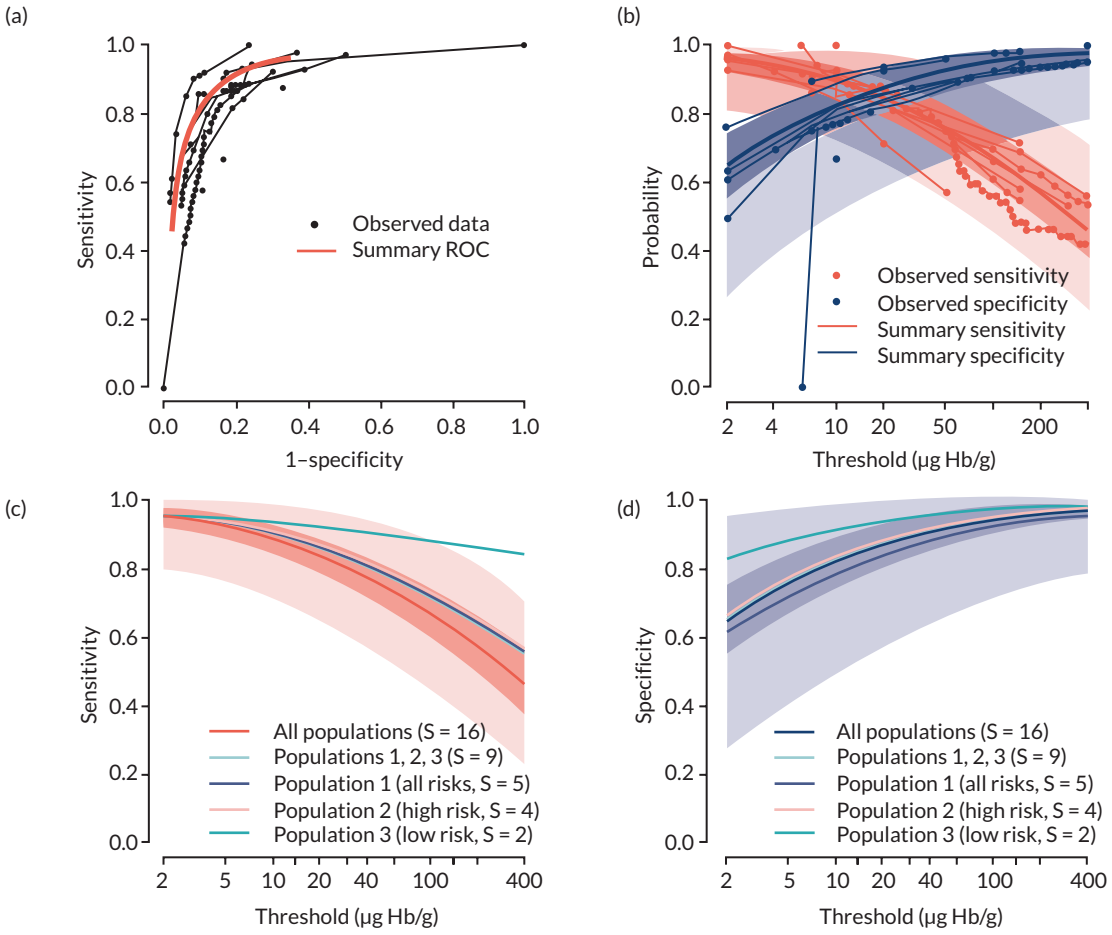

FIGURE 5 Observed data and summary sensitivity and specificity for HM JACKarc. (a) Summary ROC for all population analysis; (b) sensitivity and specificity as a function of threshold for all population analysis; (c) Pooled sensitivity for population subgroups; (d) Pooled specificity for population subgroups. Ninety-five per cent CrIs and PrIs for summary sensitivity (all population) are shown by the dark and light red regions. Ninety-five per cent CrIs and PrIs for summary specificity (all populations) are shown by the dark and light blue regions.

TABLE 4 Summary sensitivity and specificity at specific thresholds for HM-JACKarc

	All studies 1–4 (n = 16)		All 1–3 (n = 9)		Population 1 (S = 5)		Population 2 (S = 4)		Population 3 (S = 2)	
Threshold	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity(95% CrI)
2	95.9 (92.7 to 97.9)	65.1 (55.6 to 74.8)	95.5 (93 to 97.1)	66.7 (54.9 to 77.2)	95.2 (89.3 to 98.5)	61.6 (41.3 to 81.8)	95.7 (89.1 to 98.2)	66.8 (60.4 to 75.3)	95.5 (83.4 to 100)	83.3 (74.1 to 91.2)
2.5	95.3 (91.8 to 97.5)	68 (58.8 to 77.3)	95 (92.4 to 96.7)	69.3 (57.8 to 79.3)	94.6 (88.4 to 98.2)	64.2 (44.4 to 83.9)	95.2 (88.3 to 97.9)	69.5 (63.2 to 77.6)	95.3 (83 to 99.9)	84.8 (75.8 to 92.3)
3	94.7 (91.1 to 97.2)	70.3 (61.3 to 79.3)	94.5 (91.8 to 96.3)	71.3 (60.2 to 80.9)	94.1 (87.6 to 98)	66.3 (46.9 to 85.4)	94.7 (87.6 to 97.7)	71.7 (65.4 to 79.3)	95 (82.6 to 99.9)	85.9 (77.1 to 93.1)
4	93.8 (89.8 to 96.5)	73.7 (65.1 to 82.2)	93.6 (90.7 to 95.7)	74.3 (63.7 to 83.3)	93.3 (86.3 to 97.6)	69.5 (50.6 to 87.8)	93.9 (86.4 to 97.2)	74.8 (68.7 to 81.9)	94.6 (82 to 99.9)	87.6 (79.2 to 94.3)
7	91.4 (86.8 to 94.8)	79.6 (71.7 to 87.1)	91.6 (88.3 to 94.1)	79.5 (70.1 to 87.2)	91.2 (83.3 to 96.6)	75.3 (57.1 to 91.4)	91.9 (83.7 to 95.9)	80.3 (74.5 to 86.2)	93.8 (80.6 to 99.9)	90.3 (82.6 to 96)
10	89.5 (84.6 to 93.4)	82.8 (75.2 to 89.6)	90.1 (86.5 to 92.8)	82.4 (73.7 to 89.3)	89.6 (81.1 to 95.7)	78.6 (60.7 to 93.2)	90.4 (81.7 to 94.9)	83.3 (77.8 to 88.5)	93.2 (79.6 to 99.8)	91.8 (84.6 to 96.8)
20	84.7 (79.1 to 89.6)	87.9 (81.1 to 93.4)	86.3 (82.1 to 89.7)	87.1 (79.8 to 92.6)	85.7 (76.2 to 93.6)	84.1 (67.1 to 95.8)	86.7 (77.1 to 92.1)	88 (83.2 to 92.2)	NR	NR
50	75.8 (69.4 to 82.0)	92.6 (87 to 96.5)	79.5 (74.5 to 83.8)	91.7 (86 to 95.5)	78.8 (68 to 89.2)	89.5 (74 to 97.9)	79.9 (69.1 to 86.7)	92.4 (88.7 to 95.4)	NR	NR
100	67 (60 to 74.2)	94.9 (90.3 to 97.8)	73 (67.1 to 78.1)	94.1 (89.5 to 97)	72.2 (60.4 to 84.7)	92.5 (78.1 to 98.8)	73.4 (61.3 to 81.1)	94.7 (91.6 to 97)	NR	NR
120	64.5 (57.2 to 71.9)	95.4 (91 to 98.1)	71 (64.9 to 76.4)	94.6 (90.3 to 97.3)	70.2 (58.3 to 83.3)	93.1 (79.2 to 98.9)	71.4 (58.9 to 79.4)	95.2 (92.3 to 97.4)	NR	NR
150	61.3 (53.7 to 68.9)	96 (91.9 to 98.4)	68.5 (62.1 to 74.2)	95.2 (91.1 to 97.6)	67.8 (55.5 to 81.5)	93.8 (80.4 to 99.1)	68.9 (55.8 to 77.3)	95.7 (93 to 97.7)	NR	NR
200	57 (48.9 to 64.9)	96.6 (92.8 to 98.7)	65.2 (58.4 to 71.2)	95.8 (92.1 to 98)	64.4 (51.7 to 79.1)	94.6 (81.7 to 99.3)	NR	NR	NR	NR
400	46.3 (37.4 to 54.9)	97.7 (94.7 to 99.2)	56.5 (48.7 to 63.5)	97.1 (94.1 to 98.7)	55.8 (41.8 to 72.6)	96.2 (84.8 to 99.6)	NR	NR	NR	NR

Main analysis: OC-sensor

No end-to-end studies were identified. Seventeen studies reported across 18 publications[30,41,43–46,49,51,54,55,57,62,64,68,70,76–78] reported diagnostic test accuracy data for OC-Sensor (*Table 5*). One study was reported across two publications.[77,78]

Among the 17 studies, the analyser OC-Sensor iO and PLEDIA were used, as were two analysers not mentioned in the NICE scope (DIANA and MICRO). Clinical advisors to the EAG confirmed with the company that these were calibrated in the same way as the in-scope tests and can be considered equivalent. Twelve studies were in the UK[30,41,43–46,49,51,55,57,62,68] (one of which was in Scotland[62]), four (five publications) were in Spain[64,70,76–78] and one study was in Denmark.[54] Sample sizes ranged from 120[78] to 37,216[30] and CRC prevalence ranged from 0.59%[43] to 11.65%.[70] Patient characteristics (ethnicity, blood disorders, medications, anaemia) were rarely or never reported. Age was usually reported as a median, which ranged from 61 [interquartile range (IQR) 55–77] years[45] to 71.1 (IQR 62.5–78.7) years[46] among studies of types 1–4, while the proportion who were male ranged from 43%[45] to 50%.[41]

Eleven studies contributed to the main analysis:[30,41,43,45,46,49,54,55,57,62,64] three to the population type 1 analysis,[30,45,49] one to the population type 2 analysis,[44] one to the population type 3 analysis[43] and seven to the population type 4 analysis.[41,46,54,55,57,62,64] The characteristics of these studies are described in more detail in the following sections. Six studies (seven publications) reported subgroup data,[43,54,68,70,76–78] and one study reported data on dual FIT.[51] The characteristics of these studies are described in more detail in *Main analysis: dual faecal immunochemical test* and *Subgroup analyses by patient characteristics* and their associated appendices.

No diagnostic test accuracy data were identified that related to OC-Sensor Ceres. The company supplied correlation data between OC-Sensor PLEDIA and OC-Sensor Ceres, and between OC-Sensor iO and OC-Sensor Ceres, conducted in accordance with CLSI EP09-A3 (no reference given), which relates to test bias estimation using patient samples.[86] The number of samples tested was 111, and the *R* was 0.999 compared with PLEDIA and 0.998 compared with iO. This suggests a high level of correlation between the devices and that measurements are likely to be similar for most samples. However, the ERG notes that no details were given about the patient population or the methods of the analyses performed, or whether the differences would lead to different clinical decisions at specific thresholds. There was no indication of what would constitute a clinically acceptable level of disagreement. A formal recommendation of equivalence could not be made based on the evidence provided.

Main analysis

Eleven studies reported across 11 publications were included in the main analysis.[30,41,43,45,46,49,54,55,57,62,64] Thresholds ranged from 4[30,45,46,49,55,62] to 200 µg/g[55] and CRC prevalence ranged from 0.6%[43] to 6.62%.[41] There was one population type 4 study that reported a NG12 high-/medium-risk subgroup;[43] the subgroup was included instead of the population type 4 analysis to avoid enriching the sample with patients who had a positive FIT result in primary care. The Benton *et al.*[44] analysis was not included in the main analysis as the recruitment dates and locations cross over with those in Cama *et al.*,[45] but this is included in the analysis of population type 2 (NG12 high/medium risk). The reference standard was records follow-up in six studies,[30,45,49,54,57,64] while the remainder used imaging modalities, but not always 90% CTC or colonoscopy. Most studies did not report whether patients with rectal/anal masses and anal ulceration were excluded, except for the three population type 1 studies and one other that reported 0.3% with palpable masses.[62] Two studies[43,54] in the main analysis reported patient characteristics subgroup data.

Population type 1 studies (all presenting to primary care)

Three studies (all OC-Sensor ion)[30,45,49] were considered by the EAG to be population type 1 studies because they recruited a population thought to be close to all patients presenting to primary care (see *Table 5*, column 3). These reported thresholds ranging from 4 to 150 µg/g, had sample sizes ranging from 4187[49] to 37,216[30] and had CRC prevalence ranging from 1.39%[45] to 1.74%. All three used records follow-up as the reference standard. Study inclusion criteria were not uniform across studies; all exclude rectal masses, but only one excluded anal ulceration[49] and another abdominal masses.[45] One study also largely excluded IDA[45] and one excluded rectal bleeding.[45] No data relating to subgroups were reported.

TABLE 5 Study and patient characteristics of OC-Sensor studies

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status[a])	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Subgroups[b]
Population type 1 studies (all patient presenting to primary care with symptoms meeting NG12 high/medium or DG30 low risk)									
1	Crooks 2023[30] Nottingham, UK November 2017 to November 2021	• iO • Records follow-up	All referral criteria, except anorectal mass	Anorectal mass excluded	NR	• NR • NR • NR	514/37216 (1.38%)	4, 10, 20, 40, 100	0
2	Cama 2022[45] Hertfordshire, UK June 2019 to November 2021	• iO • Records follow-up	All DG30 (low risk) and most NG12 high risk (see column 6)	IDA, rectal bleeding, rectal or abdominal masses excluded	Median 61 (IQR 55–77)	• 43% (of *n* = 12,231) • NR • 2% IDA; non-IDA 4%	74/5341 (1.39%)	4, 10, 100	0
3	Georgiou Delisle 2022[49] Croydon, UK December 2019 to October 2020	• iO • Records follow-up	NICE NG12 and DG30 criteria	Rectal mass or anal ulceration referred straight to 2WW	Mean 65 (range 18–99)	• Male 44.8% • See footnote[a] • NR	61/4187 (1.46%)	4, 10, 150	0
Population type 2 studies (NG12 high risk)									
4	Benton 2022[44] 50 NHS hospitals across England, UK October 2017 to December 2019 NICE FIT	• OC Sensor PLEDIA • Colonoscopy	NG12 high risk, who had colonoscopy. Randomised to cohort 1 who were given four tests	NR	NR	• NR • NR • NR	7/233 (3.00%)	1, 10, 100	0
Population type 3 (DG30 low risk)									
5	Ball 2022[43] (additional data by personal communication) Sheffield, UK October 2019 to December 2019	• PLEDIA • Colonoscopy or CT imaging[c] and colon capsule endoscopy	DG30 low risk[c]	NR	NR for this subgroup	• NR for this subgroup • NR • NR for this subgroup	17/2892 (0.58%)	10, 20, 50, 80, 100, 120, 150	Males; females
Population type 4 (unclear/unrepresentative of all presenting to primary care)									
6	Archer 2022[41] Sheffield, UK March 2020 to July 2020	• PLEDIA • CT, colonoscopy	2WW patients	NR	*n* = 514; mean 64.5 years (SD 12.7 years)	*n* = 514 • 50% • NR • *n* = 514 IDA (23%)	11/166 (6.62%)	10, 60, 100	0

continued

TABLE 5 Study and patient characteristics of OC-Sensor studies (*continued*)

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status[a])	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Subgroups[b]
7	Chapman 2021[46] Nottingham, UK September 2016 to September 2017 Getting FIT	• DIANA • Colonoscopy and additional investigations (e.g. radiology)	2WW patients, returning both FITs	NR	Median 71.1 (IQR 62.5–78.7)	• 43.9% • NR • NR	38/732 (5.19%)	4, 10, 100	0
8	Juul 2018[54] Central Denmark September 2015 to August 2016 NCT02308384	• DIANA • Records follow-up	Patients with 'non-alarm' symptoms of CRC[c]	NR	Mean NR	• 43.9% • See footnote[c] • 12.3%	54/3462 (1.56%)	10	Unexplained anaemia
9	Laszlo 2021[55] 24 hospitals and 59 GP practices in UK April 2017 to March 2019	• iO • Colonoscopy 77.7%; CTC 14.2%; flexisig 7.5	2WW patients	NR	Median 67 (range 19–99; IQR 57–75)	• 46.6% • See footnote[c] • 19%	90/3596 (2.50%)	4, 6, 10, 20, 50, 80, 100, 120, 150, 200	0
10	Maclean 2021[57] Royal Surrey NHS Foundation Trust, UK End of March 2020 to July 2020	• PLEDIA • Assume records follow-up, as some patients were safety netted	2WW patients	NR	NR	• NR • NR • NR	12/358 (3.35%)	10, 150	0
11	Mowat 2016[62] NHS Tayside, UK October 2013 to March 2014	• iO • Colonoscopy	Symptomatic patients referred from primary care with FIT	Palpable mass 0.3%	Median 64 (range 16–90, IQR 52–73) (*n* = 755)	(*n* = 755) • 45.3% • NR • 9%	28/750 (3.73%)	4, 10	0
12	Pin Vieito 2021[64] San Sebastian, Spain January 2012 to December 2016	• NR • Records follow-up	Patients referred from primary care with FIT	NR	NR	• NR • NR • NR	73/4543 (1.61%)	10, 20	0
Subgroup data only[d]									
13	Ayling 2019[68] Derriford Hospital, Plymouth, UK March 2014 to March 2017	• NR • Endoscopy or computed tomography scan (NR what type of CT)	Population type 4 – 2WW patients	NR	NR	• NR • For *n* = 428, 99.8% White British • 100% anaemia or IDA	Low haemoglobin group: 7/178 (3.93%) IDA group: 6/137 (4.38%)	10	IDA; anaemia

TABLE 5 Study and patient characteristics of OC-Sensor studies (*continued*)

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with scope	Mean/ median age in years	Patient characteristics (male, ethnicity, anaemia status[a])	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Subgroups[b]
14	Bujanda 2018[70] Spain (assume Ourense and San Sebastian)[84] March 2012 to 2014 COLONPREDICT	• NR • Colonoscopy	Population type 4 – symptomatic patients referred from primary and secondary care	NR	Aspirin users: mean 72.5 (SD 9); aspirin non-users: mean 63.7 (SD 14)	• Aspirin users 58.10%; aspirin non-users 48.5% • NR • NR	Aspirin users: 51/485 (10.51%); aspirin non-users: 299/2567 (11.65%)	20	Aspirin users; aspirin non-users
15	Morales-Arraez 2018[76] La Laguna, Spain April 2016 to December 2017	• NR • Colonoscopy	Population type 4 – anaemic patients[e] referred from primary care	NR	Mean 71 (SD 12)	• 33.90% • NR • 100% IDA	28/245 (11.43%)	10	IDA
16	Rodriguez-Alonso 2018[77] Barcelona, Spain September 2011 to October 2012	• MICRO • Colonoscopy	Population type 4 – symptomatic patients referred from primary and secondary care	NR	PPI users: mean (SE) 64.9 ± 11.3; PPI non-users: mean (SE) 57.3 ± 14.0	• PPI users 44.20%; non-users 49.80% • NR • Users 18.6%; non-users 5.4%	PPI users: 15/525 (2.86%); PPI non-users: 15/477 (3.14%)	20	PPI use
16	Rodriguez-Alonso 2020[78] Barcelona, Spain September 2011 to October 2012	• MICRO • Colonoscopy	Population type 4 – symptomatic patients referred from primary and secondary care	NR	NR	• 48.3 • NR • 100% IDA	9/120 (7.50%)	15	IDA
Dual FIT									
17	Hunt 2022[51] Lancashire and South Cumbria Cancer Alliance (LSCCA), UK January 2019 to February 2021	• NR • Records follow-up	Population type 4 – referred to secondary care,[f] returned two FIT	NR	Median 66 (range 16–103)	• 44% • NR • NR	317/28,622 (1.11%)	10	0

a Study characteristics relating to medications that may cause GI bleeding and conditions that may affect FIT have been removed, as these data were not reported in any studies.
b 0, none; 1, IDA or anaemia; 2, male; 3, female; 4, ethnicity; 5, medications that may affect GI bleeding; 6, blood disorders that may affect the performance of the test; 7, age groups (add age).
c Ball 2022: data were also available for all patients on 2WW, the subgroup of patients who meet NG12 criteria were selected for inclusion in the review. CT imaging was a mix of CTC and other CT imaging modalities; Juul 2018: FIT aimed at those aged ≥ 30 years with non-alarm symptoms of CRC, according to GP clinical knowledge and instructions, which included change in bowel habits, abdominal pain, unexplained anaemia and unspecific symptoms (e.g. fatigue or weight loss), but not for IBS workup. Those aged ≥ 40 years with rectal bleeding, change in bowel habits > 4 weeks, abdominal pain and IDA recommended to be referred straight to secondary care; Georgiou Delisle 2022: White/White British 51.4%, Asian/Asian British 12.6%, Black/Black British 14.8%, Chinese 0.8%, other 18.6%, mixed 1.5%, not recorded 0.4%; Juul 2018: *n* (%); Danish, 3280 (94.8); immigrant Western country, 84 (2.4); immigrant non-Western country, 98 (2.8); Laszlo 2021: Black/Black British, 4.5; Asian/Asian British, 6.1; other Asian, 2.0; White, 23.5; British mixed, 17.9; multiple/other, 5.6; missing data, 40.3.
d Some type 1–4 studies also report subgroup data as indicated in the final column.
e IDA defined as Hb < 11.9 g/dl in men and Hb < 10.9 g/dl in women, and ferritin ≤ 30 g/dl.
f Unrepresentative mix of NG12 high risk and DG30 low risk due to change in referral criteria part-way through the study.

Population type 2 studies (NG12 high/medium risk)

One UK study[44] was a population type 2 study (NG12 high risk), which reported thresholds of 1, 10 and 100 using OC-Sensor PLEDIA, had a CRC prevalence of 3.00% and used colonoscopy as the reference standard. This was part of the NICE FIT study, reporting a subgroup of NICE FIT patients who met NG12 high-risk criteria, and who were invited to receive and also completed four FITs and colonoscopy (n = 233 out of 9822 recruited to NICE FIT). Recruitment dates and location for NICE FIT overlap with those in a type 1 study,[45] which was preferentially selected for inclusion in the overall OC-Sensor analysis, while the study by Benton *et al.*[44] is included in the type 2 subgroup analysis. It was not clear if rectal/abdominal mass and anal ulceration patients were included, and no diagnostic test accuracy data were reported for subgroups according to patient characteristics.

Population type 3 studies (DG30 low risk)

One UK study[43] was a population type 3 study (DG30 low risk), which reported thresholds from 10 to 150 using OC-Sensor PLEDIA, had a sample size of 2892 and a CRC prevalence of 0.6% and used imaging (colonoscopy, CT imaging and colon capsule endoscopy) as the reference standard. It was not clear if rectal/abdominal mass and anal ulceration patients were included. Male and female subgroups were reported.

Population type 4 studies (unclear/unrepresentative of patients presenting to primary care)

Seven studies[41,46,54,55,57,62,64] were population type 4 studies, comprising a mix of studies that recruited only patients referred to 2WW or were unclear or unrepresentative of patients in primary care in some other way (see *Table 5*, column 3). They used a mix of analysers (see *Table 5*), reported thresholds from 4[46,55,62] to 200,[55] had sample sizes ranging from 116 to 4543 and had CRC prevalence ranging from 1.56%[54] to 6.62%.[41] Three studies[55,57,64] used records follow-up as the reference standard, while the remainder used imaging modalities, but not always 90% CTC or colonoscopy. Studies from England[41,46,55,57] recruited patients who had been referred to 2WW since the DG30 update in 2017 and may therefore have recruited DG30 low-risk patients on the basis of a positive FIT in primary care (see *The analysis plan and rationale* for discussion of why this is problematic). Studies from elsewhere recruited patients referred to secondary care from primary care. One study[62] included a small proportion (0.3%) of patients with a palpable mass, but otherwise it was unclear if rectal/abdominal mass and anal ulceration patients were included. One study reported a subgroup of IDA/anaemia patients.[54]

Studies reporting subgroup data

In addition to the two studies[43,54] of types 3 and 4 that reported subgroup data, four additional studies (five publications)[68,70,76–78] were included that only reported subgroup data. Two (three publications)[70,77,78] of these were included in accordance with the tiered approach to study selection whereby inclusion criteria were relaxed if evidence was scarce for a given subgroup. Both studies were from Spain and included patients referred from both primary and secondary care. This may alter the patient spectrum and reduce generalisability to the primary care setting. Prevalence (see *Table 5*) varied a great deal, reflecting the highly selected nature of some of these cohorts. All studies used colonoscopy as the reference standard. Across all six studies reporting subgroup data, four reported data for ICA/anaemia,[54,68,76,78] two for medications [aspirin users[70] and proton pump inhibitor (PPI) users[77]] and one for male and female patients separately.[43] It was not clear if rectal/abdominal mass and anal ulceration patients were included. These studies are also discussed in *Subgroup analyses by patient characteristics* and its associated appendix.

Dual faecal immunochemical test using OC-Sensor

One UK type 4 study[51] reported data for dual FIT (not repeat FIT; see *Dual testing*) at a threshold of 10 µg/g, but did not state which OC-Sensor analyser was used. This study recruited both NG12 and DG30 patients, but NG12 patients were only included from June 2020 (recruitment period January 2019 to February 2021). The reference standard was records follow-up. It was not clear if rectal/abdominal mass and anal ulceration patients were included. No subgroup data were reported.

Quality assessment

Quality assessment of diagnostic test accuracy studies version 2[34] was applied to studies reporting diagnostic test accuracy data only. QUADAS-2 asks questions relating to internal validity (risk of bias) and external validity (applicability). Each is discussed separately in the following two paragraphs.

Risk of bias

Appendix 3, *Table 38*, summarises the risk-of-bias and applicability scores as assessed by the authors of this review. Full risk-of-bias scores with reasons for the scores are provided in *Appendix 3*. In terms of risk of bias, no study scored as low risk for all items, and no item scored as low risk for all studies. The index test scored as low risk most often, with only one study[62] scoring as high risk because one of the reported thresholds was selected to maximise sensitivity, and another (two publications) scoring unclear.[77,78] Where patient selection was at risk of bias this was because it was unclear whether a consecutive sample was recruited and/or because inappropriate exclusions were made, such as excluding people on the basis of not having had a colonoscopy or excluding those with rectal bleeding. Owing to the inclusion criteria for the review, all studies avoided a case–control design. The reference standard was rated as being at unclear or high risk of bias for all studies. This was usually because not all patients received a colonoscopy or CTC, or because it was unclear if the reference standard had been interpreted blind to the index test. Patient flow scored as high risk or unclear in nearly all studies. This was due to a mix of factors, including a lack of clarity about the interval between the index test and the reference standard in nearly all studies, patients receiving different reference standards depending on their FIT result or other factors, and patients being missing from the study.

Applicability

There were concerns about the representativeness of the patients recruited of all those presenting to primary care in nearly all studies due to either exclusion of some patients (study population types 2, 3 and 4), or a lack of clarity about who was included in comparison with the target population. Some studies[30,49] were classed as being population type 1 despite scoring poorly for this item as the exclusions were relatively minor (IDA and rectal bleeding), although these limitations should be noted. The index test was rated at low risk of having poor applicability in all studies. The reference standard target condition was CRC in all cases and therefore scored as low risk in all studies.

Statistical synthesis OC-Sensor

Eleven studies contributed to the meta-analysis for OC-Sensor. One study provided diagnostic accuracy at a single threshold, and the maximum number of thresholds considered within an individual study was 10.[55] The final data set comprised a total of 44 pairs of sensitivity and specificity estimates, at thresholds between 4 and 200.

Figure 6a displays the results on the ROC plane. *Figure 6b* displays the sensitivity and specificity as a function of threshold. Pooled sensitivity and specificity are shown for subgroups based on population type in *Figures 6c* and *6d*, respectively. Sensitivity and specificity for specific thresholds are summarised for all population groups in *Table 6*.

For the analysis of all studies (populations 1–4), sensitivity ranges from 94.2 (95% CrI 91.2 to 96.7; 95% PrI 84.6 to 99.0) at a threshold of 4, to 54.2 (95% CrI 48.4 to 60.2; 95% PrI 42.2 to 67.2) at a threshold of 200. Specificity ranges from 62.7 (95% CrI 47.4 to 77.2; 95% PrI 12.0 to 97.7) at a threshold of 4, to 97.3 (95% CrI 92.9 to 99.3; 95% PrI 71.9 to 100) at a threshold of 200. For the analyses of subgroups by population type, the summary estimates were similar and not statistically significant based on overlap of the 95% CrI.

A sensitivity analysis was performed to remove an outlier study.[41] This was a relatively small study with 166 patients and 11 CRC events. Removing it made almost no difference to the sensitivity at a threshold of 10 µg/g (89.9, 95% CrI 85.8 to 93.7, in the sensitivity analysis compared with 89.8, 95% CrI 85.9 to 93.3, in the main analysis), but it increased the specificity slightly (80.2, 95% CrI 69.4 to 89.9, compared with 77.6, 95% CrI 64.3 to 88.6). As this study had no clinical or methodological characteristics suggesting that it should be excluded, and as the change to specificity was small in the context of the 95% CrIs, it was retained in the main analysis.

Main analysis: FOB Gold

No end-to-end studies were identified. Three studies (three publications)[44,59,83] reported diagnostic test accuracy data for FOB-Gold (*Table 7*). Two of these studies[44,59] were comparative diagnostic test accuracy studies that reported data for more than one test-analyser and are also reported in *Comparative diagnostic test accuracy studies*. All studies were from the UK. Two studies[44,59] used FOB Gold Wide with the SENTiFIT 270 analyser, while one[83] stated the test to be FOB Gold and the analyser to be Roche Cobas c501 analyser (Roche Diagnostics, Oslo, Norway). It was not clear if the analysers would produce equivalent data. There was one type 2 study[44] (NG12 high-risk patients) and two type 4 studies.[59,83]

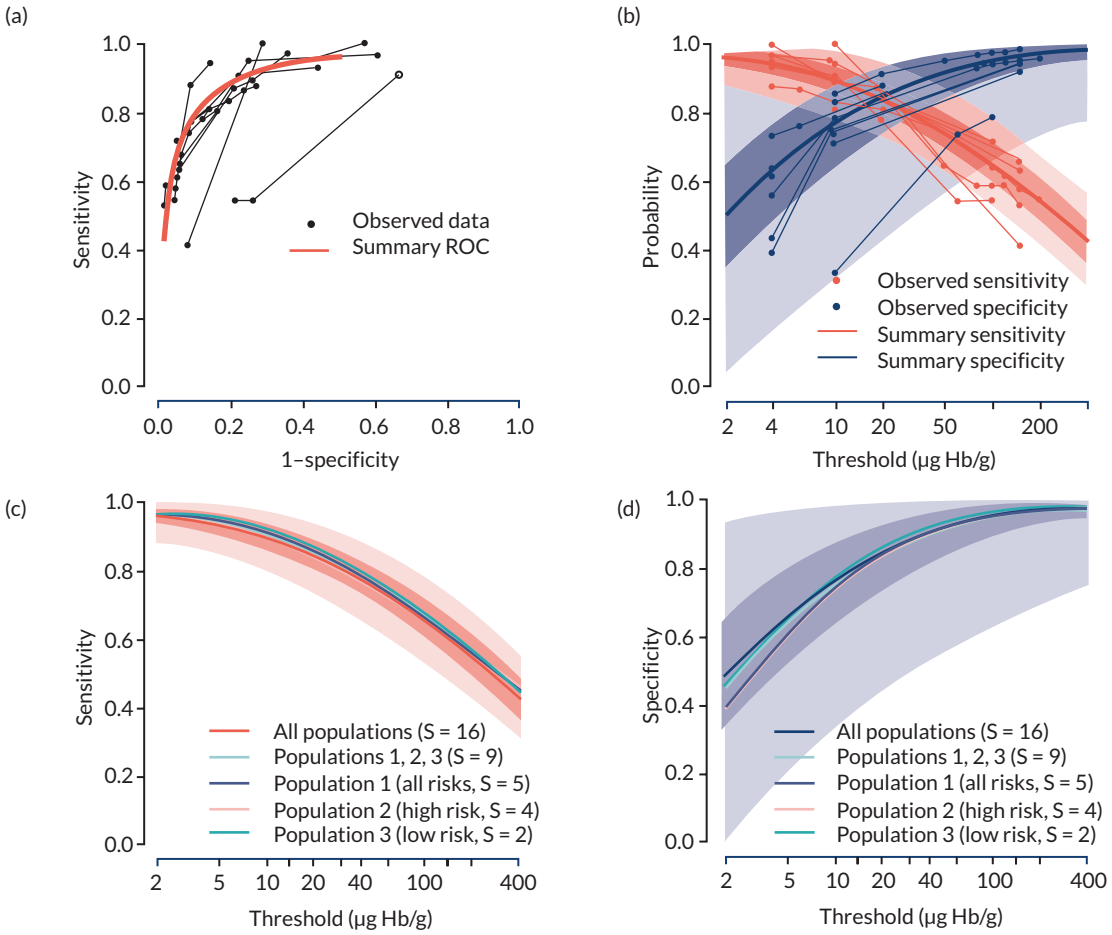

FIGURE 6 Observed data and summary sensitivity and specificity for OC-Sensor. (a) Summary ROC for all population analyses; (b) sensitivity and specificity as a function of threshold for all population analyses; (c) pooled sensitivity for population subgroups; and (d) pooled specificity for population subgroups. Ninety-five per cent CrIs and 95% PrIs for summary sensitivity (all population) are shown by the dark and light red regions. Ninety-five per cent CrIs and 95% PrIs for summary specificity (all populations) are shown by the dark and light blue regions.

TABLE 6 Summary sensitivity and specificity at specific thresholds for OC-Sensor

	All studies 1–4 (*S* = 11)		All 1–3 (*S* = 4)		Population 1 (*S* = 3)	
Threshold	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity (95% CrI)
4	94.2 (91.2 to 96.7)	62.7 (47.4 to 77.2)	95 (91.8 to 97.5)	60.8 (42 to 77.9)	95.1 (91.1 to 97.9)	55.3 (36.3 to 73.5)
7	91.8 (88.2 to 94.9)	72.3 (58.1 to 84.8)	92.8 (88.9 to 96)	72 (54.6 to 86)	92.9 (88.2 to 96.7)	67.6 (49.5 to 83.5)
10	89.8 (85.9 to 93.3)	77.6 (64.3 to 88.6)	90.9 (86.6 to 94.7)	78.1 (62.2 to 89.8)	91 (85.9 to 95.5)	74.5 (57.6 to 88.6)
20	84.7 (80.3 to 89)	85.6 (74.5 to 93.6)	86 (80.9 to 90.9)	87 (74.5 to 94.7)	86.2 (80.2 to 92.3)	84.8 (70.5 to 94.6)
50	75 (70.2 to 80)	92.5 (84.3 to 97.3)	76.3 (70.4 to 82.8)	93.9 (85.8 to 98)	76.6 (70 to 84.7)	93 (82.4 to 98.2)
100	65.3 (60.2 to 70.7)	95.5 (89.4 to 98.6)	66.3 (60.2 to 73.9)	96.6 (91.2 to 99)	66.8 (60.1 to 76.2)	96.2 (88.5 to 99.3)
120	62.5 (57.2 to 68)	96.1 (90.4 to 98.9)	63.4 (57.1 to 71.3)	97.1 (92.3 to 99.2)	64 (57.2 to 73.6)	96.8 (89.7 to 99.4)
150	58.9 (53.4 to 64.7)	96.7 (91.6 to 99.1)	59.7 (53.3 to 67.8)	97.7 (93.4 to 99.4)	60.3 (53.4 to 70.2)	97.4 (91.1 to 99.6)
200	54.2 (48.4 to 60.2)	97.3 (92.9 to 99.3)	NR	NR	NR	NR

95% CrI, 95% credible interval.

No diagnostic test accuracy data were reported for subgroups according to patient characteristics across all three studies, and there were no data on dual FIT using FOB-Gold.

Main analysis

All three studies contributed to the main analysis. Sample size ranged from 233[44] to 3349[83] and CRC prevalence ranged from 0.9%[83] to 3.0%.[44] Patient characteristics (mean age, ethnicity, blood disorders, medications, anaemia) were not reported. The proportions of male patients were 48%,[83] 48.8%[59] and not reported. Thresholds ranged from 2[44] to 150 µg/g.[59] The reference standard was imaging in two studies[44,59] and records follow-up in one.[83]

Population type 1 studies (all presenting to primary care)

There were no studies of this type.

Population type 2 studies (NG12 high/medium risk)

One UK study,[44] a subgroup of the NICE FIT study,[17,44,72,73] reported a subgroup of patients who met the NG12 high-/medium-risk criteria. It reported thresholds of 2, 10 and 100 µg/g using FOB Gold Wide and the SENTiFIT 270 analyser, had a CRC prevalence of 3.00% and used colonoscopy as the reference standard. It was not clear if rectal/abdominal mass and anal ulceration patients were included, nor was the percentage of male patients reported for this subgroup. No diagnostic test accuracy data were reported for patient characteristic subgroups for the NG12 subgroup, but they were available for the wider NICE FIT study using HM-JACKarc (see *Main analysis: HM-JACKarc*).

Population type 3 studies (DG30 low risk)

There were no studies of this type.

Population type 4 (unclear/unrepresentative of all presenting to primary care)

Two studies[59,83] were unlikely to have recruited a representative sample of the full spectrum of patients presenting to primary care. One study[59] recruited patients referred to 2WW in England and had a CRC prevalence of 2.53%, while the other[83] recruited patients from primary care in an area with a system that prompted GPs not to use FIT if a patient did not meet DG30 low-risk criteria, but did not prevent FIT requests for patients outside DG30 low-risk criteria. The study has therefore been classed as type 4 as it includes some NG12 high-risk patients, but may include predominantly DG30 low-risk patients. Thresholds ranged from 10 to 150 µg/g in one study,[59] and only data at a threshold of 10 µg/g were reported in the other. One used colonoscopy, CTC or flexible sigmoidoscopy where there were perianal symptoms or anorectal bleeding as the reference standard,[59] and the other used records follow-up.[83] It was not clear if rectal/abdominal mass and anal ulceration patients were included. One study[59] had 48.8% male patients, and the other had 48%.[83] No diagnostic test accuracy data were reported for subgroups according to patient characteristics.

Quality assessment

Quality assessment of diagnostic test accuracy studies version 2[34] was applied to studies reporting diagnostic test accuracy data only. QUADAS-2 asks questions relating to internal validity (risk of bias) and external validity (applicability). Each is discussed separately in the following two paragraphs.

Risk of bias

Appendix 3, *Table 39*, summarises the risk-of-bias and applicability scores as assessed by the authors of this review. Full risk-of-bias scores with reasons for scores are provided in *Appendix 3*. No study scored as low risk for all items. The index test conduct scored low risk for all studies. Where patient selection was at risk of bias this was because a consecutive sample was not recruited and/or because inappropriate exclusions were made, such as excluding people on the basis of not having had a colonoscopy. Owing to the inclusion criteria for the review, all studies avoided a case–control design. The reference standard was at unclear risk of bias for one[44] study because it was unclear if the reference standard had been interpreted blind to the index test and at high risk for another as not all patients received a colonoscopy or CTC.[83] Patient flow was low risk in one study,[44] high risk in two studies[59,83] due to multiple issues (interval between index test and reference standard unclear, patients receiving a different reference standard on the basis of their FIT result, and patients excluded from the analysis).

Applicability

There were concerns about whether the patients recruited were representative of all those presenting to primary care in all three studies, owing to either the exclusion of some patients (study population types 2 and 4) or there being a lack of clarity about who had been included in comparison with the target population. The index test and reference standards were rated as being at low risk of having poor applicability in all three studies.

Statistical synthesis FOB Gold

Three studies[44,59,83] contributed to the meta-analysis for FOB Gold. The number of thresholds considered in each study ranged from 1 to 4, and the final data set provided a total of eight pairs of sensitivity and specificity, at thresholds between 2 and 150.

Figure 7a displays the results on the ROC plane. Observations from the same study are joined by a line. *Figure 7b* displays the sensitivity and specificity as a function of threshold. Owing to the small number of studies evaluating FOB Gold, subgroup analyses by population type were not conducted. Sensitivity and specificity for specific thresholds are summarised for all population groups in *Table 12*.

The summary sensitivity and specificity are plotted in *Figure 8*, with information relating to the number of participants and number of positive tests in each study. For the analysis of all studies (populations 1–4), sensitivity ranged from 91.4 (95% CrI 71.6 to 99.6; 95% PrI 62.3 to 100.0) at a threshold of 2 to 73.9 (95% CrI 53.8 to 91.2; 95% PrI 51.6 to 98.0) at a threshold of 150. Specificity ranged from 78.1 (95% CrI 70.0 to 86.0; 95% PrI 62.8 to 91.8) at a threshold of 2 to 96.4 (95% CrI 92.6 to 98.9; 95% PrI 87.1 to 99.7) at a threshold of 150.

Main analysis: QuikRead go

No end-to-end studies were identified. One study[58] reporting diagnostic test accuracy data on QuikRead go recruited patients exclusively from primary care referrals. Another study[80] reported diagnostic test accuracy data on patients recruited from both primary and secondary care and was included in the analysis of dual FIT under a tiered approach,

TABLE 7 Study and patient characteristics of FOB Gold studies

	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Comparison with scope	Mean/ median age in years	Patient characteristics • Male; • Ethnicity; • Anaemia status	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Subgroups
Population type 2 studies (NG12 high risk)									
1	Benton 2022[44] 50 NHS hospitals across England, UK October 2017 to December 2019 NICE FIT	FOB Gold Wide – SENTiFIT 270 Colonoscopy	NG12 high risk, who had colonoscopy. Randomised to cohort 1 who were given four tests	NR	NR	• NR • NR • NR	7/233 (3.00%)	2, 10, 100	None
Population type 4 (unclear/unrepresentative of all presenting to primary care)									
2	MacLean 2022[59] Royal Surrey Foundation Trust, UK July 2019 and March 2020	FOB Gold Wide SENTiFIT 270 Colonoscopy or CTC or flexisig[a]	2WW referrals	NR	NR	• 48.8% • NR • NR	14/553 (2.53%)	10, 100, 150	None
3	Jordaan 2023[83] Mid-Yorkshire NHS Trust, Wakefield, UK September 2018 to the December 2019	FOB Gold, Roche Cobas c501 analyser Records follow-up	Mainly DG30 low risk, but some NG12 high risk	NR	NR	• 48% • NR • NR	30/3349 (0.90%)	10	None

a Maclean 2022:[59] flexisig if presenting with perianal symptoms or anorectal bleeding.

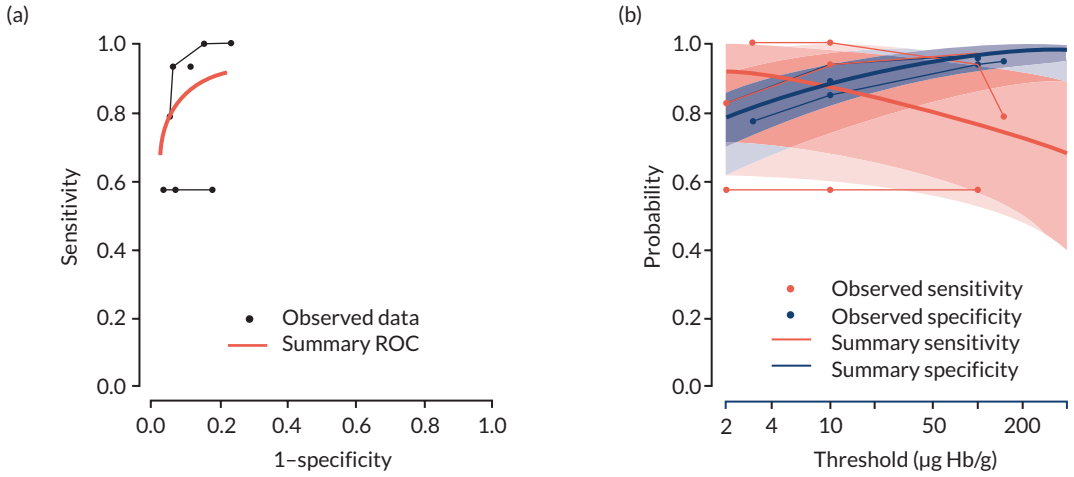

FIGURE 7 Observed data and summary sensitivity and specificity for the FOB Gold primary analysis. Ninety-five per cent CrIs and 95% PrIs for summary sensitivity are shown by the dark and light red regions. Ninety-five per cent CrIs and 95% PrIs for summary specificity are shown by the dark and light blue regions.

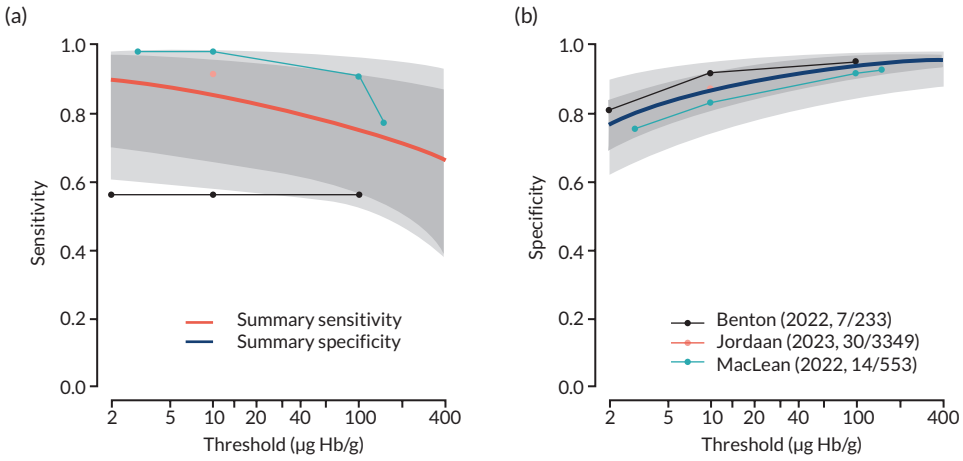

FIGURE 8 Observed data and summary sensitivity and specificity for the FOB Gold Primary analysis, with study information. Ninety-five per cent CrIs and 95% PrIs for summary sensitivity and specificity are shown by the dark and light grey regions.

as no data meeting the review inclusion criteria were identified for this test. It was not included in the analysis of single FIT as data meeting the inclusion criteria were available from MacLean *et al.*[58] Study and patient characteristics of both studies are given in *Table 8*.

Main analysis

One study[58] was included and analysed. The study was relatively small (n = 553) with a small number of CRC events (n = 14). It recruited NG12 high-/medium-risk patients referred to 2WW from primary care only. The study was conducted in the UK, had 49.9% male patients, had a CRC prevalence of 2.53%, used colonoscopy, CTC or flexible sigmoidoscopy as the reference standard and reported thresholds of 10, 100 and 150 µg/g. It was not clear if patients with rectal/anal mass or anal ulceration were included, and no diagnostic test accuracy data were reported for subgroups according to patient characteristics.

Population type 1 studies (all presenting to primary care)

There were no studies of this type.

Population type 2 studies (NG12 high/medium risk)

This analysis included the same study as the main analysis.

Population type 3 studies (DG30 low risk)

There were no studies of this type.

Population type 4 (unclear/unrepresentative of all presenting to primary care

There were no studies of this type.

Studies reporting subgroup data

There were no studies of this type.

Dual faecal immunochemical test using QuikRead go

No studies reporting diagnostic test accuracy of dual QuikRead go met the inclusion criteria for the review. One study[80] was included under a tiered approach to inclusion, which reported data for patients referred to colonoscopy from primary and secondary care. It was also relatively small (*n* = 242) with a small number of CRC events (*n* = 13). The study was conducted in Sweden, had 42.1% male patients, had a CRC prevalence of 5.37%, used colonoscopy as the reference standard and reported thresholds of 10, 12 and 20 µg/g. It was not clear if patients with rectal/anal mass or anal ulceration were included, and no diagnostic test accuracy data were reported for subgroups according to patient characteristics.

Quality assessment

Quality assessment of diagnostic test accuracy studies version 2[34] was applied to studies reporting diagnostic test accuracy data only. QUADAS-2 asks questions relating to internal validity (risk of bias) and external validity (applicability). Each is discussed separately in the following two paragraphs.

Risk of bias

Appendix 3, *Table 40* summarises the risk-of-bias and applicability scores as assessed by the authors of this review. Full risk-of-bias scores with reasons for scores are provided in *Appendix 3*. No study scored as low risk for all items. Patient selection was rated as being at risk of bias in both studies due to being unclear about or not recruiting a consecutive sample and for excluding patients without a definitive diagnosis or a colonoscopy. Owing to the inclusion criteria for the review, all studies avoided a case–control design. The index test and reference standards in both studies were rated as being at low risk of bias. Patient flow was rated as being at high risk because some patients were missing from the analysis.

Applicability

There were concerns about whether the patients recruited were representative of all those presenting to primary care in both studies, owing to the exclusion of some patients (study population types 2 and 4; those who did not have a colonoscopy). The index test and reference standard were rated as being at low risk of having poor applicability in both studies.

Statistical synthesis

As there was only one study of single FIT and one of dual FIT, no statistical synthesis was performed. The results of the single FIT study are presented in *Table 8*.

Main analysis: NS-Prime

No end-to-end studies were identified. Only one study[44] that met the inclusion criteria and reported data for NS-Prime was identified.

Main analysis

One UK study,[44] a subgroup of the NICE FIT study,[17,44,72,73] reported a subgroup of patients who met the NG12 high-/medium-risk criteria (see *Table 9*). The study was relatively small (*n* = 233) with a very small number of CRC events (*n* = 7), which may result in less precise estimates. It reported thresholds of 3, 10 and 100 µg/g using NS-Prime, had a CRC prevalence of 3.00% and used colonoscopy as the reference standard. It was not clear if rectal/abdominal mass

TABLE 8 Study and patient characteristics and diagnostic test accuracy of QuikRead go studies

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Patient characteristics (mean age in years, male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Sensitivity (95% CI)	Specificity (95% CI)
Population type 2 studies (NG12 high risk)								
1	MacLean 2021[58] Royal Surrey Foundation Trust, UK July 2019 and March 2020	QuikRead go Colonoscopy, CTC or flexisig	2WW NG12 high/ medium risk[a]	• Mean age NR • 49.9% • NR • All anaemia 12.8%; iron- or ferritin-deficient anaemia 4.5%	14/553 (2.53%)	10	92.90 (68.5 to 98.7)	70.10 (66.1 to 73.8)
						100	71.40 (45.4 to 88.3)	94.60 (92.4 to 96.2)
						150	57.10 (32.6 to 78.6)	95.90 (93.9 to 97.3)
Dual FIT								
2	Tsapournas 2020[80] Four endoscopy units in Sweden[a] November 2013 to March 2017	QuikRead go Colonoscopy	Type 4: referred for colonoscopy from primary or secondary care	• Median 65 (range 20–87) • 42.1% • NR • NR • Medications[a]	13/242 (5.37%)	See *Main analysis: dual faecal immunochemical test*		

a Tsapournas 2020:[80] Eskilstuna General district hospital, Orebro University hospital, Aleris Handen and Hotorget Endoscopy centre, Stockholm. Medications taken by participants reported as Trombyl (aspirin) 23 (9.5%), warfarin 12 (5.0%), others and combinations 8 (3.3%); Maclean 2021:[58] population confirmed with author, FIT was not being used by GPs during recruitment period, so only NG12 high-/medium-risk patients were referred.

and anal ulceration patients were included, nor was the percentage of male patients reported for this subgroup. No diagnostic test accuracy data were reported for patient characteristic subgroups for the NG12 subgroup.

Population type 1 studies (all presenting to primary care)

There were no studies of this type.

Population type 2 studies (NG12 high/medium risk)

This analysis included the same study as the main analysis.

Population type 3 studies (DG30 low risk)

There were no studies of this type.

Population type 4 (unclear/unrepresentative of all presenting to primary care)

There were no studies of this type.

Studies reporting subgroup data

There were no studies of this type.

Dual faecal immunochemical test using NS-Prime

There were no studies of this type.

Quality assessment

Quality assessment of diagnostic test accuracy studies version 2[34] was applied to studies reporting diagnostic test accuracy data only. QUADAS 2 asks questions relating to internal validity (risk of bias) and external validity (applicability). Each is discussed separately in the following two paragraphs.

Risk of bias

Appendix 3, *Table 41* summarises the risk or bias and applicability scores as assessed by the authors of this review. Full risk-of-bias scores with reasons for the scores are provided in *Appendix 3*. Only one study was included in this analysis. It scored low risk for the index test and patient flow items. Patient selection was rated as being at risk of bias because a consecutive sample was not recruited, and because patients were asked to complete four tests, which may have excluded a spectrum of patients, and only population type 2 patients were included. Owing to the inclusion criteria for the review, all studies avoided a case–control design. The reference standard was rated as being at unclear risk of bias because it was unclear if the reference standard was interpreted blind to the index test.

Applicability

There were concerns about whether the patients recruited were representative of all those presenting to primary care, owing to the problems described in the previous paragraph. The index test and reference standards were rated as being at low risk of having poor applicability in all three studies.

Statistical synthesis

As there was only one study, no statistical synthesis was performed. The results of the study are presented in *Table 9*.

Main analysis: IDK tests

IDK Turbi faecal immunochemical test

No end-to-end studies were identified and no diagnostic test accuracy data on patients presenting to primary care with symptoms of CRC were identified for this test. The company submission included an analysis that provided 2 × 2 tables for IDK TurbiFIT compared with *IDK*® Hämoglobin ELISA,[87] but no details were given of the patient population; only a simple analysis of agreement between devices in a small sample was given (*n* = 45), and this was only available at

TABLE 9 Study and patient characteristics and diagnostic test accuracy of the NS-Prime study

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Patient characteristics (mean age in years, male, ethnicity, anaemia status)	*N* with CRC/*N* analysed (%)	Threshold, µg/g	Sensitivity (95% CI)	Specificity (95% CI)
Population type 2 studies (NG12 high risk)								
1	Benton 2022[44] 50 NHS hospitals across England, UK October 2017 to December 2019 NICE FIT	NS-Prime colonoscopy	NG12 high risk, who had colonoscopy. Randomised to cohort 1 who were given four tests	• NR • NR • NR • NR	7/233 (3.00%)	3	85.70 (48.7 to 97.4)	31.90 (26.1 to 38.2)
						10	71.40 (35.9 to 91.8)	83.60 (78.2 to 87.9)
						100	57.10 (25.1 to 84.2)	97.30 (94.3 to 98.8)

N, number.

one cut-off point (10 µg/g). In this small sample, some disagreement was seen between devices in the clinical decisions that would be made at a cut-off point of 10 µg/g, and absolute values were quite different in some samples (e.g. 12.18 compared with 40.51). No assessment of agreement was performed, for example using a concordance correlation coefficient or Bland–Altman plot, and no indication was given of what a clinically acceptable level of disagreement would be. It should also be noted that IDK TurbiFIT and IDK Hämoglobin ELISA use different test methodologies (immunoturbidimetry and ELISA plates, respectively; see *Definition of the intervention*), and no evidence was provided that linked the IDK Hämoglobin ELISA or the IDK TurbiFIT to the test used in the clinical study[79] to support IDK Hämoglobin ELISA. A formal recommendation of equivalence could not be given based on the evidence provided.

IDK Haemoglobin, Hb/Hp and Hb+Hb/Hp tests

No end-to-end studies were identified. Only one study[79] reported diagnostic test accuracy data for IDK Haemoglobin (human) and haemoglobin/haptoglobin complex ELISA tests. This study was conducted in 1999 using a non-commercialised version of this test. IDK has assured the EAG that the data are generalisable to their current test, but it should be noted that no data were offered to support this assertion, and the EAG could not validate this statement. There are studies available in screening populations, but these were outside the scope of this assessment as diagnostic test accuracy is expected to differ between asymptomatic populations and symptomatic populations.

Main analysis

One German study[44] reported data on symptomatic primary care patients referred to secondary care, but the inclusion criteria were otherwise unclear. The study was relatively small (n = 621), with a small number of CRC events (n = 23). It reported data for both Hb alone and the complex Hb/Hp. Some data were also reported for Hb + Hb/Hp, but it was not possible to extract true positives, true negatives, false positives and false negatives for this test, nor the sensitivity and specificity. Immunodiagnostik proposed an equation to calculate the sensitivity and specificity of the Hb + Hb/Hp test based on the sensitivity and specificity of each test separately. The EAG notes that this equation is usually thought to be valid when the tests are independent, which is not thought to be the case with these two tests, and therefore the EAG has not used this equation. The study had a CRC prevalence of 3.70% and used colonoscopy as the reference standard. It was not clear if rectal/abdominal mass and anal ulceration patients were included, and the percentage of male patients was 45.1%. No diagnostic test accuracy data were reported for subgroups according to patient characteristics.

Population type 1 studies (all presenting to primary care)

There were no studies of this type.

Population type 2 studies (NG12 high/medium risk)

There were no studies of this type.

Population type 3 studies (DG30 low risk)

There were no studies of this type.

Population type 4 (unclear/unrepresentative of all presenting to primary care)

This analysis included the same study as the main analysis.

Studies reporting subgroup data

There were no studies of this type.

Dual faecal immunochemical test using IDK Hb, Hb/Hp

There were no studies of this type.

Quality assessment

Quality assessment of diagnostic test accuracy studies version 2[34] was applied to studies reporting diagnostic test accuracy data only. QUADAS 2 asks questions relating to internal validity (risk of bias) and external validity (applicability). Each is discussed separately in the following two paragraphs.

Risk of bias

Appendix 3, *Table 42* summarises the risk of bias and applicability scores as assessed by the authors of this review. Full risk-of-bias scores with reasons for the scores are provided in *Appendix 3*. Only one study was included in this analysis, which reported data for two tests, IDK Hb and IDK Hb/Hp complex. It scored as high risk for patient selection as it was unclear if patients were recruited consecutively, and only those referred to colonoscopy were included. It scored as low risk for the index test for the Hb test, but not for the Hb/Hp complex, as the threshold was derived to optimise accuracy, rather than being prespecified. It scored as low risk for the reference standard but as unclear risk for patient flow as it did not state the interval between index test and reference standard and did not state how many patients were recruited and analysed.

Applicability

There were concerns about whether the patients recruited were representative of all those presenting to primary care, owing to the problems described in the previous paragraph. The index test was rated as being at high risk for poor applicability because no data relating to the equivalence of this test and the commercial version were presented for symptomatic patients. The reference standard was rated as being at low risk of having poor applicability.

Statistical synthesis

As there was only one study, no statistical synthesis was performed. The results of the study are presented in *Table 10*.

Main analysis: dual faecal immunochemical test

No end-to-end studies were identified. Four studies[51,75,80,81] reported data on dual FIT, defined as using two samples from different bowel movements to guide referral rather than a single sample from one bowel movement (see *Table 11*). This is distinct from the use of a second/repeat FIT during the follow-up of patients, that is, after a decision to refer or not refer has been made.[23] Two studies used HM-JACKarc,[75,81] one study used OC-Sensor[51] and one study used QuikRead go.[80] Two studies[75,81] also provided single-FIT data for their main analyses. One dual OC-Sensor study[51] reported data only for dual FIT, and one QuikRead go study[80] did report data for single FIT, which has been included to aid comparison but was excluded from the main analysis as the cohort included patients referred from secondary care as well as primary care. It is unclear to what extent this might affect generalisability. The characteristics of the studies are presented in full in the corresponding tables in *Rationale for the analysis plan*, *The analysis plan* and *Main analysis: QuikRead go*.

TABLE 10 Study and patient characteristics of IDK Haemoglobin (human) and haemoglobin/haptoglobin complex ELISA tests

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Patient characteristics: (mean age in years; male; ethnicity; anaemia status)	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Sensitivity (95% CI)	Specificity (95% CI)
Population type 4 studies (unclear/unrepresentative of patients presenting to primary care):								
1	Sieg 1999[79] Ostringen, Germany NR, prior to publication in 1999	Immunological test for HB Colonoscopy	Referred to secondary care	• Median 59 (range 15–85) • 45.1% • NR • NR	23/621 (3.70%)	2	87.0 (84.4 to 89.6)	88.1 (85.6 to 90.6)
		Immunological test for Hb/Hp complex Colonoscopy					82.6 (79.6 to 85.6)	80.8 (77.7 to 83.9)

Dual FIT can be interpreted as positive either if both tests are positive ('both' strategy) or if either test is positive ('either' strategy). The 'both' strategy is likely to increase specificity, whereas the 'either' strategy is likely to increase sensitivity. As the test is being used as a 'rule-out' test to triage patients to secondary care, it is most useful to maximise sensitivity [the 'Sensitive test when Negative rules OUT the disease' (SNOUT) rule],[88] and this is how clinicians indicated that the test would be used during the scoping process for this assessment. Therefore, the EAG has concentrated on data relating to the 'either positive' interpretation of the test in the synthesis, but presents all data for reference.

One study[51] reported results for both interpretations at 10 µg/g, and in this analysis, sensitivity was better in the 'either' strategy than in the 'both strategy', while specificity was higher in the 'both' strategy than in the 'either' strategy. Another study[81] reported the optimal threshold for the 'either' and the 'both' strategies, showing the same pattern for sensitivity and specificity but at different thresholds (43 µg/g for the 'either' strategy and 2 µg/g for the 'both' strategy).

Two studies[75,80] reported data for both dual FIT ('either' strategy) and single FIT. In these studies, at 10 µg/g sensitivity was better in the dual FIT 'either' strategy than single FIT, and specificity was worse, and in the one study[80] that reported multiple thresholds, this general trend continued at higher thresholds.

One study[75] reported data at 10 µg/g for IBD and AA as well as for CRC in both dual FIT ('either' strategy) and single FIT. Dual FIT had similar sensitivity to but lower specificity than single FIT for IBD, and higher sensitivity but lower specificity for AAs, which equated to a 29.8% reduction in missed pathologies.

Quality assessment

A synthesis of quality assessment was not conducted for this subgroup of studies.

Statistical synthesis

No statistical synthesis was conducted, as the assessment considered each test individually, and although there were two HM-JACKarc studies, threshold points were insufficient for a meaningful synthesis to be conducted.

Additional analysis 1: synthesis of all tests together in a single analysis

This analysis was run to allow the investigation of the impact of study population type on a larger sample of studies and because these factors were unlikely to interact with test type. It was also used to inform the priors used when < 5 studies were being synthesised (see *Appendix 2*).

Studies were split into separate subgroups according to the population type recruited (type 1, 2 or 3). An analysis was also conducted for population types 1–3 together to exclude studies that may be enriched with patients recruited on the basis of a positive FIT result. The results of the analysis can be found in *Appendix 5*.

For the analyses of subgroups by population type, the summary estimates were similar and not statistically significant based on overlap of the 95% CrI. The summary specificity for population 3 is higher than for the other considered subgroups; however, this analysis was based on only three studies (HM-JACKarc, *n* = 2; OC-Sensor, *n* = 1). There is therefore considerable uncertainty in the pooled estimates, and these should be interpreted with caution.

Additional analysis 2: reference standard sensitivity analysis

Subgroup analyses were conducted that included only the studies in which at least 90% of the participants received colonoscopy as the reference standard. Subgroups analyses were performed for all FITs together, including all population types, excluding population type 4 studies, and separately for each test, where data allowed. The results of the analysis can be found in *Appendix 5*.

For all analyses the summary estimates were similar, irrespective of the reference standard grouping (all studies vs at least 90% of the participants receiving colonoscopy). The largest difference in point estimates was seen for specificity of OC-Sensor (see *Appendix 5*, *Figure 18f*); however, there were only three studies in the > 90% colonoscopy subgroup and so the apparent difference may be explained by other sources of heterogeneity between the studies. There was very little difference in specificity for the HM-JACKarc studies (see *Appendix 5*, *Figure 18d*).

TABLE 11 Study and patient characteristics of studies reporting data for dual FIT

Author, year	*N* with CRC (or IBD or AA)/*N* in analysis	Either positive			Both positive			Single FIT			
		Thr	Sensitivity (95% CI)	Specificity (95% CI)	Thr	Sensitivity (95% CI)	Specificity (95% CI)	*N* with CRC (or IBD or AA)/*N* in analysis	Thr	Sensitivity (95% CI)	Specificity (95% CI)
HM-JACKarc											
Gerrard 2023[75]	CRC: 88/2637 (3.34%)	10	96.60 (90.4 to 99.3)	71.2 (69.4 to 73.0)	NR	NR	NR	CRC: 135/3426 (3.94%)	10	93.3 (87.7 to 96.9)	78.0 (76.6 to 79.4)
	IBD: 33/2637 (1.39%)		90.90 (75.7 to 98.1)	69.70 (67.9 to 71.5)	NR	NR	NR	55/3426 (1.61%)		90.90 (80.0 to 97.0)	76.30 (74.8 to 77.7)
	AA: 97/2637 (3.68%)		68.00 (57.8 to 77.1)	70.40 (68.5 to 72.1)	NR	NR	NR	136/3426 (4.00%)		54.4 (45.6 to 63.0)	76.40 (75.0 to 77.9)
Turvill 2018[81]	27/476 (5.67%)	43	87.50 (NR)	90.70 (NR)	2	91.70	85.20	CRC: 27/505 (5.35%)	12	84.60	88.50
OC-Sensor											
Hunt 2022[51]	317/28,622 (1.11%)	10	98 (95.5 to 98.9)	66.20 (65.7 to 66.7)	10	92 (87.9 to 94.1)	81.60 (81.1 to 82.0)	NA	NA	NR	NR
QuikRead go											
Tsapournas 2020[80]	13/242 (5.37%)	10	100.00 (NR)	71.40 (65.5 to 77.3)		NR	NR	CRC: 13/242 (5.37%)	10	92.30 (77.8 to 100)	77.30 (71.9 to 82.7)
		15	92.30 (77.8 to 100)	76.80 (71.3 to 82.3)		NR	NR		15	92.30 (77.8 to 100)	81.70 76.7 to 86.7)
		20	92.30 (77.8 to 100)	81.70 (76.6 to 86.8)		NR	NR		20	84.60 (65.0 to 100)	86.50 (82.1 to 90.9)

Thr, threshold in µg/g; NA, not applicable; NR, not reported.

Main analysis: summary

As described in the previous sections, diagnostic test accuracy was similar across analyses that compared different population classifications and the reference standard used (all studies vs. at least 90% of the participants receiving colonoscopy). The analyses used to inform the model are therefore based on all included studies and undertaken individually by test type.

The key results for each test type are illustrated in *Figure 9*, with 95% CrI and PrI shown in dark and light grey, respectively, from the analysis including all FIT types. Estimated summary sensitivity and specificity at selected thresholds are presented in *Table 12* for tests that were statistically synthesised, and also for tests for which there was only one study (note that for these tests, the error is the 95% CI for a single study, rather than the 95% CrI of the summary estimate from the meta-analysis model, and these should not be directly compared). A formal comparison of the different test types was not conducted. The negative predictive value and positive predictive value of selected thresholds are given in *Report Supplementary Material 3*.

Comparative diagnostic test accuracy studies

Three studies conducted a comparison of two or more tests. Chapman *et al.*[46] reported on OC-Sensor DIANA and HM JACKarc, Benton *et al.*[44] compared HM-JACKarc, OC-Sensor PLEDIA, FOB Gold Wide/SENTiFIT 270, and NS-Prime, and MacLean *et al.*[59] compared FOB Gold Wide and QuikRead go. No one test appeared in all three comparisons.

Table 13 summarises the study characteristics and reports the sensitivity and specificity at a threshold of 10 µg/g and the conclusions drawn by the study authors. The remaining threshold data can be found in *Report Supplementary Material 2*. The largest study included 38 CRC patients among a sample of 732.[46] Both other studies[44,59] had relatively small sample sizes and CRC events (see *Table 13*).

Different sensitivities and specificities were reported across the tests, and all three studies concluded that at least one test was different from another (see column 10 of *Table 13*). Owing to the small number of CRC events in two of the trials, the small number of studies and the lack of a common comparator, it is difficult to draw any conclusions regarding the comparative performance of the tests, or what and whether different FIT cut-off values are required for each test based on these results. Benton *et al.*,[44] who performed an analysis of four tests, note that more work is required to understand the clinical impact of the use of different tests.

Subgroup analyses by patient characteristics

An exploration of the possible reasons for heterogeneity in diagnostic test accuracy across studies using meta-regression was considered. However, study-level covariates relating to patient characteristics of interest had not been reported in sufficient studies. Instead, studies that reported diagnostic test accuracy for subgroups of patients were considered in subgroup analyses. The subgroups were anaemia, age, sex, medications that might cause GI bleeding,

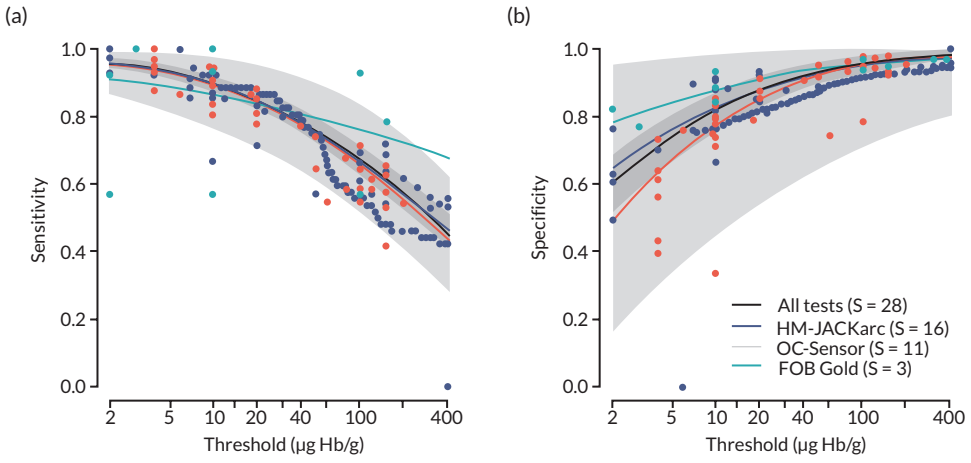

FIGURE 9 Sensitivity and specificity for all tests. Ninety-five per cent CrIs and 95% PrIs for summary sensitivity and specificity of analysis including all tests are shown by the dark and light grey regions.

TABLE 12 Summary sensitivity and specificity at selected thresholds

Threshold (µg/g)	HM-JACKarc (*S* = 16)[a]		OC-Sensor (*S* = 11)[a]		FOB Gold (*S* = 3)[a]		All tests (*S* = 28)[a]		QuikRead go (*S* = 1)[b]		NS-Prime (*S* = 1)[b]		IDK Hb (*S* = 1)[b]		IDK Hb/Hp (*S* = 1)[b]	
	Sens (95% CrI)	Spec (95% CrI)	Sens (95% CrI)	Spec (95% CrI)	Sens (95% CrI)	Spec (95% CrI)	Sens (95% CrI)	Spec (95% CrI)	Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)
2	95.9 (92.7 to 97.9)	65.1 (55.6 to 74.8)	NR	NR	91.4 (71.6 to 99.6)	78.1 (70 to 86)	96.5 (94.8 to 97.8)	58.7 (49.9 to 67.4)	NR	NR	NR	NR	87 (84.4 to 89.6)[b]	88.1 (85.6 to 90.6)[b]	82.6 (79.6 to 85.6)[b]	80.8 (77.7% to 83.9)[b]
2.5	95.3 (91.8 to 97.5)	68 (58.8 to 77.3)	NR	NR	90.9 (71.1 to 99.5)	79.9 (71.9 to 87.5)	96 (94.1 to 97.4)	62.3 (53.7 to 70.7)	NR	NR	NR	NR	NR	NR	NR	NR
3	94.7 (91.1 to 97.2)	70.3 (61.3 to 79.3)	NR	NR	90.5 (70.6 to 99.4)	81.2 (73.4 to 88.6)	95.5 (93.5 to 97.1)	65.1 (56.8 to 73.3)	NR	NR	85.70 (48.7 to 97.4)[b]	31.90 (26.1 to 38.2)[b]	NR	NR	NR	NR
4	93.8 (89.8 to 96.5)	73.7 (65.1 to 82.2)	94.2 (91.2 to 96.7)	62.7 (47.4 to 77.2)	89.8 (69.8 to 99.2)	83.2 (75.6 to 90.2)	94.6 (92.4 to 96.4)	69.4 (61.4 to 77.1)	NR	NR	NR	NR	NR	NR	NR	NR
7	91.4 (86.8 to 94.8)	79.6 (71.7 to 87.1)	91.8 (88.2 to 94.9)	72.3 (58.1 to 84.8)	88.2 (68.4 to 98.7)	86.5 (79.5 to 92.8)	92.3 (89.7 to 94.6)	76.8 (69.7 to 83.5)	NR	NR	NR	NR	NR	NR	NR	NR
10	89.5 (84.6 to 93.4)	82.8 (75.2 to 89.6)	89.8 (85.9 to 93.3)	77.6 (64.3 to 88.6)	87 (67.3 to 98.3)	88.4 (81.7 to 94.2)	90.4 (87.6 to 93)	80.8 (74.3 to 86.8)	92.90 (68.5 to 98.7)[b]	70.10 (66.1 to 73.8)[b]	71.40 (35.9 to 91.8)[b]	83.60 (78.2 to 87.9)[b]	NR	NR	NR	NR
20	84.7 (79.1 to 89.6)	87.9 (81.1 to 93.4)	84.7 (80.3 to 89)	85.6 (74.5 to 93.6)	84.5 (65.1 to 97.1)	91.3 (85.4 to 96.2)	85.6 (82.3 to 88.8)	87.1 (81.6 to 91.8)	NR	NR	NR	NR	NR	NR	NR	NR
50	75.8 (69.4 to 82)	92.6 (87 to 96.5)	75 (70.2 to 80)	92.5 (84.3 to 97.3)	80.3 (61.3 to 94.7)	94.2 (89.3 to 97.8)	76.4 (72.5 to 80.3)	92.6 (88.5 to 95.8)	NR	NR	NR	NR	NR	NR	NR	NR
100	67 (60 to 74.2)	94.9 (90.3 to 97.8)	65.3 (60.2 to 70.7)	95.5 (89.4 to 98.6)	76.4 (57.2 to 92.5)	95.7 (91.6 to 98.6)	67.1 (62.8 to 71.4)	95.3 (92.1 to 97.5)	71.40 (45.4 to 88.3)[b]	94.60 (92.4 to 96.2)[b]	57.10 (25.1 to 84.2)[b]	97.30 (94.3 to 98.8)[b]	NR	NR	NR	NR
120	64.5 (57.2 to 71.9)	95.4 (91 to 98.1)	62.5 (57.2 to 68)	96.1 (90.4 to 98.9)	75.3 (55.8 to 91.9)	96.1 (92.1 to 98.8)	64.4 (60 to 68.7)	95.8 (92.8 to 97.8)	NR	NR	NR	NR	NR	NR	NR	NR
150	61.3 (53.7 to 68.9)	96 (91.9 to 98.4)	58.9 (53.4 to 64.7)	96.7 (91.6 to 99.1)	73.9 (53.8 to 91.2)	96.4 (92.6 to 98.9)	60.9 (56.3 to 65.4)	96.4 (93.6 to 98.2)	57.10 (32.6 to 78.6)[b]	95.90 (93.9 to 97.3)[b]	NR	NR	NR	NR	NR	NR
200	57 (48.9 to 64.9)	96.6 (92.8 to 98.7)	54.2 (48.4 to 60.2)	97.3 (92.9 to 99.3)	NR	NR	56.3 (51.4 to 61)	97 (94.6 to 98.5)	NR	NR	NR	NR	NR	NR	NR	NR
400	46.3 (37.4 to 54.9)	97.7 (94.7 to 99.2)	NR	NR	NR	NR	44.8 (39.3 to 50)	98.1 (96.3 to 99.1)	NR	NR	NR	NR	NR	NR	NR	NR

Sens, sensitivity; Spec, specificity; *S*, number of studies.

a Summary estimates from the meta-analysis model.

b Individual study estimates. Estimates of error for these studies appear comparatively narrower than those from the synthesis of multiple studies due to being derived from one study only. The number of patients included in each study was QuikRead go (type 2 study), *n* = 553; NS-Prime (type 2 study), *n* = 233; IDK tests (type 4 study), *n* = 621.

ethnicity and people with blood disorders (other than anaemia) that might affect FIT. The analyses are reported in *Appendix 6*, *Subgroup analyses by patient characteristics*.

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.

Advanced adenomas and inflammatory bowel disease outcomes

Studies that reported the diagnostic test accuracy of FIT for CRC and AA or IBD were included in the review. Data relating to these analyses are reported in *Appendix 6*, *Advanced adenomas and inflammatory bowel disease outcomes*.

Other outcomes

Test uptake and repeat tests

Because these outcomes are likely to be affected by the point in the care pathway at which FIT is issued to the patient, this analysis has been restricted to studies in which FIT was issued in primary care. All dual FIT studies were conducted in secondary care, and they have been included as no other data were available. The data are summarised in *Appendix 6*, *Test uptake and repeat tests*, including *Table 57*. Additional data with lower generalisability (studies in secondary care settings and that asked patients to provide samples for multiple tests) are provided in *Report Supplementary Material 4*.

'Time to' and other outcomes

Data from the diagnostic test accuracy studies relating to time to diagnosis, time to FIT return by patient, time to FIT analysis, time to investigation in secondary care and time between samples (dual FIT) are reported in *Appendix 6*, *'Time to' outcomes*, including *Table 58*. Other outcomes listed in the scope, such as stage at diagnosis and AEs, are also reported there.

Patient perspectives

Data from diagnostic test accuracy studies that also reported outcomes including patient views on the acceptability of FIT, expressions of patient preference for FIT versus colonoscopy, and the experience of, and satisfaction with, FIT among patients with suspected CRC symptoms are reported and synthesised in *Appendix 6*, *Patient perspectives*, including *Tables 59–62*.

Sociodemographic factors

Data from diagnostic test accuracy studies that also reported on sociodemographic factors and FIT return rates are reported in *Appendix 6*, *Sociodemographic factors*, including *Tables 63* and *64*.

Data selected to enter the cost-effectiveness model

From the analyses, the EAG concluded that the assumption that tests should be considered separately was supported by the comparative diagnostic test accuracy studies, one of which notes that more work is required to understand the clinical impact of the use of different tests. The EAG also concluded that it was not necessary to exclude studies potentially enriched by FIT positives, and that it was not necessary to exclude studies in which < 90% of patients received colonoscopy as the reference standard as the estimates were largely similar. Finally, the EAG concluded that as the estimates by population type were similar and had overlapping CrIs, the same estimate of sensitivity could be assumed for FIT used in all patients presenting to primary care, compared with FIT used in DG30 low-risk patients (i.e. the current care arm of the model, where FIT is used only in DG30 patients). The impact of this assumption on DG30 low-risk patients was tested in a scenario analysis in the economic model (see *Scenario analyses*).

TABLE 13 Sensitivity and specificity reported in studies comparing different tests within the same patients

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	*N* with CRC/*N* analysed (%)	Thresholds, µg/g	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	Conclusion drawn by study authors
1	Chapman 2021[46] Nottingham University Hospitals Trust, UK September 2016 to September 2017	• HM JACKarc + HM JACKarc analyser 2WW investigations	2WW patients who returned 2 types of FIT	38/732 (5.19%)	4, 10, 22.6, 150	10	89.00 (75 to 97)	74.00 (70 to 77)	Using OC-S results in higher referrals. Consequently, OC-S detected more cancers than HM-J for the same cut-off values. Suggest that analyser-specific f-Hb cut-off values are needed, especially at lower f-Hb
		• OC-Sensor DIANA					84.00 (69 to 94)	78.00 (75 to 81)	
2	Benton 2022[44] 50 NHS hospitals across England, UK October 2017 to December 2019 NICE FIT	• HM-JACKarc colonoscopy	NG12 high risk, who had colonoscopy. Randomised to cohort 1 who were given four tests	7/233 (3.00%)	LoD, 10, 100	10	57.10 (25.1 to 84.2)	84.50 (79.2 to 88.6)	At 10 µg/g, < half the number of referrals would be made using SENTiFIT 270/FOB Gold Wide system compared to the other methods and dramatically fewer at the LoD. The calibration for the SENTiFIT 270/FOB Gold Wide gives lower f-Hb results than the other three systems. Supported by Bland–Altman difference plot. Further work is required to understand the clinical impact of these differences and to minimise them
		• OC-Sensor PLEDIA				10	71.40 (35.9 to 91.8)	85.80 (80.7 to 89.8)	
		FOB Gold Wide – SENTiFIT 270				10	57.10 (25.1 to 84.2)	93.40 (89.3 to 95.9)	
		• NS-Prime				10	71.40 (35.9 to 91.8)	83.60 (78.2 to 87.9)	
3	MacLean 2022[59] Royal Surrey Foundation Trust, UK July 2019 and March 2020	• FOB Gold Wide SENTiFIT 270 • Colonoscopy or CTC or flexisig	2WW NG12 high/medium risk	14/553 (2.53%)	10, 100, 150	10	100.00 (78.5 to 100)	84.80 (81.5 to 87.6)	Good agreement around negative threshold, but more patients would be triaged to further colonic investigation if using QuikRead go®
		• QuikRead go				10	92.90 (68.5 to 98.7)	70.10 (66.1 to 73.8)	

LoD, limit of detection; *N*, number; OC-S, OC-Sensor.

Chapter 3 Cost-effectiveness

This chapter presents a systematic review of economic analyses of FIT for symptomatic patients suspected of CRC, and the methods and results of a de novo health economic model developed by the EAG comparing the different strategies that include FIT.

Review of existing published health economic analyses

The EAG conducted systematic literature reviews to identify existing economic evaluations of the use of FIT and studies reporting utility estimates in people presenting to primary care with symptoms of CRC. The main focus of this review was to explore methodological choices made in previous economic evaluations and their potential relevance to the current decision problem and the model being developed by the EAG, rather than to assess the individual results of published economic evaluations.

The methods used by the EAG, including the searches, eligibility criteria and screening process, are reported in *Appendix 7*, *Cost-effectiveness and health-related quality of life review: methods*. The results of the review are presented in *Appendix 7*, *Cost-effectiveness and health-related quality of life review results: summary of studies identified*, which includes the PRISMA flow diagram (see *Appendix 7*, *Figure 21*), the analytic scope, modelling approaches and other relevant information from the included studies (see *Appendix 7*, *Tables 65* and *66*). Tables of excluded studies from the reviews and the results of the quality assessment of the included studies can be found in *Report Supplementary Material 4*.

Review and critical appraisal of economic analyses provided by test manufacturers

No economic analyses were provided to the EAG by the FIT manufacturers.

Independent economic evaluation

Scope of the Evidence Assessment Group economic analysis

As part of this assessment, the EAG developed a de novo model programmed in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA). The model compares different diagnostic strategies that include quantitative FIT in a primary care setting for people with symptoms of CRC. The model assesses the health outcomes and costs associated with each strategy over a lifetime horizon from the perspective of the UK NHS and PSS. The scope of the EAG model is summarised in *Table 14*.

Population

The population was adults presenting for the first time at a GP surgery (primary care) with signs and symptoms suggestive of CRC. The population in the model includes patients in both the high-/medium-risk and the low-risk groups defined in NICE DG30 and NG12.[10,11] This population excludes patients defined in the NICE scope as having 'bypass symptoms' (very high-risk symptoms: rectal or anal mass, or anal ulceration) who are assumed to be directly referred to secondary care as urgent suspected cancer referrals.

TABLE 14 Scope of the EAG economic analysis

Scope Element	Description
Population	Adults presenting to primary care with GI symptoms or signs indicating a risk of CRC, excluding people with 'bypass symptoms', such as rectal or anal mass, or anal ulceration
Interventions being compared	Three different sets of interventions that include the use of quantitative FIT in primary care were compared; two of them explored a range of different thresholds. These include: • FIT for all patients using one threshold (t) in µg/g to determine referral decisions • FIT for all patients using two thresholds (t_{high} and t_{low}) in µg/g to determine referral decisions • NICE currently recommended diagnostic pathway, with NG12 high-/medium-risk patients being directly referred to the urgent suspected cancer pathway and FIT being offered only to DG30 low-risk patients using a threshold of 10 µg/g to determine referral decisions (as defined by DG30 and NG12)[10,11]
Primary health economic outcome	• Incremental cost per QALY gained • NMB
Perspective	NHS and PSS
Time horizon	Lifetime (36 years)
Discount rate	3.5% per annum for health outcomes and costs
Price year (currency)	2021–2 (£)

CRC, colorectal cancer; EAG, Evidence Assessment Group; FIT, faecal immunochemical test(s); NMB, Net monetary benefit; QALY, quality adjusted life year; NHS, National Health Service; PSS, Personal Social Services. t, threshold; t_{high}, higher threshold; t_{low}, lower threshold.

Interventions being compared

The model compares three different sets of interventions that include the use of quantitative FIT in primary care as part of testing strategies to guide the diagnostic and clinical management of patients with suspected CRC. These are as follows.

- **Intervention 1**: FIT is offered to all patients with a single threshold (t µg/g) used to determine subsequent referral decisions, with the range of FIT thresholds considered being determined by the evidence synthesis (see Chapter 2, *Results*).
- **Intervention 2:** FIT is offered to all patients with pairs of FIT thresholds (t_{high} and t_{low} µg/g) used to determine subsequent referral decisions, with the selection of threshold pairs being determined by the output of the evidence synthesis and the clinical opinion of the EAG's advisors.
- **Intervention 3:** NICE current recommendations are used as defined in DG30 and NG12,[10,11] with all high-/medium-risk patients directly referred to the urgent suspect cancer referral (hereafter referred to as the 2WW pathway), and DG30 low-risk patients offered a FIT with subsequent referral decisions for this group based on a FIT threshold of 10 µg/g.

Perspective, time horizon and discount rate

The economic analysis was undertaken from the perspective of the NHS and PSS considering a lifetime horizon. Unit costs were valued at 2021–2 prices expressed in Great British pounds (GBP). Health outcomes and costs were discounted at a rate of 3.5% per annum as recommended by NICE.[89] The model assesses the cost-effectiveness of a range of FIT strategies using ICERs, which are reported in terms of the cost per QALY gained for the intervention strategies versus the strategy that reflects current NICE recommendations. The model also assessed NMB and other outcomes of interest listed in *Chapter 1, Outcomes*.

Conceptualisation of the model

To develop a better understanding of the diagnostic pathways of patients with symptoms and signs of suspected CRC, the EAG engaged with multiple clinical advisors before starting the conceptualisation of the model. The EAG's clinical experts included 10 healthcare professionals, including a mix of academics, GPs, GI consultants and surgeons, registrars and biochemists, all with experience in CRC. A questionnaire was sent to all advisors, with seven replies received. The questions and a summary of the responses from the clinical experts are provided in *Report Supplementary Material 5*.

Their responses were used to inform some of the parameters of the model where insufficient information was available from other evidence sources, including the EAG's review of existing economic models and targeted reviews undertaken to populate model parameters.

Model structure

The general structure of the EAG's economic model follows a similar approach to that used in the NICE NG12 and DG30 appraisals,[10,11,90] and it is also broadly consistent with a study identified in the review of published economic evaluations.[91] The model is based on a hybrid structure, with a decision tree used to model the diagnostic pathways for a cohort of patients presenting to primary care with symptoms that indicate a risk of CRC. The model is structured to capture the results of investigations reflecting the diagnosis of CRC, but also of AAs and IBD. The decision tree component of the model has a short time horizon, which reflects the assumption that the whole diagnostic pathway will cover the time between the patient's initial presentation to primary care and the confirmation of their diagnosis. Schematic representations of the decision tree part of the model are shown in *Figures 10–12* (for interventions 1, 2 and 3, respectively). The decision tree is followed by state-transition models that estimate lifetime costs, life-years and QALYs for people according to their underlying disease state (*Figure 13*). The model logic is described in *Short-term decision-tree component of the model* and *Lifetime state transition component of the model*, and the assumptions and sources of parameters used are detailed in *Key Evidence Assessment Group model assumptions* and *Evidence sources used to inform the model parameters*, respectively.

Short-term decision-tree component of the model

The EAG model simulates the diagnostic management of patients who present to primary care with symptoms suggestive of CRC and includes patients who might have underlying CRC, IBD or AAs. IBD and AAs were also included as underlying disease states as they were considered to reflect other significant bowel pathologies relevant to the decision problem according to opinion from the EAG's clinical advisors. Patients enter the decision-tree component of the model according to their underlying disease, based on the prevalence rates for CRC, AAs and IBD or with no bowel disease, and are assumed to enter the model aged 64 years. Under interventions 1 and 2, all patients are invited to complete a FIT, while under intervention 3, only those patients in the DG30 low-risk group[11] are invited to complete the test, with the NG12 high-/medium-risk group patients being directly referred to the 2WW pathway in secondary care.[10]

Patients who complete a FIT and have a result above the threshold (t µg/g in intervention 1, t_{high} µg/g in intervention 2 or 10 µg/g in intervention 3) are directly referred to the 2WW pathway. Patients who do not complete the FIT or whose test result lies below t, t_{low} or 10 µg/g are assumed to receive 'safety netting'. For the purpose of this model, safety netting was defined as the possible subsequent diagnostic decisions made by the healthcare professional in primary care following a 'negative'/low FIT result, and includes patients receiving one of the following options: (1) referral to the 2WW pathway in secondary care; (2) referral to the non-urgent referral pathway in secondary care (hereafter referred to as the 18-week wait pathway, 18WW); (3) watch and wait (also known as watchful waiting, which consists of patients being monitored in primary care with symptoms reviewed by the GP or patient re-presentation if symptoms persist or worsen); or (4) invitation to receive a second FIT. For intervention 2 only, patients who complete the FIT and receive a result that lies between t_{high} and t_{low} are assumed to follow the 'intermediate group' pathways, which the model defines as the same pathway options available in safety netting with the exception of 'watch and wait', and with a higher proportion of patients being directly referred to 2WW (see more details about the pathways and parameters included in safety netting and the 'intermediate group' pathways in *Probability of following each of the pathways following faecal immunochemical test result*). The inclusion of the direct referrals to secondary care 2WW and 18WW pathways is intended to reflect patients for whom GPs still have important clinical concerns even after the FIT result is returned and referral for further investigations in secondary care is considered necessary.

Patients who are referred to 2WW and 18WW are assumed to receive diagnostic imaging and other tests at secondary care gastroenterology visits. In particular, patients are assumed to receive one of the following imaging investigations: colonoscopy, CTC, or 'other non-invasive investigations'. The EAG opted to model explicitly only colonoscopy and CTC as the EAG's clinical advisors considered these imaging modalities the most common tests used in lower GI referrals (provided in *Report Supplementary Material 5*) and also they had been included in previous economic evaluations.[90,91] A third option denoted 'other non-invasive investigations' is also modelled to account for the group of patients who would not receive any invasive imaging investigations due to advanced disease stage, frailty, older age or patient/clinician

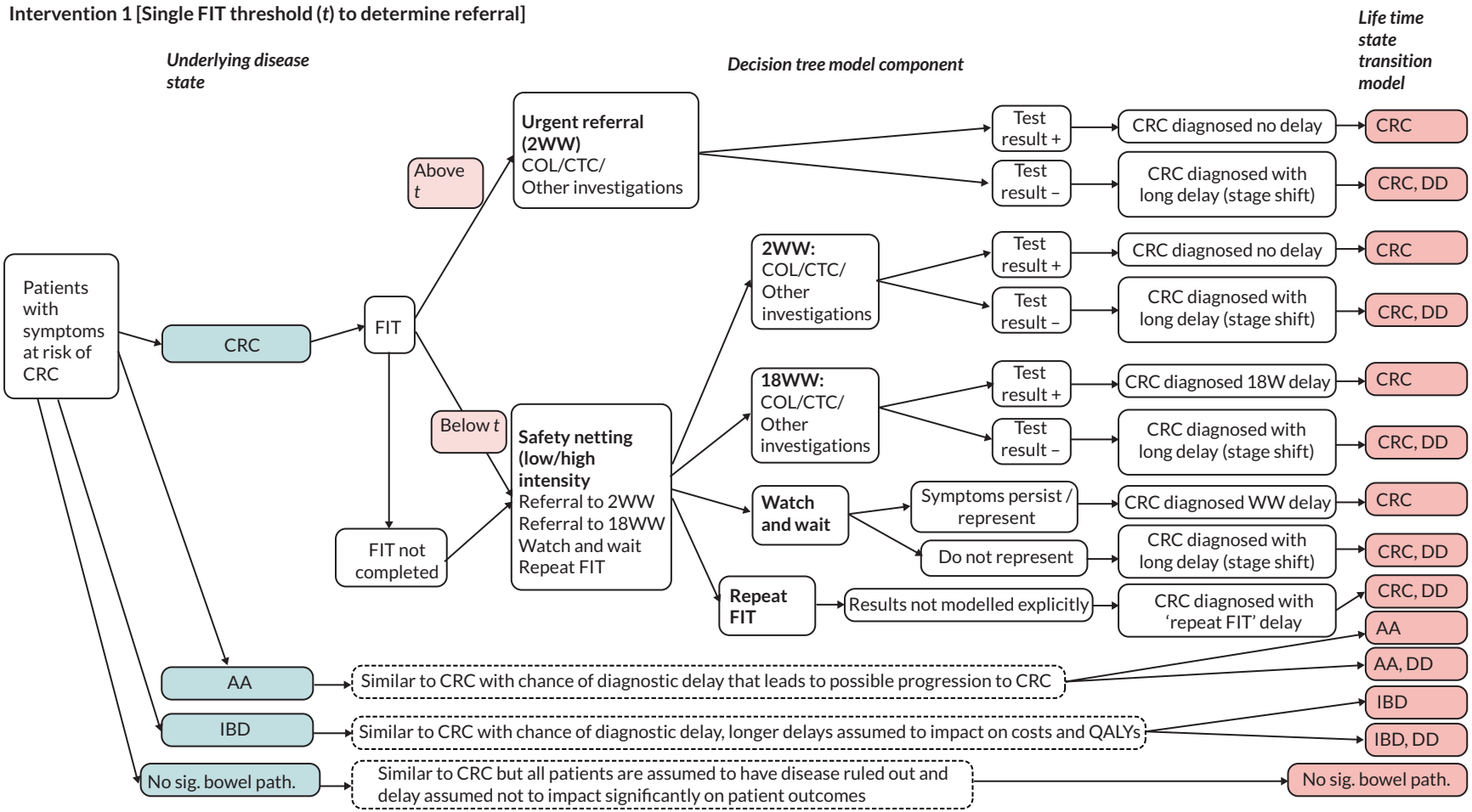

FIGURE 10 Evidence Assessment Group model: decision tree structure, intervention 1 (FIT with one threshold of *t* µg/g). *2WW*, 2 week wait; 18W, 18 weeks; 18WW, 18 week wait; AA, advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; CTC, computed tomography colonography; DD, delayed diagnosis (see *Time to diagnosis and diagnostic delays*); FIT, faecal immunochemical test; IBD, inflammatory bowel disease; *t*, threshold; WW, watch and wait conduct.

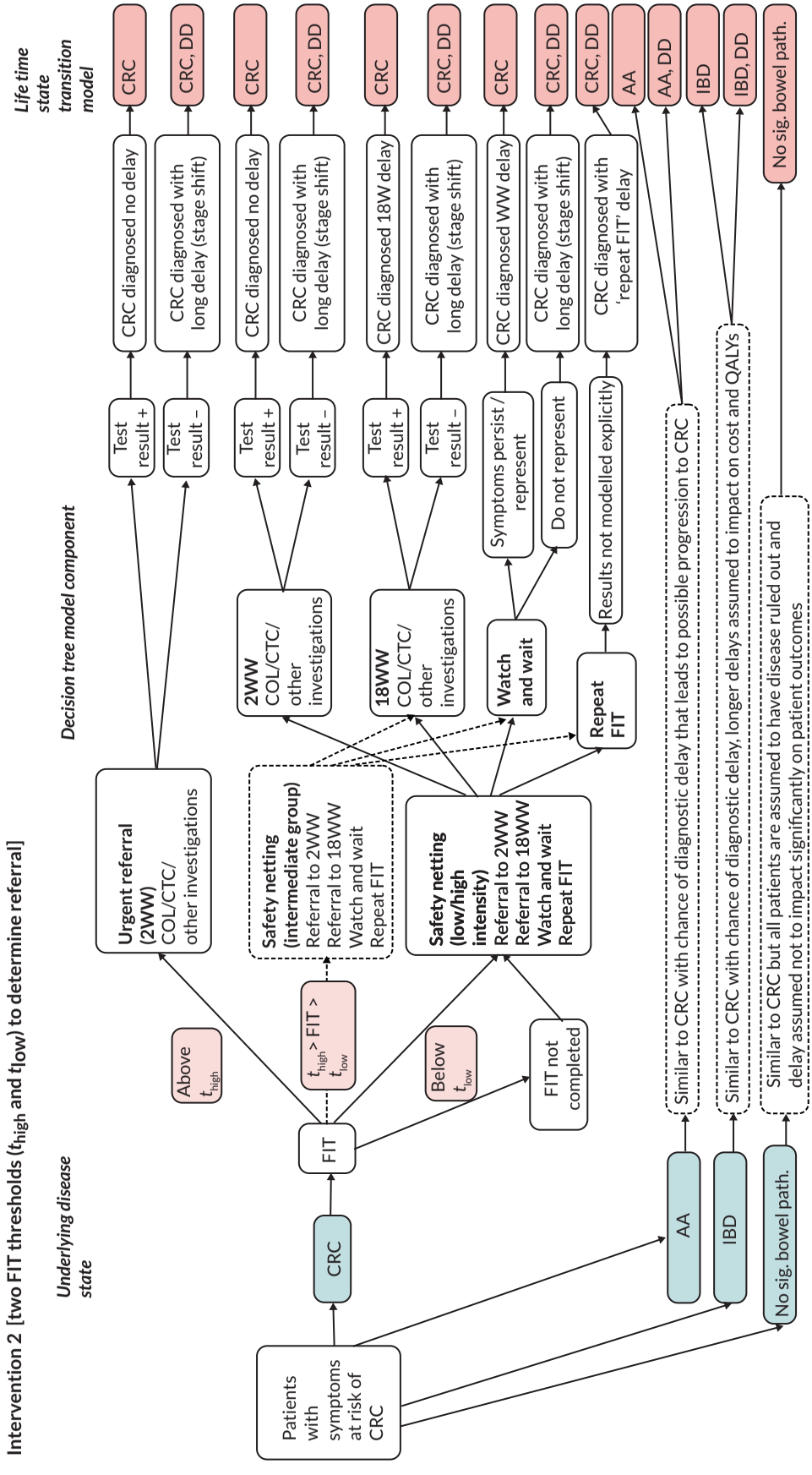

FIGURE 11 Evidence Assessment Group model: decision tree structure, intervention 2 (FIT with two thresholds of t_{high} and t_{low} µg/g). 2WW, 2 week wait;18W, 18 weeks; 18WW, 18 week wait; AA, advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; CTC, computed tomography colonography; DD, delayed diagnosis (see *Time to diagnosis and diagnostic delays*); FIT, faecal immunochemical test; IBD, inflammatory bowel disease; t_{high}, higher threshold; t_{low}, lower threshold; WW, watch and wait conduct.

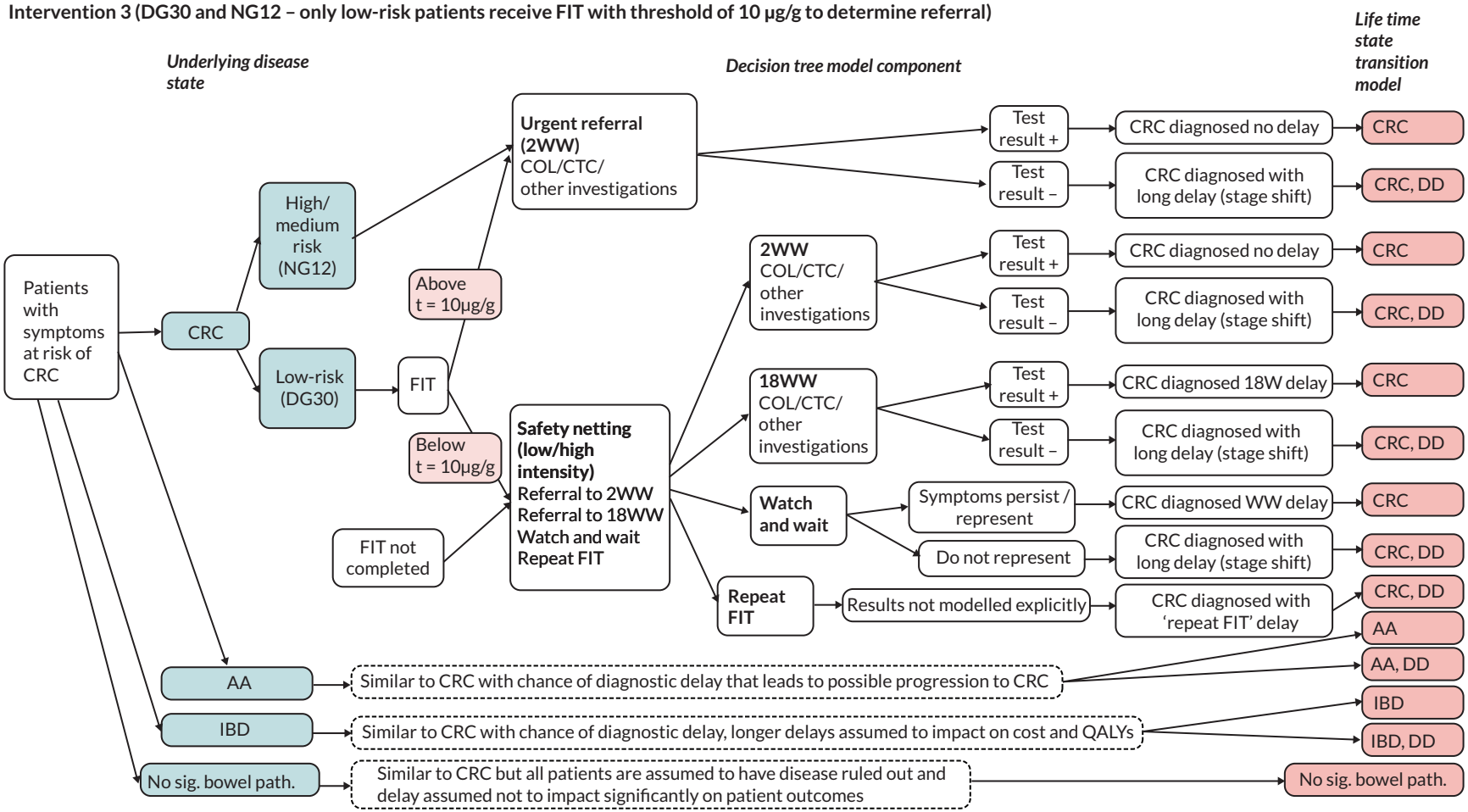

FIGURE 12 Evidence Assessment Group model: decision tree structure, intervention 3 (DG30 and NG12 recommendations, FIT for DG30 low-risk patients with threshold of 10 µg/g). 2WW, 2 week wait; 18W, 18 weeks; 18WW, 18 week wait; AA, advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; CTC, computed tomography colonography; DD, delayed diagnosis (see *Time to diagnosis and diagnostic delays*); FIT, faecal immunochemical test; IBD, inflammatory bowel disease; *t*, threshold; WW, watch and wait conduct.

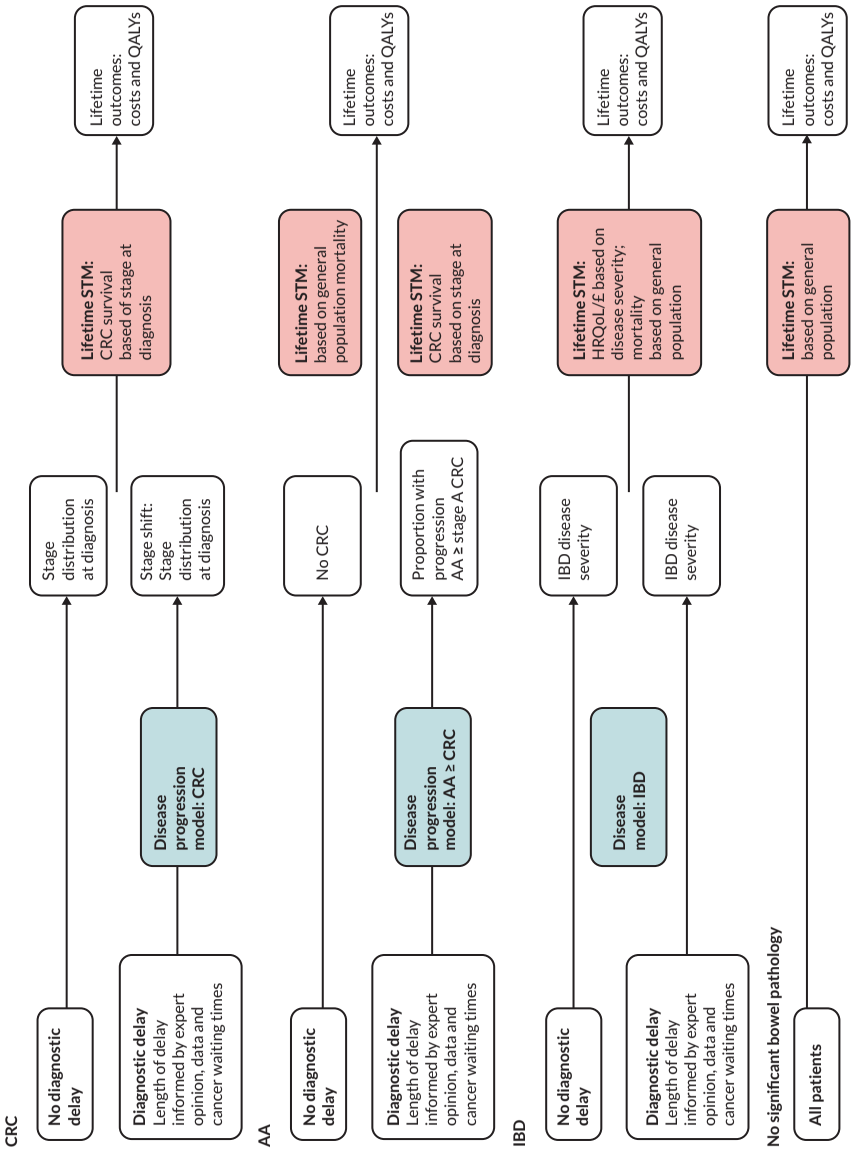

FIGURE 13 Evidence Assessment Group model: state transition model. AA, advanced adenomas; CRC, colorectal cancer; IBD, inflammatory bowel disease; QALY, quality-adjusted life year; STM, state transition model.

choice, and was intended to avoid the need to model such options separately to reduce the complexity of the model. The proportions of patients receiving each of the diagnostic imaging test in secondary care are presented in *Probability of receiving each imaging test as part of lower-gastrointestinal referral (2-week wait and 18-week wait)*. These estimates were informed by clinical opinion from the EAG's advisors and are conditional on the type of referral received (2WW or 18WW).

The model assumes that all patients with underlying CRC/IBD/AA at either 2WW or 18WW will eventually be diagnosed with underlying disease but with different times to diagnosis, based on the accuracy of the imaging tests received. Patients receiving a colonoscopy as the first imaging test are assumed not to receive any other confirmatory imaging tests, whereas patients receiving CTC with a positive result are assumed to receive a confirmatory colonoscopy. More details on the test accuracy parameters are presented in *Accuracy of the imaging tests used in 2-week wait and 18-week wait*; the EAG model differs from some previous models available for symptomatic patients in the literature in that it does not assume perfect accuracy of colonoscopy and CTC as the first imaging test. The model assumes that colonoscopy and CTC can also detect AAs and IBD, based on test accuracy data from the different literature sources (see *Accuracy of the imaging tests used in 2-week wait and 18-week wait*). The model assumes that cases missed by colonoscopy or CTC will be eventually detected by other diagnostic techniques while incurring an associated long delay in diagnosis. Patients undergoing 'other non-invasive investigations' are assumed to have their underlying disease detected by 2WW and 18WW referrals (i.e. the model assumes that the combination of other modalities have perfect accuracy, considering the group of patients for whom they are reserved); this assumption was a simplification to restrict model complexity. Patients with an underlying status of 'no significant bowel pathology' (hereafter termed NSBP) were assumed to have no disease detected at lower-GI referral; that is, the diagnostic test (or sequence of tests) used following referral has perfect specificity.

Patients in the model who follow the watch and wait pathway are assumed to be followed up in primary care, and are eventually diagnosed with their underlying condition, with an associated specific delay for this pathway. The model assumes that these patients would be diagnosed either by re-presentation to GP due to persistence or worsening of symptoms or following subsequent presentation to an accident and emergency (A&E) department. The model also includes patients who would be invited to receive a repeat FIT in primary care. In the absence of robust data identified in the EAG's review of the accuracy of a repeated FIT, the results of the second test are not modelled explicitly. Instead, the model assumes that a proportion of the patients with underlying bowel disease invited for a repeat FIT are detected via referrals and the remaining patients are detected after watch and wait, with a mean delay to diagnosis estimated for the overall group based on the time to diagnosis for each group. Patients with NSBP receiving watch and wait or repeat FIT were assumed not to re-present with persistent symptoms and/or to receive the confirmation of their underlying pathology.

After patients receive the diagnosis of their underlying disease of CRC/IBD/AA or of NSBP, they are assumed to move to each corresponding lifetime state transition model according to their true underlying pathology, where the lifetime costs, life-years gained (LYG) and QALYs and the impact of delays in the time to diagnosis for each pathology are estimated.

Impact of capacity limitations on waiting times and diagnostic delays

The EAG's model, considering the limited capacity availability of both referrals and colonoscopies in the UK NHS noted in NICE's scope,[22] estimates the impact of the use of alternative FIT thresholds on the number of referrals and colonoscopies undertaken. In the base-case analysis, capacity used is assumed to have a linear impact on waiting times (and consequently on the time to diagnosis in each pathway), based on the numbers of referrals estimated for intervention 3 (which correspond in the model to the current NICE recommendations). For example, if the demand for referrals in intervention 1 at threshold t_x results in a 10% reduction in the total number of referrals (2WW and 18WW), it is assumed that the time to obtain a diagnosis on the 2WW and 18WW pathways would be also reduced by the same proportion. Similarly, increases in referrals above the numbers experienced in intervention 3 would lead to an increase in waiting times for referrals and thus in an increase in the time to diagnosis modelled for interventions 1 and 2 for these pathways.

Lifetime state transition component of the model

This section describes the approach used to quantify survival, QALYs and costs for the long-term component of the model. The pathways followed in the decision-tree component of the model are assumed to have an impact on time to diagnosis and, as a consequence, on survival, QALYs and costs associated with each pathway followed by patients.

Evidence on the association between time to diagnosis and CRC outcomes is heterogeneous. A systematic review explored the association between shorter times to diagnosis and more favourable outcome and found that although many studies reported no associations, more studies reported a positive, rather than a negative, association.[92]
For patients with underlying CRC, the EAG's model assumes that longer delays in diagnosis may result in disease progression prior to diagnosis (stage shift) and thus have negative impacts on survival, HRQoL and costs (see *Appendix 8*).

For patients with underlying CRC and AAs, estimates of lifetime outcomes in terms of life expectancy, HRQoL and healthcare costs were generated from a separate model (the 'additional time to diagnosis impact on outcomes' model by Whyte *et al.*) by 'additional time to diagnosis' and were incorporated into the EAG's model. Full details of this separate model are provided in *Appendix 8*; a summary is provided below.

For CRC, the Whyte *et al.* model comprises two components: (1) patient outcomes (LYs, QALYs and costs) according to patient CRC stage at diagnosis and age, and (2) estimates of disease progression during the 'additional time to diagnosis' period. Patient outcomes with and without different delays are compared. CRC disease progression is estimated according to the change in stage distribution as a consequence of the additional time to diagnosis.

For AAs, disease progression is estimated based on the proportion of individuals whose AAs transform into CRC during the delay period, with those individuals who progress receiving lifetime outcomes for patients with CRC (when the delay is < 1 year, patients who progress are assumed to be diagnosed with CRC Duke's Stage A).

The proportion of individuals who experience a stage shift (a worsening in cancer stage during the delay in diagnosis or the progression from AAs to CRC stage) and the differential outcomes by stage are combined to generate estimates of expected outcomes by additional time to diagnosis. These estimates were integrated into the EAG model by applying these values as payoffs to each branch of the short-term decision tree, generating estimates for lifetime LYG, QALYs and costs for each diagnostic strategy.

Patients with IBD and no underlying disease were assumed to enter simple state transition models with two states: alive and dead. During each time interval of 1 year, patients entering these long-term models can either remain alive or die from any cause. For IBD, the model includes only patients with CD and UC, and patients with IBD are assumed to incur specific disease costs and utilities which considers the distribution of patients with each of these conditions and disease severity (see *Disease prevalence and severity/stage distribution*). A diagnosis of IBD through the 2WW and 18WW pathways is assumed to be associated with no significant delay that would have a substantial impact on outcomes, while delayed diagnosis through watch and wait, repeat FIT or long delay (false-negative patients eventually diagnosed after a long delay) is assumed to result in an increased probability of complications for 2 years after diagnosis, with associated additional costs and QALY losses due to the increase in these complications. The lifetime LYG, QALYs and cost estimates generated for each of the two groups were then integrated into each branch of the short-term decision tree, thereby generating expected estimates of lifetime LYG, QALYs and costs for each diagnostic strategy.

For patients with NSBP, any additional time to confirmation of the underlying status is assumed to have no impact on lifetime outcomes, and therefore these patients are assumed not to incur in any additional lifetime costs, and to have the same all-cause mortality risks and HRQoL as those of the same age and sex in the general population in England.

Key Evidence Assessment Group model assumptions

The EAG model makes the following structural assumptions.

- The model was designed to reflect the population of patients presenting to primary care with symptoms or signs indicating a risk of CRC. However, the model is also structured to capture incidental findings of other bowel pathologies: AAs and IBD, which includes UC and CD.
- Patients' underlying disease status (CRC, AAs, IBD and NSBP) is assumed to be mutually exclusive and exhaustive. This simplifying assumption, which does not allow more than one relevant disease to be detected at the same time, was necessary due to the structure of the model developed and was anticipated not to have a significant impact on model results. The EAG notes, however, that delays in the diagnosis of AAs could have an impact on patient outcomes in the long-term model as a result of the possibility of these progressing to CRC (but these patients are still categorised in the model as AA patients).
- Colonoscopy, CTC and 'other non-invasive' investigations in secondary care are assumed to detect only the underlying condition (e.g. they do not allow for false-positive results for IBD in patients with underlying CRC, and vice versa).
- Estimates of the accuracy of FIT for CRC, AAs and IBD were informed by the EAG's systematic review (see *Results*), while the sensitivity and specificity of colonoscopy and CTC were obtained from the literature,[93–98] with some necessary simplifications/assumptions made in line with previous published models in this disease area.[90,91,93]
- The completion rate for FIT was informed by literature.[42] Patients completing FIT are assumed to receive the test and return it to their GP practice in a timely manner so that there are no delays in processing.
- The model assumes that a small proportion of FIT samples need to be repeated because they were unsuitable, but this was assumed not to affect outcomes or time to diagnosis and was included in the model only in terms of the cost of completed FIT. The model also assumes that patients received a maximum of two FIT (the first invitation and the repeat FIT for a proportion of patients).
- The model assumes that all patients who receive a FIT result above t or t_{high} µg/g (interventions 1 and 2, respectively) will be referred to 2WW, while patients who receive a result below t or t_{low} µg/g or do not complete FIT will follow pathways under 'safety netting'. Patients in intervention 2 whose FIT result lies between t_{low} and t_{high} are assumed to follow 'intermediate group' pathways. Within these pathways, according to the GP's clinical judgement, patients are assumed to be referred to 2WW or 18WW; monitored and managed in primary care; or offered a repeat FIT.
- Patients referred to 2WW or 18WW are assumed to receive an initial appointment with a gastroenterology consultant and are offered colonoscopy or CTC as their main imaging investigation; patients are assumed to undertake the test assigned to them, with an assumed uptake rate of 100%.
- Patients who receive 'watch and wait' are assumed to incur in an additional 1.9 appointments with the GP. This is intended to reflect the patient receiving timely review or re-presenting to the GP when their symptoms persist or worsen, which would lead to a diagnosis. The majority of patients with underlying disease are assumed to be eventually diagnosed via a referral and are assumed to incur the costs of a lower bowel referral. A smaller proportion of patients are assumed to be diagnosed only after presenting at A&E; these patients are assumed to incur the costs of presenting at A&E.
- Patients who receive repeat FIT are assumed to incur the cost of the additional FIT and an additional 1.9 GP appointments; this is intended to reflect any additional appointments necessary for discussing results and options for further management. These patients are also assumed to incur the costs of 'watch and wait' or a referral to secondary care, based on accuracy estimates for FIT, which are assumed to estimate a weighted mean cost for patients receiving this pathway.
- The model assumes that patients who receive 'watch and wait' or repeat FIT would eventually be diagnosed with their true underlying condition, but their outcomes (costs, LYG and QALYs) are impacted by the delayed diagnosis.
- Where available, accuracy data for CTC and colonoscopy are informed by the literature.[94–99] The specificity of colonoscopy was assumed to be 100% for all underlying pathologies because of the nature of the test. This assumption is in line with previous models.
- Patients with a positive result for CTC are assumed to receive a confirmatory colonoscopy. This second test is assumed to have a perfect diagnostic accuracy. The EAG notes that this is a simplification to reduce model complexity.
- Patients with NSBP who are referred to 2WW or 18WW are assumed to incur costs of colonoscopy/CTC/other non-invasive investigations and are eventually ruled out of having any of the lower bowel pathologies being modelled. People with NSBP are therefore assumed not to incur any additional lifetime costs.
- The accuracy of the diagnostic tests received as part of 2WW or 18WW referrals is assumed to be independent of the underlying disease stage/severity at a patient's initial presentation to primary care.

- All patients who receive colonoscopy are assumed to be at risk of AEs associated with this imaging test (including the risk of death after perforation). AEs could result in death, QALY loss and costs.
- Patients who have NSBP or are diagnosed with IBD who remain alive following the diagnostic decision tree are assumed to have the same mortality risk as those of the same age and sex in the general population of England.
- Time to diagnosis is assumed to include the time from initial presentation at primary care to a definitive diagnosis. The model assumes that the greatest impact is derived from the additional time a patient with underlying disease has to spend in primary care in the 'watch and wait' or 'repeat FIT' pathway or by patients who were missed by imaging tests in secondary care to the point at which a correct diagnosis is obtained. Different lengths of time to diagnosis are explored in scenario analyses (see *Scenario analyses*).
- The Whyte *et al.* model assumes that a proportion of patients with underlying CRC will experience disease progression to a worse CRC stage, which depends on the length of additional time to diagnosis, and that patients with AAs may develop CRC during the additional time taken to receive diagnosis. It is assumed that patients can transition to a worse state only once per year (see *Appendix 8*) and that no disease-related deaths are incurred during the delay period.
- The Whyte *et al.* model also assumes that adenomas are asymptomatic, and therefore the impact on health outcomes and costs due to delay in diagnosis is associated only with those patients who progress to CRC in that delay period. Patients in this model diagnosed without any delays who have not progressed to CRC are assumed to accrue the same lifetime cost and health outcomes as the general population.
- The lifetime costs for CRC patients are assumed to include costs associated with the diagnosis and treatment of CRC, including hospitalisations, medications and palliative care (see *Appendix 8*).
- Patients with IBD or NSBP are assumed to have the same risk of death as the general population. NSBP patients are assumed not to incur in any additional costs from the point of diagnosis and to have the same HRQoL as the general population, while IBD patients are assumed to incur specific costs for the treatment of the underlying disease, considering disease severity and the costs and HRQoL associated with the condition. A significant delay in the IBD diagnosis is assumed to be associated with an increased probability of having disease complications and to incur additional costs and QALY losses, which are assumed to be resolved with treatment after 2 years of diagnosis.

Evidence sources used to inform the model parameters

Table 15 summarises the evidence sources used to inform the parameters of the EAG model. The individual parameter values are discussed in further detail in the subsequent sections. In addition to the review of economic evaluations and HRQoL studies, targeted literature searches were undertaken to identify studies to inform the parameters of the EAG's model, such as patients' initial characteristics, CRC stage distribution, accuracy of the imaging tests (colonoscopy/CTC), costs, morbidity including AEs associated with colonoscopy and HRQoL. These searches did not constitute a systematic review but followed the principles of NICE Decision Support Unit Technical Support Document 13.[100]

Patient characteristics

Patients' mean age was assumed to be 64 years, based on the CRC prevalence in 2013 from Public Health England (rounded down to an integer value).[101] The cohort of patients was assumed to be 54.9% female, based on D'Souza *et al.*[17]

Disease prevalence and severity/stage distribution

The data used in the EAG base-case model are summarised in *Table 16*. The probabilities of patients having underlying CRC, IBD and AAs were based on the results of the EAG's statistical analysis undertaken as part of the systematic review (see *Appendix 2*). The model assumes that in intervention 3 (based on current NICE recommendations DG30 and NG12), the proportion of patients classified as high risk was 0.537, based on the proportion of CRC patients classified as NG12 in D'Souza *et al.*[18] The EAG team considered this the more appropriate source for this parameter, and this was supported by the EAG's clinical advisors (who considered a proportion close to 0.50 reasonable). While the estimates for the CRC, AAs and IBD prevalence for the overall population and high-risk patients were available from the EAG's evidence synthesis, data for DG30 low-risk patients were only available from D'Souza *et al.*[18] The model was calibrated to ensure that the overall prevalence of each lower bowel pathology was the same in all interventions being evaluated.

The stage distribution of patients at CRC diagnosis by Duke's classification used in the Whyte *et al.* model was informed by staging data for CRC patients in England in 2019 from the National Cancer Registration and Analysis Service

TABLE 15 Evidence sources used in the model

Parameter group	Parameter group	Source
Patients' initial characteristics	Patient initial age	Based on data from Public Health England on CRC prevalence for 2013[101]
	Probability female	D'Souza *et al.*[17]
Disease prevalence and severity distribution	Disease prevalence for CRC, AAs and IBD	EAG's clinical review and synthesis (see *Appendix 2*)
	CRC stage distribution at diagnosis	Staging data in England for 2019;[102] for details see description of Whyte *et al.* model in *Appendix 8*
	Proportion of high-risk patients in all with suspected symptoms of CRC	D'Souza *et al.*[72]
	Distribution of patients with UC and CD and by disease severity in IBD	Pasvol *et al.*,[103] Ghosh and Premchand[104]
Tests' characteristics	FIT accuracy (for CRC, IBD and AAs)	EAG's clinical systematic review and analysis (see *Results*)
	COL and CTC accuracy (for CRC, AAs and IBD)	Thomas *et al.*;[99] Bressler *et al.*;[94] Atkin *et al.*;[95] Lin *et al.*;[96] Martín-López *et al.*;[97] Horsthuis *et al.*;[98] assumption
	'Other non-invasive investigations' accuracy	Assumption
	FIT return rate	Bailey *et al.*[42]
	Probabilities of having AEs related to COL	Lin *et al.*;[105] Gatto *et al.*[106]
Healthcare current pathways	Proportion of patients receiving each of the pathways following FIT result (safety netting and 'intermediate group' results pathways)	Based on EAG's clinical advisors' responses (see *Probability of following each of the pathways following faecal immunochemical test result*) and *Report Supplementary Material 5*)
	Proportion of patients receiving each of the imaging investigations in secondary care (2WW and 18WW)	Based on EAG's clinical advisors' responses [see *Probability of receiving each imaging test as part of lower-gastrointestinal referral (2-week wait and 18-week wait)* and *Report Supplementary Material 5*]
	Time to diagnosis for each pathway followed (2WW,18WW, watch and wait, repeat FIT, and patients eventually diagnosed with long delay)	Based on EAG's clinical advisors' responses (see *Time to diagnosis and diagnostic delays* and *Report Supplementary Material 5*)
Mortality	CRC and AA mortality	MiMic-Bowel model as reported in Thomas *et al.*[99] and assumption (see *Appendix 8*)
	IBD mortality	The risk of death of IBD patients were assumed to be the same as the general population
	Other-cause mortality (general population)	National life tables for England 2018–20 (ONS)[107]
Long term model probabilities of transitions	Probability of transition between CRC states (progressing)	See description of Whyte *et al.* model in *Appendix 8*
	Probability of AA progressing to CRC Duke's Stage A	See description of Whyte *et al.* model in *Appendix 8*
HRQoL	CRC	See description of Whyte *et al.* model in *Appendix 8*
	AAs	See description of Whyte *et al.* model in *Appendix 8*
	IBD	Utilities from NICE TA856[108] and TA342;[109] utility multiplier associated with delayed diagnosis from Stark *et al.*;[110] assumption of duration of impact from delayed diagnosis
	General population	Hernández Alava *et al.*[111]
	QALY losses due to colonoscopy AEs	Thomas *et al.*;[99] Dorian *et al.*;[112] Ara and Brazier[113]

TABLE 15 Evidence sources used in the model (*continued*)

Parameter group	Parameter group	Source
Costs (short term model)	FIT costs (tests)	Unit costs for FITs, GP appointment and tests from test manufacturers, PSSRU 2022[114] and NHS Reference Costs 2021/22;[115] FIT that need resampling from MacDonald *et al.*;[56] FIT return rate from Bailey *et al.*[42]
	Costs of lower GI referrals	Unit costs for appointments and imaging tests from NHS Reference Costs 2021/22;[115] proportion of patients receiving CT in 'other non-invasive interventions' estimated from clinical opinion
	Costs of watch and wait	Number of additional GP visits estimated by COLOFIT team based on data from Lyratzopoulos *et al.*;[116] proportion of patients who present to AEs from 'Routes to diagnosis for England 2018',[117] clinical visits and A&E attendance unit costs from PSSRU 2022[118] and NHS Reference Costs 2021/22[115]
	Costs of repeat FIT	Test manufacturers; MacDonald *et al.*;[56] Bailey *et al.*;[42] number of additional GP visits estimated by COLOFIT team; PSSRU 2022;[118] routes to diagnosis for England 2018[117]
	Costs of AE related to COL	Unit costs from NHS Reference Costs 2021/22[115]
Costs (long term model)	Lifetime treatment costs for CRC	MiMic-Bowel model as reported in Thomas *et al.*[99] (see *Appendix 8*)
	Lifetime treatment costs for AAs	
	Lifetime treatment costs for IBD	Annual treatment costs from Ghosh and Premchand[104] uplifted to 2022 using NHSCII index from PSSRU 2022[114]

2WW, 2 week wait; 18WW, 18 week wait; AAs, advanced adenomas; AE, adverse event; A&E, accident and emergency; BNF, *British National Formulary*; CD, Crohn's disease; COL, colonoscopy; CRC, colorectal cancer; CT, computed tomography; CTC, computed tomography colonography; EAG, Evidence assessment group; FIT, faecal immunochemical test; GI, gastrointestinal; GP, general practitioners; IBD, inflammatory bowel disease; PSSRU, Personal Social Services Research Unit; QALY, Quality-adjusted life year; TA, Technology Appraisal; UC, Ulcerative colitis.

TABLE 16 Parameters related to disease prevalence and disease stage or severity used in the base-case analysis

Parameter	Group	Mean	Source
CRC prevalence	Whole population	0.028	EAG's clinical systematic review and D'Souza *et al.*[18]
	High-risk group (NG12, intervention 3 only)	0.044	
	Low-risk group (DG30, intervention 3 only)	0.010	
IBD prevalence	Whole population	0.027	
	High-risk group (NG12, intervention 3 only)	0.032	
	Low-risk group (DG30, intervention 3 only)	0.022	
AAs prevalence	Whole population	0.023	
	High-risk group (NG12, intervention 3 only)	0.043	
	Low-risk group (DG30, intervention 3 only)	0.000	
Proportion of patients classified as NG12 high risk or DG30 low risk for CRC	High-risk group (NG12, intervention 3 only)	0.537	D'Souza *et al.*[18]
	Low-risk group (DG30, intervention 3 only)	0.463	

continued

TABLE 16 Parameters related to disease prevalence and disease stage or severity used in the base-case analysis (*continued*)

Parameter	Group	Mean	Source
CRC stage distribution at disease diagnosis	CRC Stage A	0.196	Staging data in England for 2019 (NCRAS 2021)[102]
	CRC Stage B	0.254	
	CRC Stage C	0.312	
	CRC Stage D	0.238	
IBD disease and disease severity distribution at disease diagnosis	Relative incidence of UC vs CD	0.60	Pasvol *et al.*[103]
	Proportion of patients in remission with UC	0.50	Ghosh and Premchand[104]
	Proportion of patients in relapse with UC	0.50	
	Proportion UC relapse mild-moderate	0.80	
	Proportion UC relapse severe	0.20	
	Proportion of patients in remission with CD	0.50	
	Proportion of patients in relapse with CD	0.50	

AA, advanced adenomas; CD, Crohn's disease; CRC, colorectal cancer; IBD, irritable bowel disease; UC, ulcerative colitis.

(NCRAS).[102] The distribution of patients with UC and CD was based on data from Ghosh and Premchand,[104] a detailed UK costing study in IBD, while the proportion of patients with UC from the overall population of patients with IBD in the model was taken from Pasvol *et al.*[103]

Faecal immunochemical test accuracy

Data relating to the accuracy of each of the FIT brands were informed by the EAG's evidence synthesis; a summary of the results of these analyses is presented in Chapter 2 *Main analysis: Summary*. The model used the estimates of sensitivity and specificity for CRC for selected thresholds available (between 2 µg/g and 400 µg/g; selection based on the availability of results for each FIT brand). For intervention 1, all of the thresholds were tested individually for those brands with data available. For intervention 2, due to the excessive number of possible combinations, the values for the thresholds pairs were selected based on the model results for individual thresholds and clinical interest.

For intervention 2, where the pathways followed by patients were determined by three different groups based on the results of FIT, the results of FIT were calculated as follows:

FIT result ≥ t_{high} = sensitivity of FIT for t_{high} (i.e. if t_{high} = 100, sensitivity for t_{100})

FIT result ≤ t_{low} = 1 − sensitivity of FIT for t_{low} (i.e. if t_{low} = 10, 1 − sensitivity for t_{10})

t_{low} ≤ FIT result ≤ t_{high} (intermediate group) = 1 − (sensitivity t_{high} + sensitivity t_{low})

The value for the FIT return rate of 0.91 was taken from Bailey *et al.*[42] and was assumed to be the same for the first and second FIT received.

Probability of following each of the pathways following faecal immunochemical test result

Patients who receive a FIT result above *t*, t_{high} or 10 µg/g (in interventions 1, 2 and 3, respectively), or are classified as high risk by NG12 criteria in intervention 3, are directly referred to the suspected cancer urgent pathway (2WW). However, patients who obtain FIT results below these thresholds or do not complete the test are assumed to follow one of the pathways with two possible groups: safety-netting or 'intermediate group' pathway.

Safety netting is defined in this model as the follow-up pathways for patients with FIT results below *t*, t_{low} or 10 µg/g (in interventions 1, 2 or 3, respectively) or incomplete test, while 'intermediate group pathways' is reserved for patients in

intervention 2 for whom the FIT result lies between t_{low} and t_{high}. The model represents safety netting with a proportion of patients following each of four pathways: 2WW, 18WW, watch and wait or repeat FIT, based on estimates derived from clinical opinion of the EAG's advisors (*Table 17*). The model explores two different intensities of safety-netting pathways (low or high), based on the view that clinical practice is heterogeneous across the country and has been changing in recent years, with the introduction of FIT as part of screening programmes and for triage in symptomatic DG30 low-risk patients.[11] The EAG's model generates results for both options of safety netting.

The pathways for the 'intermediate group' include the same pathway options as for safety netting, with the exception of watch and wait and with a higher proportion of patients assumed to be referred to 2WW. This approach assumes that these patients would be considered to have a higher risk of CRC and therefore would be referred to secondary care.

In the model, the watch and wait pathway consists of the GP reviewing the patient's symptoms or the patient re-presenting to the GP if their symptoms persist or worsen; these strategies are not explicitly modelled separately. The base-case model assumes that patients followed up in this pathway incur an additional 1.9 GP appointments (estimated by a member of the modelling team in the COLOFIT project based on data from Lyratzopoulos *et al.*,[116] obtained via Chloe Thomas, 22 February 2023, personal communication) and that all patients are eventually diagnosed with their true underlying condition, but a proportion of patients (22.15%) would be detected only at presentation at A&E, based on data from routes to cancer diagnosis in England in 2018 by NHS Digital.[117] Patients following this pathway are assumed to be diagnosed with their underlying condition after a significant delay (see *Time to diagnosis and diagnostic delays*), which is associated with an increased probability of CRC stage progression or a risk of AAs transforming into CRC during the delay period (see *Appendix 8*).

Patients invited to receive a second FIT in primary care (repeat FIT) do not have the results of the second test modelled explicitly, as the EAG has identified limited data on the accuracy of repeated tests (see Chapter 2 *Other outcomes* and *Appendix 6*) and on how the data on dual FIT are applicable to the population receiving the test in the primary care context. Instead, the model assumes that patients with underlying bowel disease who are invited for repeat FIT would receive the results of the second FIT and would be either diagnosed via a pathway in secondary care or eventually diagnosed by re-presenting to their GP with persistent/worsening symptoms that would be associated with a long delay in diagnosis. Therefore, patients in this pathway are assumed to obtain a diagnosis of their underlying condition with a specific delay for this pathway (see *Time to diagnosis and diagnostic delays*) and to incur additional diagnostic costs (see *Resource use and costs*).

TABLE 17 Proportion of patients receiving each of the management pathways following FIT results,[a] based on clinical advice provided to the EAG

	Results of FIT			
	Positive FIT (FIT > t, t_{high} or 10 µg/g)	'Negative' FIT (FIT < t, t_{low} or 10 µg/g) or FIT incomplete		'Intermediate' FIT (t_{low} < FIT < t_{high}) – intervention 2 only
Proportion following each pathway	Referral to 2WW directly (%)	Safety netting pathways – model base case (high intensity) (%)	Safety netting pathways – model scenario analysis (low intensity) (%)	'Intermediate group' follow-up pathways (%)
Referral to 2WW	100%	15	5	85
Referral to 18WW	–	25	10	10
Watch and wait	–	40	75	0
Repeat FIT	–	20	10	5

FIT, faecal immunochemical test; 2WW, two week wait; 18WW, 18 week wait; 18W, 18 weeks; t, single FIT threshold in Intervention 1; t_{high}, higher FIT threshold in intervention 2; t_{low}, lower FIT threshold in intervention 2.
a In intervention 3, patients classified as high risk are referred to 2WW without receiving a FIT.

Probability of receiving each imaging test as part of lower-gastrointestinal referral (2-week wait and 18-week wait)

Patients referred to secondary care under 2WW or 18WW are assumed to receive one of the following imaging investigations: colonoscopy, CTC or 'other non-invasive investigations'. The proportions of patients receiving each imaging test in the EAG's base-case analysis are presented in *Table 18* and were based on the opinion of the EAG's clinical advisors (provided in *Report Supplementary Material 5*) on the use of imaging tests in referrals in this population (answers for the overall population).

The EAG notes that responses from clinical advisors and published literature suggest that CTC capacity in England is currently very restricted[119] and this might be one of the reasons why CTC is less frequently used in clinical practice than colonoscopy. The EAG also notes that these proportions vary by age group; however, the estimates used are intended to reflect the usage by the overall population referred to secondary care with symptoms suggestive of CRC.

Accuracy of the imaging tests used in 2-week wait and 18-week wait

Data on the accuracy of colonoscopy and CTC received by patients referred to 2WW and 18WW are presented in *Table 19*. The EAG model adopts a similar approach to Thomas *et al.* in the MiMic-Bowel bowel cancer screening model.[99] Sensitivity estimates for CRC detection by colonoscopy and CTC were based on the studies by Bressler *et al.*[94] and estimates of the relative risk for the detection of CRC using CTC rather than colonoscopy from Atkin *et al.*[95] The EAG notes that the use of these estimates includes the assumption that these imaging tests would have a similar performance in symptomatic and asymptomatic patients. Sensitivity for the detection of AAs by colonoscopy was based on Martin-Lopez *et al.*,[97] while the same approach based on the relative risks from Atkin *et al.*[95] was applied to estimate the sensitivity for AAs by CTC. The specificity of colonoscopy was assumed to be 1.00 for all conditions, given the nature of the test. Specificity estimates for CRC and AAs detected by CTC were taken from Lin *et al.*[6] and were assumed

TABLE 18 Proportion of patients receiving each pathway at their lower-GI referral (2WW/18WW)

Investigation received	2WW (%)	18WW (%)
COL	90	90
CTC	7.5	7.5
Other/no investigations (e.g. CT or appointment)	2.5	2.5

COL, colonoscopy; CT, computed tomography; CTC, computed tomography colonography; 2WW, two week wait; 18WW, 18 week wait.

TABLE 19 Estimates of accuracy for imaging tests used in patients in 2WW and 18WW

Condition	Technology	Parameter	Point estimate	95% CI	Source
CRC	COL	Sensitivity	0.966	0.962 to 0.969	Thomas *et al.*;[99] Bressler *et al.*[94]
		Specificity	1.000	–	Assumption due to nature of test
	CTC	Sensitivity	0.946	0.606 to 1.473	Thomas *et al.*;[99] Atkin *et al.*[95]
		Specificity	0.881	0.873 to 0.889	Lin *et al.*[96]
AAs	COL	Sensitivity	0.925	0.894 to 0.952	Thomas *et al.*;[99] Martín-López *et al.*[97]
		Specificity	1.000	–	Assumption due to nature of test
	CTC	Sensitivity	0.759	0.465 to 1.218	Thomas *et al.*;[99] Atkin *et al.*[95]
		Specificity	0.881	0.873 to 0.889	Lin *et al.*,[96] assumption
IBD	COL	Sensitivity	1.000	–	Assumption in line with previous models for CRC symptomatic and asymptomatic patients
		Specificity	1.000	–	
	CTC	Sensitivity	0.843	0.750 to 0.918	Horsthuis *et al.*[98]
		Specificity	0.951	0.868 to 0.994	

AAs, advanced adenomas; CI - confidence interval; COL, colonoscopy; CRC, colorectal cancer; CTC, computed tomography colonography; IBD, irritable bowel disease.

to be the same for the two conditions. Sensitivity and specificity for the detection of IBD by CRC were obtained from Horsthuis *et al.*[98]

Patients receiving 'other non-invasive investigations' as part of the diagnostic pathway in 2WW and 18WW are assumed to be diagnosed and no cancer cases missed (sensitivity and specificity of 1.0 for all conditions), based on the assumption that for this small group of patients with greater frailty and possibly disease severity, a different number of non-invasive diagnostic techniques would be able to detect a patient's underlying condition.

Complications associated with colonoscopy

Complications associated with colonoscopy were included in the model for a proportion of patients receiving this imaging test. Patients receiving colonoscopy have a small probability of developing bleeding or perforation of the intestine as a consequence of the procedure (*Table 20*; these probabilities were based on Lin *et al.*[105]). Those with perforations can also die as a consequence of the complication; this probability was informed by Gatto *et al.*[106] Patients experiencing complications from colonoscopy are assumed to incur additional costs and HRQoL losses, which (with the exception of death) were assumed to be temporary and resolved without further long-term impacts on their health outcomes after disease diagnosis. Similar to the approach used by Thomas *et al.*,[99] the EAG model includes QALY losses associated with bleeding and non-fatal perforation due to colonoscopy. The utility value for serious bleeding events was taken from Dorian *et al.*[112] and was assumed to last for 2 weeks, while QALY losses due to non-fatal perforation were based on Ara and Brazier,[113] with the utility value based on the absolute difference in mean EQ-5D score between patients with 'stomach ulcer/abdominal hernia/rupture' who were not affected by the condition and those who were affected by it; this event was assumed to have an impact on HRQoL for 1 month.

The sources for costs associated with colonoscopies are presented in *Table 20*; the unit costs were taken from NHS Reference Costs 2021–2[115] and assumptions. Patients who receive colonoscopy after receiving CTC are assumed to be susceptible to the same AEs and corresponding probabilities as patients receiving colonoscopy as their main imaging investigation.

Time to diagnosis and diagnostic delays

In the model, for patients with an underlying lower bowel pathology (CRC, IBD and AA), the time to diagnosis was assumed to depend on the pathway followed. For example, in the EAG's base case, patients on 2WW were assumed to receive their diagnosis during the period informed by the clinical advisors (provided in *Report Supplementary Material 5*) and not to experience any delays in receiving their diagnosis. The time to diagnosis necessary for each pathway is presented in *Table 21* and was based on clinical input from the EAG's advisors.

TABLE 20 Complications, QALY losses and costs associated with colonoscopy included in the short-term model

Complication	Probability of having an AE	Source	QALY loss	Source	Unit cost (£)	Source
Serious bleeding	0.00175	Lin *et al.* 2021[105]	0.00579	Thomas *et al.* 2020,[99] based on data from Dorian *et al.* 2014[112]	1695.45	NHS Reference Costs 2021–2, weighted average cost of all GI bleed procedures with multiple, single or no interventions (codes FD03A to FD03H)[115]
Perforation	0.00054	Lin *et al.* 2021[105]	0.00983	Thomas *et al.* 2020,[99] based on data from Ara and Brazier 2011[113]	6299.74	NHS Reference Costs 2021–2, weighted average cost of all major large intestine procedures in adults (19 +, codes FF34A to FF34C)[115]
Death by perforation	0.05195	Gatto *et al.* 2003[106]	[a]		0.00	Assumption that the costs of perforation already capture the costs incurred before patient's death

AE, adverse event; COL, colonoscopy; QALY, quality-adjusted life year.
a Patients who die from perforation following a colonoscopy do not incur any QALY losses but are assumed not to receive any QALYs from the point of death.

TABLE 21 Estimated diagnostic delays by each type of pathway and diagnostic result[a]

Pathway followed	Estimated average time to diagnosis for patients with underlying CRC/IBD/AA (weeks)[b]		
	Model base case	Scenario analysis 1	Scenario analysis 2
Lower-GI referral (2WW) disease diagnosed at referral	2	2	3
Lower-GI referral (18WW) disease diagnosed at referral	27 (6 months)	18 (4 months)	54 (1 year)
Lower-GI referral (2WW/18WW), disease missed by COL/CTC, patient re-presents with persistent symptoms	78 (1.5 years)	52 (1 year)	157 (3 years)
Watch and wait, patient re-presents with persistent symptoms	59 (1.13 years)	35 (8 months)	104 (2 years)
Repeat FIT (weighted average of subsequent pathways)	38 (8.7 months)	23 (5.3 months)	69 (1.3 years)

2WW, 2 week wait; 18WW, 18 week wait; AA, advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; CTC, computed tomography colonography; FIT, faecal immunochemical test; GI, gastrointestinal, IBD, inflammatory bowel disease.
a The time to diagnosis does not include the time of initial investigations by the GP (initial appointment and FIT).
b For patients without underlying disease, the model includes the costs of additional investigations needed for those who have an initial positive FIT result and are referred to 2WW and 18WW and the impact on HRQoL of AEs associated with colonoscopy.

The EAG notes that time to diagnosis could be defined in several ways, such as time from symptom onset, time of presentation in primary care or time from referral. In the EAG model, a diagnosis delay is assumed to comprise the additional time to diagnosis compared with average time to diagnosis with a 2WW referral. The model assumes that the estimates of time to diagnosis do not account for small differences in diagnostic interval due to the time needed to take the FIT and receive its results. For example, patients in intervention 3 who are referred straight to 2WW would, strictly speaking, be able to receive a diagnosis faster than patients completing a FIT and, in the sequence, would be referred to 2WW with a positive result, but the model assumes that these small differences would not impact on disease progression and they were therefore not considered. Differences in times to diagnosis between FIT and repeat FIT included in the model are shown in *Table 21*.

The EAG also notes that the time to diagnosis for patients receiving a repeat FIT was estimated based on a weighted mean time for the other pathway times and the proportions receiving each of these pathways (high-intensity safety netting; see *Table 17*). The model also assumes that all time lengths include the turnaround time required to receive the results of a FIT and all delays would be in relation to the 2WW pathway.

The estimates presented in *Table 21* correspond to a reference point, used to estimate the time to diagnosis for patients in the current scenario in England, which is assumed to correspond to current NICE recommendations. To estimate the impact of the introduction of FIT for all symptomatic patients in primary care and the resulting expected impact on waiting times, the EAG included in the model structure the assumption that reductions in the number of referrals to secondary care (2WW and 18WW) would vary by the threshold applied and would have a linear impact on the waiting times for these two pathways. For example, a reduction of 10% in total referrals as a consequence of a specific threshold would reduce the time to diagnosis for patients receiving 2WW and 18WW in this strategy by the same proportion.

Three different scenarios were explored in the model:

- Base-case scenario. This scenario is intended to reflect the current situation in England.
- Scenario 1. This explores a best-case scenario with shorter times to diagnosis for all pathways.
- Scenario 2. This explores a worst-case scenario in which times to diagnosis are increased (see *Scenario analyses*).

Long-term state transition model outcomes

Appendix 8 presents the details of the estimates and source of parameters used in the Whyte *et al.* model to generate the estimates of lifetime outcomes for CRC and AA patients by 'additional time to diagnosis' in the EAG model.

The risk of death for IBD and NSBP patients was informed by the sex- and age-matched mortality estimates for the general population in the England.[107] Utilities for NSBP patients were also assumed to follow the age- and sex-matched

estimates for the UK population from Hernández Alava *et al.*[111] The health utility values used in the model for patients with IBD are summarised in *Table 22*; these were estimated based on utility values reported in NICE Technology Appraisal (TA) 856[108] for UC and in TA342[109] for CD (which were based on Woehl *et al.*[120] and the GEMINI II/III studies[121]), and the relative incidence between UC and CD and distribution of patients by disease severity in these conditions as reported in Pasvol *et al.*[103] and Ghosh and Premchand.[104] The proportions of patients by disease type and severity are summarised in *Table 23*.

Patients diagnosed with IBD were assumed to experience a QALY loss that was estimated to correspond to the impact of the increased risk of having complications, based on data from Stark *et al.*,[110] which is applied to the increase in the proportion of patients having disease complications for 2 years at the point of diagnosis.

Resource use and costs

The model includes the following cost components:

1. costs associated with the FIT (first test)
2. cost of lower GI referrals
3. costs of watch and wait
4. costs associated with 'repeat FIT'
5. costs of treating AEs related to colonoscopy
6. costs associated with treating the underlying conditions (lifetime costs for CRC, IBD and AAs).

TABLE 22 Proportion of patients by disease severity and type and health utilities applied in the EAG model for IBD patients

Underlying condition	Health state	Mean utility	Source
UC	Active UC	0.41	NICE TA856[108] based on Woehl and McEwan 2008[120]
	Remission	0.87	
	Response	0.76	
CD	Remission	0.82	NICE TA342[109] based on GEMINI II/III studies[121]
	Moderate-severe	0.57	
Estimate for IBD	All (assumption)	0.75	Estimated based on values for each condition and severity
Utility multiplier for patients having increased		0.73	Stark *et al.* 2009[110]
Increase in IBD complications		0.04	Whyte *et al.* (Sophie Whyte, 1 March 2023, personal communication)

CD, Crohn's disease; EAG, Evidence Assessment Group; IBD, inflammatory bowel disease; UC, ulcerative colitis, TA, technology appraisal.

TABLE 23 Applied in the EAG model for IBD patients

Underlying condition	Proportion of patients	Source
Relative incidence of UC vs CD	0.60	Pasvol *et al.* 2020[103]
Proportion of patients in remission with UC	0.50	Ghosh and Premchand 2015[104]
Proportion of patients in relapse with UC	0.50	
Proportion UC relapse mild-moderate	0.80	
Proportion UC relapse severe	0.20	
Proportion of patients in remission with CD	0.50	
Proportion of patients in relapse with CD	0.50	

CD, Crohn's disease; EAG, Evidence Assessment Group; IBD, inflammatory bowel disease; UC, ulcerative colitis.

Faecal immunochemical test costs

The costs for each brand of FIT were sourced from information the manufacturers provided to NICE as part of the appraisal process. The price of each test is provided in *Table 24*. The EAG notes that some of these costs are indicative and may vary depending on some factors, such as type of analyser and methodology employed by laboratories, testing volumes and capacity. The impact of these factors on the costs that will be used in the NHS is unclear.

The cost of FIT for patients who complete the test also includes the costs of samples that need to be retaken (proportion of tests that need to be resampled for technical reasons of 2.1%), based on MacDonald *et al.*[56] The model assumes that 91% of patients complete the test, based on Bailey *et al.*[42]

The model also assumes that patients would receive an appointment with the GP to discuss the FIT results (with unit costs for a surgery consultation lasting 9.22 minutes obtained from the PSSRU[118]) and receive additional blood tests, assumed to include the costs of one phlebotomy service and one blood count (DAPSS08 and DAPS05 codes from NHS Reference Costs 2021–2[115]). Under intervention 3, the weighted mean of the price of all FIT was used (price = £4.24), with the weights based on the number of studies included in the statistical analysis of all tests (see *Chapter 2 Additional analysis 1: synthesis of all tests together in a single analysis*), that is, the analysis that informed the accuracy of FIT in intervention 3. The total cost per patient of the first FIT completed is estimated to be between £47.44 and £50.26, depending on the FIT brand received. Patients who do not return the FIT are assumed to incur only the cost of the test (£4.24 for intervention 3 and between £3.70 and £6.46 for interventions 1 and 2).

Costs of lower gastrointestinal referrals

In the model, patients referred to 2WW or 18WW are assumed to receive, regardless of the type of referral received, an initial appointment with a gastroenterologist consultant (£186.48, based on NHS Reference Costs – weighted average of first attendances face to face and non-face to face with gastroenterologist – codes 301, WF01B and WF01D).[115] Patients will also incur the costs of the imaging test received:

TABLE 24 Test costs assumed in EAG analysis

Test[a]	Total cost per test (£)[b]	Comments
NS-Prime	6.00	Cost per test provided by the company includes the cost of analyser and all consumables
QuikRead go	4.40	The manufacturer provided the cost for 50 tests, and the costs of 'sampling test', 'control quantitative' and 'instrument' separately. The total cost per test was estimated based on 50 tests
HM-JACKarc	4.10	The cost per test includes rental of the analyser, reagents, consumables, training and servicing, and patient packs
IDK Hemoglobin	6.46	The manufacturer provided the cost for different quantities of tests, and the costs of sampling test and extraction tubes separately for 100 tests. The total cost per test was estimated based on the lowest cost per test (96 tests)
IDK Hemoglobin/ Haptoglobin	6.46	The manufacturer provided the cost for different quantities of tests, and the costs of sampling test and extraction tubes separately for 100 tests. The total cost per test was estimated based on the lowest cost per test (96 tests)
OC Sensor	4.53	Total cost per test was based on the 'total costs including materials' from DG30.[11] The manufacturer also clarified that this cost included reagent rental of the analyser and that the cost per test is indicative, as it varies by testing volume and methodology employed by the testing laboratory
FOB Gold	3.70	Total cost per test based on the midpoint of the range of costs provided by the manufacturer. It is unclear if it includes the costs of other required consumables, or the analyser

a Only tests for which there were diagnostic test accuracy data have been included.
b The EAG notes that it is unclear if the prices provided for NS-Prime, HM-JACKarc and OC Sensor include VAT. For the other brands, the EAG was informed by the manufacturers that the prices do not include VAT.

- Colonoscopy: The model assumes to include the costs of a colonoscopy, based on the weighted average costs of colonoscopy with biopsy, therapeutic colonoscopy and diagnostic colonoscopy (codes FE30Z, FE31Z and FE32Z) from NHS Reference Costs,[115] plus the costs of same-day diagnostic imaging admission or attendance (code RD97Z) and a follow-up appointment with a gastroenterologist (based on weighted average cost for face-to-face and non-face-to-face attendances with a gastroenterologist – codes 301, WF01A and WF01C). The total cost of a colonoscopy was estimated to be £1003.34.
- CTC: The model includes the costs of a CTC scan (code RD61Z) in addition to the costs of a same-day diagnostic imaging admission or attendance (code RD97Z) and a follow-up appointment (based on weighted average cost for face-to-face and non-face-to-face attendances with a gastroenterologist – codes 301, WF01A and WF01C).[115] The total cost of a CTC was estimated to be £341.17.
- Other non-invasive investigations: the model includes the cost of a CT (based on the weighted average cost of CT scans of one or more areas with and without contrast available for adults – codes RD20A to RD27Z)[115] for 80% of patients (assumption made based on opinion of the EAG's clinical advisors, provided in *Report Supplementary Material 5*), and one additional appointment with a consultant gastroenterologist to discuss treatment action for all patients (weighted average of face-to-face and non-face-to-face attendances with a gastroenterologist, follow-up codes R301, WF01A and WF01C). The total cost of other non-invasive investigations was estimated to be £256.29 per patient.

Costs of watch and wait

Patients who receive a FIT result lower than t or t_{high} or who do not complete the test are followed up in primary care under the 'watch and wait' pathway. These patients are assumed to receive an additional 1.9 GP appointments based on estimated data from COLOFIT modelling team based on Lyratzopoulos *et al.*,[116] costing £68.40.[118] Additional costs during watchful waiting were estimated by calculating the weighted mean of potential costs.

Based on NCRAS data on routes to diagnosis for England 2018,[117] 22% of patients with underlying bowel disease (which was estimated to be 7.8% of patients presenting to the GP with symptoms of CRC) are assumed to incur the costs of A&E attendance (£296.88) based on the weighted average of all A&E attendances excluding dental and dead on arrival (codes VB01Z to VB11Z from NHS Reference Costs[115]). The remaining 78% of patients with underlying bowel disease are assumed to be eventually detected and referred by the GP to receive an initial appointment with a consultant gastroenterologist (£186.48) and a colonoscopy (£1003.34).[115] Patients with NSBP are assumed not to re-present to their GP and not to incur costs further to the additional 1.9 GP appointments.

The weighted mean cost per patient on the 'watch and wait' pathway was estimated to be £145.97.

Costs associated with 'repeat FIT'

Patients invited to complete a second FIT (under the 'repeat FIT' pathway) are assumed to incur the costs of the test, including the costs of the additional samples needed for those who complete it. The proportion of completed FIT that need to be resampled was assumed to be the same as for the first test. The total cost of repeat FIT per patient also includes 1.9 additional GP appointments (unit cost taken from PSSRU 2022[118]). Patients who do not complete the second FIT or who receive a 'negative' result are assumed to incur in the costs of 'watch and wait', while patients who complete it and receive a 'positive' result (based on the accuracy estimates of the test for CRC) are assumed to incur the cost of a colonoscopy and an appointment with a consultant gastroenterologist (both based on NHS Reference Costs 2021–2[115] using the same codes listed in the costs of referrals). The cost per patient of the 'repeat FIT' pathway varies depending on the FIT brand received and the threshold used, and is estimated to be, for example, between £373.82 and £535.35 for the threshold 10 µg/g (for the same brand, the cost decreases at higher thresholds).

Costs of treating complications related to colonoscopy

The costs of treating complications related to colonoscopy are summarised in *Table 20*.

Costs associated with treating the underlying conditions (lifetime costs for colorectal cancer, inflammatory bowel disease and advanced adenomas)

Appendix 8 presents details of the cost estimates used in the Whyte *et al.* model for CRC and AA patients.

For IBD patients, the lifetime costs were estimated from annual costs related to the disease treatment based on the proportion of patients by disease severity and type of disease reported in *Table 23* and unit costs from Ghosh and Premchand.[104] The annual cost of treatment was estimated to be £3083.94 for UC and £6156.44 for CD. Based on the relative incidence of UC versus CD of 0.6 from Pasvol *et al.*,[103] the annual cost of IBD was estimated to be £4297.70, which was uplifted to £5015.75 in 2022 prices using the NHS Cost Inflation Index (NHSCII).[114] The impact of delay on diagnosis of IBD was estimated to be £399.66, based on the difference in costs between severe relapse and milder forms of UC and CD, applied for 2 years since diagnosis, to those patients diagnosed within 'watch and wait' or 'repeat FIT' or missed by the diagnostic tests. Patients with NSBP are assumed not to incur in any long-term costs.

Methods of model evaluation

The health outcomes and costs of each testing strategy were generated for each brand of FIT based on each threshold and pair of thresholds being evaluated (with exception of intervention 3, where the threshold of 10 µg/g currently in place under DG30 was used). The total outcomes were evaluated against each other in full incremental analyses. The cost-effectiveness of each test brand was also compared against that of each of the others for selected thresholds. Results based on NMB were also generated. Central estimates of cost-effectiveness were based on the expectation of the mean. Uncertainty was evaluated using probabilistic sensitivity analysis and deterministic sensitivity analyses. Probabilistic sensitivity analysis was undertaken using simple Monte Carlo sampling methods (1000 samples). The choice of distribution assumed for each group of parameters in the model is summarised in *Appendix 9*, *Table 72*.

Scenario analyses

The EAG undertook a number of deterministic scenario analyses to test the robustness of the results generated by the model to changes in key assumptions. These included:

- Testing the influence of the estimates of time to diagnosis on the results. The EAG tested two different scenarios in which the lengths of time to diagnosis for each possible pathway in the model were assumed to be shorter (Scenario 1) and longer (Scenario 2) than in the model's base case. The estimates were based on the opinion of the EAG's clinical advisors (see *Table 21*).
- Inclusion of a QALY loss equivalent of 1 day of full health for the same age and sex match, to estimate the patient's anxiety about and inconvenience associated with receiving a colonoscopy. This QALY loss was applied to all patients who received a colonoscopy in the model.
- Inclusion of a QALY loss equivalent of 1 day of full health for each month of diagnostic delay. This QALY loss was applied to all patients with underlying disease (CRC, IBD, AAs) and also to those without underlying disease who had been referred to 2WW and 18WW.
- Testing the use of dual FIT instead of a single FIT. This scenario assumed to have accuracy data for dual FIT taken from the EAG's clinical review for HM JACKarc (see *Main analysis: dual faecal immunochemical test*) and the unit costs of the first FIT for intervention 1 to be the double for that brand.
- Removing IBD and AAs from the model. Given the uncertainty around the parameters for the other bowel pathologies included in the model, the EAG tested the removal of IBD and AA by assuming that they have zero prevalence.
- Lower return rate for FIT. Given that the value of the return rate of FIT in the base case is 0.91 and it might be considered high in the primary care context, the EAG tested using a second source from Moss *et al.*[122] of 0.664.
- Alternative assumption about diagnostic accuracy of FIT in the DG30 low-risk patients in intervention 3 based on the EAG's systematic review of studies that recruited only DG-30 low-risk patients. The values of sensitivity and specificity for DGD30 low-risk patients in intervention 3 in this scenario were 0.910 (95% CI 0.815 to 0.978) for sensitivity and 0.911 (95% CI 0.769 to 0.983) for specificity (see *Appendix 5*, *Table 49*).
- Alternative assumption of increased resource use in terms of GP appointments for patients with NSBP undertaking watch and wait and repeat FIT pathways. In this alternative scenario, the model assumes that patients without underlying disease who are not directly referred would receive one additional GP appointment.
- Alternative assumption of the cost of the test at current recommendations. In this scenario, the EAG changed the unit cost of FIT for intervention 3 from the weighted average to the lowest cost available of the unit costs informed by the manufacturers (the unit cost was changed from £4.24 to £3.70).

- The EAG also ran a scenario in which the FIT was assumed to have perfect accuracy (sensitivity and specificity = 1.0) and where all patients return the test (return rate = 1) to test an extreme scenario in which no patients are missed by test or wrongly sent to 2WW.
- The EAG also performed two additional analyses in which the prevalences of CRC, AAs and IBDs were reduced by 50% (Scenario 12) and increased by 50% (Scenario 13).

Model verification and validation

The EAG undertook a number of measures to ensure the validity of the model:

- peer review of the economic analysis by a modeller not involved in the assessment
- verification and scrutiny of the executable model by two model developers
- double-checking of the accuracy of all model inputs against sources
- comparison of model results using point estimates of parameters and the expectation of the mean
- comparison of mean of all probabilistic parameter samples against point estimates of parameters
- examination of all identified sources of discrepancy
- model testing using sensitivity analysis and use of extreme parameter values.

Cost-effectiveness results

Four key sets of results have been produced that include high or low safety-netting intensity (see *Probability of following each of the pathways following faecal immunochemical test result*) and assuming a willingness-to-pay threshold of £20,000 or £30,000 per QALY gained.[89] To allow multiple results to be shown on figures, an iNMB approach has been used by the EAG that requires specification of the assumed threshold. iNMB is defined as the cost per QALY gained threshold multiplied by the incremental QALY gain minus the incremental cost;[123] under this framework the largest estimated iNMB is deemed to be the most cost-effective strategy, which could be zero if the benchmark intervention is most cost-effective. The absolute loss (valued in terms of cost) of moving to a different strategy is calculated by comparing the estimated iNMBs.

Net monetary benefit also has the advantage that if the assumed costs are believed to be imprecise then the level of additional or reduced costs (e.g. the additional costs of GP appointments incurred over the base case) can be directly applied to the NMB values. It is for this reason that NMB is preferred to net health benefits, although the conclusions are identical whichever metric is used. The NMB values presented are per person.

The conclusions from all four analyses are similar and therefore only one set of results are presented in the main text, with the results of the remaining three analyses in *Report Supplementary Material 6*. The chosen combination uses a low safety-netting approach and a threshold of £20,000 per QALY gained as this is most different from current standard of care and the lower threshold as there is likely to be considerable uncertainty in the ICER given the small differences in QALYs between strategies.

The structure for presenting (and interpreting) results is as follows:

1. A figure depicting the iNMBs for each of the seven tests at selected thresholds when only one threshold is used (denoted intervention 1).
2. A figure depicting the iNMBs for the five tests with sufficient data at selected thresholds when two thresholds are used, t_{low} and t_{high} (denoted intervention 2).
3. Tables for each test that display summarised data relating to the clinical effectiveness and cost-effectiveness of each test compared with current care (denoted intervention 3). The EAG has selected the data that it considers to be most pertinent for decision-making; other data can be provided by the EAG on request.

The EAG ran probabilistic sensitivity analysis for selected tests and thresholds; these indicated that the NMB values differed by less than £10 on average. Given the linearity of the results and the timescales of the project, the EAG deemed that presenting only deterministic results would not influence decision-making.

The incremental net monetary benefits of the seven tests using one threshold assuming a threshold of £20,000 per quality-adjusted life-year gained and low-intensity safety-netting threshold

Figure 14 shows the iNMB for the diagnostic strategies for FIT when using one threshold (intervention 1). The iNMBs for higher threshold values were in the region of £300–350 for all tests. With the exception of NS-Prime at a low threshold, all tests have positive iNMBs compared with current practice. The reason for the negative iNMB for NS-Prime at a threshold of 3 µg/g is the poor estimated specificity of the test at this threshold (0.319), which came from one study, Benton *et al.*,[44] that had a very small number of events. This results in a large number of patients being referred to colonoscopy. iNMB values in *Figure 14* have been interpolated resulting in straight lines where there is a distance between thresholds (e.g. 10 µg/g and 100 µg/g). While the highest iNMB values appear to be around a threshold of 100 µg/g, the iNMB loss using a threshold of 50 µg/g is slight. Although for the majority of tests there is a noticeable reduction in iNMB at a threshold of 10 µg/g, the results show that all tests used at this threshold have a higher iNMB than current practice. Given the uncertainty in the model input parameters, the EAG notes that the generated comparisons between thresholds for a particular test, or between tests themselves, may not be robust, although broad conclusions are likely to be robust.

The incremental net monetary benefits of the five tests using two thresholds assuming a threshold of £20,000 per quality-adjusted life-year gained and low-intensity safety-netting threshold

Figure 15 shows the iNMB when using FIT strategies with two thresholds. The iNMBs for higher threshold values appear to be in the region of £200–£350 for all tests. These iNMBs are lower than when one threshold is used (see *Figure 14*). All iNMBs are positive except for when the value for t_{low} is set to 3 µg/g for NS-Prime, which is due to the low specificity of this test at this threshold. Many pairs of combinations have reasonably similar iNMB values given the underlying uncertainty, for instance using paired values of 7 µg/g and 50 µg/g compared with using 10 µg/g and 100 µg/g.

Tabulated results for each test assuming a threshold of £20,000 per quality-adjusted life-year gained and low intensity safety netting threshold

Tables 25–34 show tabulated results for each test (at selected thresholds). The tests are presented in alphabetical order. The vast majority of ICERs presented in these tables lie in the south-west quadrant of the cost-effectiveness plane, which is due to the tests generating marginally fewer QALYs and lower costs than current practice. The explanation for this phenomenon, using FOB Gold at a threshold of 10 µg/g as an example (as it is first alphabetically), follows.

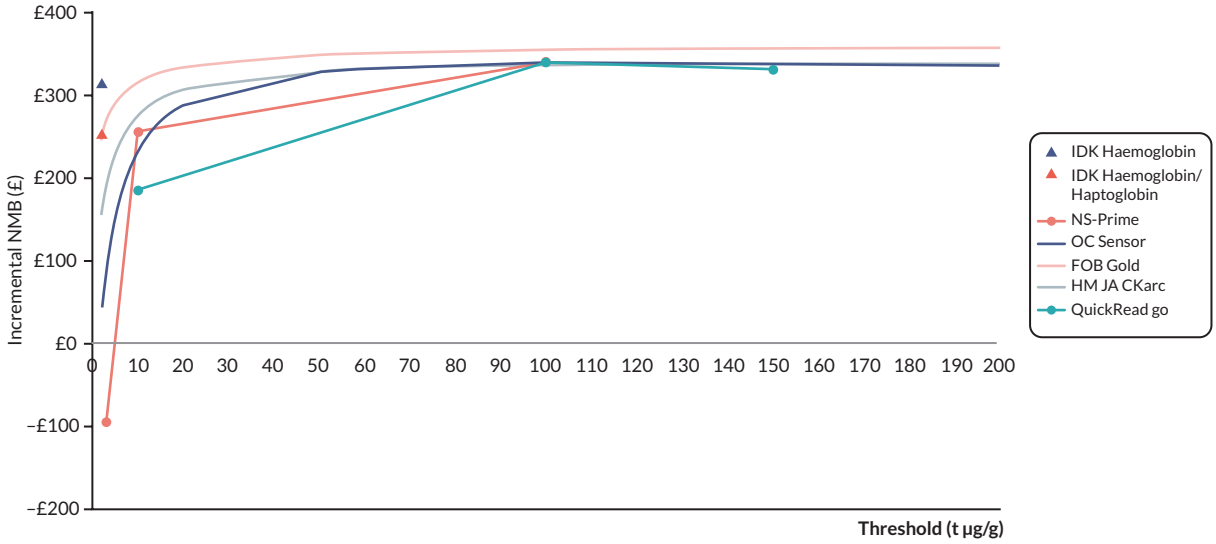

FIGURE 14 Net monetary benefit for intervention 1 assuming a threshold of £20,000 per QALY gained and low-intensity safety netting.

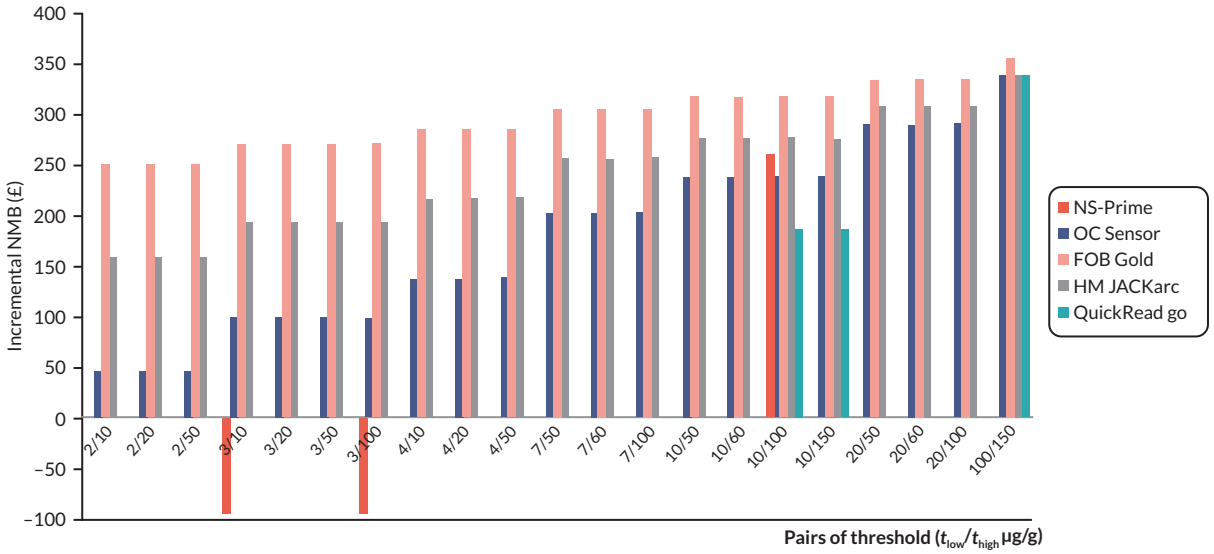

FIGURE 15 Net monetary benefit for intervention 2 assuming a threshold of £20,000 per QALY gained and low-intensity safety netting.

Illustrated example of why quality-adjusted life-years marginally decrease using faecal immunochemical tests

Under current practice, a large number of patients are referred to receive colonoscopy (0.620 colonoscopies per patient) (see *Table 25*). For FOB Gold at a threshold of 10 µg/g, some patients with bowel disease will be missed due to the comparatively lower sensitivity of the test, and fewer patients will receive colonoscopy, with an estimated 0.255 colonoscopies per patient (see *Table 25*). Undertaking more colonoscopies increases the probability that bowel disease will be detected and patients will receive appropriate treatment. Patients with underlying CRC who do not receive colonoscopy experience delays that result in worse outcomes and lower QALYs for the cohort.

Although there is a benefit in quicker time to diagnosis for those in the 2WW and 18WW pathways using FIT due to lower demand for colonoscopy resources, this is not sufficient to outweigh the losses associated with delayed diagnosis. The mean time to a diagnosis of CRC is 1.388 months in current practice and 3.014 months with FOB Gold at a threshold of 10 µg/g (see *Table 25*). The later average diagnosis of CRC (and similarly for AAs and IBD) means that QALYs for the cohort are decreased from 10.895 in current practice to 10.892 for FOB Gold at a threshold of 10 µg/g (see *Table 25*), which is < 1 day of full health for all patients in the cohort.

The different proportions of patients with CRC diagnosed in the 2WW pathway, the 18WW pathway and the watch and wait and repeat FIT combined by current practice and FOB Gold at a threshold of 10 µg/g are shown in *Figure 16*.

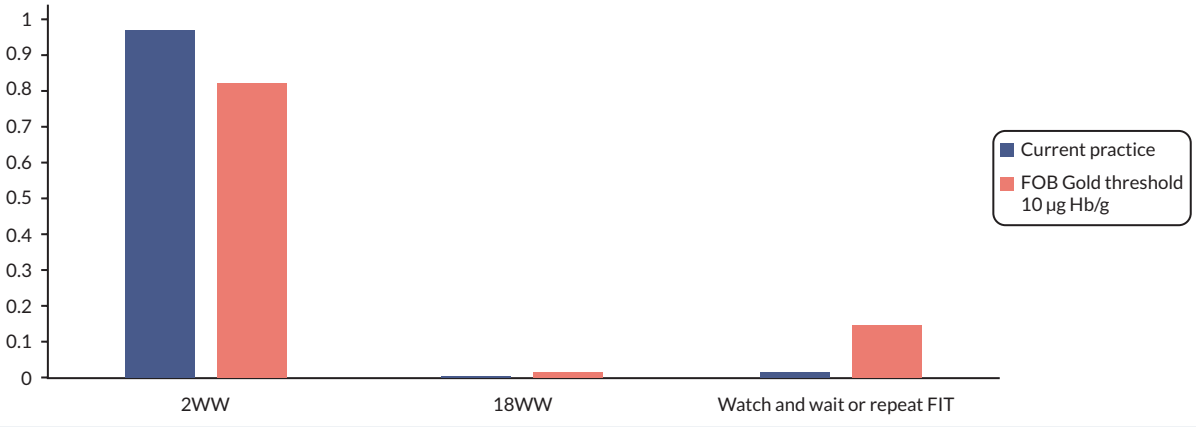

FIGURE 16 The proportion of patients diagnosed by category.

Illustrated example of why costs are lower when using faecal immunochemical tests

While there will be an increase in the costs associated with the use of FIT, there is a consequential decrease in the number of colonoscopies undertaken. As noted in the previous section, the estimated average number of colonoscopies undertaken per person was 0.620 for current practice and 0.255 for FOB Gold at a threshold of 10 µg/g. This reduction in colonoscopy usage generates a considerable saving, which drives an overall reduction of costs from £3138 in current care to £2773 for FOB Gold at a threshold of 10 µg/g (see *Table 25*).

Combining the estimated implications for QALYs accrued and costs incurred, the ICER for FOB Gold at a threshold of 10 µg/g is calculated to be £149,794 (see *Table 25*), although this is in the south-west quadrant, indicating that for every QALY yielded there would be a saving of £149,794. Alternatively, this could be viewed as current care having an ICER of £149,794 compared with FOB Gold at a threshold of 10 µg/g, which is higher than the thresholds of £20,000 or £30,000 published by NICE.[89]

TABLE 25 Tabulated results for FOB Gold using one threshold

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.165	14.165	14.165	14.164	14.164	14.164	14.168
QALYs	10.893	10.893	10.893	10.893	10.892	10.892	10.891	10.891	10.891	10.895
Costs (£)	2857	2817	2805	2789	2773	2747	2720	2715	2704	3138
ICER (pairwise, vs. intervention 3)[a] (£)	169,015	165,767	162,824	157,162	149,794	137,172	120,371	117,255	108,935	–
NMB λ = 20,000 (vs. intervention 3) (£)	248	283	292	305	316	334	349	351	355	–
NMB λ = 30,000 (vs. intervention 3) (£)	231	263	272	283	292	306	314	315	315	–
Number of 2WW referrals (total)	0.282	0.238	0.226	0.208	0.192	0.164	0.135	0.130	0.118	0.639
Number of 18WW referrals (total)	0.076	0.080	0.081	0.083	0.085	0.088	0.091	0.092	0.093	0.038
Number of repeat FIT (total)	0.076	0.080	0.081	0.083	0.085	0.088	0.091	0.092	0.093	0.038
Number of watch and wait (total)	0.567	0.601	0.611	0.625	0.638	0.660	0.683	0.687	0.696	0.285
Number of COLs (total)	0.328	0.293	0.283	0.268	0.255	0.232	0.208	0.204	0.194	0.620
Reduction in number of referrals (total – 2WW + 18WW)	47.2%	53.0%	54.6%	56.9%	59.1%	62.8%	66.6%	67.2%	68.8%	–
Reduction in number of referrals (2WW only)	55.9%	62.7%	64.7%	67.4%	70.0%	74.4%	78.9%	79.6%	81.5%	–
Increase in number of referrals (18WW only)[b]	99.1%	111.1%	114.6%	119.4%	124.1%	131.8%	139.7%	141.1%	144.4%	–
Reduction in number of COLs	47.0%	52.7%	54.4%	56.7%	58.9%	62.6%	66.4%	67.0%	68.7%	–
Mean time to diagnosis – CRC	2.661	2.789	2.837	2.917	3.014	3.235	3.613	3.698	3.964	1.388
Mean time to diagnosis – AAs	4.384	5.073	5.341	5.780	6.307	7.185	8.319	8.538	9.098	1.956
Mean time to diagnosis – IBD	2.873	3.300	3.463	3.735	4.058	4.777	5.811	6.020	6.588	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

TABLE 26 Tabulated results for FOB Gold using two thresholds

	Int 2: FIT 2 thresholds												Intervention 3: DG30 and NG12
t_{low}/t_{high} (µg/g)	2/10	2/20	2/50	4/10	4/20	4/50	7/50	7/100	10/50	10/100	20/50	20/100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.166	14.165	14.165	14.165	14.165	14.165	14.165	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.893	10.893	10.892	10.892	10.892	10.892	10.892	10.895
Costs (£)	2854	2853	2852	2815	2814	2813	2786	2786	2771	2771	2746	2745	3138
ICER (pairwise, vs. intervention 3)[a] (£)	166,148	164,264	161,281	164,219	162,749	160,365	153,606	151,995	147,284	145,916	136,022	135,003	–
NMB λ = 20,000 (vs. intervention 3) (£)	250	250	251	284	284	285	306	306	317	317	335	335	–
NMB λ = 30,000 (vs. intervention 3) (£)	233	233	233	264	264	264	283	283	292	292	306	306	–
Number of 2WW referrals (total)	0.267	0.263	0.259	0.231	0.227	0.222	0.197	0.194	0.183	0.180	0.159	0.157	0.639
Number of 18WW referrals (total)	0.085	0.088	0.091	0.085	0.088	0.091	0.091	0.093	0.091	0.093	0.091	0.093	0.038
Number of repeat FIT (total)	0.080	0.082	0.083	0.083	0.084	0.086	0.087	0.088	0.088	0.089	0.090	0.090	0.038
Number of watch and wait(total)	0.567	0.567	0.567	0.601	0.601	0.601	0.625	0.625	0.638	0.638	0.660	0.660	0.285
Number of COLs (total)	0.324	0.323	0.321	0.291	0.289	0.288	0.265	0.264	0.252	0.251	0.231	0.230	0.620
Reduction in number of referrals (total – 2WW + 18WW)	47.9%	48.2%	48.4%	53.3%	53.5%	53.8%	57.5%	57.6%	59.6%	59.7%	63.0%	63.2%	–
Reduction in number of referrals (2WW only)	58.2%	58.9%	59.6%	63.9%	64.6%	65.3%	69.2%	69.6%	71.4%	71.8%	75.1%	75.5%	–
Increase in number of referrals (18WW only)	124.1%	131.8%	139.7%	124.1%	131.8%	139.7%	139.7%	144.4%	139.7%	144.4%	139.7%	144.4%	–

continued

TABLE 26 Tabulated results for FOB Gold using two thresholds (*continued*)

	Int 2: FIT 2 thresholds												Intervention 3: DG30 and NG12
t_{low}/t_{high} (µg/g)	2/10	2/20	2/50	4/10	4/20	4/50	7/50	7/100	10/50	10/100	20/50	20/100	10
Increase in number of repeat FIT	111.6%	115.4%	119.4%	117.6%	121.5%	125.4%	129.6%	131.9%	131.9%	134.3%	135.8%	138.1%	–
Increase in number of watch and waits	99.1%	99.1%	99.1%	111.1%	111.1%	111.1%	119.4%	119.4%	124.1%	124.1%	131.8%	131.8%	–
Reduction in number of COLs	47.7%	47.9%	48.2%	53.1%	53.3%	53.5%	57.2%	57.4%	59.3%	59.5%	62.8%	62.9%	–
Mean time to diagnosis – CRC	2.682	2.695	2.718	2.802	2.814	2.835	2.955	2.974	3.046	3.065	3.254	3.272	1.388
Mean time to diagnosis – AAs	4.497	4.548	4.615	5.141	5.190	5.252	5.914	5.955	6.410	6.450	7.241	7.280	1.956
Mean time to diagnosis – IBD	2.944	2.988	3.050	3.343	3.384	3.443	3.849	3.892	4.152	4.194	4.830	4.870	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.

TABLE 27 Tabulated results for HM JACKarc using one threshold

	Intervention 1: FIT 1 threshold (µg/g)									Intervention 3: DG30and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.162	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.892	10.891	10.891	10.890	10.895
Costs (£)	2955	2890	2870	2843	2817	2774	2730	2723	2705	3138
ICER (pairwise, vs. intervention 3)[a] (£)	132,329	148,950	149,562	147,281	141,344	127,155	105,147	100,914	90,028	–
NMB λ = 20,000 (vs. intervention 3) (£)	156	215	232	255	276	307	330	333	337	–
NMB λ = 30,000 (vs. intervention 3) (£)	142	199	214	235	253	278	292	292	289	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FITs (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038

TABLE 27 Tabulated results for HM JACKarc using one threshold (*continued*)

	Intervention 1: FIT 1 threshold (µg/g)									Intervention 3: DG30and NG12
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	2.310	2.461	2.529	2.653	2.815	3.237	4.045	4.242	4.849	1.388
Mean time to diagnosis – AAs	4.456	5.128	5.390	5.822	6.341	7.207	8.329	8.546	9.102	1.956
Mean time to diagnosis – IBD	2.946	3.355	3.513	3.777	4.092	4.798	5.821	6.027	6.592	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

TABLE 28 Tabulated results for HM JACKarc using two thresholds

	Intervention 2: FIT 2 thresholds												Intervention 3: DG30 and NG12
t_{low}/t_{high} (µg/g)	2/10	2/20	2/50	4/10	4/20	4/50	7/50	7/100	10/50	10/100	20/50	20/100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.166	14.166	14.166	14.165	14.165	14.165	14.165	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.893	10.893	10.893	10.892	10.892	10.892	10.892	10.895
Costs (£)	2951	2950	2948	2887	2886	2884	2839	2838	2814	2813	2772	2771	3138
ICER (pairwise, vs. intervention 3) (£)	129,593	126,912	122,103	147,094	144,787	140,505	141,529	138,380	137,280	134,675	125,399	123,641	–
NMB λ = 20,000 (vs. intervention 3) (£)	159	159	159	217	218	218	257	257	277	277	308	308	–

continued

TABLE 28 Tabulated results for HM JACKarc using two thresholds (*continued*)

	Intervention 2: FIT 2 thresholds												Intervention 3: DG30 and NG12
t_{low}/t_{high} (µg/g)	2/10	2/20	2/50	4/10	4/20	4/50	7/50	7/100	10/50	10/100	20/50	20/100	10
NMB λ = 30,000 (vs. intervention 3) (£)	144	144	143	200	200	200	236	235	254	253	279	278	–
Number of 2WW referrals (total)	0.363	0.355	0.348	0.302	0.295	0.288	0.246	0.242	0.222	0.219	0.184	0.180	0.639
Number of 18WW referrals (total)	0.080	0.085	0.090	0.080	0.085	0.090	0.090	0.092	0.090	0.092	0.090	0.092	0.038
Number of repeat FIT (total)	0.072	0.075	0.077	0.076	0.079	0.081	0.084	0.085	0.085	0.086	0.087	0.089	0.038
Number of watch and wait (total)	0.485	0.485	0.485	0.541	0.541	0.541	0.581	0.581	0.603	0.603	0.638	0.638	0.285
Number of COLs (total)	0.406	0.404	0.402	0.351	0.349	0.347	0.308	0.307	0.287	0.286	0.252	0.251	0.620
Reduction in number of referrals (total – 2WW + 18WW)	34.6%	34.9%	35.3%	43.5%	43.9%	44.2%	50.4%	50.6%	53.9%	54.1%	59.5%	59.7%	–
Reduction in number of referrals (2WW only)	43.3%	44.4%	45.5%	52.7%	53.8%	54.9%	61.5%	62.1%	65.2%	65.8%	71.2%	71.8%	–
Increase in number of referrals (18WW only)	111.6%	124.2%	136.5%	111.6%	124.2%	136.5%	136.5%	143.4%	136.5%	143.4%	136.5%	143.4%	–
Increase in number of repeat FIT	90.9%	97.1%	103.3%	100.8%	107.1%	113.3%	120.2%	123.6%	124.1%	127.5%	130.3%	133.8%	–
Increase in number of watch and waits	70.1%	70.1%	70.1%	90.0%	90.0%	90.0%	103.9%	103.9%	111.6%	111.6%	124.2%	124.2%	–
Reduction in number of COLs	34.4%	34.8%	35.1%	43.3%	43.7%	44.0%	50.2%	50.4%	53.7%	53.9%	59.3%	59.5%	–
Mean time to diagnosis – CRC	2.345	2.373	2.428	2.484	2.510	2.560	2.734	2.781	2.884	2.930	3.280	3.323	1.388
Mean time to diagnosis – AAs	4.582	4.639	4.714	5.202	5.255	5.323	5.965	6.009	6.451	6.493	7.265	7.305	1.956
Mean time to diagnosis – IBD	3.025	3.074	3.143	3.402	3.446	3.510	3.898	3.943	4.191	4.235	4.853	4.895	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.

a South-west quadrant ICER

TABLE 29 Tabulated results for IDK Haemoglobin using one threshold

	Intervention 1: FIT 1 threshold	Intervention 3: DG30 and NG12
t (µg/g)	2	10
LYs	14.165	14.168
QALYs	10.893	10.895
Costs (£)	2783	3138
ICER (pairwise, vs. intervention 3)[a] (£)	180,462	–
NMB λ = 20,000 (vs. intervention 3) (£)	316	–
NMB λ = 30,000 (vs. intervention 3) (£)	296	–
Number of 2WW referrals (total)	0.201	0.639
Number of 18WW referrals (total)	0.084	0.038
Number of repeat FIT (total)	0.084	0.038
Number of watch and wait (total)	0.631	0.285
Number of COLs (total)	0.263	0.620
Reduction in number of referrals (total – 2WW + 18WW)	57.9%	–
Reduction in number of referrals (2WW only)	68.5%	–
Increase in number of referrals (18WW only)[b]	121.5%	–
Reduction in number of COLs	57.6%	–
Mean time to diagnosis – CRC	3.029	1.388
Mean time to diagnosis – AAs	4.330	1.956
Mean time to diagnosis – IBD	2.816	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

TABLE 30 Tabulated results for IDK Haemoglobin/haptoglobin using one threshold

	Intervention 1: FIT 1 threshold	Intervention 3: DG30 and NG12
t (µg/g)	2	10
LYs	14.164	14.168
QALYs	10.892	10.895
Costs (£)	2836	3138
ICER (pairwise, vs. intervention 3)[a] (£)	125,502	–
NMB λ = 20,000 (vs. intervention 3) (£)	254	–
NMB λ = 30,000 (vs. intervention 3) (£)	230	–
Number of 2WW referrals (total)	0.258	0.639
Number of 18WW referrals (total)	0.078	0.038

continued

TABLE 30 Tabulated results for IDK Haemoglobin/haptoglobin using one threshold (*continued*)

	Intervention 1: FIT 1 threshold	Intervention 3: DG30 and NG12
Number of repeat FIT (total)	0.078	0.038
Number of watch and wait (total)	0.585	0.285
Number of COLs (total)	0.309	0.620
Reduction in number of referrals (total – 2WW + 18WW)	0.503	–
Reduction in number of referrals (2WW only)	59.6%	–
Increase in number of referrals (18WW only)[b]	105.6%	–
Reduction in number of COLs	50.1%	–
Mean time to diagnosis – CRC	3.479	1.388
Mean time to diagnosis – AAs	4.369	1.956
Mean time to diagnosis – IBD	2.856	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

TABLE 31 Tabulated results for NS-Prime

	Intervention 1: FIT 1 threshold			Intervention 2: FIT 2 thresholds			Intervention 3: DG30 and NG12 (FIT T= 10)
Threshold – t or t_{low}/t_{high} (µg/g)	**3**	**10**	**100**	**3/10**	**3/100**	**10/100**	**10**
LYs	14.164	14.162	14.160	14.164	14.164	14.162	14.168
QALYs	10.892	10.891	10.889	10.892	10.892	10.891	10.895
Costs (£)	3183	2804	2684	3177	3175	2800	3138
ICER (pairwise, vs. intervention 3) (£)	Dominated	87,155[a]	80,479[a]	Dominated	Dominated	86,080[a]	–
NMB λ = 20,000 (vs. intervention 3) (£)	−97	258	341	−94	−95	260	–
NMB λ = 30,000 (vs. intervention 3) (£)	−124	219	285	−121	−124	221	–
Number of 2WW referrals (total)	0.645	0.226	0.101	0.579	0.559	0.206	0.639
Number of 18WW referrals (total)	0.037	0.081	0.095	0.081	0.095	0.095	0.038
Number of repeat FIT (total)	0.037	0.081	0.095	0.059	0.066	0.088	0.038
Number of watch and wait (total)	0.280	0.611	0.710	0.280	0.280	0.611	0.285
Number of COLs (total)	0.624	0.282	0.180	0.604	0.598	0.276	0.620
Reduction in number of referrals (total – 2WW + 18WW)	−0.8%	54.6%	71.1%	2.5%	3.4%	55.6%	–
Reduction in number of referrals (2WW)	−0.9%	64.7%	84.2%	9.4%	12.5%	67.8%	–
Increase in number of referrals (18WW)	−1.7%	114.6%	149.2%	114.6%	149.2%	149.2%	–
Increase in number of repeat FIT	−1.7%	114.6%	149.2%	56.5%	73.8%	131.9%	–
Increase in number of watch and waits	−1.7%	114.6%	149.2%	−1.7%	−1.7%	114.6%	–
Reduction in number of COLs	−0.7%	54.5%	71.0%	2.6%	3.5%	55.4%	–

TABLE 31 Tabulated results for NS-Prime (*continued*)

	Intervention 1: FIT 1 threshold			Intervention 2: FIT 2 thresholds			Intervention 3: DG30 and NG12 (FIT T= 10)
Threshold – t or t_{low}/t_{high} (µg/g)	3	10	100	3/10	3/100	10/100	10
Mean time to diagnosis – CRC	3.454	4.514	5.763	3.554	3.664	4.584	1.388
Mean time to diagnosis – AAs	5.038	6.333	9.083	5.157	5.396	6.482	1.956
Mean time to diagnosis – IBD	3.389	4.084	6.573	3.457	3.675	4.224	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.

TABLE 32 Tabulated results for OC-Sensor using one threshold

	Intervention 1: FIT 1 threshold (µg/g)									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.161	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.892	10.891	10.891	10.890	10.895
Costs (£)	3066	2970	2940	2898	2857	2791	2731	2722	2701	3138
ICER (pairwise, vs. intervention 3)[a] (£)	48,900	97,821	108,234	118,602	123,019	120,367	103,122	99,109	88,384	–
NMB λ = 20,000 (vs. intervention 3) (£)	43	134	162	200	236	290	328	332	339	–
NMB λ = 30,000 (vs. intervention 3) (£)	28	117	143	180	213	261	289	290	289	–
Number of 2WW referrals (total)	0.510	0.402	0.369	0.322	0.278	0.209	0.147	0.138	0.117	0.639
Number of 18WW referrals (total)	0.052	0.063	0.066	0.071	0.076	0.083	0.090	0.091	0.093	0.038
Number of repeat FIT (total)	0.052	0.063	0.066	0.071	0.076	0.083	0.090	0.091	0.093	0.038
Number of watch and wait (total)	0.387	0.472	0.498	0.535	0.570	0.625	0.673	0.680	0.697	0.285
Number of COLs (total)	0.514	0.426	0.399	0.361	0.325	0.269	0.218	0.211	0.193	0.620
Reduction in number of referrals (total – 2WW + 18WW)	17.1%	31.4%	35.8%	41.9%	47.7%	56.9%	65.0%	66.2%	69.0%	–
Reduction in number of referrals (2WW only)	20.2%	37.2%	42.3%	49.6%	56.5%	67.3%	77.0%	78.4%	81.6%	–
Increase in number of referrals (18WW only)[b]	35.9%	65.9%	75.0%	87.9%	100.1%	119.3%	136.4%	138.9%	144.7%	–

continued

TABLE 32 Tabulated results for OC-Sensor using one threshold (*continued*)

	Intervention 1: FIT 1 threshold (µg/g)									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
Reduction in number of COLs	17.1%	31.3%	35.6%	41.7%	47.5%	56.7%	64.8%	66.0%	68.8%	–
Mean time to diagnosis – CRC	2.350	2.477	2.538	2.654	2.814	3.249	4.120	4.336	5.005	1.388
Mean time to diagnosis – AAs	4.540	5.190	5.445	5.865	6.374	7.222	8.330	8.545	9.097	1.956
Mean time to diagnosis – IBD	3.034	3.418	3.567	3.820	4.124	4.812	5.821	6.026	6.587	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Conclusions from the cost-effectiveness analyses undertaken in the Evidence Assessment Group's base case

The results generated by the EAG indicate that in the vast majority of analyses, the use of FIT has a positive NMB compared with current care. This is produced not by an increase in patient health, as there is a very small decrease in estimated QALYs per person (< 0.005 QALYs per person), but instead by the moderate cost savings per person (in the region of £300). This conclusion holds across a wide range of thresholds, with the EAG noting that the complex real-world process has been simplified in the model and that uncertainty in parameter inputs results in large uncertainty when directly comparing thresholds for the same test or comparing tests directly. The EAG has undertaken sensitivity analysis to explore the robustness of the broad conclusions when using alternative assumptions and data inputs.

The EAG notes that where the use of FIT results in a positive NMB compared with current care, this additionally reduces demand for colonoscopies. The EAG also notes that the economic analyses were developed based on the UK setting (NHS/PSS perspective), and the results generated by the model may not be transferable to other healthcare systems.

Deterministic scenario analyses

The results of the 13 scenario analyses run by the EAG can be found in *Appendix 10*, *Tables 73–86*. For illustrative purposes, all of these analyses have been conducted only on the comparison between HM JACKarc using one threshold (intervention 1) and current recommendations (intervention 3) using the lower intensity option for safety netting.

TABLE 33 Tabulated results for OC-Sensor using two thresholds

	Intervention 2: FIT 2 thresholds												Intervention 3: DG30 and NG12
t_{low}/t_{high} (µg/g)	2/10	2/20	2/50	4/10	4/20	4/50	7/50	7/100	10/50	10/100	20/50	20/100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.166	14.166	14.166	14.165	14.165	14.165	14.165	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.893	10.893	10.893	10.892	10.892	10.892	10.892	10.895
Costs (£)	3061	3060	3059	2966	2965	2963	2892	2891	2852	2851	2789	2788	3138
ICER (pairwise, vs. intervention 3)[a] (£)	49,968	49,430	47,803	97,494	96,302	93,440	114,504	111,697	119,798	117,313	118,801	117,031	–
NMB λ = 20,000 (vs. intervention 3) (£)	46	47	46	137	138	138	204	203	239	239	291	291	–
NMB λ = 30,000 (vs. intervention 3) (£)	31	31	30	119	120	119	182	181	215	214	261	261	–
Number of 2WW referrals (total)	0.473	0.462	0.453	0.382	0.371	0.361	0.295	0.290	0.257	0.253	0.199	0.194	0.639
Number of 18WW referrals (total)	0.076	0.083	0.090	0.076	0.083	0.090	0.090	0.093	0.090	0.093	0.090	0.093	0.038
Number of repeat FIT (total)	0.064	0.067	0.071	0.069	0.073	0.076	0.081	0.082	0.083	0.084	0.087	0.088	0.038
Number of watch and wait (total)	0.387	0.387	0.387	0.472	0.472	0.472	0.535	0.535	0.570	0.570	0.625	0.625	0.285
Number of COLs (total)	0.503	0.499	0.496	0.420	0.416	0.413	0.353	0.351	0.319	0.317	0.266	0.264	0.620
Reduction in number of referrals (total – 2WW + 18WW)	18.9%	19.4%	19.9%	32.4%	32.9%	33.4%	43.2%	43.5%	48.7%	49.0%	57.3%	57.6%	–
Reduction in number of referrals (2WW only)	26.0%	27.7%	29.2%	40.2%	41.9%	43.5%	53.9%	54.6%	59.7%	60.5%	68.8%	69.6%	–
Increase in number of referrals (18WW only)	100.1%	119.3%	136.4%	100.1%	119.3%	136.4%	136.4%	144.7%	136.4%	144.7%	136.4%	144.7%	–
Increase in number of repeat FIT	68.0%	77.6%	86.1%	83.0%	92.6%	101.1%	112.1%	116.3%	118.2%	122.4%	127.8%	132.0%	–
Increase in number of watch and waits	35.9%	35.9%	35.9%	65.9%	65.9%	65.9%	87.9%	87.9%	100.1%	100.1%	119.3%	119.3%	–
Reduction in number of COLs	18.9%	19.4%	19.9%	32.3%	32.8%	33.3%	43.1%	43.3%	48.6%	48.8%	57.1%	57.4%	–
Mean time to diagnosis – CRC	2.388	2.422	2.489	2.501	2.531	2.591	2.747	2.803	2.892	2.944	3.296	3.344	1.388
Mean time to diagnosis – AAs	4.681	4.746	4.831	5.271	5.328	5.403	6.017	6.064	6.487	6.532	7.280	7.321	1.956
Mean time to diagnosis – IBD	3.121	3.174	3.252	3.468	3.516	3.586	3.947	3.996	4.226	4.272	4.868	4.910	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.
a South-west quadrant ICER.

TABLE 34 Tabulated results for QuikRead go

	Intervention 1: FIT 1 threshold			Intervention 2: FIT 2 thresholds			Intervention 3: DG30 and NG12
t (µg/g)	10	100	150	10/100	10/150	100/150	10
LYs	14.166	14.163	14.160	14.166	14.166	14.162	14.168
QALYs	10.893	10.890	10.889	10.893	10.892	10.890	10.895
Costs (£)	2913	2710	2692	2906	2905	2710	3138
ICER (pairwise, vs. intervention 3)[a] (£)	108,338	96,527	77,581	103,941	100,442	95,336	–
NMB λ = 20,000 (vs. intervention 3) (£)	184	339	331	188	187	339	–
NMB λ = 30,000 (vs. intervention 3) (£)	163	295	274	165	163	294	–
Number of 2WW referrals (total)	0.338	0.126	0.110	0.305	0.302	0.123	0.639
Number of 18WW referrals (total)	0.070	0.092	0.094	0.092	0.094	0.094	0.038
Number of repeat FIT (total)	0.070	0.092	0.094	0.081	0.082	0.093	0.038
Number of watch and wait (total)	0.523	0.690	0.703	0.523	0.523	0.690	0.285
Number of COLs (total)	0.374	0.200	0.187	0.364	0.363	0.200	0.620
Reduction in number of referrals (total – 2WW + 18WW)	39.8%	67.8%	69.9%	41.4%	41.6%	68.0%	–
Reduction in number of referrals (2WW only)	47.1%	80.3%	82.8%	52.4%	52.8%	80.7%	–
Increase in number of referrals (18WW only)[b]	83.5%	142.3%	146.7%	142.3%	146.7%	146.7%	–
Reduction in number of COLs	39.7%	67.7%	69.8%	41.3%	41.4%	67.8%	–
Mean time to diagnosis – CRC	2.558	4.439	5.771	2.679	2.764	4.504	1.388
Mean time to diagnosis – AAs	6.420	9.105	9.486	6.589	6.613	9.123	1.956
Mean time to diagnosis – IBD	4.169	6.595	7.028	4.327	4.355	6.616	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; NMB, net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year; *t*, threshold.

a South-west quadrant ICER.

b Also the value for increased repeat FIT and increased number of watch and waits.

Chapter 4 Discussion and conclusions

Statement of principal findings

Clinical effectiveness: principal findings

The systematic review found no end-to-end RCT studies for any of the tests. The number of diagnostic test accuracy data differed across tests. Seventeen studies reported across 21 publications[17,18,27,46–48,50,52,56,60,61,63,65,66,71–75,81,82] reported data for HM-JACKarc; 17 studies reported across 18 publications[30,41,43–46,49,51,54,55,57,62,64,68,70,76–78] reported data for OC-Sensor; three studies[44,59,83] reported data for FOB-Gold; one study[58] reported data on QuikRead go; one study[44] reported data for NS-Prime; and one study[79] reported data for both IDK Hb and for IDK Hb/Hp complex. No diagnostic test accuracy data were found for the combined use of IDK Hb + Hb/Hp or for IDK TurbiFIT tests.

Population types

Studies were categorised according to the recruitment criteria as either population type 1 (studies closest to being a representative spectrum of all patients presenting to primary care with symptoms of CRC who meet NG12 or DG30 criteria); population type 2 (studies closest to being a representative spectrum of NG12 high-/medium-risk patients); population type 3 (studies closest to being a representative spectrum of DG30 low-risk patients); or population type 4 (unclear/likely to be unrepresentative spectrum). This last category contained a mix of studies that had different reasons for being 'unrepresentative', the most common being that the study had recruited patients who had been referred to the 2WW secondary care referral pathway. This may be a mix of NG12 high-/medium-risk patients and DG30 low-risk patients who were referred from primary care to secondary care based on a positive FIT result in primary care, alongside other patients referred for a variety of reasons. 'Enrichment' with patients who were referred on the basis of a positive FIT was thought by the EAG to be a source of heterogeneity between studies that may affect estimates of diagnostic test accuracy (see *Population types among included studies*).

Main analysis

As the NICE scope indicated that tests and test–analyser combinations should be considered separately, and as it could not be assumed that all tests were equivalent, the main analysis synthesised data on each test separately. There were only a small number of head-to-head comparative studies and so comparative test accuracy was not formally quantified. Data for dual FIT were considered separately from single FIT. For each test separately, the main analysis included all population types 1 to 4 together, assuming that the symptoms a patient presents with (i.e. population type) do not affect sensitivity and specificity. Sensitivity analyses were then conducted excluding type 4 studies, which may be enriched with patients who were referred on the basis of a positive FIT or which are otherwise unrepresentative or unclear. Additional sensitivity analyses were conducted for each of the population types 1, 2 and 3 separately to see if patient symptoms/population type affected test accuracy. The analyses were possible only for HM-JACKarc and OC-Sensor owing to the small number of studies for the other tests.

The meta-analysis included data at all reported thresholds and provided summary estimates at all possible thresholds. Considering a threshold of 10 µg/g, the results were as follows, for sensitivity and specificity, respectively: HM-JACKarc (n = 16 studies), 89.5% (95% 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%); FOB Gold (n = 3 studies), 87.0% (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 and IDK tests, as there was only one study for each. For these studies, the estimates of sensitivity and specificity at 10 µg/g respectively were QuikRead go, 92.90% (95% 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 by IDK to be 87% (95% CI 84.4% to 89.6%) and 88.1% (95% CI 85.6% to 90.6%); IDK Hb/Hp, 82.6% (95% CI 79.6% to 85.6%) and 80.8% (95% CI 77.7% to 83.9%). As is usual for diagnostic test accuracy, sensitivity was higher at lower thresholds, and specificity was higher at higher thresholds.

The sensitivity analyses showed that the exclusion of type 4 studies did not have significant impact on the pooled estimates, with differences in the point estimates not consistent across the tests, and small in magnitude compared with the uncertainty (as quantified by the CrIs and PrIs). From these analyses, the EAG concludes that it is not necessary to exclude population type 4 studies from the analyses. In the analyses by population types 1, 2 and 3 separately, for HM-JACKarc, the summary sensitivity and specificity for population 3 were higher than for the other population type subgroups; however, this analysis was based on only two studies that contributed data at two thresholds (2 and 10 µg/g) and was not statistically significantly different based on the overlap of the 95% CrIs across the subgroups. For the analyses of subgroups by population type for OC-Sensor, the summary estimates were similar and not statistically significant based on overlap of the 95% CrIs.

Additional analyses 1 and 2

The main analysis was supplemented with two additional analyses. In both additional analyses, all tests were synthesised together to allow the investigation of the impact of study population type and reference standards on a larger sample of studies and because these factors were thought unlikely to be affected by the test type. Additional analysis 1 conducted the same subgroup analyses by population type as described in the previous paragraph, while additional analysis 2 restricted to studies where > 90% of the patients received a colonoscopy or CTC as the reference standard, to investigate the effect of the reference standard on estimates of diagnostic test accuracy. In additional analysis 2, studies were also subgrouped by test.

In additional analysis 1 (the analysis of subgroups by population type), the summary estimates were similar to those of the main analysis and not statistically significantly different based on overlap of the 95% CrI.

In additional analysis 2, the summary estimates were similar, irrespective of the reference standard grouping (all studies vs at least 90% of the participants receiving colonoscopy) in all analyses.

Risk-of-bias assessment

There were risk-of-bias and/or applicability concerns with all the studies included in the review. Studies mostly fell into two types: (1) those that recruited patients referred to secondary care and who had a colonoscopy/CTC/other imaging as the reference standard, as this was part of their routine diagnostic workup (these are usually population type 2 or 4 studies); and (2) those that recruited patients in primary care and for whom the reference standard was either colonoscopy/CTC/other imaging where this was received as part of their diagnostic workup, or was records follow-up where a secondary care referral was not made (these are usually population type 1 or 3 studies). Studies of type (1) generally scored as being at high risk of bias for patient selection, as some primary care patients were not recruited (see *Rationale for the analysis plan* for a discussion of the different populations), while studies of type (2) generally scored as being at high risk of bias for the reference standard (see *Comparators* for discussion of the different reference standards), although there were occasional exceptions. Both of these factors have been investigated in the statistical synthesis as they could theoretically affect estimates of sensitivity and specificity. Various other sources of bias were noted; for example, the interval between reference standard and index test was poorly reported overall, but due to the 'real world' nature of many of the studies this is likely to have been within weeks or months rather than years of the index test and is not thought to be a concerning source of bias.

Dual faecal immunochemical test

Four studies reported data using a dual FIT strategy: two using HM-JACKarc, and one each using OC-Sensor and QuikRead go. In studies that reported estimates for both, sensitivity was higher and specificity was lower when using dual FIT (test positive if either FIT positive) than that achieved when using only the first FIT result to interpret the test.

Comparative diagnostic test accuracy studies

Three studies compared two or more tests with each other in the same sample of patients. All three concluded that there were some differences between tests, but none was able to conclude whether (and what) different FIT cut-off values would be required for each test. In accordance with this uncertainty about test performance characteristics, the EAG's base-case analysis uses data for each test separately.

Patient characteristics subgroup analyses

Eleven studies reported diagnostic test accuracy for anaemic patients, three reported data according to age group, three reported data according to sex, and three reported data for people taking medications that may affect FIT results. No studies were identified according to ethnicity or for people with blood disorders that may affect FIT results. Across these subgroup analyses, evidence was generally limited and sometimes inconsistent. It was not possible to conclude what or whether different FIT thresholds are required according to the patient characteristics specified in the NICE scope.

Advanced adenomas and inflammatory bowel disease

Eight studies reported data for the test accuracy of FITs for AA and IBD. Uncertainty was high in these analyses, with a large amount of heterogeneity between studies.

Test failures, uptake and repeat tests

Eleven studies reported test failure rates, and these were largely between 2% and 5%. Only two studies reported test uptake in primary care and only one reported this where return of FIT was part of the diagnostic pathway. In this instance, the non-return rate was 9.4%. For dual FIT, non-return rates appeared generally higher; all dual FIT studies were in secondary care.

'Time to' outcomes

Data on the time to different points within the diagnostic pathway for patients receiving single or dual FIT were reported in six diagnostic test accuracy studies, but these were largely non-comparative data and difficult to interpret. One further study also reported other outcomes relating to referral rates and emergency presentations and reported reductions in referral rates since the introduction of FIT.

Patient perspectives

Two studies reported patient perspectives. The authors' conclusions were that most patients found FIT acceptable, but strategies are needed to engage patients who have more negative views of FIT, and shared decision-making by patient and clinician should be considered for patients dissatisfied with relying on FIT results to decide whether or not further investigation is needed. Generalisability of these findings may have been affected by the fact that all patients included had been referred to secondary care.

Sociodemographic factors

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, older patients aged ≥ 65 years than for those aged < 65 years, White patients than for Asian, Black and mixed/other ethnicity groups, and the least socioeconomically deprived quintile than for all other quintiles. Suggested strategies for addressing demographic differences in FIT return rate, which may reflect strategies for engagement with services as a whole, included following up after FIT non-return, using multiple languages, having shared decision-making and providing patient counselling to address concerns.

Cost-effectiveness: principal findings

The EAG developed a de novo health economic model to assess the cost-effectiveness of FIT for people with suspected symptoms of CRC. The model compares three sets of interventions that include the use of quantitative FIT in a primary care setting, exploring a range of different thresholds to determine whether a person would be referred to the 2WW pathway or follow alternative further management pathways (a safety-netting pathway or an intermediate group pathway). These latter pathways could result in people being referred to the 2WW pathway due to ongoing clinical concerns, the 18WW pathway, or the watch and wait pathway, or being offered a repeat FIT. The health economic analysis was undertaken from the perspective of the NHS and PSS and was consistent with previous models retrieved by the EAG's review of economic studies, including the one developed to inform NICE DG30.[11] The EAG model adopts a hybrid decision tree–state transition structure. The model parameters were informed by a number of sources, including the EAG's clinical review and synthesis, NCRAS and ONS data sets, previous NICE TAs (TA856 and TA342), expert clinical opinion, the Whyte *et al.* model, and standard costing sources.

The EAG's base-case model suggests that for all FIT brands there are strategies that have a positive iNMB compared with current care regardless of the cost-effectiveness threshold used, or whether one or two thresholds were used. This was due to cost savings associated with fewer colonoscopies, although this was at the expense of a slight reduction in patient health caused by patients who previously would have had a colonoscopy receiving a false-negative FIT result. These conclusions produced by the EAG's base-case analysis were robust to the sensitivity analyses undertaken.

The exact brand 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.

Strengths and limitations of the assessment

Strengths and limitations of the clinical evidence base

Strengths and limitations of the evidence base

Although the evidence base was large, it was also complicated and incomplete with respect to the scope issued by NICE. Key evidence gaps were for the IDK tests. No diagnostic test accuracy data were found for IDK Hb + Hb/Hp, as the EAG was of the opinion that an assumption of independence between the two tests that make up this test could not be made. No diagnostic test accuracy data were identified for the IDK TurbiFIT tests, and the EAG was of the opinion that the analysis comparing IDK TurbiFIT with IDK Hb was not sufficient for the EAG to make a formal recommendation on equivalence. Additionally, data for IDK Hb and Hb/Hp complex came from the one small study with limited details about patient recruitment, and no data were provided to show that the test used in that study was equivalent to the current commercial test. Similarly, evidence was limited for NS-Prime (1 study, *n* recruited = 233, *n* of CRC events = 7) and QuikRead go (*n* recruited = 553, *n* of CRC events = 14). The NS-Prime study was conducted in a subgroup of patients from the NICE FIT study who returned all four tests, which may have introduced additional generalisability concerns if the non-return of FIT meant that the patient spectrum had been altered in such a way that may affect the estimates of sensitivity or specificity, for example excluding older patients. No diagnostic test accuracy data were found relating to OC-Sensor Ceres, and the correlation data provided by the company comparing OC-Sensor Ceres with OC-Sensor iO and OC-Sensor PLEDIA were not sufficient for the EAG to make a formal recommendation on equivalence. Reporting of the inclusion or exclusion of patients with 'bypass' symptoms (rectal or anal mass or anal ulceration) was often missing from the studies, and some studies excluded rectal bleeding, factors that may affect the patient spectrum in comparison with the scope. Data on test failures, test uptake and repeat tests were largely available only for HM-JACKarc and OC-Sensor. There were also no diagnostic test data according to ethnicity or for people with blood disorders that may affect FIT, and the available data on other patient characteristics were not conclusive. Data on patient outcomes such as HRQoL and anxiety were not available in the studies that reported diagnostic test accuracy, and it was beyond the scope of this assessment and time available to review these data in studies of other designs.

Strengths and limitations of the systematic review

The systematic review was conducted to high standards and used two reviewers to validate data extractions and risk-of-bias assessments. However, there were limitations due to the limited time available to complete the work combined with a large and complicated evidence base. Among the limitations was the use of one reviewer to conduct most of the study selection process, which may have resulted in studies being missed, although potential errors in misunderstanding the inclusion criteria were mitigated by concordance between two reviewers being established on the first 200 records. Clinical advisors and Specialist Committee Members were also consulted about potentially missed studies. Studies may also have been missed if these had been excluded from the original DG30 review, although this is thought unlikely as that review had wide inclusion criteria. An additional element of the search for this assessment was also added to mitigate missed studies by including search terms for each of the tests, without date limits. In addition, some quality assessment work could not be completed in the time available, although all studies that contributed to the three analyses this affected (comparative diagnostic test accuracy studies; dual FIT studies; AA and IBD studies) were assessed in the context of the other analyses to which they contributed, except for one.[80]

Because of the emphasis in this project to identify the optimal way of using FIT to reduce the number of people without significant bowel pathology being referred to the suspected CRC pathway, taking into consideration the threshold used to define a positive test, the synthesis used an advanced statistical model that accommodates estimates of sensitivity and specificity at multiple thresholds from each study. This has several advantages over the more commonly used approach of performing separate bivariate meta-analyses at selected thresholds, including making use of all available data, increasing precision, ensuring consistency of pooled results, and producing summary estimates at all thresholds of interest to be considered in the cost-effectiveness model. However, some of the analyses are subject to considerable uncertainty owing to the small number of studies and should be interpreted with caution. There were several potential sources of between-study heterogeneity. Although these were explored using subgroup analyses, it was not possible to make conclusive recommendations on any of these factors. Although it would be possible to extend the presented synthesis to include covariates that may explain the heterogeneity between studies (e.g. population type, reference standard, population characteristics), this was not conducted due to time constraints and the challenges presented by small numbers of studies in certain subgroups.

It was challenging to assign studies to population-type categories, and although authors were contacted to clarify inclusion criteria, this did not always resolve ambiguities. As a result, it is possible that some studies were wrongly categorised by the EAG reviewers. This may also have affected the sensitivity analyses carried out to test the effect of population type, as if studies were miscategorised, this may have altered the effect of removing them and obscured real differences.

It is thought likely by clinical advisors to the EAG that FITs will be used in a wider spectrum of patients (including those with less serious symptoms) in primary care than only those with NG12 high-/medium-risk or DG30 low-risk symptoms. It is unclear if test accuracy would be similar in a wider spectrum of patients with less serious symptoms. It was not the focus of this assessment to consider this issue, as the scope was limited to NG12 and DG30 patients. Three studies[27,42,63] were highlighted by clinical advisors to the EAG as having potentially recruited a wider population than just NG12 or DG30 patients. In our analysis, two of these studies[27,63] contributed to the type 4 subgroup, as they were not exclusively of DG30 low-risk patients, while the other[42] was categorised as a type 1 study. It is possible that some of the other studies, in particular those in Scotland, also recruited wider populations.

Owing to time constraints, it was not possible to consider the impact of distribution and sample return methods on return rates. Methods encountered in the literature included distribution by GPs in a face-to-face appointment, distribution by post and distribution in secondary care. Other methods may be used across the country. It was also not possible to consider the causes of test failures. Causes reported in the literature include labelling errors, incorrect containers, no date of collection or sample being too old, volume errors and laboratory accidents. These may be amenable to improvement through training of GPs (e.g. in how to describe the test to patients), patient information leaflets (e.g. to avoid overfilling and labelling/date errors) and laboratory personnel (e.g. in how to avoid accidents), or other interventions to avoid test failures.

Comparison with other analyses

This analysis has some differences in estimates from the BSG/ACPGBI review and the DG30 review. At a threshold of 10 µg/g, the BSG/ACPGBI review found a pooled sensitivity and specificity, respectively, for HM-JACKarc of 95.2% (95% CI 86.5% to 99.0%) and 78.2% (95% CI 69.2% to 85.2%) compared with 89.5% (95% CrI 84.6% to 93.4%) and 82.8% (95% CrI 75.2% to 89.6%) in the EAG's analysis, and 100% (95% CI 71.5% to 100%) and 76.6% (95% CI 72.6% to 80.3%) in the DG30 analysis. For OC-Sensor the ACPGBI/BSG analysis pooled estimates were 90.2% (95% CI 86.2% to 93.1%) and 74.5% (95% CI 68.1% to 79.9%), compared with 89.8% (95% CrI 85.9% to 93.3%) and 77.6% (95% CrI 64.3% to 88.6%) in the EAG's analysis, and 92.1% (95% CI 86.9% to 95.3%), and specificity was 85.8% (95% CI 78.3% to 91.0%) in the DG30 analysis. For FOB Gold the ACPGBI/BSG analysis pooled estimates were 95.2% (95% CI 86.5% to 99.0%) and 71.3% (95% CI 68.0% to 74.3%), compared with 87.0% (95% CrI 67.3% to 98.3%) and 88.4% (95% CrI 81.7% to 94.2%) in the EAG's analysis (there were no FOB Gold data in the DG30 review).

The differences in estimates for HM-JACKarc and OC-Sensor are generally small and may be due to the relatively large number of additional studies and patients included in the review for this assessment compared with the DG30 and ACPGBI/BSG review, even though less than 1 year had elapsed since the ACPGBI/BSG review searches were

conducted. A bivariate meta-analysis including 14 HM-JACKarc studies that report diagnostic test accuracy at a threshold of 10 was conducted by the EAG and found a pooled sensitivity and specificity of 89.2 (95% CrI 85.7 to 92.0) and 79.4 (95% CrI 75.0 to 83.3), respectively. This suggests that the higher sensitivity reported in previous reviews may be largely explained by the difference in studies contributing to the analysis, rather than the different statistical methods used (stratified bivariate model vs multiple thresholds model). The difference in the specificity is within the CIs reported across the analyses and may be due to the different studies that have entered the analysis and/or the methods of the multiple threshold model.

Strengths and limitations relating to the health economic analysis

The EAG model has a number of strengths; in particular, (1) the model structure builds on other published models that evaluate FIT in people with symptoms of CRC; (2) the model includes colonoscopy capacity to impact on the waiting times for patients following the 2WW and 18WW pathways; (3) AAs and IBD are included in the mathematical model; and (4) the uncertainty in the model inputs and assumptions has been explored in sensitivity analyses.

However, the model is also subject to several limitations and uncertainties relating to the cost-effectiveness analysis, which include (1) the uncertainty in data inputs, particularly diagnostic accuracy data and those reliant on expert opinion; (2) the structure of the model, which may have omitted aspects of the complex real-world problem; (3) the relative similarity in iNMB values for FIT strategies, which meant that no robust estimate of the FIT brand or the threshold(s) that generated the greatest iNMB could be made and (4) the potential uncertainty associated with the modelled gains in health outcomes for those patients with quicker time to diagnosis. The limitations of the cost-effectiveness analysis are outlined in *Key Evidence Assessment Group model assumptions model assumptions*, *The incremental net monetary benefits of the seven tests using one threshold assuming a threshold of £20,000 per quality-adjusted life-year gained and low-intensity safety-netting threshold*, Chapter 5 *Conclusions from the cost-effectiveness analyses undertaken in the Evidence Assessment Group's base case*, *Cost-effectiveness – principal findings*, *Generalisability* and *Appendix 8*.

Uncertainties

It was beyond the scope of this assessment to conduct cost-effectiveness analyses for the patient characteristics subgroups defined in the scope, and clinical data limitations would have prevented such analyses had they been planned. The project deadlines and a lack of evidence also prevented more in-depth analyses of the inputs informed by clinical opinion.

Evidence for the accuracy of IDK TurbiFit was lacking; therefore, this test could not be analysed. Diagnostic test accuracy data for four other tests (QuikRead go, NS-Prime, IDK Hb and IDK Hb/Hp) relied on a single, fairly small study for each.

The EAG also notes that the standard care of the model may not reflect current use of the test in some locations in England, as it is known that there is some heterogeneity between diagnostic and clinical management of patients with suspected CRC. However, current care was intended to reflect NICE current recommendations as defined in DG30 and NG12, in accordance with the NICE scope.

Generalisability

The assessment included only studies conducted in patients who presented to primary care with symptoms of CRC, except where insufficient evidence necessitated the use of studies that also recruited from secondary care. This affected the dual FIT QuikRead go analysis and the analysis of medications that might affect FIT results. As noted, the assessment of the evidence base has considered potential sources of heterogeneity among the populations recruited to the included studies, and although analyses did not indicate that population type affected estimates, the limitations of the analyses (e.g. difficulties categorising studies) mean that these were not conclusive.

Data on test failures, test uptake and repeat tests were largely only available for HM-JACKarc and OC-Sensor; the generalisability of these data to other tests has been assumed in the model. For dual FIT for these outcomes, data were only available from studies conducted in secondary care, which may affect generalisability.

As noted, heterogeneity in clinical practice across England may affect the generalisability of some modelling assumptions, for example safety netting. The model was robust to all scenario analyses.

Implications for service provision

The model makes assumptions about the effects of safety netting, which may not be consistent with the safety netting offered across the country at present. Standardisation of and improvements in safety-netting practice may be required. Interventions may be required to increase FIT return rates, especially in some socioeconomic groups, and to improve the experience of a minority of patients who have negative views about FIT, and dissatisfaction with reliance on FIT for diagnostic purposes. The optimal way to distribute FIT should be considered, for example via post or from the GP. Ways to avoid test failures should also be considered, for example GP training, patient information leaflets and laboratory staff training. The implementation of FIT among patients with symptoms defined in NG12 and DG30 may lead to the use of FIT in a wider group, and this possibility may need to be monitored and/or mitigated.

Suggested research priorities

In order of priority, our research recommendations are as follows.

It is unclear if test accuracy would be similar in a wider spectrum of patients with less serious symptoms, and future primary studies or a careful consideration of the existing evidence base, as noted above in *Strengths and limitations of the systematic review*, could address this issue.

The comparative diagnostic test accuracy between tests remains uncertain, largely due to the limited evidence base comparing tests with one another; new primary research studies may be required.

It remained unclear whether and what different thresholds are required for patients with characteristics that may affect FIT accuracy; new primary research studies may be required, although the publication of the COLOFIT study[124] may address some of these issues.

While the analysis was not able to detect an effect of population type, enrichment with FIT positives or the reference standard used, these are all issues that should be considered in future primary studies and evidence syntheses, as the analyses conducted here were not conclusive. Efforts could be made to include the relevant patient spectrum with the best possible reference standard. Many studies used a differential reference standard, whereby some patients received colonoscopy/imaging and some received only long-term follow-up, based on the results of their index test. This type of reference standard brings with it a high risk of bias as it allows false negatives to be missed by relying on following up patient records, or to be generated by the emergence of interval cancers. However, the low prevalence of disease and cost and ethical concerns make it difficult and impractical to give all patients colonoscopies. Alternative designs that use sampling methods[125] could be used to reduce the number of patients who need to receive a colonoscopy, and potentially avoid the differential reference standard design.

Although some studies reviewed the evidence demonstrating the potential impact of time to diagnosis or time from symptom development to diagnosis in patients with CRC,[92,126] there is still uncertainty around the effect of time to and delays in receiving a diagnosis on outcomes such as survival or cancer stage, with most studies included in these reviews suggesting no association. However, these reviews have not performed quantitative syntheses of these time intervals. Future studies could be conducted that examine the effects of time to diagnosis on health outcomes in this population, using more standardised time intervals and examining specifically delays in diagnosis on outcomes such as survival, cancer stage and HRQoL.

Use of patient and public involvement

There was no patient and public involvement in producing this report; this was not considered possible within the timescales of the project. However, the EAG is aware that there will be patient and public involvement and representation at the NICE Technology Appraisal Committee for discussing this topic, and this may result in the EAG changing model parameters and generating revised results.

Equality, diversity and inclusion

As this report is secondary research, no patient participation was involved and the EAG did not need to consider the equality, diversity and inclusion of participants. The primary research team was part of the SCHARR Technology Assessment Group contracted by the Department of Health and Social Care, and this team is a group representing a range of protected characteristics in terms of seniority, ages, ethnicity and religious beliefs, and including both male and female researchers. The lead author is not the most senior member of the team.

Additional information

CRediT contribution statement

Sue Harnan (https://orcid.org/0000-0002-9318-9206**):** Conceptualisation, Data curation, Methodology, Project administration, Investigation (systematic review), Writing – original draft, Writing – review and editing, Supervision.

Aline Navega Biz (https://orcid.org/0000-0002-6923-4556**):** Conceptualisation, Methodology, Data curation, Formal analysis (health economic analysis), Investigation (modelling), Writing – original draft, Writing – review and editing.

Jean Hamilton (https://orcid.org/0000-0003-3326-9842**):** Conceptualisation, Methodology (statistical aspects), Data curation, Formal analysis (synthesis), Writing – original draft, Writing – review and editing, Supervision, Visualisation.

Sophie Whyte (https://orcid.org/0000-0002-7963-2523**):** Conceptualisation, Methodology, Data curation, Formal analysis (health economic analysis), Investigation (modelling), Writing – original draft (see *Appendix 8*), Writing – review and editing, Visualisation, Supervision.

Emma Simpson (https://orcid.org/0000-0001-7353-5979**):** Methodology, Data curation, Investigation (systematic review), Writing – original draft, Writing – review and editing.

Shijie Ren (https://orcid.org/0000-0003-3568-7124**):** Methodology, Data curation, Formal analysis (synthesis), Writing – original draft, Writing – review and editing.

Katy Cooper (https://orcid.org/0000-0002-7702-8103**):** Conceptualisation, Writing – review and editing.

Mark Clowes (https://orcid.org/0000-0002-5582-9946**):** Methodology (search strategies), Data curation, Writing – original draft, Writing – review and editing.

Muti Abulafi (https://orcid.org/0000-0001-7088-4292**):** Investigation (systematic review), Resources (expert clinical opinion), Writing – review and editing.

Alex Ball (https://orcid.org/0000-0003-2631-9629**):** Resources (expert clinical opinion), Writing – review and editing.

Sally C Benton (https://orcid.org/0000-0001-9230-9088**):** Resources (expert clinical opinion), Writing – review and editing.

Richard Booth (https://orcid.org/0000-0001-9058-385X**):** Investigation (systematic review), Resources (expert clinical opinion), Writing – review and editing.

Rachel Carten (https://orcid.org/0000-0002-3362-8721**):** Investigation (systematic review), Resources (expert clinical opinion), Writing – review and editing.

Stephanie Edgar: Resources (expert clinical opinion), Writing – review and editing.

Willie Hamilton (https://orcid.org/0000-0003-1611-1373**):** Resources (expert clinical opinion), Writing – review and editing.

Matt Kurien (https://orcid.org/0000-0002-4227-9500**):** Resources (expert clinical opinion), Writing – review and editing.

Louise Merriman (https://orcid.org/0009-0000-7465-3555**):** Resources (expert clinical opinion), Writing – review and editing.

Kevin Monahan (https://orcid.org/0000-0002-7918-4003**):** Investigation (systematic review), Resources (expert clinical opinion), Writing – review and editing.

Laura Heathcote (https://orcid.org/0000-0001-8063-7447**):** Investigation (modelling), Writing – original draft (see *Appendix 8*), Writing – review and editing, Visualisation.

Matt Stevenson (https://orcid.org/0000-0002-3099-9877**):** Conceptualisation, Methodology, Data curation, Formal analysis (health economic analysis), Investigation (modelling), Writing – original draft, Writing – review and editing, Visualisation, Supervision.

Acknowledgements

We would like to thank Chloe Thomas, Research Fellow, SCHARR, and Olena Mandrik, Research Fellow, SCHARR, for sharing their work (see *Appendix 8*) and advising on the modelling. We would like to thank Hayley Jones (Population Health Sciences, University of Bristol) for advice on the statistical synthesis. We would like to thank Paul Tappenden, Professor of Health Technology Assessment, SCHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, SCHARR, for providing administrative support and in preparing and formatting the report.

About Sheffield Centre for Health and Related Research

The Sheffield Centre for Health and Related Research (SCHARR) (previously the School for Health and Related Research, SCHARR) is part of the Division of Populational Health within the School of Medicine and populational Health that comprise the Faculty of Health at the University of Sheffield. SCHARR brings together a wide range of medical- and health-related disciplines including public health, general practice, mental health, epidemiology, health economics, management sciences, medical statistics, operational research and information science. It includes the Sheffield unit of the Trent Institute for Health Services Research, which is funded by NHS R&D to facilitate high-quality health services research and capacity development.

The SCHARR Technology Assessment Group (SCHARR-TAG) synthesises research on the clinical effectiveness and cost-effectiveness of healthcare interventions for the NIHR Health Technology Assessment Programme on behalf of a range of policy makers, including the National Institute for Health and Care Excellence (NICE). SCHARR-TAG is part of a wider collaboration of a number of units from other regions including Southampton Health Technology Assessment Centre (SHTAC), University of Southampton; Aberdeen Health Technology Assessment Group (Aberdeen HTA Group), University of Aberdeen; Liverpool Reviews & Implementation Group (LRiG), University of Liverpool; Peninsular Technology Assessment Group (PenTAG), University of Exeter; the NHS Centre for Reviews and Dissemination, University of York; Warwick Evidence, The University of Warwick; the BMJ Group and Kleijnen Systematic Reviews.

Data-sharing statement

All data from the systematic review synthesis are available in an online supplement to this report. All other data requests should be submitted to the corresponding author for consideration.

Ethics statement

This review did not involve the collection or analysis of any data that was not included in previously published research in the public domain. Therefore, it was exempt from formal ethical review.

Information governance statement

No personal information was collected or stored during this project.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/AHPE4211.

Primary conflicts of interest: Stephanie Edgar is the Deputy Primary Care Lead for South Yorkshire and Bassetlaw (SYB) Cancer Alliance and is the Cancer Research UK GP for SYB. Willie Hamilton has received funding from Bowel Cancer UK for research into faecal immunochemical tests, is a co-applicant on an NIHR RfPB grant on early diagnosis of colorectal cancer (award ID NIHR203526) and co-PI on an NIHR HTA on use of FIT in primary care (award ID NIHR133852). Matt Kurien is in receipt of two research awards from SYB Cancer Alliance. Laura Heathcote worked on a project awarded to SCHARR on colorectal cancer diagnosis, funded by Origin Sciences Limited.

Publication

Harnan S, Hamilton J, Simpson E, Clowes M, Navega Biz A, Whyte S, *et al*. *PP09. Fecal Immunochemical Tests for Patients with Symptoms Suggestive of Colorectal Cancer: A Systematic Review and Multiple-Threshold Meta-Analysis.* Poster presented at Health Technology Assessment International 2024 European meeting, 17 June 2024.

References

1. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010;**138**:2044–58.
2. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* 2019;**16**:713–32.
3. Bretthauer M, Kalager M, Adami HO. Do's and don'ts in evaluation of endoscopic screening for gastrointestinal cancers. *Endoscopy* 2015;**48**:75–80.
4. Cancer Research UK. *Bowel Cancer Statistics*. 2022. URL: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer (accessed 22 July 2024).
5. IHME GHDx. *GBD Results*. 2023. URL: https://vizhub.healthdata.org/gbd-results/ (accessed 1 June 2023).
6. Vuik FE, Nieuwenburg SA, Bardou M, Lansdorp-Vogelaar I, Dinis-Ribeiro M, Bento MJ, *et al*. Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut* 2019;**68**:1820–6.
7. Oliphant R, Brewster D, Morrison D. The changing association between socioeconomic circumstances and the incidence of colorectal cancer: a population-based study. *Br J Cancer* 2011;**104**:1791–6.
8. National Cancer Intelligence Network. *Cancer by Deprivation in England: Incidence, 1996–2010 & Mortality, 1997–2011*. Oxford: National Cancer Intelligence Network; 2014.
9. Tweed E, Allardice G, McLoone P, Morrison D. Socio-economic inequalities in the incidence of four common cancers: a population-based registry study. *Public Health* 2018;**154**:1–10.
10. National Institute for Health and Care Excellence. *Suspected Cancer: Recognition and Referral*. London: NICE; 2015.
11. National Institute for Health and Care Excellence. *Quantitative Faecal Immunochemical Tests to Guide Referral for Colorectal Cancer in Primary Care*. Diagnostics guidance [DG30]. London: NICE; 2017.
12. Monahan KJ, Davies MM, Abulafi M, Banerjea A, Nicholson BD, Arasaradnam R, *et al*. Faecal immunochemical testing (FIT) in patients with signs or symptoms of suspected colorectal cancer (CRC): a joint guideline from the Association of Coloproctology of Great Britain and Ireland (ACPGBI) and the British Society of Gastroenterology (BSG). *Gut* 2022;**71**:1939–62.
13. NHS England. *Using Faecal Immunochemical Testing (FIT) in the Lower Gastrointestinal (GI) Pathway (Primary Care)*. 2022. URL: www.england.nhs.uk/publication/supporting-the-use-of-faecal-immunochemical-testing/ (accessed 22 July 2024).
14. NHS England. *Using Faecal Immunochemical Testing (FIT) in the Lower Gastrointestinal (GI) Pathway (Secondary Care)*. 2022. URL: www.england.nhs.uk/publication/supporting-the-use-of-faecal-immunochemical-testing/ (accessed 22 July 2024).
15. Mozdiak E, Weldeselassie Y, McFarlane M, Tabuso M, Widlak MM, Dunlop A, *et al*. Systematic review with meta-analysis of over 90 000 patients. Does fast-track review diagnose colorectal cancer earlier? *Aliment Pharmacol Ther* 2019;**50**:348–72.
16. Booth R, Carten R, D'Souza N, Westwood M, Kleijnen J, Abulafi M. Role of the faecal immunochemical test in patients with risk-stratified suspected colorectal cancer symptoms: a systematic review and meta-analysis to inform the ACPGBI/BSG guidelines. *Lancet Reg Health Eur* 2022;**23**:100518.
17. D'Souza N, Delisle TG, Chen M, Benton S, Abulafi M. Faecal immunochemical test is superior to symptoms in predicting pathology in patients with suspected colorectal cancer symptoms referred on a 2WW pathway: a diagnostic accuracy study. *Gut* 2021;**70**:1130–8.
18. D'Souza N, Hicks G, Benton S, Abulafi M. The diagnostic accuracy of the faecal immunochemical test for colorectal cancer in risk-stratified symptomatic patients. *Ann R Coll Surg Eng* 2020;**102**:174–9.

19. Fraser CG, Allison JE, Young GP, Halloran SP, Seaman HE. Improving the reporting of evaluations of faecal immunochemical tests for haemoglobin. *Eur J Cancer Prev* 2015;**24**:24–6.
20. Carroll M, Piggott C, Pearson S, Seaman H, Bruce H, Halloran S. An evaluation of quantitative faecal immunochemical tests for haemoglobin. *Gut* 2014;**63**:A129–A30.
21. National Institute for Health and Care Excellence. *Quantitative Faecal Immunochemical Tests to Guide Referral in Primary Care for People with High Risk Symptoms. Final scope*. London: NICE; 2020.
22. National Institute for Health and Care Excellence. *Quantitative Faecal Immunochemical Tests to Guide Colorectal Cancer Pathway Referral in Primary Care*. Final scope. London: NICE; 2022.
23. Farkas NG, Fraser CG, Maclean W, Jourdan I, Rockall T, Benton SC. Replicate and repeat faecal immunochemical tests in symptomatic patients: a systematic review. *Ann Clin Biochem* 2023;**60**:27–36.
24. Westwood M, Corro Ramos I, Lang S, Luyendijk M, Zaim R, Stirk L, *et al*. Faecal immunochemical tests to triage patients with lower abdominal symptoms for suspected colorectal cancer referrals in primary care: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2016;**21**.
25. Burr NE, Derbyshire E, Taylor J, Whalley S, Subramanian V, Finan PJ, *et al*. Variation in post-colonoscopy colorectal cancer across colonoscopy providers in English National Health Service: population based cohort study. *BMJ* 2019;**367**:l6090.
26. Obaro AE, Plumb AA, Fanshawe TR, Torres US, Baldwin-Cleland R, Taylor SA, *et al*. Post-imaging colorectal cancer or interval cancer rates after CT colonography: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2018;**3**:326–36.
27. Nicholson BD, James T, Paddon M, Justice S, Oke JL, East JE, Shine B. Faecal immunochemical testing for adults with symptoms of colorectal cancer attending English primary care: a retrospective cohort study of 14 487 consecutive test requests. *Aliment Pharmacol Ther* 2020;**52**:1031–41.
28. Centre for Reviews and Dissemination. *CRD's Guidance for Undertaking Reviews in Health Care*. York: Centre for Reviews and Dissemination; 2008.
29. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2010. URL: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://methods.cochrane.org/sdt/sites/methods.cochrane.org.sdt/files/uploads/Chapter%2010%20-%20Version%201.0.pdf (accessed 1 June 2023).
30. Crooks CJ, Banerjea A, Jones J, Chapman C, Oliver S, West J, *et al*. Assessing empirical thresholds for investigation in people referred on a symptomatic colorectal cancer pathway: a cohort study utilising faecal immunochemical and blood tests in England. *medRxiv* 2023. https://doi.org/10.1101/2023.03.29.23287919:2023.03.29.23287919
31. Bailey J, Morton A, Jones J, Chapman C, Oliver S, Morling J, *et al*. Sociodemographic variations in the uptake of faecal immunochemical tests in primary care: a retrospective study. *Br J Gen Pract* 2023;**73**:e843–9.
32. Georgiou Delisle T, D'Souza N, Davies B, Benton S, Chen M, Ward H, Abulafi M; NICE FIT Steering Committee. Faecal immunochemical test for suspected colorectal cancer symptoms: patient survey of usability and acceptability. *BJGP Open* 2022;**6**:BJGPO.2021.0102.
33. Maclean W, Whyte MB, Farkas N, Benton SC, Rockall T, Jourdan I. Patient-reported outcome measures show FIT as an acceptable investigation to rule out colorectal cancer in the two-week wait cohort. *Ann R Coll Surg Engl* 2022;**105**:336–41.
34. Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36.
35. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, *et al*.; QUADAS-C Group†. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med* 2021;**174**:1592–9.
36. Jones HE, Gatsonsis CA, Trikalinos TA, Welton NJ, Ades AE. Quantifying how diagnostic test accuracy depends on threshold in a meta-analysis. *Stat Med* 2019;**38**:4789–803.

37. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2022.

38. Plummer M. *RJAGS: Bayesian Graphical Models Using MCMC. R Package Version 4–10*. 2019. URL: https://CRAN.R-project.org/package=rjags (accessed 22 July 2024).

39. Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998;**7**:434–55.

40. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol* 2002;**64**:583–639.

41. Archer T, Aziz I, Kurien M, Knott V, Ball A. Prioritisation of lower gastrointestinal endoscopy during the COVID-19 pandemic: outcomes of a novel triage pathway. *Front Gastroenterol* 2022;**13**:225–30.

42. Bailey JA, Weller J, Chapman CJ, Ford A, Hardy K, Oliver S, *et al*. Faecal immunochemical testing and blood tests for prioritization of urgent colorectal cancer referrals in symptomatic patients: a 2-year evaluation. *BJS Open* 2021;**5**:zraa056.

43. Ball AJ, Aziz I, Parker S, Sargur RB, Aldis J, Kurien M. Fecal immunochemical testing in patients with low-risk symptoms of colorectal cancer: a diagnostic accuracy study. *J Natl Compr Canc Netw* 2022;**20**:989–96.e1.

44. Benton SC, Piggott C, Zahoor Z, O'Driscoll S, Fraser CG, D'Souza N, *et al*. A comparison of the faecal haemoglobin concentrations and diagnostic accuracy in patients suspected with colorectal cancer and serious bowel disease as reported on four different faecal immunochemical test systems. *Clin Chem Lab Med* 2022;**60**:1278–86.

45. Cama R, Kapoor N, Sawyer P, Patel B, Landy J. Evaluation of 13,466 fecal immunochemical tests in patients attending primary care for high- and low-risk gastrointestinal symptoms of colorectal cancer. *Dig Dis Sci* 2022;**68**:2023–9.

46. Chapman CJ, Banerjea A, Humes DJ, Allen J, Oliver S, Ford A, *et al*. Choice of faecal immunochemical test matters: comparison of OC-Sensor and HM-JACKarc, in the assessment of patients at high risk of colorectal cancer. *Clin Chem Lab Med* 2021;**59**:721–8.

47. Elbeltagi A, Salama M, Boxall P, Roos J, Lim M. The yield of faecal immunochemical test in the detection of colorectal cancer within a fast-track pathway at York, United Kingdom. *Turkish J Colorectal Dis* 2022;**32**:178–85.

48. Faux JW, Cock K, Bromley R, Feldman M. Colorectal two-week wait service and quantitative FIT: it's not just about colon cancer. *Ann R Coll Surg Engl* 2022;**104**:257–60.

49. Georgiou Delisle T, D'Souza N, Tan J, Najdawi A, Chen M, Ward H, Abulafi M. Introduction of an integrated primary care faecal immunochemical test referral pathway for patients with suspected colorectal cancer symptoms. *Colorectal Dis* 2022;**24**:1526–34.

50. Godber IM, Todd LM, Fraser CG, MacDonald LR, Younes HB. Use of a faecal immunochemical test for haemoglobin can aid in the investigation of patients with lower abdominal symptoms. *Clin Chem Lab Med* 2016;**54**:595–602.

51. Hunt N, Rao C, Logan R, Chandrabalan V, Oakey J, Ainsworth C, *et al*. A cohort study of duplicate faecal immunochemical testing in patients at risk of colorectal cancer from North-West England. *BMJ Open* 2022;**12**:e059940.

52. Johnstone MS, Burton P, Kourounis G, Winter J, Crighton E, Mansouri D, *et al*. Combining the quantitative faecal immunochemical test and full blood count reliably rules out colorectal cancer in a symptomatic patient referral pathway. *Int J Colorectal Dis* 2022;**37**:457–66.

53. Johnstone MS, MacLeod C, Digby J, Al-Azzawi Y, Pang G, Watson AJM, *et al*. Prevalence of repeat faecal immunochemical testing in symptomatic patients attending primary care. *Colorectal Dis* 2022;**01**:01.

54. Juul JS, Hornung N, Andersen B, Laurberg S, Olesen F, Vedsted P. The value of using the faecal immunochemical test in general practice on patients presenting with non-alarm symptoms of colorectal cancer. *Br J Cancer* 2018;**119**:471–9.

55. Laszlo HE, Seward E, Ayling RM, Lake J, Malhi A, Stephens C, *et al*. Faecal immunochemical test for patients with 'high-risk' bowel symptoms: a large prospective cohort study and updated literature review. *Br J Cancer* 2022;**126**:736–43.

56. MacDonald S, MacDonald L, Godwin J, Macdonald A, Thornton M. The diagnostic accuracy of the faecal immunohistochemical test in identifying significant bowel disease in a symptomatic population. *Colorectal Dis* 2022;**24**:257–63.

57. Maclean W, Limb C, Mackenzie P, Whyte MB, Benton SC, Rockall T, Jourdan I. Adoption of faecal immunochemical testing for 2-week-wait colorectal patients during the COVID-19 pandemic: an observational cohort study reporting a new service at a regional centre. *Colorectal Dis* 2021;**23**:1622–9.

58. Maclean W, Mackenzie P, Limb C, Zahoor Z, Whyte MB, Rockall T, *et al*. Diagnostic accuracy of point of care faecal immunochemical testing using a portable high-speed quantitative analyser for diagnosis in 2-week wait patients. *Colorectal Dis* 2021;**23**:2376–86.

59. MacLean W, Zahoor Z, O'Driscoll S, Piggott C, Whyte MB, Rockall T, *et al*. Comparison of the Quikread go point-of-care faecal immunochemical test for haemoglobin with the FOB Gold wide laboratory analyser to diagnose colorectal cancer in symptomatic patients. *Clin Chem Lab Med* 2022;**60**:101–8.

60. Mowat C, Digby J, Strachan JA, McCann R, Hall C, Heather D, *et al*. Impact of introducing a faecal immunochemical test (FIT) for haemoglobin into primary care on the outcome of patients with new bowel symptoms: a prospective cohort study. *BMJ Open Gastroenterol* 2019;**6**:e000293.

61. Mowat C, Digby J, Strachan JA, McCann RK, Carey FA, Fraser CG, Steele RJ. Faecal haemoglobin concentration thresholds for reassurance and urgent investigation for colorectal cancer based on a faecal immunochemical test in symptomatic patients in primary care. *Ann Clin Biochem* 2021;**58**:211–19.

62. Mowat C, Digby J, Strachan JA, Wilson R, Carey FA, Fraser CG, Steele RJC. Faecal haemoglobin and faecal calprotectin as indicators of bowel disease in patients presenting to primary care with bowel symptoms. *Gut* 2016;**65**:1463–9.

63. Nicholson BD, James T, East JE, Grimshaw D, Paddon M, Justice S, *et al*. Experience of adopting faecal immunochemical testing to meet the NICE colorectal cancer referral criteria for low-risk symptomatic primary care patients in Oxfordshire, UK. *Front Gastroenterol* 2019;**10**:347–55.

64. Pin-Vieito N, Garcia Nimo L, Bujanda L, Roman Alonso B, Gutierrez-Stampa MA, Aguilar-Gama V, *et al*. Optimal diagnostic accuracy of quantitative faecal immunochemical test positivity thresholds for colorectal cancer detection in primary health care: a community-based cohort study. *United European Gastroenterol J* 2021;**9**:256–67.

65. Tang A, Chandler S, Torkington J, Harris DA, Dhruva Rao PK. Adapting the investigation of patients on urgent suspected cancer pathway with lower gastrointestinal symptoms across Wales during COVID-19. *Ann R Coll Surg Engl* 2022;**105**:S35–41.

66. Withrow DR, Shine B, Oke J, Tamm A, James T, Morris E, *et al*. Combining faecal immunochemical testing with blood test results for colorectal cancer risk stratification: a consecutive cohort of 16,604 patients presenting to primary care. *BMC Med* 2022;**20**:116.

67. National Institute for Health and Care Excellence. *Quantitative Faecal Immunochemical Testing to Guide Colorectal Cancer Pathway Referral in Primary Care. Diagnostics Assessment Programme Draft Guidance*. London: NICE; 2023.

68. Ayling RM, Lewis SJ, Cotter F. Potential roles of artificial intelligence learning and faecal immunochemical testing for prioritisation of colonoscopy in anaemia. *Br J Haematol* 2019;**185**:311–16.

69. Bailey J, Morton A, Jones J, Chapman C, Oliver S, Morling J, *et al*. 'Low FIT' Colorectal cancer: a four-year comparison of the Nottingham '4F' protocol with FIT10 in symptomatic patients. *Colorectal Dis* 2024;**26**:309–16.

70. Bujanda L, Sarasqueta C, Vega P, Salve M, Quintero E, Alvarez-Sánchez V, *et al*. Effect of aspirin on the diagnostic accuracy of the faecal immunochemical test for colorectal advanced neoplasia. *United European Gastroenterol J* 2018;**6**:123–30.

71. Cunin L, Khan AA, Ibrahim M, Lango A, Klimovskij M, Harshen R. FIT negative cancers: a right-sided problem? Implications for screening and whether iron deficiency anaemia has a role to play. *Surgeon* 2021;**19**:27–32.

72. D'Souza N, Delisle TG, Chen M, Benton SC, Abulafi M; the NICE FIT Steering Committee. Faecal immunochemical testing in symptomatic patients to prioritize investigation: diagnostic accuracy from NICE FIT Study. *Br J Surg* 2021;**108**:804–10.

73. D'Souza N, Monahan K, Benton SC, Wilde L, Abulafi M, Group NFS, *et al*. Finding the needle in the haystack: the diagnostic accuracy of the faecal immunochemical test for colorectal cancer in younger symptomatic patients. *Colorectal Dis* 2021;**23**:2539–49.

74. Farrugia A, Widlak M, Evans C, Smith SC, Arasaradnam R. Faecal immunochemical testing (FIT) in symptomatic patients: what are we missing? *Front Gastroenterol* 2020;**11**:28–33.

75. Gerrard AD, Maeda Y, Miller J, Gunn F, Theodoratou E, Noble C, *et al*.; Edinburgh Colorectal Group. Double faecal immunochemical testing in patients with symptoms suspicious of colorectal cancer. *Br J Surg* 2023;**110**:471–80.

76. Morales Arraez D, Carrillo G, Adrian M, Gimeno Z, Quintero A. Role of faecal immunochemical testing in the diagnostic workup of patients with iron deficiency anaemia. *United European Gastroenterol J* 2018;**6**:A403–4.

77. Rodriguez-Alonso L, Rodriguez-Moranta F, Arajol C, Gilabert P, Serra K, Martin A, *et al*. Proton pump inhibitors reduce the accuracy of faecal immunochemical test for detecting advanced colorectal neoplasia in symptomatic patients. *PLOS ONE* 2018;**13**:e0203359.

78. Rodriguez-Alonso L, Rodriguez-Moranta F, Ruiz-Cerulla A, Arajol C, Serra K, Gilabert P, *et al*. The use of faecal immunochemical testing in the decision-making process for the endoscopic investigation of iron deficiency anaemia. *Clin Chem Lab Med* 2020;**58**:232–9.

79. Sieg A, Thoms C, Lüthgens K, John MR, Schmidt-Gayk H. Detection of colorectal neoplasms by the highly sensitive hemoglobin-haptoglobin complex in feces. *Int J Colorectal Dis* 1999;**14**:267–71.

80. Tsapournas G, Hellström PM, Cao Y, Olsson LI. Diagnostic accuracy of a quantitative faecal immunochemical test vs. symptoms suspected for colorectal cancer in patients referred for colonoscopy. *Scand J Gastroenterol* 2020;**55**:184–92.

81. Turvill J, Mellen S, Jeffery L, Bevan S, Keding A, Turnock D. Diagnostic accuracy of one or two faecal haemoglobin and calprotectin measurements in patients with suspected colorectal cancer. *Scand J Gastroenterol* 2018;**53**:1526–34.

82. Turvill J, Turnock D, Cottingham D. The Fast Track FIT study: diagnostic accuracy of faecal immunochemical test for haemoglobin in patients with suspected colorectal cancer. *Br J Gen Pract* 2021;**71**:E643–51.

83. Jordaan M, Welbourn H, Tyldsley K, Mustafa A, Mevada C, Harrison M, *et al*. *Development of a Primary Care Pathway for Using a Faecal Immunochemical Test (FIT) to Triage Patients Presenting with Bowel Symptoms.* Medicine PL; 2022.

84. Cubiella J, Salve M, Díaz-Ondina M, Vega P, Alves MT, Iglesias F, *et al*. Diagnostic accuracy of the faecal immunochemical test for colorectal cancer in symptomatic patients: comparison with NICE and SIGN referral criteria. *Colorectal Dis* 2014;**16**:O273–82.

85. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 2021;**88**:105906.

86. Eiken Chemical Co. Ltd. *Analytical Performance Report*. *OC-Sensor Ceres*. Tokyo: Eiken Chemical Co. Ltd; 2022.

87. Immunodiagnostik. *Comparison IDK® TurbiFIT® and IDK® Hemoglobin ELISA*. 2022.

88. Baeyens JP, Serrien B, Goossens M, Clijsen R. Questioning the 'SPIN and SNOUT' rule in clinical testing. *Arch Physiother* 2019;**9**:4.

89. National Institute for Health and Care Excellence. *NICE Health Technology Evaluations: The Manual. Process and methods [PMG36]*. London: NICE; 2022.

90. Westwood M, Corro Ramos I, Lang S, Luyendijk M, Zaim R, Stirk L, *et al*. Faecal immunochemical tests to triage patients with lower abdominal symptoms for suspected colorectal cancer referrals in primary care: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2017;**21**.

91. Medina-Lara A, Grigore B, Lewis R, Peters J, Price S, Landa P, *et al*. Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis. *Health Technol Assess* 2020;**24**:1–332.

92. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, *et al*. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 2015;**112**:S92–107.

93. Thomas C, Mandrik O, Whyte S. *Development of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies*. Sheffield: ScHARR; 2020.

94. Bressler B, Paszat LF, Chen Z, Rothwell DM, Vinden C, Rabeneck L. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. *Gastroenterology* 2007;**132**:96–102.

95. Atkin W, Dadswell E, Wooldrage K, Kralj-Hans I, von Wagner C, Edwards R, *et al*.; SIGGAR investigators. Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. *Lancet* 2013;**381**:1194–202.

96. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, *et al*. Screening for colorectal cancer: updated evidence report and systematic review for the US preventive services task force. *JAMA* 2016;**315**:2576–94.

97. Martín-López JE, Beltrán-Calvo C, Rodríguez-López R, Molina-López T. Comparison of the accuracy of CT colonography and colonoscopy in the diagnosis of colorectal cancer. *Colorectal Dis* 2014;**16**:O82–9.

98. Horsthuis K, Bipat S, Bennink RJ, Stoker J. Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: meta-analysis of prospective studies. *Radiology* 2008;**247**:64–79.

99. Thomas C, Mandrik O, Whyte S. *Development of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies*. Sheffield: School of Health and Related Research (ScHARR), University of Sheffield; 2020.

100. Kaltenthaler E, Tappenden P, Paisley S, Squires H. *NICE DSU Technical Support Document 13: Identifying and Reviewing Evidence to Inform the Conceptualisation and Population of Cost-effectiveness Models*. Sheffield: Decision Support Unit, University of Sheffield; 2011.

101. Public Health England, Macmillan. *UK Complete Cancer Prevalence for 2013. Macmillan-NCRAS Cancer Prevalence Project. Updated UK Complete Cancer Prevalence for 2013 Workbook*. London: Public Health England, Macmillan; 2017.

102. NHS Digital, National Cancer Registration and Analysis Service (NCRAS). *Staging Data in England. Stage Group by Clinical Commissioning Group (CCG), Sustainability and Transformation Partnership (STP) or Cancer Alliance (CA) by Cancer Type for 21 Cancer Types 2019*. London: NHS Digital, NCRAS; 2021.

103. Pasvol T, Horsfall L, Bloom S, Segal A, Sabin C, Field N, *et al*. Incidence and prevalence of inflammatory bowel disease in UK primary care: a population-based cohort study. *BMJ Open* 2020;**10**:e036584.

104. Ghosh N, Premchand P. A UK cost of care model for inflammatory bowel disease. *Front Gastroenterol* 2015;**6**:169–74.

105. Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Agency for healthcare research and quality. *JAMA* 2021;**5**:1461.

106. Gatto NM, Frucht H, Sundararajan V, Jacobson JS, Grann VR, Neugut AI. Risk of perforation after colonoscopy and sigmoidoscopy: a population-based study. *J Natl Cancer Inst* 2003;**95**:230–6.

107. Office for National Statistics. *National Life Tables: England 2018–2020*. Newport: Office for National Statistics; 2021.

108. National Institute for Health and Care Excellence. *Upadacitinib for Treating Moderately to Severely Active Ulcerative Colitis [TA856]. Committee Papers*. London: NICE; 2022.

109. National institute for Health and Care Excellence. *Vedolizumab for Treating Moderately to Severely Active Ulcerative Colitis [TA342]*. London: NICE; 2015.

110. Stark RG, Reitmeir P, Leidl R, König HH. Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. *Inflamm Bowel Dis* 2010;**16**:42–51.

111. Hernández Alava M, Pudney S, Wailoo A. *Estimating EQ-5D by Age and Sex for the UK*. Sheffield: Decision Support Unit, University of Sheffield; 2022.

112. Dorian P, Kongnakorn T, Phatak H, Rublee DA, Kuznik A, Lanitis T, *et al*. Cost-effectiveness of apixaban vs. current standard of care for stroke prevention in patients with atrial fibrillation. *Eur Heart J* 2014;**35**:1897–906.

113. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value Health* 2011;**14**:539–45.

114. Jones K, Burns A. *Unit Costs of Health and Social Care*. Canterbury: PSSRU, University of Kent; 2021.

115. NHS England. *National Cost Collection: National Schedule of NHS Costs – Year 2021/22 – NHS Trusts and NHS Foundation Trust*. London: NHS England; 2023.

116. Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *Br J Cancer* 2013;**108**:686–90.

117. NHS Digital. *Routes to Diagnosis, 2018*. Leeds: NHS Digital; 2022.

118. Jones KC, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A, *et al*. *Unit Costs of Health and Social Care 2022 Manual. Technical Report*. Canterbury: Personal Social Services Research Unit, University of Kent; 2023.

119. Arasaradnam RP, Bhala N, Evans C, Greenaway J, Logan R, Penman I, *et al*.; BSG Endoscopy COVID working group. Faecal immunochemical testing in the COVID-19 era: balancing risk and costs. *Lancet Gastroenterol Hepatol* 2020;**5**:717–9.

120. Woehl A, McEwan P. The relation between disease activity, quality of life and health utility in patients with ulcerative colitis. *Gut* 2008;**57**:A153.

121. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, *et al*.; GEMINI 1 Study Group. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2013;**369**:699–710.

122. Moss S, Mathews C, Day TJ, Smith S, Seaman HE, Snowball J, Halloran SP. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. *Gut* 2017;**66**:1631–44.

123. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;**18**:S68–80.

124. Crooks CJ, West J, Jones J, Hamilton W, Bailey S, Abel G, *et al*. COLOFIT: development and internal-external validation of models using age, sex, faecal immunochemical and blood tests to optimise diagnosis of colorectal cancer in symptomatic patients. *medRxiv* 2025. https://doi.org/10.1111/apt.18459

125. Holtman GA, Berger MY, Burger H, Deeks JJ, Donner-Banzhoff N, Fanshawe TR, *et al*. Development of practical recommendations for diagnostic accuracy studies in low-prevalence situations. *J Clin Epidemiol* 2019;**114**:38–48.

126. Drosdowsky A, Lamb KE, Karahalios A, Bergin RJ, Milley K, Boyd L, *et al*. The effect of time before diagnosis and treatment on colorectal cancer outcomes: systematic review and dose–response meta-analysis. *Br J Cancer* 2023;**129**:993–1006.

127. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2012;**33**:607–17.

128. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. *J Stat Software* 2005;**12**:1–16.

129. Gies A, Gruner LF, Schrotz-King P, Brenner H. Effect of imperfect compliance with instructions for fecal sample collection on diagnostic performance of 9 fecal immunochemical tests. *Clin Gastroenterol Hepatol* 2019;**17**:1829–39.e4.

130. James T, Nicholson BD, Marr R, Paddon M, East JE, Justice S, *et al*. Faecal immunochemical testing (FIT): sources of result variation based on three years of routine testing of symptomatic patients in English primary care. *Br J Biomed Sci* 2021;**78**:211–7.

131. Mellen S, de Ferrars M, Chapman C, Bevan S, Turvill J, Turnock D. Evaluation of sample stability for a quantitative faecal immunochemical test and comparison of two sample collection approaches. *Ann Clin Biochem* 2018;**55**:657–64.

132. O'Driscoll S, Carroll M, Maclean W, Piggott C, Jourdan I, Benton SC. Assessment of the analytical performance of point-of-care faecal immunochemical tests for haemoglobin. *Ann Clin Biochem* 2021;**58**:181–9.

133. O'Driscoll S, Piggott C, Bruce H, Benton SC. An evaluation of ten external quality assurance scheme (EQAS) materials for the faecal immunochemical test (FIT) for haemoglobin. *Clin Chem Lab Med* 2021;**59**:307–13.

134. Piggott C, Carroll MRR, John C, O'Driscoll S, Benton SC. Analytical evaluation of four faecal immunochemistry tests for haemoglobin. *Clin Chem Lab Med* 2020;**59**:173–8.

135. Piggott C, Shugaa Z, Benton SC. Independent internal quality control (IQC) for faecal immunochemical tests (FIT) for haemoglobin: use of FIT manufacturers' IQC for other FIT systems. *Clin Chem Lab Med* 2021;**59**:e41–3.

136. Zubero MB, Arana-Arri E, Pijoan JI, Portillo I, Idigoras I, Lopez-Urrutia A, *et al*. Population-based colorectal cancer screening: comparison of two fecal occult blood test. *Front Pharmacol* 2014;**4**:175.

137. Ayling RM, Machesney M. Service evaluation of faecal immunochemical testing introduced for use in North East London for patients at low risk of colorectal cancer. *J Clin Pathol* 2021;**74**:163–6.

138. Bailey J, Chapman C, Jones J, Oliver S, Morling J, Banerjea A, *et al*. Sociodemographic variations in the uptake of faecal immunochemical tests (FIT) in a primary care symptomatic pathway for colorectal cancer. *Colorectal Dis* 2022;**24**:8–9.

139. Bailey J, Ibrahim H, Bunce J, Chapman C, Morling J, Simpson J, *et al*. Quantitative FIT stratification is superior to NICE referral criteria NG12 in a two-week wait colorectal cancer population. *Tech Coloproctol* 2021;**25**:1151–54.

140. Bailey JA, Khawaja A, Andrews H, Weller J, Chapman C, Morling JR, *et al*. GP access to FIT increases the proportion of colorectal cancers detected on urgent pathways in symptomatic patients in Nottingham. *Surgeon* 2021;**19**:93–102.

141. Bailey SER, Abel GA, Atkins A, Byford R, Davies SJ, Mays J, *et al*. Diagnostic performance of a faecal immunochemical test for patients with low-risk symptoms of colorectal cancer in primary care: an evaluation in the South West of England. *Br J Cancer* 2021;**124**:1231–6.

142. Chapman C, Bunce J, Oliver S, Ng O, Tangri A, Rogers R, *et al*. Service evaluation of faecal immunochemical testing and anaemia for risk stratification in the 2-week-wait pathway for colorectal cancer. *BJS Open* 2019;**3**:395–402.

143. Chapman C, Thomas C, Morling J, Tangri A, Oliver S, Simpson JA, *et al*. Early clinical outcomes of a rapid colorectal cancer diagnosis pathway using faecal immunochemical testing in Nottingham. *Colorectal Dis* 2020;**22**:679–88.

144. Ibrahim HAH, Chapman C, Morling J, Weller J, Tangri A, Simpson JA, *et al*. Keeping FIT: faecal haemoglobin measurement with FIT has stratification value in the diagnosis of colorectal cancer in all symptom and age groups. *Colorectal Dis* 2019;**21**:10.

145. Khasawneh F, Osborne R, Stephenson J , Barnes D, Seehra J, Danaher P, *et al.* Faecal immunochemical testing is a cost-effective way to stratify symptomatic patients for urgent straight to test investigation. *Colorectal Dis* 2020;**22**:6–12.

146. McSorley ST, Digby J, Clyde D, Cruickshank N, Burton P, Barker L, *et al*. Yield of colorectal cancer at colonoscopy according to faecal haemoglobin concentration in symptomatic patients referred from primary care. *Colorectal Dis* 2021;**23**:1615–21.

147. Steele R, Fraser C. *Haemoglobin for Timely Assessment of Patients with Symptoms of Colorectal Disease in Olsen Timely Diagnosis of Colorectal Disease*. New York: Springer; 2018.

148. Widlak MM, Neal M, Daulton E, Thomas CL, Tomkins C, Singh B, *et al*. Risk stratification of symptomatic patients suspected of colorectal cancer using faecal and urinary markers. *Colorectal Dis* 2018;**20**:O335–42.

149. Widlak MM, Thomas CL, Thomas MG, Tomkins C, Smith S, O'Connell N, *et al*. Diagnostic accuracy of faecal biomarkers in detecting colorectal cancer and adenoma in symptomatic patients. *Aliment Pharmacol Ther* 2017;**45**:354–63.

150. Burke CA, Lieberman D, Feuerstein JD. AGA clinical practice update on approach to the use of noninvasive colorectal cancer screening options: commentary. *Gastroenterology* 2022;**162**:952–6.

151. Craig M, Turner J, Torkington J, Crosby T. Faecal immunochemical test: challenges and opportunities for cancer diagnosis in primary care. *Br J Gen Pract* 2022;**72**:366–7.

152. D'Souza N, Anthony B, Muti A. Faecal immunochemical testing in general practice. *Br J Gen Pract* 2019;**69**:60.

153. Mowat C, Digby J, Strachan JA, Steele RJC, Fraser CG. Low sensitivity of fecal immunochemical tests (FIT) for detection of sessile serrated adenomas/polyps confirmed over clinical setting, geography, and FIT system. *Dig Dis Sci* 2019;**64**:3024–6.

154. Ray K. Prognostic potential of repeated faecal haemoglobin levels in CRC detection. *Nat Rev Gastroenterol Hepatol* 2022;**19**:416.

155. Rees CJ, Hamilton W. BSG guidelines on faecal immunochemical testing: are they 'FIT' for purpose? *Gut* 2022;**72**:1805–6.

156. Trivedi M, Gupta S. Is Promotion of fecal immunochemical testing 'FIT' to address COVID-19 disruptions to colorectal cancer screening? *Gastroenterology* 2022;**162**:1761–2.

157. Ali O, Gupta S, Brain K, Lifford KJ, Paranjothy S, Dolwani S. Acceptability of alternative technologies compared with faecal immunochemical test and/or colonoscopy in colorectal cancer screening: a systematic review. *J Med Screen* 2022;**30**:14–27.

158. Auge JM, Fraser CG, Rodriguez C, Roset A, Lopez-Ceron M, Grau J, *et al*. Clinical utility of one versus two faecal immunochemical test samples in the detection of advanced colorectal neoplasia in symptomatic patients. *Clin Chem Lab Med* 2016;**54**:125–32.

159. Auge JM, Rodriguez C, Espanyol O, Rivero L, Sandalinas S, Grau J, *et al*. An evaluation of the SENTiFIT 270 analyser for quantitation of faecal haemoglobin in the investigation of patients with suspected colorectal cancer. *Clin Chem Lab Med* 2018;**56**:625–33.

160. Cahill C, Lipson ME, Afzal AR, Maclean AR, Wong CK, Roen S, *et al*. Improved survival in a cohort of patients 75 years and over with FIT-detected colorectal cancer. *Dis Colon Rectum* 2022;**28**:28.

161. Chandrapalan S, Hee SW, Widlak MM, Farrugia A, Alam MT, Smith S, Arasaradnam RP. Performance of the faecal immunochemical test for the detection of colorectal neoplasms and the role of proton pump inhibitors in their diagnostic accuracy. *Colorectal Dis* 2021;**23**:1649–57.

162. Cubiella J, Vega P, Salve M, Díaz-Ondina M, Alves MT, Quintero E, *et al*.; COLONPREDICT study investigators. Development and external validation of a faecal immunochemical test-based prediction model for colorectal cancer detection in symptomatic patients. *BMC Med* 2016;**14**:128.

163. Digby J, Cleary S, Gray L, Datt P, Goudie DR, Steele RJC, *et al*. Faecal haemoglobin can define risk of colorectal neoplasia at surveillance colonoscopy in patients at increased risk of colorectal cancer. *United European Gastroenterol J* 2020;**8**:559–66.

164. Digby J, Strachan JA, McCann R, Steele RJC, Fraser CG, Mowat C. Measurement of faecal haemoglobin with a faecal immunochemical test can assist in defining which patients attending primary care with rectal bleeding require urgent referral. *Ann Clin Biochem* 2020;**57**:325–7.

165. Digby J, Strachan JA, Mowat C, Steele RJC, Fraser CG. Appraisal of the faecal haemoglobin, age and sex test (FAST) score in assessment of patients with lower bowel symptoms: an observational study. *BMC Gastroenterol* 2019;**19**:213.

166. Eskelinen M, Meklin J, Guimaraes DP, Selander T, Tiusanen T, Syrjanen K, Eskelinen M. The ColonView (CV) quick test for fecal occult blood shows significantly higher diagnostic accuracy in detecting distal than proximal colorectal cancer. *Anticancer Res* 2022;**42**:1879–91.

167. Gies A, Cuk K, Schrotz-King P, Brenner H. Direct comparison of ten quantitative fecal immunochemical tests for hemoglobin stability in colorectal cancer screening. *Clin Translational Gastroenterol* 2018;**9**:e168.

168. Gies A, Cuk K, Schrotz-King P, Brenner H. Direct comparison of diagnostic performance of 9 quantitative fecal immunochemical tests for colorectal cancer screening. *Gastroenterology* 2018;**154**:93–104.

169. Gies A, Cuk K, Schrotz-King P, Brenner H. Combination of different fecal immunochemical tests in colorectal cancer screening: any gain in diagnostic performance? *Clin Chem Lab Med*. 2019;**11**:120.

170. Gies A, Niedermaier T, Alwers E, Hielscher T, Weigl K, Heisser T, *et al*. Consistent major differences in sex- and age-specific diagnostic performance among nine faecal immunochemical tests used for colorectal cancer screening. *Cancers* 2021;**13**:3574.

171. Habbu PP, Ananthi N, Shaikh AK. Study of specificity, sensitivity, efficiency and clinical correlation between Timp-1 and Mif Protein as biochemical markers in colorectal cancer patients. *Int J Pharm Sci Res* 2022;**13**:3298–303.

172. Herrero JM, Vega P, Salve M, Bujanda L, Cubiella J. Symptom or faecal immunochemical test based referral criteria for colorectal cancer detection in symptomatic patients: a diagnostic tests study. *BMC Gastroenterol* 2018;**18**:155.

173. Hicks G, D'Souza N, Georgiou Delisle T, Chen M, Benton SC, Abulafi M; NICE FIT Steering Group. Using the faecal immunochemical test in patients with rectal bleeding: evidence from the NICE FIT study. *Colorectal Dis* 2021;**23**:1630–8.

174. Jin P, You P, Fang J, Kang Q, Gu F, Cai Y, *et al*. Comparison of performance of two stool DNA tests and a fecal immunochemical test in detecting colorectal neoplasm: a multicenter diagnostic study. *Cancer Epidemiol Biomarkers Prev* 2022;**31**:654–61.

175. Jin P, You P, Fang J, Kang Q, Gu F, Cai Y, *et al*. Comparison of performance of two stool DNA tests and a fecal immunochemical test in detecting colorectal neoplasm: a multicenter diagnostic study. *Cancer Epidemiol Biomarkers Prev* 2022;**31**:654–61.

176. Kapidzic A, van Roon AH, van Leerdam ME, van Vuuren AJ, van Ballegooijen M, Lansdorp-Vogelaar I, *et al*. Attendance and diagnostic yield of repeated two-sample faecal immunochemical test screening for colorectal cancer. *Gut* 2017;**66**:118–23.

177. Kaul A, Shah A, Magill F, Hawkins S, Skaife P. Immunological faecal occult blood testing: a discriminatory test to identify colorectal cancer in symptomatic patients. *Int J Surg* 2013;**11**:329–31.

178. Lincoln A, Benton S, Piggott C, North BV, Rigney J, Young C, *et al*. Exploring the utility and acceptability of faecal immunochemical testing (FIT) as a novel intervention for the improvement of colorectal cancer (CRC) surveillance in individuals with Lynch syndrome (FIT for Lynch study): a single-arm, prospective, multi-centre, non-randomised study. *BMC Cancer* 2022;**22**:1144.

179. Lu DC, Zhang QF, Li L, Luo XK, Liang B, Lu YH, *et al*. Methylated Septin9 has moderate diagnostic value in colorectal cancer detection in Chinese population: a multicenter study. *BMC Gastroenterol* 2022;**22**:232.

180. Lué A, Hijos G, Sostres C, Perales A, Navarro M, Barra MV, *et al*. The combination of quantitative faecal occult blood test and faecal calprotectin is a cost-effective strategy to avoid colonoscopies in symptomatic patients without relevant pathology. *Ther Adv Gastroenterol* 2020;**13**:1756284820920786.

181. Luthgens K, Maier A, Kampert I, Sieg A, Schmidt-Gayk HH. A highly sensitive assay for the detection of fecal occult blood. *Clin Lab* 1998;**44**:543–51.

182. Mattar R, Marques SB, Minata MK, Silva-Etto J, Sakai P, EGH DEM. Diagnostic accuracy of one sample or two samples quantitative fecal immunochemical tests for intestinal neoplasia detection. *Arq Gastroenterol* 2020;**57**:316–22.

183. Meester RGS, van de Schootbrugge-Vandermeer HJ, Breekveldt ECH, de Jonge L, Toes-Zoutendijk E, Kooyker A, *et al*. Faecal occult blood loss accurately predicts future detection of colorectal cancer. A prognostic model. *Gut* 2022;**72**:101–8.

184. Navarro M, Hijos G, Sostres C, Lue A, Puente-Lanzarote JJ, Carrera-Lasfuentes P, Lanas A. Reducing the cut-off value of the fecal immunochemical test for symptomatic patients does not improve diagnostic performance. *Front Med* 2020;**7**:410.

185. Parente F, Marino B, Ilardo A, Fracasso P, Zullo A, Hassan C, *et al*. A combination of faecal tests for the detection of colon cancer: a new strategy for an appropriate selection of referrals to colonoscopy? A prospective multicentre Italian study. *Eur J Gastroenterol Hepatol* 2012;**24**:1145–52.

186. Rodríguez-Alonso L, Rodríguez-Moranta F, Ruiz-Cerulla A, Lobatón T, Arajol C, Binefa G, *et al*. An urgent referral strategy for symptomatic patients with suspected colorectal cancer based on a quantitative immunochemical faecal occult blood test. *Dig Liver Dis* 2015;**47**:797–804.

187. Small S, Coulson R, Spence R, McAllister I. Is qFIT a useful tool in prioritising symptomatic patients referred with suspect colorectal cancer in the COVID-19 era? *Ulster Med J* 2022;**91**:79–84.

188. Suehiro Y, Zhang Y, Hashimoto S, Takami T, Higaki S, Shindo Y, *et al*. Highly sensitive faecal DNA testing of TWIST1 methylation in combination with faecal immunochemical test for haemoglobin is a promising marker for detection of colorectal neoplasia. *Ann Clin Biochem* 2018;**55**:59–68.

189. Switalski J, Tatara T, Wnuk K, Miazga W, Karauda D, Matera A, *et al*. Clinical Effectiveness of faecal immunochemical test in the early detection of colorectal cancer-an umbrella review. *Cancers* 2022;**14**:4391.

190. van Turenhout ST, Oort FA, van der Hulst RWM, Visscher AP, Terhaar sive Droste JS, Scholten P, *et al*. Prospective cross-sectional study on faecal immunochemical tests: sex specific cut-off values to obtain equal sensitivity for colorectal cancer? *BMC Gastroenterol* 2014;**14**:217.

191. Xu H, Chen H, Hu J, Xiong Z, Li D, Wang S, Yu J. Feasibility of quantification based on novel evaluation with stool DNA and fecal immunochemical test for colorectal cancer detection. *BMC Gastroenterol* 2022;**22**:384.

192. Xu J, Rong L, Gu F, You P, Ding H, Zhai H, *et al*. Asia-Pacific Colorectal Screening Score combined with stool DNA test improves the detection rate for colorectal advanced neoplasms. *Clin Gastroenterol Hepatol* 2022;**14**:14.

193. Young GP, Woodman RJ, Symonds E. Detection of advanced colorectal neoplasia and relative colonoscopy workloads using quantitative faecal immunochemical tests: an observational study exploring the effects of simultaneous adjustment of both sample number and test positivity threshold. *BMJ Open Gastroenterol* 2020;**7**:e000517.

194. Zhao S, Wang S, Pan P, Xia T, Wang R, Cai Q, *et al*.; Gastrointestinal Early Cancer Prevention and Treatment Alliance of China (GECA). FIT-based risk-stratification model effectively screens colorectal neoplasia and early-onset colorectal cancer in Chinese population: a nationwide multicenter prospective study. *J Hematol Oncol* 2022;**15**:162.

195. Navarro M, Omella I, Carrera P, Hijos G, Roncales P, Sopena F, *et al*. Fecal hemoglobin concentration, a good predictor of risk of advanced colorectal neoplasia in symptomatic and asymtomatic patients. *Gastroenterology* 2017;**152**:S548.

196. Schwettmann L, Lied A, Eriksen R. Evaluation of the Sentinel-FOB gold faecal immunochemical test for the presence of haemoglobin using the automated Roche Cobas 8000 system. *Prac Lab Med* 2022;**29**:e00263.

197. Calanzani N, Pannebakker MM, Tagg MJ, Walford H, Holloway P, de Wit N, *et al*. Who are the patients being offered the faecal immunochemical test in routine English general practice, and for what symptoms? A prospective descriptive study. *BMJ Open* 2022;**12**:e066051.

198. Carroll MRR, John C, Mantio D, Djedovic NK, Benton SC. An assessment of the effect of haemoglobin variants on detection by faecal immunochemical tests. *Ann Clin Biochem* 2018;**55**:706–9.

199. Chen CC, Chang PY, Chang YS, You JF, Chan EC, Chen JS, *et al*. MicroRNA-based signature for diagnosis and prognosis of colorectal cancer using residuum of fecal immunochemical test. *Biomedical J* 2022;**46**:144–53.

200. Fernandez de Castro JD, Baiocchi Ureta F, Fernandez Gonzalez R, Pin Vieito N, Cubiella Fernandez J. Faecal immunochemical test impact on prognosis of colorectal cancer detected in symptomatic patients. *Diagnostics* 2022;**12**:1013.

201. Hunt N, Allcock R, Myers M. Faecal immunochemical testing (FIT) for colorectal cancer in symptomatic primary care patients. *Clin Chim Acta* 2019;**493**:S153.

202. Lee JM, Park MJ, Heo W, Park KG, Park YG, Han SB, *et al*. Clinical utility of fecal immunochemical transferrin test in gastrointestinal bleeding detection. *ACM* 2018;**21**:51–7.

203. Maria Theresa R, Matthew R, Richard MM, Fareeda C, Mona J, Julia W. Rapid diagnostic pathways for suspected colorectal cancer: views of primary and secondary care clinicians on challenges and their potential solutions. *BMJ Open* 2015;**5**:e008577.

204. Navarro M, Hijos G, Ramirez T, Omella I, Carrera-Lasfuentes P, Lanas A. Fecal hemoglobin concentration, a good predictor of risk of advanced colorectal neoplasia in symptomatic and asymptomatic patients. *Front Med* 2019;**6**:91.

205. Niedermaier T, Alwers E, Chen X, Heisser T, Hoffmeister M, Brenner H. A single measurement of fecal hemoglobin concentration outperforms polygenic risk score in colorectal cancer risk assessment. *medRxiv* 2022;**22**:353.

206. Wilson N, Baker-Beal L, Kyaw WW, Sonabend R, Logan R. Real world experience of faecal immunochemical testing (FIT) in UK primary care to support the referral and diagnosis of colorectal cancer (CRC). *United European Gastroenterol J* 2020;**8**:579–80.

207. Wisse P, Moen S, Marijnissen F, Raab H, van Vuuren A, Lansdorp Vogelaar I, *et al*. *Use of Fecal Immunochemical Test in Symptomatic Patients*. WEO The Voice of World Endoscopy, abstract no. 4245. URL: www.worldendo.org/assets-craft/pdf/committee/colorectal-cancer/220520_Wisse_Pieter_WEO_Use-of-FIT-in-symptomatic-population.pdf (accessed 22 July 2024).

208. Krivec S, Konda G, Šibli R, Marc J. *Assessment of the Diagnostic Applicability of Quantitative Immunochemical Faecal Occult Blood Tests. Clin Chem Lab Med*. 2011;49:S1–S874. https://doi.org/10.1515/CCLM.2011.518

209. Sandhu K, Naik S, Ayling RM. Use of faecal immunochemical testing as an alternative to faecal calprotectin in children. *Ann Clin Biochem* 2021;**58**:230–5.

210. Zacharopoulou L, Cama R, Kapoor N, Mebarek L, Bhatti H, Sawyer P, *et al*. PTH-99 Faecal immunochemical tests for younger patients presenting with bowel symptoms. *Gut* 2021;**70**:A1–62-A.

211. Zhu M, Fan L, Han M, Zhu S, Zhang S, Shi H. The usefulness of fecal hemoglobin and calprotectin tests in diagnosing significant bowel diseases: a prospective study. *Scand J Gastroenterol* 2022;**58**:368–74.

212. Rodriguez-Alonso L, Rodriguez-Moranta F, Maisterra S, Botargues JM, Berrozpe A, Ruiz-Cerulla A, *et al*. The EPAGE guidelines are not an effective strategy for managing colonoscopies during the COVID-19 pandemic. *Gastroenterol Hepatol* 2022;**45**:9–17.

213. Almoneef NM, Alkhenizan AH, Mahmoud AS, Alsoghayer SA, Aldheshe AA. The yield of fecal occult blood testing as a screening tool for colon cancer in a primary care setting. *J Family Med Prim Care* 2022;**11**:4435–9.

214. Cubiella J, Digby J, Rodríguez-Alonso L, Vega P, Salve M, Díaz-Ondina M, *et al*.; COLONPREDICT study investigators. The fecal hemoglobin concentration, age and sex test score: development and external validation of a simple prediction tool for colorectal cancer detection in symptomatic patients. *Int J Cancer* 2017;**140**:2201–11.

215. Khan AA, Klimovskij M, Harshen R. Accuracy of faecal immunochemical testing in patients with symptomatic colorectal cancer. *BJS Open* 2020;**4**:1180–8.

216. Maclean W, Benton SC, Whyte MB, Rockall T, Jourdan I. Efficacy and accuracy of faecal sampling by a digital rectal examination for FIT. *Ann Clin Biochem* 2023;**60**:169–76.

217. McDonald P, Digby J, Innes C, Strachan J, Carey F, Steele R, *et al*. Low faecal haemoglobin concentration potentially rules out significant colorectal disease. *Colorectal Dis* 2013;**15**:e151–e9.

218. Region Örebro County, Region Jämtland Härjedalen, Region Östergötland, Uppsala County Council Sweden, Dalarna County Council Sweden, Landstinget i Värmland. *Accuracy and Predictive Values for Colorectal Cancer of Quantitative FIT in Symptomatic Patients in Primary Care*. 2021. URL: https://ClinicalTrials.gov/show/NCT05156307 (accessed 14 April 2025).

219. Ghada M, Bohra A, Segal J. *Sensitivity and Specificity of Stool Markers in Inflammatory Bowel Disease: A Systematic Review and Meta-analysis*. PROSPERO 2022 CRD42022354526. URL: www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022354526 (accessed 22 July 2024).

220. Jen-Hao Yeh W-LW. Efficacy of fecal immunochemical test in young patients (aged ≤ 50 years). URL: www.crd.york.ac.uk/PROSPERO/view/CRD42022333124 (accessed 22 July 2024).

221. Pham J, Laven-Law G, Winter JM, Wassie MM, Cock C, Symonds EL. The diagnostic accuracy of a fecal immunochemical test in detecting colorectal cancer and advanced precancerous colorectal neoplasia in patients with iron deficiency: a protocol for systematic review and meta-analysis. *Gastroenterol Res Pract* 2023;2023:5982580.

222. Loov A, Hogberg C, Lilja M, Theodorsson E, Hellstrom P, Metsini A, Olsson L. Diagnostic accuracy for colorectal cancer of a quantitative faecal immunochemical test in symptomatic primary care patients: a study protocol. *Diagn Progn Res* 2022;**6**:16.

223. Tian Zhi Lim JLAUJLGW. Uncovering the barriers and facilitators towards undergoing follow-up colonoscopy among individuals with FIT positive results: a systematic review. URL: www.crd.york.ac.uk/PROSPERO/view/CRD42022358651 (accessed 22 July 2024).

224. Tian Zhi Lim JLJLGW. What can we do to improve compliance to follow-up colonoscopy among individuals with FIT positive results? a systematic review. URL: www.crd.york.ac.uk/PROSPERO/view/CRD42022358655 (accessed 22 July 2024).

225. Mahidol University. *FIT and Fecal Calprotectin in Patients with Chronic Lower GI Symptoms*. Bangkok: Mahidol University; 2020.

226. Poulsen W, Xu W, Theodoratou E, Timofeeva M. *Factors Associated with the Quality of Faecal Occult Blood Test (FOBT) for Early Detection of Colorectal Cancer: A Systematic Review and Meta-analysis Protocol*. University of Southern Denmark; 2022.

227. Zhen Junhai CB. Value of faecal immunochemical tests in significant bowel disease screening among patients with lower bowel symptoms in primary care: a meta analysis. URL: www.crd.york.ac.uk/PROSPERO/view/CRD42022353328 (accessed 22 July 2024).

228. Farkas NG, Fraser CG, Maclean W, Jourdan I, Rockall T, Benton SC. Replicate and repeat faecal immunochemical tests in symptomatic patients: a systematic review. *Ann Clin Biochem* 2022;**60**:27–36.

229. Jung YS, Im E, Park CH. Impact of antiplatelet agents and anticoagulants on the performance of fecal immunochemical tests: a systematic review and meta-analysis. *Surg Endosc* 2022;**36**:4299–311.

230. Nasir Kansestani A, Zare ME, Tong Q, Zhang J. Comparison of faecal protein biomarkers' diagnostic accuracy for colorectal advanced neoplasms: a systematic review and meta-analysis. *Sci Rep* 2022;**12**:2623.

231. Pang SJ, Lin ZP, Sun Z, Zhang Y, Yuan ZG, Yang N. Impact of antithrombotic drugs on the accuracy of fecal occult blood testing for advanced colorectal neoplasia screening: a meta-analysis and systematic review. *Z Gastroenterol* 2022;**17**:17.

232. Pin Vieito N, Zarraquinos S, Cubiella J. High-risk symptoms and quantitative faecal immunochemical test accuracy: systematic review and meta-analysis. *World J Gastroenterol* 2019;**25**:2383–401.

233. Pin-Vieito N, Tejido-Sandoval C, de Vicente-Bielza N, Sanchez-Gomez C, Cubiella J. Faecal immunochemical tests safely enhance rational use of resources during the assessment of suspected symptomatic colorectal cancer in primary care: systematic review and meta-analysis. *Gut* 2022;**71**:950–60.

234. Quyn AJ, Steele RJC, Digby J, Strachan JA, Mowat C, McDonald PJ, *et al*. Application of NICE guideline NG12 to the initial assessment of patients with lower gastrointestinal symptoms: not FIT for purpose? *Ann Clin Biochem* 2018;**55**:69–76.

235. Saw KS, Liu C, Xu W, Varghese C, Parry S, Bissett I. Faecal immunochemical test to triage patients with possible colorectal cancer symptoms: meta-analysis. *Br J Surg* 2022;**109**:182–90.

236. Westwood M, Lang S, Armstrong N, van Turenhout S, Cubiella J, Stirk L, *et al*. Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. *BMC Med* 2017;**15**:189.

237. Zou J, Xiao Z, Wu Y, Yang J, Cui N. Noninvasive fecal testing for colorectal cancer. *Clin Chimica Acta* 2022;**524**:123–31.

238. Cilona A, Zullo A, Hassan C, Ridola L, Annese M. Is faecal-immunochemical test useful in patients with iron deficiency anaemia and without overt bleeding? *Digestive Liver Dis* 2011;**43**:1022–4.

239. Díaz O, Blanco V, Ceballos O, Salve M, Macía P, Cubiella J. *Clinical or Analytical Criteria for Colorectal Cancer (CRC) Detection in Symptomatic Patients? A Diagnostic Tests Study*. Paper presented at IFCC WorldLab, Istanbul, Türkiye, June 22–27, 2014.

240. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* 1996;**313**:275–83.

241. NHS England. *Waiting Times for Suspected and Diagnosed Cancer Patients: 2020–21 Annual Report*. London: NHS England; 2021.

242. NHS England. *Waiting Times for Suspected and Diagnosed Cancer Patients: 2019–20 Annual Report*. London: NHS England; 2020.

243. National Disease Registration Service. *Cancer Waiting Times (CWT) Urgent Suspected Cancer Referrals: Referral, Conversion and Detection Rates*. CancerData; 2019.

244. National Disease Registration Service. *Staging Data in England*. CancerData; 2019.

245. Jones E, Epstein D, Garcia-Mochon L. A procedure for deriving formulas to convert transition rates to probabilities for multistate Markov models. *Med Decis Making* 2017;**37**:779–89.

246. Office for National Statistics. *National Life Tables: England*. Newport: Office for National Statistics; 2021.

247. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**:509–18.

Appendix 1 Literature search strategies

Clinical review search strategy

Ovid MEDLINE(R) ALL 1946 to 6 December 2022

1 f?ecal immunochemical test.mp. 1259
2 f?ecal occult blood.mp. 4447
3 f?ecal h?emoglobin.mp. 269
4 ((immunochromatographic or immuno-chromatographic or immunochem$ or immuno-chem$ or immunohisto-chem$ or immuno-histochem$ or immunol$ or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) adj4 (f?ecal or f?eces or stool or stools or FIT)).mp. 3598
5 (iFOBT or qFIT).mp. 208
6 or/1–5 7290
7 F?ecal h?emoglobin.ti,ab,ot,hw. 256
8 H?emoccult.ti,ab,ot,hw. 728
9 FOBT.ti,ab,ot,hw. 1429
10 7 or 8 or 9 2335
11 (f?ecal or f?eces or stool or stools).ti,ab,ot,hw. 211,912
12 occult blood/or occult blood.ti,ab,ot,hw. 8924
13 (test$ or measur$ or screen$ or exam$).ti,ab,ot,hw. 10,705,030
14 11 and 12 and 13 6023
15 6 or 10 or 14 8736
16 exp colorectal neoplasms/ 231,240
17 exp cecal neoplasms/ 6041
18 ((colorect$ or rectal$ or rectum$ or colon$ or sigma$ or sigmo$ or rectosigm$ or bowel$ or anal or anus) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot,hw. 324,030
19 CRC.ti,ab,ot. 43,421
20 ((cecum or cecal or caecum or caecal or il?eoc?ecal or il?eoc?ecum) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 2755
21 (large intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 1839
22 (lower intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 34
23 16 or 17 or 18 or 19 or 20 or 21 or 22 335,879
24 15 and 23 5937
25 limit 24 to yr="2022 -Current" 426
26 (FOB gold$ or FOBgold$ or SENTiFIT).mp. 38
27 (JACK-arc$ or JACKarc$ or HM-JACK$ or HM JACK$ or HMJACK$).mp. 23
28 (OC Sensor$ or OC-Sensor$ or OCSensor$ or Ceres).mp. 371
29 (OC Pledia$ or OC-Pledia$ or OCPledia or OC-iO).mp. 0
30 (NS-Prime or NSPrime or NS-Plus).mp. 37
31 (POC FIT QRG or POCFITQRG).mp. 0
32 (immundiagnostik or IDK or turbifit or turbitube).mp. 125
33 quikread.mp. 19
34 or/25–33 994
35 limit 34 to yr="2016 -Current" 740
36 exp animals/not (exp animals/and humans/) 5,072,762
37 35 not 36 729

EMBASE 1974 to week 49 2022 (searched 6 December 2022)

1 f?ecal immunochemical test.mp. 2253
2 f?ecal occult blood.mp. 6908
3 f?ecal h?emoglobin.mp. 436
4 ((immunochromatographic or immuno-chromatographic or immunochem$ or immuno-chem$ or immunohistochem$ or immuno-histochem$ or immunol$ or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) adj4 (f?ecal or f?eces or stool or stools or FIT)).mp. 6067
5 (iFOBT or qFIT).mp. 392
6 or/1–5 11,749
7 F?ecal h?emoglobin.ti,ab,ot,hw. 422
8 H?emoccult.ti,ab,ot,hw. 987
9 FOBT.ti,ab,ot,hw. 2786
10 7 or 8 or 9 4077
11 (f?ecal or f?eces or stool or stools).ti,ab,ot,hw. 279,057
12 occult blood/or occult blood.ti,ab,ot,hw. 18,102
13 (test$ or measur$ or screen$ or exam$).ti,ab,ot,hw. 13,879,472
14 11 and 12 and 13 10,328
15 6 or 10 or 14 14,766
16 exp colorectal cancer/or colon cancer/or rectum cancer/ 316,446
17 exp cecum tumor/ 2471
18 ((colorect$ or rectal$ or rectum$ or colon$ or sigma$ or sigmo$ or rectosigm$ or bowel$ or anal or anus) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot,hw. 506,633
19 CRC.ti,ab,ot. 70,056
20 ((cecum or cecal or caecum or caecal or il?eoc?ecal or il?eoc?ecum) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 3559
21 (large intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 1807
22 (lower intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 44
23 16 or 17 or 18 or 19 or 20 or 21 or 22 515,018
24 15 and 23 9914
25 limit 24 to yr="2022 -Current" 649
26 (FOB gold$ or FOBgold$ or SENTiFIT).mp. 107
27 (JACK-arc$ or JACKarc$ or HM-JACK$ or HM JACK$ or HMJACK$).mp. 73
28 (OC Sensor$ or OC-Sensor$ or OCSensor$ or Ceres).mp. 774
29 (OC Pledia$ or OC-Pledia$ or OCPledia or OC-iO).mp. 0
30 (NS-Prime or NSPrime or NS-Plus).mp. 75
31 (POC FIT QRG or POCFITQRG).mp. 0
32 (immundiagnostik or IDK or turbifit or turbitube).mp. 411
33 quikread.mp. 52
34 or/25-331997
35 limit 34 to yr="2016 -Current" 1406
36 limit 35 to embase732
37 limit 35 to conference abstracts 500
38 limit 35 to "preprints (unpublished, non-peer reviewed)" 7

The Cochrane Library (searched 12 December 2022)

Search name: DAP50 final

Date run: 12 December 2022 18:29:15

ID Search Hits

#1 (fecal immunochemical test* or faecal immunochemical test*):ti,ab,kw (Word variations have been searched) 497
#2 (fecal occult blood or faecal occult blood):ti,ab,kw (Word variations have been searched) 1087
#3 (fecal hemoglobin or faecal hemoglobin or fecal haemoglobin or faecal haemoglobin):ti,ab,kw (Word variations have been searched) 298
#4 ((immunochromatographic or immuno-chromatographic or immunochem* or immuno-chem* or immunohistochem* or immuno-histochem* or immunol* or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) near/4 (fecal or faecal or feces or faeces or stool or stools or FIT)):ti,ab,kw (Word variations have been searched) 1041
#5 (iFOBT or qFIT):ti,ab,kw (Word variations have been searched) 37
#6 (Hemoccult or haemoccult):ti,ab,kw (Word variations have been searched) 129
#7 (FOBT):ti,ab,kw (Word variations have been searched) 411
#8 ((fecal or feces or faecal or faeces or stool or stools)):ti,ab,kw AND (occult blood):ti,ab,kw AND (test* or measur* or screen* or exam*):ti,ab,kw (Word variations have been searched) 1153
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 2191
#10 MeSH descriptor: [Colorectal Neoplasms] explode all trees 9373
#11 MeSH descriptor: [Cecal Neoplasms] explode all trees 21
#12 ((colorect* or rectal* or rectum* or colon* or sigma* or sigmo* or rectosigm* or bowel* or anal or anus) near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 26,503
#13 (CRC):ti,ab,kw (Word variations have been searched) 5111
#14 ((cecum or cecal or caecum or caecal or ileocecal or ileocecum or ileocaecal or ileocaecum) near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 246
#15 (large intestin* near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 172
#16 (lower intestin* near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 182
#17 #10 or #11 or #12 or #13 or #14 or #15 or #16 27,348
#18 #9 and #17 with Cochrane Library publication date Between Jan 2022 and Dec 2022 103
#19 (FOB gold* or FOBgold* or SENTiFIT):ti,ab,kw OR (JACK-arc* or JACKarc* or HM-JACK* or HM JACK* or HM-JACK*):ti,ab,kw OR (OC Sensor* or OC-Sensor* or OCSensor* or Ceres or OC Pledia* or OC-Pledia* or OCPledia or OC-iO):ti,ab,kw OR (POC FIT QRG or POCFITQRG or immundiagnostik or IDK or turbifit or turbitube or quikread):ti,ab,kw OR (NS-Prime or NSPrime or NS-Plus):ti,ab,kw (Word variations have been searched) 175
#20 #18 or #19 with Cochrane Library publication date Between Jan 2016 and Dec 2022 224

INAHTA (searched 13 December 2022)

Single word strings:

Faecal/fecal/colorectal/colon/cecal = 0 results

NIHR HTA programme website (searched 13 December 2022)

Searched website – only found a few blogs including references to the RECEDE study.

PROSPERO (searched 13 December 2022)

This website only allows for simple searches:

Colorectal AND faecal (records added to PROSPERO since 1 January 2022) = 15 results

Colorectal AND fecal = 20 results (including the 15 above)

fecal immunochemical test = 6 results

faecal immunochemical test = 6 results

faecal occult blood = 7 results

fecal occult blood = 11 results

FOBT = 8 results

MeSH Colorectal Neoplasms/= 41 results

Faecal and test* = 51 results

Fecal and test* and cancer = 27 results

ClinicalTrials.gov (searched 13 December 2022)

(CTgov automatically expands the search to include synonyms and alternate spellings)

Colorectal cancer AND faecal = 341 results since 1/1/2016

Colorectal cancer AND FIT = 159 results since 1/1/2016

Colon cancer AND faecal = 90 results since 1/1/2016

Colon cancer AND FIT = 32 results “”

Rectal cancer AND FIT = 11 results “”

Rectal cancer AND faecal = 92 results “”

EU Trials Register (searched 13 December 2022)

0 results

WHO ICTRP (searched 13 December 2022)

colon cancer OR colorectal cancer OR rectal cancer OR cecal cancer

AND

faecal OR fecal OR FIT or FOBT or iFOBT

32 results

Economic modelling search strategies

Cost-effectiveness and quality of life studies of faecal immunochemical test in patients with symptoms suggestive of colorectal cancer

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions 1946 to 22 February 2023

1 f?ecal immunochemical test.mp. 1292
2 f?ecal occult blood.mp. 4479
3 f?ecal h?emoglobin.mp. 272
4 ((immunochromatographic or immuno-chromatographic or immunochem$ or immuno-chem$ or immunohisto-chem$ or immuno-histochem$ or immunol$ or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) adj4 (f?ecal or f?eces or stool or stools or FIT)).mp. 3658
5 (iFOBT or qFIT).mp. 214
6 or/1-57377
7 F?ecal h?emoglobin.ti,ab,ot,hw. 258
8 H?emoccult.ti,ab,ot,hw. 728
9 FOBT.ti,ab,ot,hw. 1440
10 7 or 8 or 9 2348
11 (f?ecal or f?eces or stool or stools).ti,ab,ot,hw. 214,450
12 occult blood/or occult blood.ti,ab,ot,hw. 8990
13 (test$ or measur$ or screen$ or exam$).ti,ab,ot,hw. 10,821,494
14 11 and 12 and 13 6077
15 6 or 10 or 14 8824
16 exp colorectal neoplasms/ 233,170
17 exp cecal neoplasms/ 6078
18 ((colorect$ or rectal$ or rectum$ or colon$ or sigma$ or sigmo$ or rectosigm$ or bowel$ or anal or anus) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot,hw. 327,674
19 CRC.ti,ab,ot. 44,474
20 ((cecum or cecal or caecum or caecal or il?eoc?ecal or il?eoc?ecum) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 2778
21 (large intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 1841
22 (lower intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 34
23 16 or 17 or 18 or 19 or 20 or 21 or 22 339,635
24 15 and 23 6007
25 (FOB gold$ or FOBgold$ or SENTiFIT).mp. 39
26 (JACK-arc$ or JACKarc$ or HM-JACK$ or HM JACK$ or HMJACK$).mp. 23
27 (OC Sensor$ or OC-Sensor$ or OCSensor$ or Ceres).mp. 373
28 (OC Pledia$ or OC-Pledia$ or OCPledia or OC-iO).mp. 0
29 (NS-Prime or NSPrime or NS-Plus).mp. 37
30 (POC FIT QRG or POCFITQRG).mp. 0
31 (immundiagnostik or IDK or turbifit or turbitube).mp. 126
32 quikread.mp. 20
33 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 6459
34 Economics/ 27,492
35 exp "costs and cost analysis"/ 262,760
36 Economics, Dental/ 1920
37 exp economics, hospital/ 25,681
38 Economics, Medical/ 9240
39 Economics, Nursing/ 4013

40 Economics, Pharmaceutical/ 3095
41 (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. 1,010,903
42 (expenditure$ not energy).ti,ab. 36,017
43 value for money.ti,ab. 2078
44 budget$.ti,ab. 34,677
45 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 1,174,425
46 ((energy or oxygen) adj cost).ti,ab. 4690
47 (metabolic adj cost).ti,ab. 1676
48 ((energy or oxygen) adj expenditure).ti,ab. 28,548
49 46 or 47 or 48 33,866
50 45 not 49 1,166,602
51 letter.pt. 1,208,094
52 editorial.pt. 636,950
53 historical article.pt. 369,079
54 or/51-53 2,193,150
55 50 not 54 1,126,959
56 exp animals/not humans/ 5,094,439
57 55 not 56 1,053,612
58 bmj.jn. 87,108
59 "cochrane database of systematic reviews".jn. 16,141
60 health technology assessment winchester england.jn. 1496
61 or/58–60 104,745
62 57 not 61 1,046,870
63 33 and 62 1042
64 limit 63 to yr="2016 -Current" 388
65 quality-adjusted life years/or quality of life/ 272,427
66 (sf36 or sf 36 or sf-36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or short-form thirty six or short form thirty six or short form thirtysix or short form thirty six).ti,ab,ot. 29,855
67 (sf6 or sf 6 or sf-6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).ti,ab,ot. 2562
68 (sf12 or sf 12 or sf-12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).ti,ab,ot. 7345
69 (sf6D or sf 6D or sf-6D or short form 6D or shortform 6D or sf six D or sfsixD or shortform six D or short form six D).ti,ab,ot. 980
70 (sf20 or sf 20 or sf-20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).ti,ab,ot. 448
71 (sf8 or sf 8 or sf-8 or short form 8 or shortform 8 or sf eight or sfeight or shortform eight or short form eight).ti,ab,ot. 728
72 "health related quality of life".ti,ab,ot. 55,787
73 (Quality adjusted life or Quality-adjusted-life).ti,ab,ot. 16,616
74 "assessment of quality of life".ti,ab,ot. 2154
75 (euroqol or euro qol or eq5d or eq 5d).ti,ab,ot. 15,788
76 (hql or hrql or hqol or h qol or hrqol or hr qol).ti,ab,ot. 26,757
77 (hye or hyes).ti,ab,ot. 75
78 health$ year$ equivalent$.ti,ab,ot. 40
79 (hui or hui1 or hui2 or hui3 or hui4 or hui-4 or hui-1 or hui-2 or hui-3).ti,ab,ot. 1904
80 (quality time or qwb or quality of well being or "quality of wellbeing" or "index of wellbeing" or "index of well being").ti,ab,ot,hw. 1127
81 (Disability adjusted life or Disability-adjusted life or health adjusted life or health-adjusted life or "years of healthy life" or healthy years equivalent or "years of potential life lost" or "years of health life lost").ti,ab,ot. 5764
82 (QALY$ or DALY$ or HALY$ or YHL or HYES or YPLL or YHLL or qald$ or qale$ or qtime$ or AQoL$).ti,ab,ot. 19,557
83 (timetradeoff or time tradeoff or time trade-off or time trade off or TTO or Standard gamble$ or "willingness to pay").ti,ab,ot. 10,702
84 15d.ti,ab,ot. 1923

85 (HSUV$ or health state$ value$ or health state$ preference$ or HSPV$).ti,ab,ot. 493
86 (utilit$ adj3 ("quality of life" or valu$ or scor$ or measur$ or health or life or estimat$ or elicit$ or disease$)).ti,ab,ot. 15,634
87 (utilities or disutili$).ti,ab,ot. 9413
88 (Functional Assessment of Cancer Therapy$ or FACT-G).ti,ab,ot. 3041
89 (QLQ-C30 or QLQ-C-30 or EORTC QLQ$ or "European Organization for Research and Treatment of Cancer Quality of Life Questionnaire" or "EORTC Quality of Life Questionnaire").ti,ab,ot. 6633
90 or/65-89342,017
91 33 and 90132
92 91 not 6468

EMBASE 1974 to week 7 2023

1 f?ecal immunochemical test.mp. 2444
2 f?ecal occult blood.mp. 7022
3 f?ecal h?emoglobin.mp. 470
4 ((immunochromatographic or immuno-chromatographic or immunochem$ or immuno-chem$ or immunohisto-chem$ or immuno-histochem$ or immunol$ or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) adj4 (f?ecal or f?eces or stool or stools or FIT)).mp. 6393
5 (iFOBT or qFIT).mp. 411
6 or/1-5 12,159
7 F?ecal h?emoglobin.ti,ab,ot,hw. 454
8 H?emoccult.ti,ab,ot,hw. 989
9 FOBT.ti,ab,ot,hw. 2830
10 7 or 8 or 9 4155
11 (f?ecal or f?eces or stool or stools).ti,ab,ot,hw. 287,846
12 occult blood/or occult blood.ti,ab,ot,hw. 18,594
13 (test$ or measur$ or screen$ or exam$).ti,ab,ot,hw. 14,258,724
14 11 and 12 and 13 10,644
15 6 or 10 or 14 15,206
16 exp colorectal cancer/or colon cancer/or rectum cancer/ 377,137
17 exp cecum tumor/ 7397
18 ((colorect$ or rectal$ or rectum$ or colon$ or sigma$ or sigmo$ or rectosigm$ or bowel$ or anal or anus) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot,hw. 521,146
19 CRC.ti,ab,ot. 73,761
20 ((cecum or cecal or caecum or caecal or il?eoc?ecal or il?eoc?ecum) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 3624
21 (large intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 1821
22 (lower intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).ti,ab,ot. 44
23 16 or 17 or 18 or 19 or 20 or 21 or 22 537,004
24 15 and 23 10,244
25 (FOB gold$ or FOBgold$ or SENTiFIT).mp. 115
26 (JACK-arc$ or JACKarc$ or HM-JACK$ or HM JACK$ or HMJACK$).mp. 78
27 (OC Sensor$ or OC-Sensor$ or OCSensor$ or Ceres).mp. 798
28 (OC Pledia$ or OC-Pledia$ or OCPledia or OC-iO). mp.0
29 (NS-Prime or NSPrime or NS-Plus).mp. 79
30 (POC FIT QRG or POCFITQRG).mp. 0
31 (immundiagnostik or IDK or turbifit or turbitube).mp. 431
32 quikread.mp. 56
33 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 11,170
34 health-economics/ 35,319

35 exp economic-evaluation/ 349,167
36 exp health-care-cost/ 332,717
37 exp pharmacoeconomics/ 227,487
38 34 or 35 or 36 or 37 739,026
39 (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab. 1,357,302
40 (expenditure$ not energy).ti,ab. 49,546
41 (value adj2 money).ti,ab. 2941
42 budget$.ti,ab. 46,255
43 39 or 40 or 41 or 42 1,401,000
44 38 or 43 1,760,393
45 letter.pt .1,291,591
46 editorial.pt. 764,786
47 note.pt. 926,965
48 45 or 46 or 47 2,983,342
49 44 not 48 1,627,860
50 (metabolic adj cost).ti,ab. 1840
51 ((energy or oxygen) adj cost).ti,ab. 5006
52 ((energy or oxygen) adj expenditure).ti,ab. 36,873
53 50 or 51 or 52 42,520
54 49 not 53 1,619,172
55 exp animal/ 30,254,802
56 exp animal-experiment/ 3,050,020
57 nonhuman/ 7,373,530
58 (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh. 6,395,872
59 55 or 56 or 57 or 58 32,471,017
60 exp human/ 25,120,688
61 exp human-experiment/ 638,565
62 60 or 61 25,123,083
63 59 not (59 and 62) 7,349,104
64 54 not 63 1,465,032
65 33 and 64 1817
66 limit 65 to yr="2016 -Current" 704
67 quality adjusted life year/or quality of life index/ 37,657
68 Short Form 12/or Short Form 20/or Short Form 36/or Short Form 8/ 47,910
69 "International Classification of Functioning, Disability and Health"/or "ferrans and powers quality of life index"/or "gastrointestinal quality of life index"/ 4281
70 (sf36 or sf 36 or sf-36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or short-form thirty six or short form thirty six or short form thirtysix or short form thirty six).ti,ab,ot. 49,160
71 (sf6 or sf 6 or sf-6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).ti,ab,ot. 2885
72 (sf12 or sf 12 or sf-12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).ti,ab,ot. 11,999
73 (sf6D or sf 6D or sf-6D or short form 6D or shortform 6D or sf six D or sfsixD or shortform six D or short form six D).ti,ab,ot. 1827
74 (sf20 or sf 20 or sf-20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).ti,ab,ot. 512
75 (sf8 or sf 8 or sf-8 or short form 8 or shortform 8 or sf eight or sfeight or shortform eight or short form eight).ti,ab,ot. 1186
76 "health related quality of life".ti,ab,ot. 82,918
77 (Quality adjusted life or Quality-adjusted-life).ti,ab,ot. 26,027
78 "assessment of quality of life".ti,ab,ot. 3479
79 (euroqol or euro qol or eq5d or eq 5d).ti,ab,ot. 29,306
80 (hql or hrql or hqol or h qol or hrqol or hr qol).ti,ab,ot. 44,711

81 (hye or hyes).ti,ab,ot. 162
82 health$ year$ equivalent$.ti,ab,ot. 41
83 (hui or hui1 or hui2 or hui3 or hui4 or hui-4 or hui-1 or hui-2 or hui-3).ti,ab,ot. 3863
84 (quality time or qwb or "quality of well being" or "quality of wellbeing" or "index of wellbeing" or index of well being).ti,ab,ot,hw. 1521
85 (Disability adjusted life or Disability-adjusted life or health adjusted life or health-adjusted life or "years of healthy life" or healthy years equivalent or "years of potential life lost" or "years of health life lost").ti,ab,ot. 7121
86 (QALY$ or DALY$ or HALY$ or YHL or HYES or YPLL or YHLL or qald$ or qale$ or qtime$ or AQoL$).ti,ab,ot. 33,625
87 (timetradeoff or time tradeoff or time trade-off or time trade off or TTO or Standard gamble$ or "willingness to pay").ti,ab,ot. 16,459
88 15d.ti,ab,ot. 2918
89 (HSUV$ or health state$ value$ or health state$ preference$ or HSPV$).ti,ab,ot. 760
90 (utilit$ adj3 ("quality of life" or valu$ or scor$ or measur$ or health or life or estimat$ or elicit$ or disease$)).ti,ab,ot. 25,437
91 (utilities or disutili$).ti,ab,ot. 15,477
92 (Functional Assessment of Cancer Therapy$ or FACT-G).ti,ab,ot. 5504
93 (QLQ-C30 or QLQ-C-30 or EORTC QLQ$ or "European Organization for Research and Treatment of Cancer Quality of Life Questionnaire" or "EORTC Quality of Life Questionnaire").ti,ab,ot. 13,823
94 or/67-93 259,151
95 33 and 94 239
96 95 not 66 112

EconLit 1886 to 9 February 2023

1 f?ecal immunochemical test.mp. 0
2 f?ecal occult blood.mp. 12
3 f?ecal h?emoglobin.mp. 0
4 ((immunochromatographic or immuno-chromatographic or immunochem$ or immuno-chem$ or immunohistochem$ or immuno-histochem$ or immunol$ or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) adj4 (f?ecal or f?eces or stool or stools or FIT)).mp.5
5 (iFOBT or qFIT).mp. 3
6 or/1-5 15
7 F?ecal h?emoglobin.mp. 0
8 H?emoccult.mp. 2
9 FOBT.mp. 7
10 7 or 8 or 9 9
11 (f?ecal or f?eces or stool or stools).mp. 104
12 occult blood.mp.12
13 (test$ or measur$ or screen$ or exam$).mp. 516,772
14 11 and 12 and 13 2
15 6 or 10 or 14 16
16 ((colorect$ or rectal$ or rectum$ or colon$ or sigma$ or sigmo$ or rectosigm$ or bowel$ or anal or anus) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).mp. 151
17 CRC.mp. 93
18 ((cecum or cecal or caecum or caecal or il?eoc?ecal or il?eoc?ecum) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).mp. 0
19 (large intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).mp. 0
20 (lower intestin$ adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or adenom$ or lesion$)).mp. 0

21 16 or 17 or 18 or 19 or 20 214
22 15 and 21 15
23 (FOB gold$ or FOBgold$ or SENTiFIT).mp. 0
24 (JACK-arc$ or JACKarc$ or HM-JACK$ or HM JACK$ or HMJACK$).mp. 0
25 (OC Sensor$ or OC-Sensor$ or OCSensor$ or Ceres).mp. 17
26 (OC Pledia$ or OC-Pledia$ or OCPledia or OC-iO).mp. 0
27 (NS-Prime or NSPrime or NS-Plus).mp. 0
28 (POC FIT QRG or POCFITQRG).mp. 0
29 (immundiagnostik or IDK or turbifit or turbitube).mp. 0
30 quikread.mp.0
31 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 32
32 15 or 31 33

Cochrane search – already run as part of clinical SLR in December 2022

Re-run to find new records added between December 2022 and February 2023.

Date run: 23 February 2023 16:18:46

ID Search Hits

#1 (fecal immunochemical test* or faecal immunochemical test*):ti,ab,kw (Word variations have been searched) 500
#2 (fecal occult blood or faecal occult blood):ti,ab,kw (Word variations have been searched) 1091
#3 (fecal hemoglobin or faecal hemoglobin or fecal haemoglobin or faecal haemoglobin):ti,ab,kw (Word variations have been searched) 299
#4 ((immunochromatographic or immuno-chromatographic or immunochem* or immuno-chem* or immunohistochem* or immuno-histochem* or immunol* or immunoassay or immuno* assay or immunoturbidimetric or immunosorbent or elisa) near/4 (fecal or faecal or feces or faeces or stool or stools or FIT)):ti,ab,kw (Word variations have been searched) 1047
#5 (iFOBT or qFIT):ti,ab,kw (Word variations have been searched) 37
#6 (Hemoccult or haemoccult):ti,ab,kw (Word variations have been searched) 129
#7 (FOBT):ti,ab,kw (Word variations have been searched) 412
#8 ((fecal or feces or faecal or faeces or stool or stools)):ti,ab,kw AND (occult blood):ti,ab,kw AND (test* or measur* or screen* or exam*):ti,ab,kw (Word variations have been searched) 1158
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 2202
#10 MeSH descriptor: [Colorectal Neoplasms] explode all trees 10,857
#11 MeSH descriptor: [Cecal Neoplasms] explode all trees 27
#12 ((colorect* or rectal* or rectum* or colon* or sigma* or sigmo* or rectosigm* or bowel* or anal or anus) near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 26,712
#13 (CRC):ti,ab,kw (Word variations have been searched) 5145
#14 ((cecum or cecal or caecum or caecal or ileocecal or ileocecum or ileocaecal or ileocaecum) near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 247
#15 (large intestin* near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 175
#16 (lower intestin* near/3 (cancer* or neoplas* or oncolog* or malignan* or tumo* or carcinoma* or adenocarcinoma* or sarcoma* or adenom* or lesion*)):ti,ab,kw (Word variations have been searched) 183
#17 #10 or #11 or #12 or #13 or #14 or #15 or #16 27,565
#18 #9 and #17 with Cochrane Library publication date Between Jan 2022 and Dec 2022 106
#19 (FOB gold* or FOBgold* or SENTiFIT):ti,ab,kw OR (JACK-arc* or JACKarc* or HM-JACK* or HM JACK* or HM-JACK*):ti,ab,kw OR (OC Sensor* or OC-Sensor* or OCSensor* or Ceres or OC Pledia* or OC-Pledia* or OCPle-

dia or OC-iO):ti,ab,kw OR (POC FIT QRG or POCFITQRG or immundiagnostik or IDK or turbifit or turbitube or quikread):ti,ab,kw OR (NS-Prime or NSPrime or NS-Plus):ti,ab,kw (Word variations have been searched) 177
#20 #18 or #19 with Cochrane Library publication date Between Jan 2016 and Dec 2022 229
#21 #18 or #19 with Cochrane Library publication date Between Dec 2022 and Feb 2023 5

Cost-Effectiveness Analysis (CEA) Registry (internet)

https://research.tufts-nemc.org/cear4/Home.aspx

Date searched: 23 February 2023

Records found: 195

Search term (basic search)	Records found
Colonoscopy	50
Computed tomographic colonography	2
CT colonography	2
Coloscopy	0
Sigmoidoscopy	4
Magnetic resonance imaging	64
MRI	60
CT scan	12
CAT scan	0
Total	195

Appendix 2 Statistical methods and additional results

Conversion of sensitivity and specificity to true positive, true negative, false positive, false negative

When the absolute number of diagnostic counts (true positive, true negative, false negative, false positive) were not reported in a study, but data for the total number of patients, the total number of positive cases, sensitivity and specificity were available, the count data were calculated using the following equations:

TP = sensitivity × number of positive cases

FN = (1 – sensitivity) × number of positive cases

FP = (1 – specificity) × (total number – number of positive × cases)

TN = specificity × (total number – number of positive cases)

Statistical methods for the evidence synthesis

The statistical model is briefly described following the notation in Jones *et al.*[36] True disease status is assumed to be known through the application of a perfect gold-standard test. Populations without and with CRC are indexed by j = 1, 2 respectively. Each study, i, reports estimates of sensitivity and specificity, or directly reports count data, at Ti distinct thresholds. Test results above a given threshold are considered positive.

The observed count data are modelled using multinomial likelihoods, reparametrised as conditional binomial distributions for computational convenience. The model assumes that some transformation, $g()$, of the continuous test results in population j of study i has a logistic distribution with mean μij and scale parameter σij. In our analyses we prespecify $g()$=log$e()$. Jones *et al.* describe the more flexible (but computationally intensive) case where $g()$ is in the set of Box-Cox transformation, defined by a parameter that is estimated alongside other model parameters.

Within-study model

The probability of a positive test result at threshold Cit in population j of study i is:

$$\text{logit}(pr_{ijt}) = \frac{(\mu_{ij} - g(C_{it}))}{\sigma_{ij}} \quad (2)$$

For j = 1 we have pri_1t (the false-positive rate, FPR = 1 – sensitivity) and for j=2 we have pri_2t (the true-positive rate, TPR=sensitivity).

TABLE 35 Parameters used to inform priors for syntheses with < 5 studies

	CRC outcomes (S = 28)			AA outcomes (S = 9)			IBD outcomes (S = 9)		
Parameter	Mean	SD	Truncation	Mean	SD	Truncation	Mean	SD	Truncation
$\tau_{\mu 1}$	0.2698	0.1543	1.703	0.2859	0.3432	2.478	0.4532	0.3561	3.027
$\tau_{\mu 2}$	−0.8489	0.2995	0.67	−0.4517	0.8583	2.174	0.3827	0.8695	3.92
$\tau_{\sigma 1}$	−0.8863	0.1559	0.538	−0.5228	0.415	1.247	−0.4273	0.4077	1.345
$\tau_{\sigma 2}$	−1.4215	0.2693	0.368	0.152	0.5711	3.079	−0.8792	0.9243	1.339

Between-study model

The study-specific location (μij) and scale (σij) parameters are modelled as random effects. Across studies, μij has mean mij and standard deviation (SD) $\tau\mu j$, while log(σij) has mean $m\sigma j$ and SD $\tau\sigma j$.

Different options for the correlation structure between these four sets of random effects are described in Jones *et al.*: (1) full correlation matrix, (2) structured correlations matrix, and (3) independence model with the four sets of random effects assumed to be independent of each other. Models with a structured correlation matrix and assuming independence were explored. Including additional parameters for between-study correlations did not improve the model fit according to the DIC (see *Table 36*), and therefore the simpler independence model was used for all main analyses.

Prior distributions are required for the four hyperparameters: $m\mu j$, $\tau\mu j$, $m\sigma j$, $\tau\sigma j$ for j = 1, 2. For analyses with sufficient sample data, standard reference priors as used in Jones *et al.* were used. Normal (0, 102) prior distribution were given to all means ($m\mu j$, $m\sigma j$), and Uniform (0, 5) prior for between-study SDs ($\tau\mu j$, $\tau\sigma j$).

For analyses with small numbers of contributing studies, informative priors were used for the between-study SDs. These were informed by fitting log-normal distributions to posterior samples from the analyses of all test types together. This was considered to be a conservative option. A truncation was also applied, based on the 95th centile of the posterior distribution. Parameter values for all analyses are provided in *Table 35*.

Model fit

Model fit for all analyses is shown in *Table 36*. Differences in DIC across the two correlations structures (structured and independent) were minimal. Because including additional parameters for between-study correlations did not improve the model fit according to the DIC, the simpler model structure was preferred, and all analyses presented in the report use the model of Jones *et al.*[36] with the four sets of random effects assumed to be independent of each other (independence model).

TABLE 36 Meta-analysis model fit statistics

Tests in analysis	Populations	Studies	Correlation structure	Model fit		
					pD	DIC
CRC outcomes						
All	All	28	S	6718.40	72.91	6791.31
	All	28	I	6711.25	71.78	6783.03
	1, 2, 3	13	I	1780.45	31.63	1812.08
	1	8	I	1541.81	21.73	1563.54
	2	5	I	173.87	NaN	173.87
	2	5	I	175.77	10.26	186.02
	3	3	I	107.51	NaN	107.51
HM JACKArc	All	16	S	5297.88	41.45	5339.32
	All	16	I	5296.11	40.46	5336.56
	1, 2, 3	9	I	648.54	19.29	667.83
	1	5	I	495.25	12.69	507.93
	2	4	I	143.50	8.11	151.60
	3	2	I	39.00	Nan	39.00

TABLE 36 Meta-analysis model fit statistics (*continued*)

Tests in analysis	Populations	Studies	Correlation structure	Model fit		
					pD	DIC
OC Sensor	All	11	S	1395.86	28.05	1423.92
	All	11	I	1394.99	28.71	1423.70
	1, 2, 3	4	I	1112.37	11.18	1123.55
	1	3	I	1044.86	9.35	1054.21
FOB Gold	All	3	I	114.56	NaN	Nan
AA outcomes						
All tests	All	9	I	308.81	23.41	332.21
HM-JACKarc	All	6	I	240.87	13.86	254.74
OC Sensor	All	2	I	38.67	6.02	44.69
IBD outcomes						
All tests	All	9	I	286.38	23.97	310.36
HM-JACKarc	All	6	I	220.37	14.64	235.01
OC Sensor	All	2	I	38.27	6.12	44.39
DUAL FIT						
All tests	All	4	I	63.87	8.03	71.90

I, independence model; S, structured correlation matrix.

Methods for pooling prevalence data from the Evidence Assessment Group clinical review

Data on prevalence were pooled using a Bayesian Markov chain Monte Carlo approach based on a random-effects meta-analysis.[127] The prevalences for the overall population and by population type (NG12 high-/medium- and DG30 low-risk groups) were analysed separately. The random-effects model allowed for heterogeneity in prevalence across studies within each population type. The model assumed that the log-odds of study-specific prevalence are from a normal distribution, where the mean represents the overall population prevalence, and the variance represents heterogeneity among the studies.

Vague priors were assumed for all model parameters for the meta-analysis of CRC as the outcome using population type 1 studies because this was the analysis with the greatest number of included studies that allowed for an appropriate estimation of the heterogeneity parameter. For all other meta-analyses, a vague prior was assumed for the mean and an informative prior generated using the posterior distribution of the heterogeneity parameter for the meta-analysis of CRC as the outcome using population type 1 studies was assumed for the heterogeneity parameter due to limited studies to inform the estimation of the heterogeneity parameter.

All analyses were conducted in the freely available software package WinBUGS and R, using the R2Winbugs interface package.[128] Convergence to the target posterior distributions was assessed using the Gelman–Rubin statistic.[39] The chains converged within 50,000 iterations so a burn-in of 50,000 iterations was used. A further 30,000 iterations of the Markov chain was retained to estimate parameters using one chain and thinning every five iterations. The absolute goodness of fit was checked by comparing the number of data points (which is the number of included studies) with the total residual deviance.

Appendix 3 Quality assessment of diagnostic test accuracy studies version 2[34] scoring scheme and scores with reasons for all studies

Scoring scheme

Domain 1: Patient selection

Was a consecutive or random sample of patients enrolled?
- Score yes if states consecutive or random
- Score no if states another method of patient sampling/selection
- Score unclear if unclear

Was a case–control design avoided?
- Score yes if not case–control
- Score no if case control
- Score unclear if unclear

(there should be no case–control studies in the included studies, but please double check)

Did the study avoid inappropriate exclusions?
Score yes if the study only excluded bypass symptom patients
Score no if the study has made inappropriate exclusions for example on basis of having had a colonoscopy, taking certain medications, having blood disorders, or on the basis of eventually being diagnosed with IBD (list not exhaustive)
Score unclear if it is unclear

Risk of bias summary score: Could the selection of patients have introduced bias?
Low/high/unclear
THIS IS A SUMMARY SCORE BASED ON ANSWERS TO QUESTIONS 1 TO 3
Score Low if all domains are Yes
Score High if one or more domain is No
Anything in between score Unclear

Applicability summary score: Is there concern that the included patients and settings do not match the review question?
Score low if the study selected all patients presenting to primary care with symptoms of CRC as listed in DG30 and NG12. If the study recruited a wider population, i.e. patients who do not meet these criteria, please state unclear risk (wider)
Score High if the study missed some of the primary care patients, for example if only those referred to colonoscopy were recruited (unless all primary care NG12/DG30 are referred to secondary care)
Low/high/unclear

Domain 2: Index test(s)

Were the index test results interpreted without knowledge of the results of the reference standard?
o Score yes if index test was interpreted blind to the reference standard or the index test was clearly interpreted before the reference standard was known, for example FIT before colonoscopy
o Score no if results of reference standard were already known for example FIT done after colonoscopy
o Score unclear if unclear

If a threshold was used, was it prespecified?
o Score yes if prespecified cut-off values were used (validation study), for example one or a range of cut-offs reported such as 10, 20, 50, 100 ug/g and these were not chosen on the basis of having the highest accuracy
o Score no if cut-off values were fitted to the data (derivation study), for example cut-off with highest accuracy reported
o Score unclear if unclear
NB if study reports both the highest precision cut-off, and several other 'round number' cut-offs, score yes/no

Risk of bias summary score: Could the conduct or interpretation of the index test have introduced bias?
Low/high/unclear
THIS IS A SUMMARY SCORE BASED ON ANSWERS TO QUESTIONS 1 TO 2
Score Low if all domains are Yes
Score High if one or more domain is No
Anything in between score Unclear

Applicability summary score: Is there concern that the index test, its conduct, or interpretation differ from the review question?
Low/high/unclear

Domain 1: Patient selection

Domain 3: Reference standard

Is the reference standard likely to correctly classify the target condition?
Please note limitations/test type against score
Score Yes if all patients received either colonoscopy or CT colonography (CTC)
Score No if the reference standard was not full colonic imaging (see yes criteria) for all patients
Score Unclear if it is unclear

Were the reference standard results interpreted without knowledge of the results of the index test?
In the case of tiered testing, this is likely not to be the case
o Score yes if the reference standard was interpreted blind to the index test or the reference standard was clearly interpreted before the index test was known
o Score no if the results of the index test were known, for example where patients were referred on the basis of a FIT result
o Score unclear if unclear

Risk of bias summary score: Could the reference standard, its conduct, or its interpretation have introduced bias?
Low/high/unclear
(THIS IS A SUMMARY SCORE BASED ON ANSWERS TO QUESTIONS 1 TO 2)

Applicability summary score: Is there concern that the target condition as defined by the reference standard does not match the review question?
Score low risk if the target condition is CRC
Score high risk if the target is not just CRC
Score unclear if the target condition is unclear
NB: all studies should score low risk
NB: we are not scoring for AA and IBD
Low/high/unclear
The reference standard may be free of bias, but the target condition that it defines may differ from the target condition specified in the review question

Domain 4: Flow and timing

Was there an appropriate interval between index test(s) and reference standard?
Score low risk if all patients received colonoscopy and this was conducted within 3 months of the index test, or if some patients received records follow-up, this should be for a minimum of 3 months
Score high risk if colonoscopies were not conducted within 3 months, or follow-up is for < 3 months but more than 12 months
Score unclear if the time intervals were unclear
NB: likely most will not report time interval for colonoscopy
Yes/no/unclear

Did all patients receive a reference standard?
Score yes if all patients got a reference standard, even if these were different (see next question)
Score no if a partial verification reference standard: only some participants get any reference standard, for example those who test negative at FIT don't get followed up or any further tests (these studies should be excluded)
Score unclear if it is unclear who received the reference standard

Did patients receive the same reference standard?
The following score 'no':
Complete index test-dependent differential verification reference standard: participants get a different reference standard according to the index test result, for example FIT positive get colonoscopy, FIT negative get records follow-up
Differential verification dependent on other known or unknown factors: participants get a different reference standard according to some known or unknown factors, for example those with clinical signs or symptoms proceed to colonoscopy regardless of FIT, whilst reminder get records follow-up
The following score 'yes':
All received the same reference standard, for example all get colonoscopy

Were all patients included in the analysis?
Score yes if all patients who were recruited/enrolled into the study were included in the analysis or if an acceptable explanation (i.e. missing at random) is provided for any discrepancy
Score no if there are participants excluded from the analysis and no/concerning explanation is given for any discrepancy
Score unclear if insufficient information is given to assess whether any patients were excluded from the analysis

Risk of bias summary score: Could the patient flow have introduced bias?
Low/high/unclear
(THIS IS A SUMMARY SCORE BASED ON ANSWERS TO QUESTIONS 1 TO 4)
Score Low if all domains are Yes
Score High if one or more domain is No
Anything in between score Unclear

Summary of scores

TABLE 37 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability

		RoB items				Applicability items		
	Analyses[a]	RoB: patient selection	RoB: index test	RoB: reference standard	RoB: patient flow	Applicability risk: patients and setting	Applicability risk: index test	Applicability risk: reference standard
Benton 2022[44]	2	High	Low	Unclear	Low	High	Low	Low
Chapman 2021[46]	4	High	Low	Unclear	Unclear	High	Low	Low
Cunin 2020[71]	4; anaemia	Low	Low	High	High	High	Low	Low
D'Souza 2020[18]	1, 2, 3	High	Low	Unclear	Low	Unclear	Low	Low
D'Souza 2021[72]	2; anaemia	Unclear	Low	Unclear	Unclear	High	Low	Low
D'Souza 2021,[17] D'Souza 2021[73]	2; age	Unclear	Low	Unclear	Unclear	High	Low	Low
Elbeltagi 2022[47]	4	Unclear	Low	Unclear	High	High	Low	Low
Farrugia 2020[74]	2	High	Low	Unclear	Unclear	High	Low	Low
Faux 2022[48]	4	High	Low	High	High	High	Low	Low
Gerrard 2023[75]	1±; single and dual FIT; anaemia	High	Low	High	High	High	Low	Low
Godber 2016[50]	4	High	Low	Unclear	Low	High	Low	Low
Johnstone 2022[52]	1; anaemia	Low	Low	Unclear	High	Low	Low	Low
MacDonald 2022[56]	1±	Low	Low	High	High	Unclear	Low	Low
Mowat 2021 and 2019[60,61]	1	Unclear	Low	High	High	Unclear	Low	Low
Nicholson 2018[63a]	4	Low	Low	High	High	High	High	Low
Nicholson 2020[27]	4; sex	Low	Low	High	High	High	High	Low
Tang 2022[65]	4; anaemia	High	Low	Unclear	Unclear	High	Low	Low
Turvill 2021[81]	4; anaemia; age, medications, sex	High	High/ Low	High	High	High	High/Low	Low
Turvill 2018[81]	2; single and dual FIT	High	High	High	High	High	High	Low
Withrow 2022[66b]	3; anaemia; age	High	Low	High	High	High	Low	Low

RoB, risk of bias.
a Numbers relate to population-type analyses.
b Nicholson 2020 and Withrow 2022 include some of the same patients, but it was not clear if the same methodology was used in both studies to select patients and conduct follow-up, so scores are provided for each study based on the information given for that study.

TABLE 38 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability

		RoB items				Applicability items		
	Analyses[a]	Patient selection	Index test	Reference standard	Patient flow	Patients and setting	Index test	Reference standard
Archer 2022[41]	4	High	Low	Unclear	Unclear	High	Low	Low
Ayling 2019[68]	Anaemia	High	Low	Low	High	High	Low	Low
Ball 2022[43] (personal communication)	3, sex	High	Low	High	High	High	Low	Low
Ball 2022[43]	4	Low	Low	High	High	Low	Low	Low
Benton 2022[44]	2	High	Low	Unclear	Low	High	Low	Low
Bujanda 2018[70]	Aspirin	High	Low	Low	Unclear	High	Low	Low
Cama 2022[45]	1	High	Low	High	High	High	Low	Low
Chapman 2021[46]	4	High	Low	Unclear	Unclear	High	Low	Low
Crooks 2023,[30] Bailey 2021[42]	1	High	Low	High	High	High	Low	Low
Georgiou Delisle 2022[49]	1	High	Low	Unclear	Low	High	Low	Low
Hunt 2022[51]	Dual FIT	High	Low	Unclear	Unclear	High	Low	Low
Juul 2018[54]	4, anaemia	High	Low	High	High	High	Low	Low
Laszlo 2021[55]	4	High	Low	Low	High	High	Low	Low
Maclean 2021[57]	4	Low	Low	High	High	High	Low	Low
Morales-Arraez 2018[76]	Anaemia	High	Low	Unclear	Unclear	High	Low	Low
Mowat 2016[62]	4	High	High (LoD); low (10 µg/g)	Low	High	High	Low	Low
Pin Vieto 2020[64]	4	Unclear	Low	High	High	Unclear	Low	Low
Rodriguez-Alonso 2018,[77] Rodriguez-Alonso 2019[78]	Anaemia, PPIs	High	Unclear	Unclear	Unclear	High	Low	Low

LoD, limit of detection; RoB, risk of bias.
a Numbers relate to population-type analyses.

TABLE 39 FOB-Gold studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	Analyses[a]	RoB: patient selection	RoB: index test	RoB: reference standard	RoB: patient flow	Applicability risk: patients and setting	Applicability risk: index test	Applicability risk: reference standard
Benton 2022[44]	2	High	Low	Unclear	Low	High	Low	Low
MacLean 2022[59]	2	High	Low	Low	High	High	Low	Low
Jordaan 2023[83]	4	Unclear	Low	High	High	High	Low	Low

a Numbers relate to population-type analyses.

TABLE 40 QuikRead go studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	RoB: patient selection	RoB: index test	RoB: reference standard	RoB: patient flow	Applicability risk: patients and setting	Applicability risk: index test	Applicability risk: reference standard
Maclean 2021[58]	High	Low	Low	High	High	Low	Low
Tsapournas 2020[80]	High	Low	Low	High	High	Low	Low

RoB, risk of bias.

TABLE 41 NS-Prime study: SCHARR's assessment of risk of bias and applicability

	Analyses[a]	RoB: patient selection	RoB: index test	RoB: reference standard	RoB: patient flow	Applicability risk: patients and setting	Applicability risk: index test	Applicability risk: reference standard
Benton 2022[44]	2	High	Low	Unclear	Low	High	Low	Low

RoB, risk of bias.
a Numbers relate to population-type analyses.

TABLE 42 IDK studies: SCHARR's assessment of risk of bias and applicability

	RoB: patient selection	RoB: index test	RoB: reference standard	RoB: patient flow	Applicability risk: patients and setting	Applicability risk: index test	Applicability risk: reference standard
Sieg 1999	High	Hb: low Hb/Hp: high	Low	Unclear	High	High	Low

RoB, risk of bias.

Reasons for scores

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Benton 2022[44]	No (participation voluntary)	Yes	No (subgroup, participation voluntary)	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes (some missing data, explanations given, no reason to assume not random)	Low
Chapman 2021[46]	Yes	Yes	No	High	High	Yes	Yes	Low	Low	Yes (follow-up, implied all colonoscopy or imaging)	Unclear	Unclear	Low	Unclear	Yes	Unclear – because some may have received more than just colonoscopy	Yes (missing data lack of follow-up)	Unclear
Cunin 2020[71]	Yes	Yes	Yes	Low	High	Yes	Yes	Low	High	No	Unclear	High	Low	Unclear	Yes	No	Yes	High
D'Souza 2020[18]	Yes	Yes	No	High	Unclear – some may be missed	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Low
D'Souza 2021,[72] D'Souza 2021[17]	Unclear	Yes	Yes	Unclear	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Unclear

continued

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
D'Souza 2021[73]	Unclear	Yes	No (colonoscopy)	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Unclear
Elbeltagi 2022[47]	Yes – states 'all'	Yes	Unclear	Unclear	High – patients on 2WW who had COL/CTC so mix of pts	Yes	Yes	Low	Low	Yes (colonoscopy/ CTC)	Unclear	Unclear	Low	Unclear	Yes	No, unclear how determined	No, only those who had col/CTC	High
Farrugia 2020[74]	Unclear	Yes	No	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Unclear
Faux 2022[48]	Unclear	Yes	No, rectal bleeding and anaemia excluded	High	High	Yes	Yes	Low	Low	No	Unclear	High	Low	Unclear	Yes	No, dependent on FIT and presenting symptom	Yes	High
Gerrard 2023[75]	Yes, states consecutive	Yes	No, IDA on its own was not a referral criterion; only those with colonoscopy/CTC	High	High	Yes	Yes	Low	Low	No	Unclear, but secondary care investigation received not based on FIT	High	Low	Yes – states median time to diagnosis	Yes	Yes all either col or CTC	No – some excluded as triaged away from imaging	High

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Godber 2016[50]	Yes	Yes	No	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Low
Johnstone 2022[52]	Yes (database)	Yes	Yes	Low	Low	Yes	Yes	Low	Low	No	Unclear	Unclear	Low	Unclear	Yes (some imaging, some follow-up)	No (some imaging, some follow-up)	Yes	High
Johnstone 2022[53]	Na	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MacDonald 2022[56]	Yes – states consecutive	Yes	Yes	Low	Unclear risk (wider and narrower)	Yes	Yes	Low	Low	No – not all had colonoscopy/CTC, unclear what 1113 had	Unclear – FIT used to define test in secondary care but unclear if interpreted blind	High	Low	High – records follow-up for > 2 years could allow CRC to emerge; unclear how long between FIT and imaging/ secondary care appointment	Yes	No – Differential verification dependent on other known or unknown factors, as 1113 unknown	Yes	High

continued

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Mowat 2021[61] and 2019[60]	Unclear – uptake of FIT by GPs was slow, unclear who was missed in early adoption period	Yes	Yes (CRC results); No (IBD/HRA results)	Unclear	Unclear (it is not entirely clear that ALL NG12 and DG30 patients were included, but it is likely the case. Though use of FIT was rolled out slowly, so not clear if this was due to more GPs adopting, or gps becoming more confident to use in all symptoms. If the latter, may skew distribution towards more or less 'serious' symptoms)	Yes	Yes	Low	Low	No – some did not get full imaging	No	High	Low	Unclear – time interval for colonoscopy not stated	Yes	No – differential verification dependent on other known or unknown factors: some patients with FIT< 10 referred for COL, some with FIT> 10 not referred	Yes	High

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Nicholson 2018[63]	Yes, states consecutive	Yes	Yes	Low	High, only DG30	Yes	Yes	Low	High – some patients had two test results, and if either was positive the test was judged as positive	No – records follow up and imaging	Unclear – not stated	High	Low – states adenocarcinoma, but no other types of CRC were reported in the 'significant bowel pathology' category	No – follow-up was for 21 and 23 months that may have allowed for the emergency of CRC since the index test	Yes	No – unclear if completely dependent on index test	Unclear – not clear if more patients were offered FIT but did not complete it	High

continued

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Nicholson 2020[27a]	Low	Yes	Yes	Low	High – only DG30	Yes	Yes	Low	High – some patients had two test results, and if either was positive the test was judged as positive	No – records follow-up and imaging	Unclear – not stated	High	Low	Yes – minimum of 6 months, analyses show that longer follow-up did not significantly alter the sens/spec	Yes	No – unclear if completely dependent on index test	Unclear – some excluded due to not long enough follow-up but this is unlikely to introduce bias, but not clear if patients missing for other reasons also	High
Tang 2022[65]	Yes	Yes	No (needed both fit and colonoscopy)	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Unclear

TABLE 43 HM-JACKarc studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Turvill 2021[82]	No – convenience sample	Yes	Yes – all patients recruited accounted for and reasons for exclusions not likely to skew selection	High	High – some patients outside of NG12 criteria	Yes	Yes/No – some are derived optimal cut-off, some are round numbers	High/low	Low	No – some did not get full imaging	Yes – states blind	High	Low	Unclear – not reported	Yes	No – Differential verification dependent on other known or unknown factors	Yes – some missing but unlikely to skew results	High
Turvill 2018[81]	No (consent to study, convenience series)	Yes	No (colonoscopy or imaging)	High	High	Yes	No	High	High	No	Yes	High	Low	Unclear	Yes	No	Yes (reasons for exclusion given, no reason to assume not random)	High
Withrow 2022[66a]	Yes – states all	Yes	No – only patients with five most common core blood tests available were included	High	High	Yes	Yes	Low	Low	No – follow-up	Unclear	High	Low	Unclear	Yes (follow-up)	No (some no imaging)	No – unclear why some missing	High

COL, colonoscopy.

a Nicholson 2020 and Withrow 2022 include some of the same patients, but it was not clear if the same methodology was used in both studies to select patients and conduct follow-up, so scores are provided for each study based on the information given for that study.

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Archer 2022	Yes (all referred)	Yes	No – excluded those without colonoscopy or FIT	High	High – those referred to 2WW, + colonoscopy + FIT	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	No	Unclear
Ayling 2019	Unclear	Yes	No, only colonoscopy/CT included	High	High	Yes	Yes	Low	Low	Yes (assume CT scan is CTC)	Yes ('faecal sample for immunological measurement of haemoglobin and their blood count parameters were analysed using an artificial intelligence (AI) flagging system; these results were not made available for patient management')	Low	Low	Unclear	Yes	No	No	high

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Ball 2022 (personal communication)	Yes	Yes	Yes	High	High	Yes	Yes	Low	Low	No – records follow-up	Unclear	High	High	No – unclear if colonoscopy done within 3 months; records follow up was at least 18 months, which could allow for CRC to emerge after the index test	Yes	No – unclear if complete or partial differential bias	Yes	High
Ball 2022	Yes	Yes	Yes	Low	Low	Yes	Yes	Low	Low	No – records follow-up	Unclear	High	High	No – unclear if colonoscopy done within 3 months; records follow up was at least 18 months, which could allow for CRC to emerge after the index test	Yes	No – unclear if complete or partial differential bias	Yes	High

continued

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Bujanda 2018	Yes	Yes	No (all colonoscopy; occasional aspirin users)	High	High	Yes	Yes	Low	Low	Yes	Yes	Low	Low	Unclear	Yes	Yes	Yes	unclear
Cama 2022	Yes – states consecutive	Yes	No – excluded IDA and rectal bleeding	High	High	Yes	Yes	Low	Low	No	Unclear	High	Low	Yes – 12 months' follow-up	Yes	No – some will have had imaging, some will have had records	Yes	High
Chapman 2021	Yes	Yes	No	High	High	Yes	Yes	Low	Low	Yes (follow-up, implied all colonoscopy or imaging)	Unclear	Unclear	Low	Unclear	Yes	Unclear – because some may have received more than just colonoscopy	Yes (missing data lack of follow-up)	unclear

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Georgiou Delisle 2022	Yes, states all	Yes	No, some patients sent to alternative pathways (rectal bleeding or CLASP) which may not be available in all areas of England	High	High	Yes	Yes	Low	Low	No – follow-up for some patients	Unclear	Unclear	Low	Yes – no patients excluded on basis of colonoscopy not having been performed at 6 months' follow-up	Yes	Yes – Complete index test-dependent differential verification reference standard: participants get a different reference standard according to the index test result	Yes	Low
Hunt 2022	Yes – states 'all'	Yes	No – only some areas tested those with rectal bleeding	High	High – rectal bleeding not representative; due to covid there is an unrepresentative mix of patients in the total cohort	Yes	Yes	Low	Low	Unclear – not clear how many patients had colonoscopy/CTC	Unclear	Unclear	Low	Unclear – does not state time period between FIT and imaging; doesn't state how long records follow-up for	Yes	Unclear – does not state how many had col/CTC or not	Yes	Unclear

continued

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Juul 2018	Yes – states 'all'	Yes	No, excluded some patients that meet NG12 high risk, but not all	High	High	Yes	Yes	Low	Low	No – not all had colonoscopy/CTC	No – states doctors performing colonoscopy not blind	High	Low	Unclear	Yes	No (some did not get diagnostic investigation)	Yes	High
Laszlo 2021	Unclear – does not state 'all' or 'consecutive'	Yes	No – excluded those without a definitive diagnosis	High	High	Yes, states blind	Yes	Low	Low	Yes	Yes – states blind	Low	Low	Unclear, does not state interval	Yes	No – differential verification dependent on other known or unknown factors	No – excluded patients without definitive diagnosis, which may have introduced bias if these patients are systematically different	High
Maclean 2021[57]	Yes – states all	Yes	Yes	Low	High – 2WW	Yes – supplement	Yes	Low	Low	No – not all col	No – used to triage patients to reference standard	High	Low	Unclear	Unclear (not clear if those ref back to GP were followed up)	No	No – some excluded due to frailty, cancellation, etc.	High

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Morales-Arraez 2018	Yes	Yes	No (needed both fit and colonoscopy)	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes	Unclear
Mowat 2016	Yes	Yes	No – only patients who had a colonoscopy were included, based on secondary care triage	High	High	Yes	Yes/ no – LoD selected as >/= 10 had fns – therefore lod high risk, 10 low risk	High (LoD); low (10 µg/g)	Low	Yes	Yes – all clinicians and endoscopists were blind to the faecal test results	Low	Low	Yes – patients triaged to endoscopy were investigated within 6 weeks	Yes	Yes	No – patients w/o colonoscopy excluded and five patients missing from analysis without explanation	High

continued

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Pin Vieto	Unclear – just states inclusion criteria, but not that were consecutive or 'all'	Yes	Yes – excluded pts with CRC in prior 2 years	Unclear	Unclear – not enough information provided to score	Yes	Yes	Low	Low	No – some did not get full imaging	No – states performed as 'part of their medical treatment'	High	Low	Unclear – not clear how quickly colonoscopy was done; Follow-up of records for 2 years	Yes	No – complete index test-dependent differential verification reference standard: participants get a different reference standard according to the index test result, for example FIT positive get colonoscopy, FIT negative get records follow-up (2 years)	Unclear	High

TABLE 44 OC-Sensor studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores (*continued*)

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Rodriguez-Alonso 2018, Rodriguez-Alonso 2019	Unclear	Yes	No, excluded premenopausal women, some others	High	High – only secondary care referrals, some referred from secondary care not primary care (included due to reporting anaemia data)	Unclear	Yes	Unclear	Low (NB OC-Sensor MICRO)	Yes (colonoscopy)	Unclear	Unclear	Low	Unclear – does not state interval	Yes	Yes	Unclear who was excluded based on incomplete tests etc	Unclear
Crooks 2023, Bailey J 2021[42]	Yes (all logged)	Yes	No	High	High	Yes	Yes	Low	Low	No	Unclear	High	Low	Unclear	Yes	No	No	high

TABLE 45 FOB Gold studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis? .	RoB: Could the patient flow have introduced bias?
Benton 2022[44]	No (participation voluntary)	Yes	No (subgroup, participation voluntary)	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes (some missing data, explanations given, no reason to assume not random)	Low
MacLean 2022[59]	Unclear – does not state consecutive or all	Yes	No – only included those with definitive diagnosis, which may exclude a spectrum of patients	High	High	Yes, states blind	Yes	Low	Low	Yes	Yes, states blind	Low	Low	Unclear, does not state interval	Yes	No – differential verification dependent on other known or unknown factors	No – excluded patients without definitive diagnosis, non-return of sample, pack not received, dnas all excluded	High
Jordaan 2023[83]	Unclear, only samples submitted to lab included, use of FIT increased over time, not clear why	Yes	Yes	Unclear	Low	Yes	Yes	Low	Low	No	Unclear	High	Low	Unclear	Yes	No	Yes	High

TABLE 46 QuikRead go studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis? .	RoB: Could the patient flow have introduced bias?
Maclean 2021[58]	Unclear – does not state consecutive or all	Yes	No – only included those with definitive diagnosis, which may exclude a spectrum of patients	High	High	Yes, states blind	Yes	Low	Low	Yes	Yes, states blind	Low	Low	Unclear, does not state interval	Yes	No – Differential verification dependent on other known or unknown factors	No – excluded patients without definitive diagnosis, non-return of sample, pack not received, DNAs all excluded	High
Tsapournas 2020[80]	No – convenience sample	Yes	No, only included those with colonoscopy	High	High – includes referrals from primary and secondary care	Yes – states blind	Yes	Low	Low	Yes	Yes – states blind	Low	Low	Yes, < 30 days (states tests > 30 days excluded)	Yes	Yes	No, excluded patients who did not have a colonoscopy but doesn't detail why	High

TABLE 47 NS-Prime study: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis?	RoB: Could the patient flow have introduced bias?
Benton 2022[44]	No (participation voluntary)	Yes	No (subgroup, participation voluntary)	High	High	Yes	Yes	Low	Low	Yes	Unclear	Unclear	Low	Unclear	Yes	Yes	Yes (some missing data, explanations given, no reason to assume not random)	Low

TABLE 48 FOB Gold studies: SCHARR's assessment of risk of bias and applicability, with reasons for scores

	1. Was a consecutive or random sample of patients enrolled?	2. Was a case–control design avoided?	3. Did the study avoid inappropriate exclusions?	RoB: Could the selection of patients have introduced bias?	Applicability: Is there concern that the included patients and settings do not match the review question?	1. Were the index test results interpreted without knowledge of the results of the reference standard?	2. If a threshold was used, was it prespecified?	RoB: Could the conduct or interpretation of the index test have introduced bias?	Applicability: Is there concern that the index test, its conduct, or interpretation differ from the review question?	1. Is the reference standard likely to correctly classify the target condition?	2. Were the reference standard results interpreted without knowledge of the results of the index test?	RoB: Could the reference standard, its conduct, or its interpretation have introduced bias?	Applicability: Is there concern that the target condition as defined by the reference standard does not match the review question?	1. Was there an appropriate interval between index test(s) and reference standard?	2. Did all patients receive a reference standard?	3. Did patients receive the same reference standard?	4. Were all patients included in the analysis? .	RoB: Could the patient flow have introduced bias?
Sieg 1999	Unclear	Yes	No, those going to colonoscopy	High	High	Yes – states blind	Hb: yes Hb/Hp: no	Hb: low Hb/Hp: high	Yes – no data relating to the equivalence of this test and the commercial version has been presented in symptomatic patients	Yes	Yes – states blind	Low	Low	Unclear	Yes	Yes	Unclear (*n* recruited not reported)	Unclear

Appendix 4 Clinical review: table of excluded studies with rationale

Studies excluded from the review of diagnostic test accuracy, with reasons

Reason	Number of studies excluded	References
Analytical performance	9	87,129–136
Crossover, no new data or superseded	14	137–149
Editorial, comment, letter	7	150–156
Incorrect population	41	80,157–196
Insufficient data to calculate DTA/data ambiguous/not DTA study	11	197–207
Outcome not CRC or CRC only	4	208–211
Not English language	1	212
Not FIT or in-scope test	6	84,213–217
Ongoing study or systematic review	10	218–227
Systematic review or review	13	12,16,90,228–237
Threshold not reported	2	238,239

Appendix 5 Additional analyses

Additional analysis 1: synthesis of all tests together in a single analysis

This analysis was run to allow the investigation of the impact of study population type and reference standards on a larger sample of studies and because these factors were unlikely to interact with test type. It was also used to inform the priors used when < 5 studies were being synthesised (see *Appendix 2*, *Statistical methods for the evidence synthesis*).

Twenty-eight studies contributed to the meta-analysis for all tests (OC-Sensor:11, HM JACKarc: 15, FOB Gold: 2). This total is different from a simple addition of all studies contributing to each test individually because some studies were excluded from the analysis to avoid double counting of patients. This was the case for some studies that reported data for multiple tests and for some studies where the recruitment dates and locations overlapped with other studies.

Nine studies provided diagnostic accuracy at a single threshold, and the maximum number of thresholds considered within a single study was 103. The final data set provided a total of 194 pairs of sensitivity and specificity estimates, at thresholds between 2 and 401.

Figure 17a displays the results on the ROC plane. *Figure 17b* displays the sensitivity and specificity as a function of threshold. Pooled sensitivity and specificity are shown for subgroups based on population type in *Figures 17c* and *17d*, respectively. Sensitivity and specificity for specific thresholds are summarised for all population groups in *Table 49*.

For the analysis of all studies (populations 1–4), sensitivity ranges from 96.4% (95% CrI 94.7% to 97.7%; 95% PrI 86.9% to 99.6%) at a threshold of 2, to 45.2% (95% CrI 39.7% to 50.6%; 95% PrI 27.7% to 62.3%) at a threshold of 400. Specificity ranges from 60.3% (95% CrI 51.6% to 68.8%; 95% PrI 16.6% to 95.0%) at a threshold of 2, to 98.3% (95% CrI 96.6% to 99.2%; 95% PrI 80.9% to 100%) at a threshold of 400. For the analyses of subgroups by population type, the summary estimates were similar and not statistically significant based on overlap of the 95% CrI. The summary specificity for population 3 are higher than for the other considered subgroups; however, this analysis was based on only three studies (2 HM-JACKarc studies and 1 OC-Sensor). There is, therefore, considerable uncertainty in the pooled estimates, and these should be interpreted with caution.

Additional analysis 2: reference standard sensitivity analysis

Subgroup analyses were conducted that included only the studies where at least 90% of the participants received colonoscopy as the reference standard. Subgroups analyses were considered for all FIT together, including all population types, excluding population type 4 studies, and separately for each test (where data allowed).

Summary estimates of sensitivity and specificity are illustrated in *Figure 18*. For all analyses the summary estimates were similar, irrespective of the reference standard grouping (all studies vs at least 90% of the participants receiving colonoscopy). The largest difference in point estimates was seen for specificity of OC-Sensor (see *Figure 18f*); however, there were only three studies in the > 90% colonoscopy subgroup and so the apparent difference may be explained by other sources of heterogeneity between the studies. There was very little difference in specificity for the HM-JACKarc studies (see *Figure 18d*).

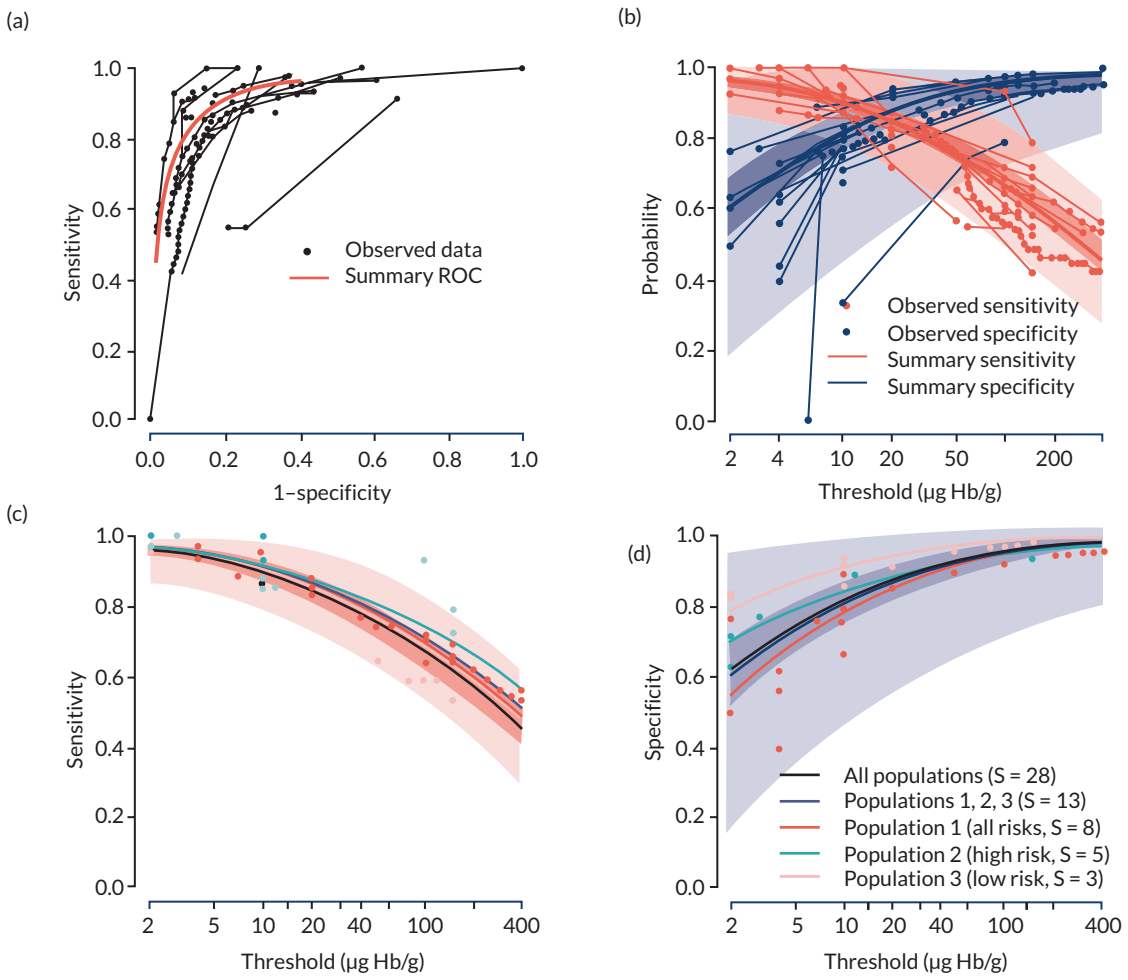

FIGURE 17 Observed data and summary sensitivity and specificity for all tests together.

TABLE 49 Summary sensitivity and specificity at specific thresholds for all tests together

Threshold, µg/g	All studies 1–4 (*n* = 28)		All 1–3 (*n* = 13)		Population 1 (*S* = 8)		Population 2 (*S* = 5)		Population 3 (*S* = 3)	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
2	96.4 (94.7 to 97.7)	60.3 (51.6 to 68.8)	96.3 (94.6 to 97.5)	61.8 (49.2 to 73.6)	96.2 (93.5 to 97.9)	54.4 (37.6 to 71)	96.3 (83.1 to 99.9)	69.8 (58.9 to 81.7)	96.7 (89.9 to 99.6)	78.6 (59.5 to 93.1)
2.5	95.8 (94 to 97.2)	63.8 (55.5 to 72.1)	95.8 (93.9 to 97.1)	65.1 (52.9 to 76.4)	95.6 (92.8 to 97.5)	58.2 (41.6 to 74.4)	95.8 (82.1 to 99.9)	72.1 (61.3 to 83.4)	96.2 (89.1 to 99.5)	80.9 (62.4 to 94.3)
3	95.3 (93.4 to 96.9)	66.7 (58.6 to 74.7)	95.3 (93.3 to 96.7)	67.7 (55.8 to 78.6)	95.1 (92.1 to 97.2)	61.3 (44.9 to 77.1)	95.4 (81.1 to 99.8)	74 (63.3 to 84.8)	95.7 (88.4 to 99.4)	82.6 (64.7 to 95.1)
4	94.4 (92.3 to 96.2)	70.9 (63.2 to 78.4)	94.4 (92.3 to 96.1)	71.6 (60.4 to 81.8)	94.2 (91 to 96.6)	65.9 (50.1 to 80.8)	94.6 (79.6 to 99.7)	76.7 (66.4 to 86.7)	94.9 (87.1 to 99.2)	85.1 (68.1 to 96.1)
7	92.1 (89.6 to 94.3)	78.1 (71.3 to 84.6)	92.4 (89.8 to 94.4)	78.3 (68.3 to 87)	92 (88.2 to 95)	74.1 (59.6 to 86.9)	92.7 (76.4 to 99.5)	81.3 (71.7 to 89.8)	92.8 (84 to 98.5)	89.1 (73.8 to 97.7)
10	90.2 (87.4 to 92.7)	82 (75.7 to 87.8)	90.7 (87.8 to 93)	81.9 (72.7 to 89.7)	90.3 (86.1 to 93.7)	78.6 (65 to 89.9)	91.2 (73.9 to 99.2)	83.9 (74.8 to 91.5)	91 (81.5 to 97.8)	91.1 (76.9 to 98.3)
20	85.4 (82.2 to 88.5)	88 (82.7 to 92.5)	86.4 (83.1 to 89.4)	87.6 (79.9 to 93.6)	85.8 (81 to 90.3)	85.7 (73.7 to 94.2)	87.5 (68.6 to 98.3)	88.1 (79.9 to 94.1)	86.6 (75.5 to 95.7)	94.1 (81.9 to 99.1)
50	76.3 (72.5 to 80.1)	93.2 (89.3 to 96.2)	78.5 (74.3 to 82.4)	92.7 (86.8 to 96.7)	77.4 (72 to 83.4)	91.9 (82.4 to 97.4)	80.5 (59.3 to 95.8)	92.2 (85.2 to 96.5)	78 (62.5 to 90.8)	96.6 (87.1 to 99.6)
100	67.2 (62.8 to 71.5)	95.7 (92.7 to 97.7)	70.5 (65.7 to 75.3)	95.2 (90.6 to 98.1)	69 (63.2 to 76.3)	94.9 (87.2 to 98.6)	73.7 (50.1 to 92.2)	94.4 (88.4 to 97.7)	69.3 (48.4 to 85.7)	97.8 (90 to 99.8)
120	64.5 (60 to 68.9)	96.2 (93.4 to 98)	68.1 (63.2 to 73.2)	95.7 (91.4, 98.3)	66.5 (60.6, 74.1)	95.5 (88.3, 98.8)	71.6 (47.4, 90.9)	94.8 (89.1 to 97.9)	66.7 (44.3 to 84.2)	98 (90.7 to 99.8)
150	61.1 (56.4 to 65.7)	96.7 (94.1 to 98.4)	65.1 (59.9 to 70.5)	96.3 (92.3, 98.6)	63.4 (57.3, 71.4)	96.1 (89.4, 99)	69 (43.4, 89.2)	95.4 (89.9 to 98.1)	63.4 (39.1 to 82.4)	98.3 (91.5 to 99.9)
200	56.5 (51.6 to 61.4)	97.3 (95 to 98.7)	61 (55.5 to 66.8)	96.9 (93.3, 98.9)	59.2 (52.8, 67.6)	96.8 (90.7, 99.3)	NR	NR	NR	98.6 (92.3 to 99.9)
400	45.2 (39.7 to 50.6)	98.3 (96.6 to 99.2)	50.7 (44.6 to 57.2)	98 (95.3, 99.3)	48.7 (41.7, 57.9)	98 (93.4, 99.6)	NR	NR	NR	99.1 (94.2 to 100)

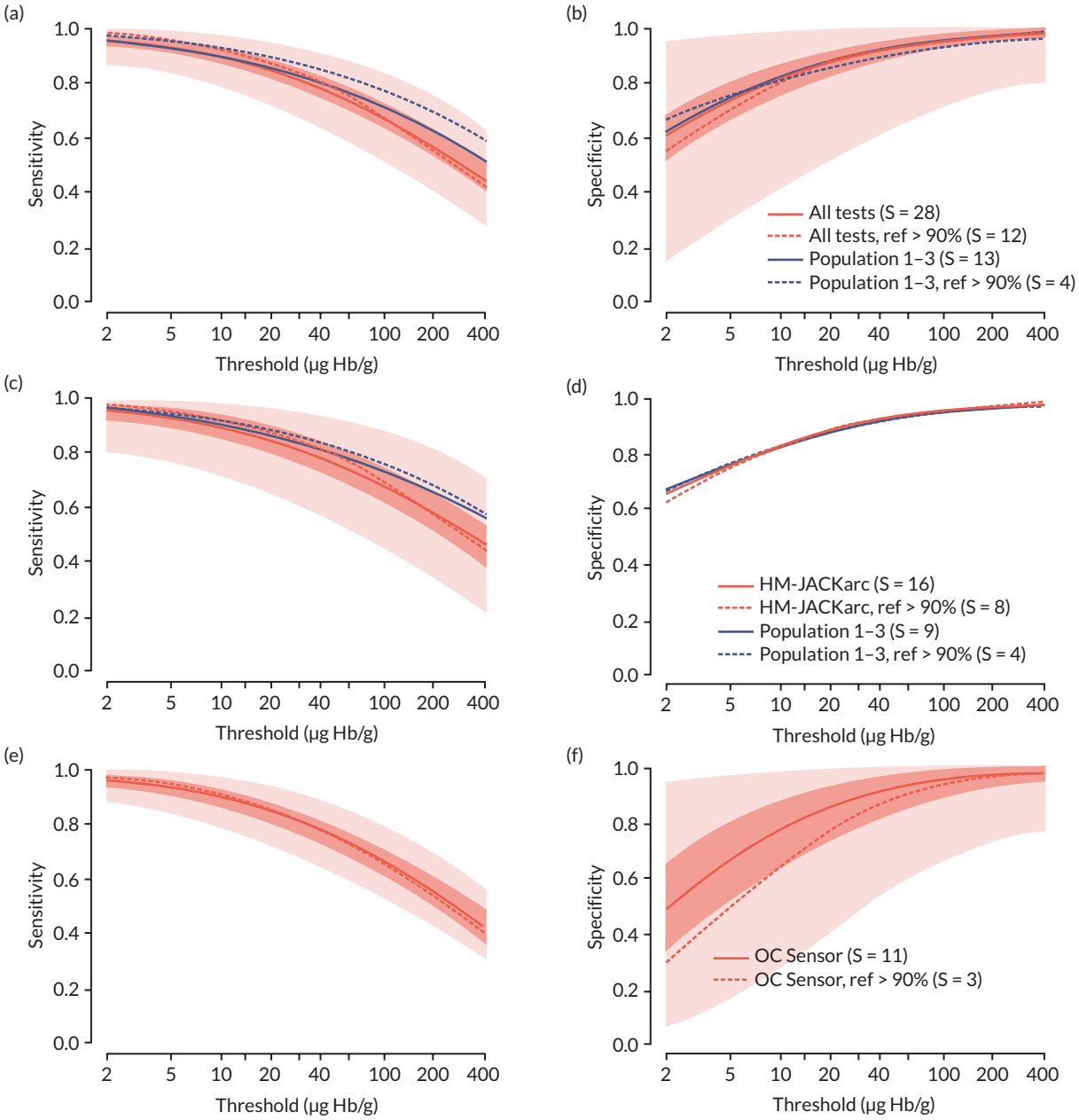

FIGURE 18 Summary sensitivity and specificity, reference standard sensitivity analysis.

Appendix 6 Subgroup analyses and other outcomes

Subgroup analyses by patient characteristics

An exploration of the potential reasons for heterogeneity in diagnostic test accuracy across studies using meta-regression was considered. However, study-level covariates relating to patient characteristics of interest were not reported in all studies. Instead, studies that reported diagnostic test accuracy for subgroups of patients were considered in subgroup analyses.

Anaemia

Studies reporting data on anaemia are summarised in *Table 50*. Considering all the available data on anaemia regardless of the test used, population type and reference standard, 11 studies[52,54,65,66,68,71,72,75,76,78,82] reported data on anaemia or IDA. The studies can be broadly categorised as comparative, comparing those with anaemia with those without; comparative, comparing those with anaemia with the study population unselected on the basis of anaemia (whole cohort); or non-comparative.

When considering studies that compare those with anaemia with those without, both[52,71] reported lower sensitivity and specificity at a threshold of 10 for those with anaemia. One further study[82] reported that the optimal threshold (defined as the point on the ROC curve that maximises sensitivity and specificity) for those with anaemia is higher than for those without. It should be noted that the definition of 'optimal' in this study is not necessarily the same as optimising the threshold for cost-effectiveness or clinical decision-making, where it may be preferable to optimise either sensitivity (where the test is used to rule out disease) or specificity (where the test is used to rule in disease).

Among studies that compared those with anaemia with the whole study population, the results were more mixed. One study[75] showed the same trend of lower sensitivity and specificity, three[65,66,72] showed higher sensitivity and lower specificity, and one[54] showed lower sensitivity and higher specificity.

Of particular note is a study by Withrow *et al.*,[66] which shows that sensitivity increases as the threshold for anaemia is increased (i.e. more anaemic), while the specificity decreases in both men and women.

It should be noted that the definition of anaemia varied across studies, with some considering IDA to be anaemia and others considering other types of anaemia as well or instead.

Age

Three studies[66,73,82] reported data according to age group (see *Table 51*). All were large studies with > 5000 patients, the largest comprising 16,604 patients.[66] All studies used HM JACKarc.

One study[82] reported the optimal cut-off point (the point that maximises both sensitivity and specificity) based on the ROC curves for those aged < 60 years and aged ≥ 60 years separately, and reported that the optimal threshold was lower in the ≥ 60 years age group (19 µg/g) than in the younger age group (37 µg/g). This study concluded that FIT could be incorporated into a risk score based on sex, age, symptoms and signs, drug history and blood parameters, but did not conduct the analyses required to produce such a score in that publication. It should be noted that the definition of 'optimal' in this study is not necessarily the same as optimising the threshold for cost-effectiveness or clinical decision-making, where it may be preferable to optimise either sensitivity (where the test is used to rule out disease) or specificity (where the test is used to rule in disease).

Another study[73] reported a limited range of thresholds (2, 10 and 150 µg/g) for those aged < 50 years and aged ≥ 50 years. Sensitivity was lower in the younger age group at any given threshold, although CIs overlapped. This trend was less evident at the highest threshold and the number of events in the younger age group was small (n = 16). In this study the authors noted that in younger patients it may be appropriate to interpret any detectable faecal Hb as a positive test.

TABLE 50 Sensitivity and specificity of studies reporting data for patients with anaemia

Author, year Test	Population type	Anaemia type	Threshold, µg/g	*N* with CRC/*N* analysed	Sensitivity (95% CI)	Specificity (95% CI)	Summary of IDA/anaemia vs comparator
Studies comparing those with to those without anaemia/IDA							
Cunin 2020[71a] HM-JACKarc	2	Whole cohort	10	48/928 (5.17%)	85.4 (NR)	86.9 (NR)	Vs. 'no IDA'; sensitivity and specificity lower
		No IDA	10	28/739 (3.79%)	89.00 (70 to 97.1)	84.00 (81.1 to 86.6)	
		IDA	10	20/189 (10.58%)	80.00 (55.7 to 93.3)	81.60 (74.8 to 87)	
Johnstone 2022[52] HM-JACKarc	1	No anaemia	10	32/3238 (0.99%)	96.9 (96.3 to 97.5)[a]	81.3 (80 to 82.6)[a]	Sensitivity and specificity lower
		Anaemia	10	26/793 (3.28%)	84.6 (82.1 to 87.1)[a]	72.9 (69.8 to 76)[a]	
Turvill 2021[82] HM-JACKarc	4	No IDA	19	101/3582 (2.82%)	88.1 (80.2 to 93.7)	85.3 (84.0 to 86.4)	Optimal threshold higher Optimal FIT threshold was ≥ 21 vs. ≥ 19 µg/g in anaemic vs non-anaemic
		IDA	21	34/559 (6.08%)	82.40 (65.5 to 93.2)	81.50 (77.9 to 84.8)	
Turvill 2021[82] HM-JACKarc	4	No non-ID anaemia	19	110/3597 (3.06%)	84.5 (76.4 to 90.7)	85.0 (83.7 to 86.1)	Optimal threshold higher Optimal FIT threshold was ≥ 30 vs. ≥ 19 µg/g in anaemic vs non-anaemic
		Non-ID Anaemia	30	25/544 (4.60%)	92.00 (74.0 to 99.0)	85.50 (82.2 to 88.5)	
Studies comparing those with anaemia/IDA to patients unselected on the basis of anaemia (whole cohort)							
D'Souza 2021[72] HM-JACKarc	4	Whole cohort	10	12/298 (4.03%)	92.20% (88.2 to 95.2)	82.30% (81.3 to 83.2)	Sensitivity higher, specificity similar
		IDA	10	16/479 (3.34%)	100% (89.4 to 100)	81.60% (77.7 to 85.1)	
Tang 2022[65] HM-JACKarc	4	Whole cohort	10	20/603 (3.32%)	90.00 (68.3 to 98.77)	83.20 (79.9 to 86.14)	Sensitivity higher, specificity lower (low events in IDA)
		IDA	10	1/78 (1.28%)	100 (NE to NE)[a]	76.6 (67.2 to 86)[a]	
Juul 2018[54] OC-Sensor	4	Whole cohort	10	54/3462 (1.56%)	94.4 (93.6 to 95.2)[a]	85.7 (84.5 to 86.9)[a]	Sensitivity lower, specificity higher/similar
		Unexplained anaemia	10	54/3462 (1.56%)	20.4 (16.6 to 24.2)[a]	79.5 (75.7 to 83.3)[a]	
Gerrard 2023[75] Single FIT, HM-JACKarc	1	Whole cohort	10	69/2260 (3.05%)	84.10 (73.3 to 91.8)	77.4 (75.6 to 79.1)	Sensitivity similar/lower, specificity lower
		Anaemia	10	38/567 (6.70%)	81.6 (78.4 to 84.8)[a]	68.6 (64.8 to 72.4)[a]	
Gerrard 2023[75] Dual FIT, HM-JACKarc	1	Whole cohort	10	88/2637 (3.34%)	96.60 (90.4 to 99.3)[a]	71.2 (69.4 to 73.0)[a]	Sensitivity similar/lower, specificity lower
		Anaemia	10	29/480 (6.04%)	93.1 (90.8 to 95.4)[a]	60.1 (55.7 to 64.5)[a]	

TABLE 50 Sensitivity and specificity of studies reporting data for patients with anaemia (*continued*)

Author, year Test	Population type	Anaemia type	Threshold, µg/g	*N* with CRC/*N* analysed	Sensitivity (95% CI)	Specificity (95% CI)	Summary of IDA/anaemia vs comparator
Withrow 2022[66] HM-JACKarc	4	Whole cohort (both sexes)	10	139/16,604 (0.84%)	92.1 (91.7 to 92.5)[a]	91.5 (91.1 to 91.9)[a]	Sensitivity higher, specificity lower
		Low haemoglobin (< 130 g/l in men, < 120g/l in women)	10	72/507 (1.42%)	95.8 (95.2 to 96.4)[a]	88 (87.1 to 88.9)[a]	
		Whole cohort (men)	10	83/7019 (1.18%)	92.8 (92.2 to 93.4)[a]	90.3 (89.6 to 91)[a]	Sensitivity same or higher, specificity lower with increasing anaemia
		Men, < 130 g/l	10	46/2091 (2.20%)	93.5 (92.4 to 94.6)[a]	85.5 (84 to 87)[a]	
		Men, < 120 g/l	10	36/1141 (3.16%)	91.7 (90.1 to 93.3)[a]	82.7 (80.5 to 84.9)[a]	
		Men, < 110 g/l	10	23/494 (4.66%)	95.7 (93.9 to 97.5)[a]	79 (75.4 to 82.6)[a]	
		Men, < 100 g/l	10	14/216 (6.48%)	100 (NE to NE)[a]	72.3 (66.3 to 78.3)[a]	
		Men, < 90 g/l	10	9/89 (10.11%)	100 (NE to NE)[a]	71.2 (61.8 to 80.6)[a]	
		Whole cohort (women)	10	57/9585 (0.59%)	91.1 (90.5 to 91.7)[a]	92.4 (91.9 to 92.9)[a]	Sensitivity higher, specificity lower with increasing anaemia
		Women, < 120 g/l	10	25/2758 (0.91%)	100 (NE to NE)[a]	89.4 (88.3 to 90.5)[a]	
		Women, < 110 g/l	10	13/1297 (1.00%)	100 (NE to NE)[a]	88 (86.2 to 89.8)[a]	
		Women, < 100 g/l	10	6/491 (1.22%)	100 (NE to NE)[a]	84.5 (81.3 to 87.7)[a]	
		Women, < 90 g/l	10	3/189 (1.59%)	100 (NE to NE)[a]	79.6 (73.9 to 85.3)[a]	
Non-comparative studies							
Ayling 2019[68] OC-Sensor	4	Anaemia	10	7/178 (3.93%)	71.4 (64.8 to 78)[a]	95.9 (93 to 98.8)[a]	
Ayling 2019[68] OC-Sensor	4	IDA	10	6/137 (4.38%)	68.7 (60.9 to 76.5)[a]	95.4 (91.9 to 98.9)[a]	
Morales-Arraez 2018[76] OC-Sensor	4	Moderate-severe IDA	10	28/245 (11.43%)	92.9 (89.7 to 96.1)[a]	57.1 (50.9 to 63.3)[a]	
Rodriguez-Alonso 2020[78]	4	IDA	10	9/120 (7.50%)	100 (NE to NE)[a]	77.5 (70 to 85)[a]	

N, number; NE, not estimable.
a Calculated by EAG reviewer.

In the third study,[66] the thresholds were 2 µg/g and 10 µg/g and were reported for those aged < 40 years and then for those aged ≥ 40, ≥ 50, ≥ 60, ≥ 70 and ≥ 80 years. This study performed multivariable modelling including FIT, blood tests, age, and sex and concluded that that age-specific thresholds for FIT positivity would not improve test performance.

Overall, there is some indication that FIT thresholds may need to be lower in younger patients to achieve the same sensitivity as for older patients. The available data do not provide conclusive evidence that different FIT thresholds should be used or what they should be.

Sex

Three studies[27,43,82] reported data for men and women separately (see *Table 52*). All were studies with > 3000 patients, with the largest comprising 9899 patients.[27] One study used OC-Sensor PLEDIA[43] and two used HM-JACKarc.[27,82]

One study[82] reported the optimal cut-off value (the point that maximises both sensitivity and specificity), based on the ROC curves for men and women separately, and reported that the optimal threshold was lower for women (16 µg/g) than for men (21 µg/g). This study concluded that FIT could be incorporated into a risk score based on sex, age, symptoms and signs, drug history and blood parameters. It should be noted that the definition of 'optimal' in this study is not necessarily the same as optimising the threshold for cost-effectiveness or clinical decision-making, where it may be preferable to optimise either sensitivity (where the test is used to rule out disease) or specificity (where the test is used to rule in disease).

The two other studies[43,66] reported a range of thresholds (from 10 to 150 µg/g), and generally showed that at thresholds above 10 µg/g, sensitivity and specificity are higher in women than in men. This difference was more pronounced in one study[43] than in the other,[27] but the small number of studies meant that it was not possible to tell if this was due to the use of different analysers or some other factor. At 10 µg/g, one study showed roughly equivalent sensitivity and specificity,[27] while the other study showed numerically lower sensitivity in men, but stated that no significant difference in FIT sensitivity was found.[43] Withrow *et al*. conducted a multivariable analysis including sex and showed that the probability of CRC reached 3% at 17 µg/g and 25 µg/g for males and females, respectively.

If sensitivity and specificity are different in women and men at a given threshold, a different threshold in women may be required to achieve equivalent sensitivity and specificity in the two sexes. However, it was not possible on the basis of the available data to conclude what and whether different FIT cut-off values are required according to sex.

Medications that might cause gastrointestinal bleeding

The scope issued by NICE states the assessment should consider whether the FIT threshold should be different for 'People taking medications or with conditions which increase the risk of gastrointestinal bleeding'. In a slight widening of the scope, NICE confirmed an additional study that looked at the effect of taking proton pump inhibitors, which may decrease the risk of GI bleeding, and may be of interest to the committee and has therefore been included.

Consequently, three studies[70,77,82] were included in this subgroup analysis. The studies are summarised in *Table 53*. All studies included more than 1000 patients, with the largest comprising 5040 in total. Two studies (three references) used OC-Sensor analysers,[70,77,84] and one used HM-JACKarc.[82]

One study[82] reported the optimal cut-off value (the point that maximises both sensitivity and specificity) based on the ROC curves for those using antiplatelet, anticoagulants, non-steroidal anti-inflammatory drugs (NSAIDs) and those not using these drugs. The optimal threshold was 19 µg/g in both cases, although the sensitivity and specificity were superior in those not using the drugs than in those who were. This study concluded that FIT could be incorporated into a risk score based on sex, age, symptoms and signs, drug history and blood parameters.

Another study (part of the 'colonpredict' study)[70,84] reported test accuracy data for those using aspirin and those not using aspirin. It should be noted that this study recruited symptomatic patients from secondary as well as primary care and was therefore excluded from the main analysis. The analysis of aspirin users was included because of the sparsity of data in this subgroup, but it is unclear how generalisable these results will be to the primary care setting. Only one threshold was included (20 µg/g). As with the previous study, the sensitivity and specificity were superior in those not

TABLE 51 Sensitivity and specificity by age

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Population type	*N* with CRC/*N* analysed (%)	Threshold, µg/g	Age group in years	Sensitivity (95% CI)	Specificity (95% CI)	Conclusion drawn by study authors
1	D'Souza 2021[73] NICE FIT October 2017 to December 2019	HM JACKarc analytical system Colonoscopy	4	16/1103 (1.45%)	2	< 50	87.50 (61.7 to 98.4)	70.40 (67.6 to 73.1)	Detectable f-Hb on FIT in symptomatic younger patients may indicate referral for investigation of CRC and serious bowel disease
				313/8719 (3.59%)	2	≥ 50	97.40 (95.0 to 98.9)	64.10 (63.1 to 65.2)	
				16/1103 (1.45%)	10	< 50	81.30 (54.4 to 96.0)	83.60 (81.3 to 85.5)	
				313/8719 (3.59%)	10	≥ 50	91.40 (87.7 to 94.2)	83.50 (82.7 to 84.3)	
				16/1103 (1.45%)	150	< 50	68.80 (41.3 to 89.0)	92.20 (90.4 to 93.7)	
				313/8719 (3.59%)	150	≥ 50	70.90 (65.6 to 75.9)	94.90 (94.4 to 95.3)	
2	Turvill 2021[82] Yorkshire and Humber, UK April 2018 to December 2019 Fast track FIT	HM JACKarc Full colonoscopy or CT colonography, or a lesser investigation (such as CT abdomen/pelvis with contrast or flexible sigmoidoscopy)	4	30/1217 (2.47%)	37	< 60	90.00 (73.5 to 97.9)	87.40 (85.4 to 89.3)	The optimal cut-off value for people aged ≥ 60 years (19 µg/g faeces) is lower than for those aged < 60 years (37 µg/g faeces). FIT could be incorporated into a risk score based on sex, age, symptoms and signs, drug history and blood parameters
				19/3823 (0.49%)	19	≥ 60	83.50 (75.6 to 89.6)	85.40 (84.2 to 86.5)	
3	Withrow 2022[66] (same study as Nicholson 2020)[27] Oxfordshire, UK March 2017 to 21 December 2020 CSS-BIO-3 4730	HM JACKarc Records follow-up	4	9/1390 (0.65%)	2	< 40	100 (70.1 to 100)	89.1 (87.4 to 90.7)	The lack of an apparent age effect after taking into account FIT suggests that age-specific thresholds for FIT positivity would not improve test performance
				130/15,214 (0.85%)	2	> 40	96.2 (91.3 to 98.3)	83.0 (82.4 to 83.6)	
				118/12,936 (0.91%)	2	> 50	95.8 (90.5 to 98.2)	81.8 (81.1 to 82.4)	
				98/8755 (1.12%)	2	> 60	94.9 (88.6 to 97.8)	78.8 (77.9 to 79.7)	
				77/3043 (2.53%)	2	> 70	94.8 (87.4 to 98)	51.8 (50 to 53.6)	
				41/2527 (1.62%)	2	> 80	95.1 (83.9 to 98.7)	68.4 (66.6 to 70.2)	
				9/1390 (0.65%)	10	< 40	88.9 (56.5 to 98)	93.4 (92 to 94.6)	
				130/15,214 (0.85%)	10	> 40	92.3 (86.4 to 95.8)	91.3 (90.9 to 91.8)	
				118/12,936 (0.91%)	10	> 50	91.5 (85.1 to 95.3)	90.7 (90.2 to 91.2)	
				98/8755 (1.12%)	10	> 60	89.8 (82.2 to 94.4)	89.0 (88.3 to 89.6)	
				77/5863 (1.31%)	10	> 70	89.6 (80.8 to 94.6)	87.1 (86.2 to 87.9)	
				41/2533 (1.62%)	10	> 80	87.8 (74.5 to 94.7)	83.1 (81.6 to 84.5)	

N, number.

TABLE 52 Sensitivity and specificity by sex

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	*N* with CRC/*N* analysed (%)	Threshold, µg/g	Men: sensitivity (95% CI)	Men: specificity (95% CI)	Women: sensitivity (95% CI)	Women: specificity (95% CI)	Conclusion drawn by study authors
1	Ball 2022[43] (additional data by personal communication) Sheffield, UK October 2019 to December 2019	OC-Sensor PLEDIA Colonoscopy or CT imaging and colon capsule endoscopy	4	Men: 25/1566 (1.60%) Women: 20/1940 (1.03%)	10	84.00 (63.1 to 94.7)	79.20 (77.0 to 81.2)	100.00 (80 to 100)	82.00 (80.2 to 83.7)	Sex did not significantly influence FIT sensitivity on subgroup analysis
					20	80.00 (58.7 to 92.4)	85.40 (83.5 to 87.1)	95.00 (73.1 to 99.7)	88.80 (87.2 to 90.2)	
					50	68.00 (46.4 to 84.3)	91.60 (90.0 to 92.9)	80.00 (55.7 to 93.3)	94.10 (92.9 to 95.1)	
					80	64.00 (42.6 to 81.3)	93.90 (92.6 to 95.0)	70.00 (45.7 to 87.2)	95.80 (94.8 to 96.6)	
					100	64.00 (42.6 to 81.3)	94.60 (93.3 to 95.7)	70.00 (45.7 to 87.2)	96.70 (95.8 to 97.4)	
					120	60.00 (38.9 to 78.2)	95.20 (94.0 to 96.2)	65.00 (40.9 to 83.7)	97.00 (96.1 to 97.7)	
					150	52.00 (31.8 to 71.7)	96.40 (95.3 to 97.2)	55.00 (32.0 to 76.2)	97.30 (96.5 to 98.0)	
2	Turvill 2021[82] Yorkshire and Humber, UK April 2018 to December 2019 Fast-track FIT	HM JACKarc Full colonoscopy or CT colonography, or a lesser investigation (such as CT abdomen/pelvis with contrast or flexible sigmoidoscopy)	4	Men: 89/2242 (3.97%) Women: 62/2798 (2.22%)	Men: 21 Women: 16 NB: optimal threshold was derived	85.40 (76.3 to 92.0)	83.70 (82.0 to 85.2)	87.10 (76.1 to 94.3)	85.60 (84.2 to 86.9)	The optimal cut-off value is lower for females (16 µg/g faeces) than for males (21 µg/g faeces). FIT could be incorporated into a risk score based on sex, age, symptoms and signs, drug history and blood parameters
3	Nicholson 2020 (same study as Withrow 2022)[27a] Oxfordshire, UK March 2017 to December 21, 2020 CSS-BIO-3 4730	HM JACKarc Records follow-up	4	Men: 65/4104 (1.58%) Women: 40/5795 (0.69%)	7	92.30 (85.8 to 98.8)	87.90 (86.9 to 88.9)	90.00 (80.7 to 99.3)	91.10 (90.3 to 91.8)	The area under the curve for all adults did not change substantially by gender. From Withrow 2022: The probability of CRC reached 3% at 17 µg/g and 25 µg/g, for males and females, respectively
					10	90.80 (83.7 to 97.8)	89.80 (88.8 to 90.7)	90.00 (80.7 to 99.3)	92.40 (91.8 to 93.1)	
					20	83.10 (74.0 to 92.2)	92.30 (91.5 to 93.2)	87.50 (77.3 to 97.7)	94.60 (94.1 to 95.2)	
					50	73.80 (63.2 to 84.5)	95.50 (94.9 to 96.2)	75.00 (61.6 to 88.4)	96.90 (96.5 to 97.4)	
					100	60.00 (48.1 to 71.9)	96.80 (96.3 to 97.3)	62.50 (47.5 to 77.5)	98.10 (97.8 to 98.5)	
					120	55.40 (43.3 to 67.5)	97.20 (96.7 to 97.7)	60.00 (44.8 to 75.2)	98.30 (98.0 to 98.6)	
					150	50.80 (38.6 to 62.9)	97.50 (97.1 to 98.0)	60.00 (44.8 to 75.2)	98.50 (98.2 to 98.8)	

N, number.

a Nicholson 2020 (*n* = 9896) is an earlier data cut of the same study as Withrow 2022. (n = 11,142). The data from Nicholson *et al.* have been included in this analysis in preference to the Withrow *et al.* data as the former report more thresholds, even though the study population is smaller. However, the Withrow *et al.* study conducted a multivariable analysis including sex, and the conclusions relating to this have been reported.

using the drug. This study concluded that aspirin use did not change the diagnostic accuracy of FIT in patients with GI symptoms.

The third study compared PPI users with PPI non-users. It should be noted that this study recruited symptomatic patients from secondary as well as primary care and was therefore excluded from the main analysis. The analysis of PPI users was included because of the sparsity of data in this subgroup, but it is unclear how generalisable these results will be to the primary care setting. At a threshold of 20 µg/g sensitivity was similar, and specificity was slightly higher in non-users. This study did not conclude anything for the detection of CRC, but it concluded that there was impaired FIT performance in PPI users for the detection of advanced neoplasia.

The evidence base is currently small, and it was not possible on the basis of the available data to conclude what and whether different FIT cut-off values are required according to medications being taken by a patient.

Ethnicity

No studies reporting the diagnostic test accuracy of any of the in-scope tests according to ethnicity were identified.

People with blood disorders

No studies were identified that reported the diagnostic test accuracy of any of the in-scope tests in a subgroup of people with blood disorders (e.g. beta thalassemia) that could affect the performance of the test.

Advanced adenomas and inflammatory bowel disease outcomes

Nine studies (10 publications)[17,18,54,56,58,60–62,75,79] reported data on AA and IBD. These are summarised in *Table 54*. One study reported data for the IDK Hb and Hb/Hp complex ELISA tests, while the remainder reported data for immunoturbidimetry tests. The synthesis focused on the immunoturbidimetry tests, but the IDK data were used in the model for the IDK tests.

Statistical synthesis of advanced adenoma outcomes

Nine studies[17,18,54,56,58,60–62,75] contributed to the meta-analysis for AA outcomes (HM-JACKarc, *n* = 6; OC-Sensor, *n* = 2; QuikRead Go, *n* = 1). Five studies provided diagnostic accuracy at a single threshold and the maximum number of thresholds considered by a single study was three. The full data set (all studies) provided a total of 15 pairs of sensitivity and specificity, at thresholds between 2 and 150. *Figures 19a* and *19b* illustrate the results for all studies, irrespective of test type. Separate syntheses are also provided for HM-JACKarc (see *Figure 19c*) and OC-Sensor (see *Figure 19d*).

One of the studies[75] also reported data for AA and IBD when using dual FIT and is reported in *Main analysis: dual faecal immunochemical test*.

For the analysis of all test types together, sensitivity ranges from 80.4% (95% CrI 55.8 to 98.3; 95% PrI 50.0 to 100.0%) at a threshold of 2, to 20.4 (95% CrI 0.6 to 47.5; 95% PrI 0 to 57.4) at a threshold of 150. Specificity ranges from 51.6 (95% CrI 31.6 to 71.1; 95% PrI 3.2 to 98) at a threshold of 2, to 95.7 (95% CrI 82.5 to 99.5; 95% PrI 58.5 to 100) at a threshold of 150. There is a large amount of heterogeneity between studies, as illustrated by the wide 95% CrI and PrI. Point estimates of summary sensitivity and specificity changed considerably for the separate analyses by test type (see *Figures 19c* and *19d* , and *Table 55*), emphasising the large amount of uncertainty.

Statistical synthesis of inflammatory bowel disease outcomes

Nine studies contributed to the meta-analysis for IBD outcome (HM-JACKarc, *n* = 6; OC-Sensor, *n* = 2; QuikRead go, *n* = 1). 5 provided diagnostic accuracy at a single threshold, and the maximum number of thresholds considered by a single study was three. The full data set (all studies) provided a total of 15 pairs of sensitivity and specificity, at thresholds between 2 and 150.

Figures 20a and *20b* illustrate the results for all studies, irrespective of test type. Separate syntheses are also provided for HM-JACKarc (see *Figure 20c*) and OC-Sensor (see *Figure 20d*).

TABLE 53 Sensitivity and specificity for patients taking medications that may affect the risk of GI bleeding

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Inclusion criteria	Group	*N* with CRC/*N* analysed (%)	Threshold, µg/g	Sensitivity (95%CI)	Specificity (95%CI)	Conclusion drawn by study authors
1	Bujanda 2018[70] Spain (assume Ourense and San Sebastian)[84] March 2012 to 2014 COLONPREDICT	OC-Sensor[84] Colonoscopy	Population type 4 – symptomatic patients referred from primary and secondary care	Aspirin users	51/485 (10.51%)	20	88.00 (75 to 95)	66.97 (62 to 71)	Aspirin use did not change the diagnostic accuracy of FIT in patients with GI symptoms
				Aspirin non-users	299/2567 (11.65%)	20	92.00 (88 to 95)	71.00 (69 to 73)	
2	Turvill 2021[82] Yorkshire and Humber, UK April 2018 to December 2019 Fast track FIT	HM JACKarc Full colonoscopy or CT colonography, or a lesser investigation (such as CT abdomen/pelvis with contrast or flexible sigmoidoscopy)	Population type 4 – 2WW patients	Antiplatelets, anticoagulants, NSAIDs	19/1356 (1.40%)	19	82.40 (69.1 to 91.6)	80.50 (78.2 to 82.6)	The specificity differed according to use of antiplatelets, anticoagulants, NSAIDs. FIT could be incorporated into a risk score based on sex, age, symptoms and signs, drug history and blood parameters
				No use of antiplatelets, anticoagulants, NSAIDs	100/3684 (2.71%)	19	87.0 (78.8 to 92.9)	86.9 (85.7 to 88.0)	
3	Rodriguez-Alonso 2018[77] Barcelona, Spain September 2011 to October 2012	OC-Sensor MICRO Colonoscopy	Population type 4 – symptomatic patients referred from primary and secondary care	PPI users	15/525 (2.86%)	20	93.3	85.1	No conclusion drawn for the identification of CRC in PPI users, concluded impaired FIT performance in PPI users for detection of advanced neoplasia.
				PPI non-users	15/477 (3.14%)	20	93.3	87.4	

N, number

For the analysis of all test types together, sensitivity ranges from 85.7% (95% CrI 70% to 96.7%; 95% PrI 42.3% to 100.0%) at a threshold of 2, to 41.7% (95% CrI 15.9% to 66.1%; 95% PrI 0.9% to 91.4%) at a threshold of 150. Specificity ranges from 53.8% (95% CrI 33.1% to 75.5%; 95% PrI 2.6% to 99.3%) at a threshold of 2, to 95.0% (95% CrI 80.2% to 99.5%; 95% PrI 55.0% to 100%) at a threshold of 150. As with AA, there is a large amount of heterogeneity between studies, as illustrated by the wide 95% CrIs and PrIs. Point estimates of summary sensitivity and specificity changed considerably for the separate analyses by test type (see *Figures 20c* and *20d* and *Table 56*), emphasising the large amount of uncertainty.

Other outcomes

Test uptake and repeat tests

Test failures, uptake and repeat tests

As these outcomes are likely to be affected by the point within the care pathway at which FIT is issued to the patient, this analysis has been restricted to studies where FIT was issued in primary care. All dual FIT studies were conducted in secondary care, and they have been included as no other data were available. The data are summarised in *Table 57*. Additional data with less generalisability (studies that were in secondary care settings and asked patients to provide samples for multiple tests) are provided in *Report Supplementary Material 3*.

Test failure rates

Test failure rates were reported in 11 studies (12 references)[42,43,45,49,52,54,56,60–62,68,83] ranged from 0.2% in two[42,62] separate studies to 18.8%.[49] Data were available for OC-Sensor, HM-JACKarc and FOB Gold only. The majority of studies reported rates between 2% and 5%,[45,52,54,56,60,61,68] although three studies reported rates > 14%.[43,49,83] It was not possible to tell whether test failure rates differed by test due to other sources of heterogeneity in the study designs, such as sending both a FIT and calprotectin test at the same time. It was also not clear if all studies defined this outcome consistently. Two studies provided the most details about the test failures, which included problems such as buffer loss, labelling errors, incorrect containers, no date of collection, volume errors and laboratory accidents.[49,83] Other studies tended to report spoiled or unsuitable samples, which may represent a narrower definition of test failure, although a precise definition was often missing.

One study[75] in dual FIT reported that FIT was inappropriate for 4.5% of patients, or that emergency presentation predated FIT postage.

Uptake

Only two[42,62] studies in primary care explicitly reported non-return of FIT, both with OC-Sensor. One study had an extremely high non-return rate (52%),[62] but this may be confounded by the fact that a referral had already been made and did not depend on the return of the FIT sample. The other study reported non-return rate of 9.4%, where FIT was being used as part of the diagnostic pathway. A later update[69] of the same study reported that 3631 out of 38,920 (9.3%) first FIT requests were not returned.

One study[75] of dual FIT showed that 10.7% of patients returned no FIT, and a further 20.5% returned only one FIT. Another study[51] reported that 4.9% of patients only returned one FIT, and one further study[80] noted that a stool sample was missing for 16.1% of patients. All studies took place in secondary care.

Repeat tests

Five studies (six references)[42,43,53,60,61,83] reported data on repeat FIT. The largest of these was a study pooling data from three Scottish regions. Of 135,396 tests, 12,359 (9.1%) were repeat FIT. This study also reported how many times repeat FIT were ordered for some patients, as can be seen in column 7 of *Table 57*. The other four studies report that 0.7%,[43] 1.7%,[42] 2.07%[60,61] and 9.9%[83] repeat FIT were ordered, although a later update[69] of one study[42] reported that 8349 (17.0%) requests were repeat tests in 6640 patients.

TABLE 54 Studies reporting data on AAs and IBD

#	Author, year; location; recruitment dates; study name (if available)	Analyser; reference standard	Population types	*N* with AA/*N* analysed (%) *N* with IBD/*N* analysed (%)	Thresholds, µg/g
1	Sieg 1999[79] Ostringen, Germany NR, prior to publication in 1999	Immunological test for Hb/Hp complex Colonoscopy Immunological test for HB Colonoscopy	4	AA: 37/621 (5.95%) IBD: 22/621 (3.54%)	2
2	D'Souza 2020[18] Croydon, UK November 2016 to October 2017	HM JACKarc analytical system Colonoscopy	1,2,3	AA (population 1): 4/298 (1.34%) IBD (population 1): 12/298 (4.03%)	2, 10
3	D'Souza 2021[72] and D'Souza 2021[17] NICE FIT October 2017 to December 2019	HM JACKarc analytical system Colonoscopy	4	AA: 421/982 (4.29%) IBD: 427/9822 (4.35%)	2, 10, 150
4	Gerrard 2023[75] Lothian, Scotland, UK January 2019 to February 2020	HM-JACKarc Endoscopy or CT with colorectal protocol	1	AA: 105/2260 (4.65%) and 136/3426 (3.97%) IBD: 59/226 (2.61%) and 55/3426 (1.61%)	10
5	Juul 2018[54] Central Denmark September 2015 to August 2016 NCT02308384	OC-Sensor DIANA Records follow-up	4	AA: 68/3462 (1.96%) IBD: 31/3462 (0.90%)	10
6	MacDonald 2022[56] NHS Lanarkshire, Scotland, UK October 2016 to February 2019	HM-JACKarc Records follow-up	1	AA: 47/5250 (0.90%) IBD: 131/5250 (2.50%)	10
7	MacLean 2021[58] Royal Surrey Foundation Trust, UK July 2019 and March 2020	QuikRead go Colonoscopy, CTC or flexisig	2	AA: 29/553 (5.24%) IBD: 9/553 (1.63%)	10, 100, 150
8	Mowat 2016[62] NHS Tayside, UK October 2013 to March 2014	OC-Sensor iO Colonoscopy	4	AA: 40/750 (5.33%) IBD: 34/750 (4.53%)	4, 10
9	Mowat 2021[61] and 2019[60] NHS Tayside, UK December 2015 to December 2016	HM JACKarc Records follow-up	4	AA: 133/1447 (9.19%) IBD: 68/1447 (4.70%)	10

N, number.

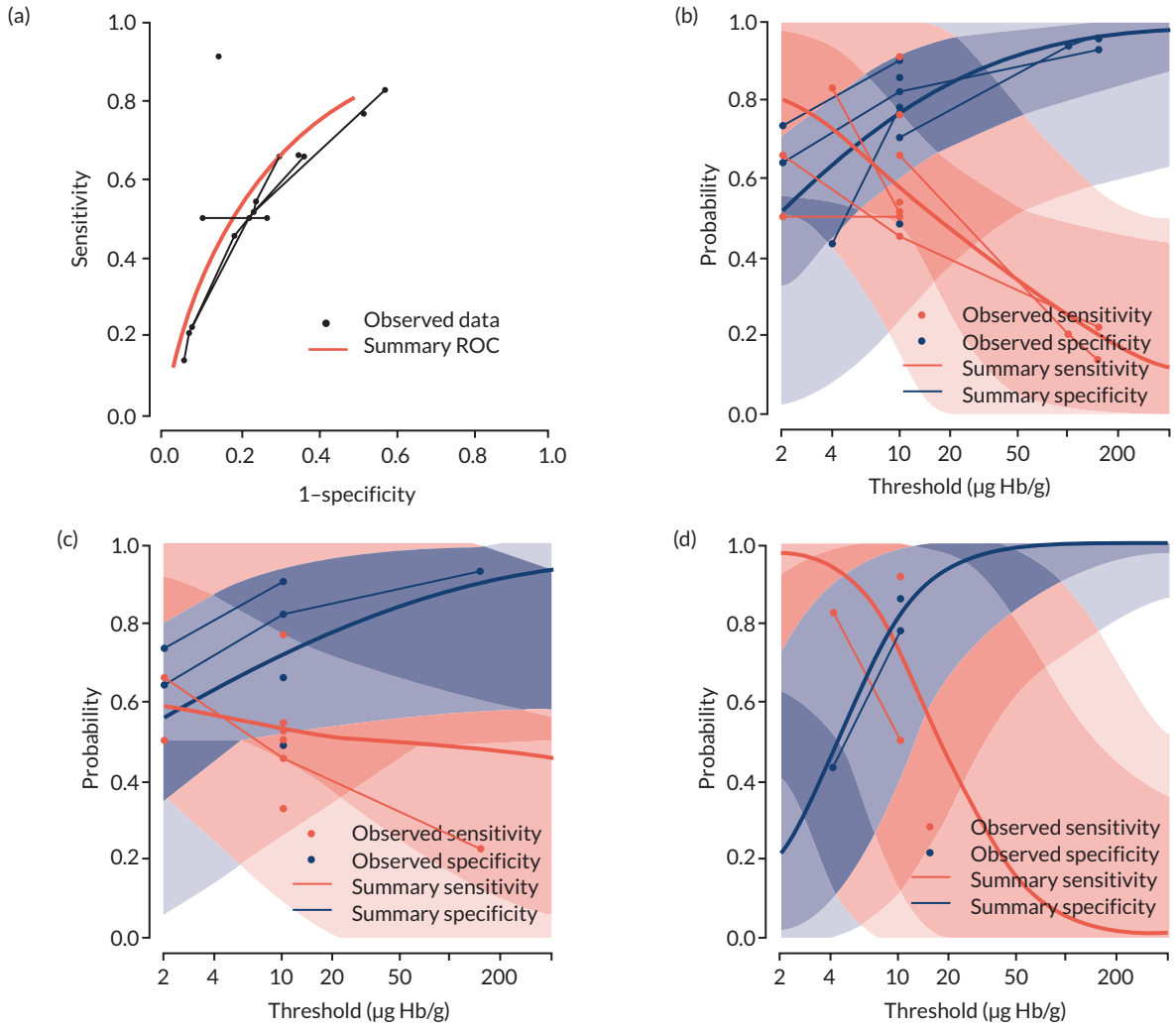

FIGURE 19 Observed data and summary sensitivity and specificity for AA outcomes. (a) All tests, ROC; (b) all tests as a function of threshold; (c) HM-JACKarc; and (d) OC-Sensor. Results for OC-Sensor use an informative prior, based on the synthesis of all tests together.

TABLE 55 Summary sensitivity and specificity at selected thresholds for AA outcome

	All tests (*S* = 9)		HM-JACKarc (*S* = 6)		OC-sensor (*S* = 2)	
Threshold, µg/g	Sensitivity, % (95% CrI)	Specificity, % (95% CrI)	Sensitivity, % (95% CrI)	Specificity, % (95% crI)	Sensitivity, % (95% CrI)	Specificity, % (95% crI)
2	80.4 (55.8 to 98.3)	51.6 (31.6 to 71.1)	59.1 (50 to 92)	55.9 (35.1 to 80.6)		
2.5	78 (55.2 to 97.4)	55.4 (36.1 to 74.4)	58.4 (50 to 90.5)	58.1 (38.4 to 82.9)		
3	75.9 (54.7 to 96.4)	58.4 (39.6 to 77.1)	57.7 (50 to 89.3)	59.9 (41 to 84.6)		
4	72.2 (53.7 to 93.8)	63.1 (45.6 to 81)	56.7 (49.8 to 86.8)	62.8 (45.1 to 87.3)	93.9 (51.5 to 100)	46.8 (9.5 to 90.3)
7	63.9 (51.4 to 84.6)	71.7 (55.3 to 87.7)	54.7 (48.2 to 81.2)	68.5 (50.8 to 91.7)	84.6 (27.8 to 100)	70.7 (27.2 to 96.7)
10	57.7 (48.6 to 76.7)	76.5 (60.3 to 90.9)	53.2 (45.9 to 77.6)	71.9 (52 to 93.8)	73.2 (10.1 to 99.9)	82.2 (41.6 to 98.7)
20	47.4 (26.1 to 64.4)	84.2 (68.1 to 95.3)	50.9 (37.3 to 71.6)	77.9 (53.7 to 96.5)		
50	34.1 (5.6 to 53.2)	91.1 (75.7 to 98.2)	49.8 (24.3 to 65.7)	84.4 (55.4 to 98.5)		
100	25 (1.4 to 48.9)	94.4 (80.2 to 99.2)	48.7 (16 to 61.9)	88.2 (56.3 to 99.2)		
120	22.8 (1 to 48.3)	95 (81.3 to 99.3)	48.3 (14.2 to 61.1)	89.1 (56.6 to 99.4)		
150	20.4 (0.6 to 47.5)	95.7 (82.5 to 99.5)	47.8 (12.3 to 60.1)	90.1 (56.9 to 99.5)		

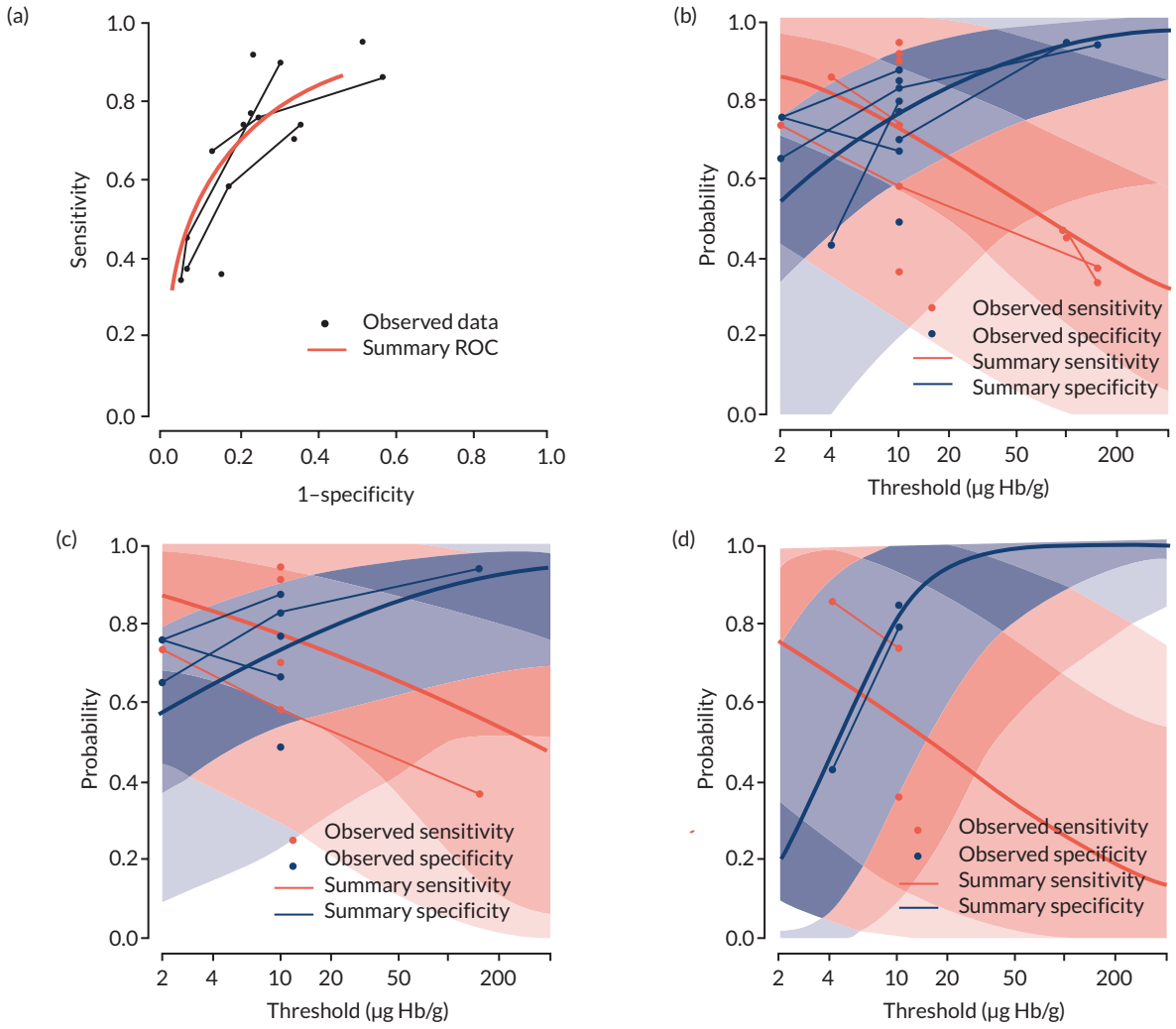

FIGURE 20 Observed data and summary sensitivity and specificity for all tests. IBD outcomes.

TABLE 56 Summary sensitivity and specificity at selected thresholds for IBD outcomes

	All tests (*S* = 9)		HM-JACKarc (*S* = 6)		OC-sensor (*S* = 2)	
Threshold, µg/g	Sensitivity, % (95% CrI)	Specificity, % (95% CrI)	Sensitivity, % (95% CrI)	Specificity, % (95% CrI)	Sensitivity, % (95% CrI)	Specificity, % (95% CrI)
2	85.7 (70 to 96.7)	53.8 (33.1 to 75.5)	86.8 (68.4 to 98.5)	57 (36.4 to 78.7)		
2.5	84.3 (68.5 to 96)	57.2 (37.2 to 78.4)	85.8 (67.3 to 98.1)	59.4 (39.2 to 80.6)		
3	83.1 (67.2 to 95.3)	60 (40.5 to 80.6)	84.9 (66.4 to 97.8)	61.3 (41.3 to 82.1)		
4	81 (65.1 to 94)	64.2 (45.8 to 84)	83.4 (64.7 to 97.1)	64.3 (44.8 to 84.4)	67 (24.7 to 97.9)	46.4 (7.4 to 92)
7	76.3 (60.4 to 90.7)	72 (54.7 to 89.4)	80.1 (61.2 to 95.3)	69.9 (50.8 to 88.1)	59.8 (16.4 to 95.5)	70.3 (22.3 to 97.5)
10	72.9 (57.1 to 88.2)	76.4 (59.2 to 92.1)	77.6 (58.6 to 94)	73.3 (53.3 to 90.2)	55.1 (12.2 to 93.1)	81.9 (35.3 to 99)
20	65.3 (49.2 to 82.9)	83.6 (66.3 to 95.8)	72.3 (52.2 to 91.1)	79.2 (57.3 to 93.4)		
50	54.4 (33.6 to 75.5)	90.3 (73.7 to 98.3)	64.9 (35.1 to 86.9)	85.4 (61.7 to 96.2)		
100	46.3 (21.7 to 69.7)	93.6 (78 to 99.2)	59.2 (21.3 to 83.4)	89.1 (64.3 to 97.5)		
120	44.2 (19 to 68.1)	94.3 (79 to 99.3)	57.7 (18.3 to 82.4)	89.9 (64.9 to 97.8)		
150	41.7 (15.9 to 66.1)	95 (80.2 to 99.5)	55.7 (15.2 to 81.2)	90.8 (65.8 to 98.1)		

'Time to' outcomes

Eight studies (nine publications)[18,45,49,60,61,65,69,75,81] reported other outcome data listed in the NICE scope. It should be noted that, in accordance with the protocol, data relating to these outcomes were sought only from studies included in the diagnostic test accuracy review. The data are summarised in *Table 58*.

'Time to' outcomes

Six studies[18,45,65,69,75,81] reported data on the time to different points in the diagnostic pathway for patients receiving FIT. Among four studies[18,45,65,75] relating to single FIT, one[45] reported time to return FIT result (median 7 days, IQR 4–11 days), another[18] reported time to analysis of FIT (averaged 10.1 days), one[75] reported time to investigation (median 21 days, IQR 11–43 days) and one[65] reported time to diagnosis (median 59 days, range 8–114 days). One of these also reported that 12 of the 15 patients who had a negative FIT but who had CRC were referred within 2 months, 9 of whom were diagnosed within 2 months, and that the median time to diagnosis for the 15 patients was 51 days (IQR 36.5–174.5 days), indicating that some patients experience a relatively long delay before receiving a diagnosis. Another study[69] using single FIT reported a number of outcomes (see *Table 58*) for patients who tested negative using FIT (in this study the threshold was < 20 µg/g), but who were eventually diagnosed with CRC. Three categories were reported: FIT < 4 µg/g, FIT 4–9.9 µg/g and FIT 10–19.9 µg/g. Median time to diagnosis was < 90 days in all categories, although the IQR was as high as 456.5 in the < 4 µg/g subgroup and time to diagnosis was extremely long (> 1000 days) for a minority of patients and especially among those with FIT < 10 µg/g. This study also reported stage at diagnosis for those with missed diagnoses, which is difficult to interpret without comparative data. This study also reported diagnoses in those who failed to return their FIT, and this rate was 1%.

Two studies[75,81] reported time to outcomes for dual FIT. One[75] reported a small increase in the median number of days to investigation for dual FIT (median 26,[a] IQR 17–45 days) versus single FIT (median 21 days, IQR 11–43 days; $p < 0.050$). The other study[81] reported a median interval of 6 days (IQR 5–8 days) between FIT samples.

Other outcomes

One study[60,61] reported a number of outcomes after introducing FIT into their diagnostic pathway using a threshold of 10 µg/g (see *Table 58*). Notably, they reported a 9.2% reduction in referrals to colorectal services from 4303 in previous year to 3905 after the introduction of FIT, and similarly a 24.1% reduction in gastroenterology outpatient referrals from 2796 in previous years to 2121 after the introduction of FIT. They also report one emergency presentation out of 5372 who had FIT.

Patient perspectives

Articles identified by the searches described earlier (in *Search strategy*) were sifted for patient-reported outcomes of patient acceptability. Patient-reported outcomes sought were patient views on the acceptability of FIT, expressions of patient preference for FIT versus colonoscopy, and the experience of, and satisfaction with, FIT among patients with suspected CRC symptoms.

Two studies were identified that investigated patient acceptability: Georgiou Delisle *et al.*[32] and MacLean *et al.*[33] (see *Table 59*). Both recruited a subset of patients from studies included in this report as diagnostic test accuracy studies (in *Results*). Georgiou Delisle *et al.*[32] recruited participants from the NICE FIT study. MacLean *et al.*[33] recruited participants from the POC FIT study. Both of these studies included UK patients referred under the 2WW pathway with suspected CRC symptoms (population type 4).

Both studies designed surveys for their study, rather than using pre-existing surveys. Both studies used a Likert scale of 1–5. Georgiou Delisle *et al.*[32] designed the survey based on a literature review of previous questionnaires, with input from study authors, experts and a patient panel, and MacLean *et al.*[33] designed the survey with input from study authors and expert academics.

The two studies did not investigate the same themes. Georgiou Delisle *et al.*[32] investigated the feasibility of FIT; patient feelings of faecal aversion related to using FIT; knowledge in relation to bowel cancer; and future test. Twenty-one statements were included in the questionnaire. The themes investigated by MacLean *et al.*[33] were expectations, satisfaction that colonoscopy/CTC would rule out CRC C and satisfaction if FIT results had meant avoiding colonic investigation, and patient experience. There were five questions in this survey.

TABLE 57 Studies issuing FIT in primary care or issuing dual FIT, and reporting test failure rates, test uptake and number of repeat tests

Author, year	Analyser	FIT provided in	*N* with CRC/*N* analysed (%)	Invalid/test failure rates	Test uptake/ non-return	Repeat tests
Johnstone 2022[52]	HM-JACKarc (personal communication)	Primary care	61/4737 (1.29%)	231/4968 (4.6%)	NR	NR
MacDonald 2022[56]	HM-JACKarc	Primary care, those undergoing referral	151/5250 (2.88%)	Rejected for technical reasons 115 (2.1%)	NR	NR
Mowat 2021[61] and 2019[60]	HM JACKarc	Primary care	105/5381 (1.95%)	Unsuitable for analysis, *n* = 152/5422 (2.8%)	NR	*n* = 112/5422 (2.07%) repeat tests
Johnstone 2022[53] Symptomatic patients who had two FITs between 1 week and 1 year apart	HM-JACKarc	Primary care	42/5761 individuals (0.73%)	NR	NR	12,359/135 396 (9.1%) repeat FITS in total, from 5761 individuals. FITs between 1 week and 1 year apart: 2 FITs: *n* = 5027 3 FITs: *n* = 649 4 FITs: *n* = 71 5 FITs: *n* = 10 6 FITs: *n* = 4
Bailey 2021[42] and Bailey 2024[69]	OC-Sensor iO	Primary care	15,589 FIT requests (CRC NR)	34/15,589 (0.2%) spoiled or not suitable for analysis	Kit not returned 1393/14,788 (9.4%) Updated analysis: 3631/38,920 (9.3%)[69]	229/13,361 (1.7%) Updated analysis: 8349 (17.0%) requests were repeat tests in 6640 patients from 40,817 patients[69]
Cama 2022[45]	OC-Sensor iO	Primary care	74/5341 (1.39%)	No result returned in 2% of samples (*n* = 13,466)	NR	NR
Georgiou Delisle 2022[49]	OC-Sensor iO	Primary care	61/4187 (1.46%)	Could not be processed: 948/5050 (18.8%)[a]	NR	NR
Ball 2022[43]	OC-Sensor PLEDIA	Primary care	17/2892 (0.59%)	*n* = 599/4219 (14.2%) due to insufficient clinical details, sample errors, insufficient ID/ labelling[b]	NR	*n* = 29/4219 (0.7%)
Juul 2018[54]	OC Sensor DIANA	Primary care	54/3462 (1.56%)	Invalid FIT = 91/3745 (2.4%)	NR	NR
Mowat 2016[62]	OC Sensor iO	Primary care	28/750 (3.73%)	*n* = 5/2789 (0.2%) spoiled/ unsuitable samples	FIT not returned: 1130/2173[c] (52.0%)	

TABLE 57 Studies issuing FIT in primary care or issuing dual FIT, and reporting test failure rates, test uptake and number of repeat tests (*continued*)

Author, year	Analyser	FIT provided in	*N* with CRC/*N* analysed (%)	Invalid/test failure rates	Test uptake/ non-return	Repeat tests
Jordaan 2023[83]	FOB Gold	Primary care	30/3349 (0.90%)	*N* = 610/3959 (15.4%) could not be analysed Reasons: 55% – buffer loss;[d] 2.15% wrong container;[d] 4.7% – no label; 4.0% overfilled; 4.2% other reasons	NR	392/3959 (9.9%) were test failures; 64.2% (*n* = 392) of test failures completed their retest
Subgroups						
Ayling 2019[68]	OC Sensor	Secondary care	Low haemoglobin group: 7/178 (3.93%) IDA group: 6/137 (4.38%)	6/184 (3.3%) FIT unusable	NR	NR
Dual FIT						
Gerrard 2023[75]	HM-JACKarc	Secondary care	88/2637 (3.34%)	Clinician considered FIT inappropriate, or emergency presentation predated FIT postage: 205/4559 (4.5%)	FIT not returned: 464/4354 (10.7%) Only one FIT returned: 891/4354 (20.5%)	NR
Hunt 2022[51]	OC-Sensor	Secondary care	317/28622 (1.11%)	NR	Only returned one FIT: 1482/30104 (4.9%)	NR
Tsapournas 2020[80]	QuikRead go	Secondary care	13/242 (5.37%)	NR	Stool sample missing *n* = 57/355 (16.1%)	NR

N, number.

a Reason for incorrect FIT processing: sample labelling errors, *n* = 223 (5.3%); wrong sample type, *n* = 142 (2.8%); sample not processed, *n* = 102 (2%); wrong container type, *n* = 94 (1.9%); sample delivery error (no date of collection), *n* = 105 (2.1%); sample unlabelled, *n* = 97 (1.9%); sample volume error, *n* = 2 (0.04%); laboratory accident, *n* = 1 (0.02%); other, *n* = 97 (1.9%).

b Unclear what proportion due to each problem. Not all problems were inherent to FIT; for example, missing clinical details was important to study, but not to the processing of FIT in clinical care.

c Note that in this study, patients had already been referred, so there was less incentive to return the FIT if referral depended on FIT sample. In addition, two tests had to be done on one sample (one FIT, one faecal calprotectin).

d Buffer loss thought to be due to opening the tube at wrong end. This was thought by the Jordaan *et al.* authors not to be a problem with tubes used for a dedicated Sentifit analyser, but no data were supplied to support this view; the use of wrong container is thought likely to be due to mix-up with the calprotectin tube given at the same time.

TABLE 58 Studies reporting other outcomes listed in the NICE scope

Author, year	Analyser	FIT provided in	*N* with CRC/*N* analysed (%)	'Time to' outcomes	Other outcomes
Bailey *et al.* 2024[69]	OC-Sensor iO	Primary care	561/35,289 (1.59%)	Time to diagnosis for false-negative FIT, median (IQR) • FIT < 4 µg/g, with CRC (*n* = 26): 83.5 days (39.5–456.5), max 1023 days • FIT 4–9.9 µg/g, with CRC (*n* = 37): 83.0 days (44.5–192.5), *n* = 3 > 1000 days • FIT 10–19.9 µg/g (*n* = 25): 41.0 days (26.5–78.0) • FIT < 20 µg/g (*n* = 88): 64.0 (34.5–212.5), 23/88 > 180 days	Stage at diagnosis: In the delayed group, 8 (34.8%) patients had Stage I disease at diagnosis, 4 (17.4%) Stage II, 6 (26.1%) Stage III, 4 (17.4%) Stage I and in 1 cancer staging was unavailable CRC in patients who did not return FIT: 38/3631 (1%) CRC in patients with repeat test: 62/6640 (0.9%)
Cama 2022[45]	OC-Sensor iO	Primary care	74/5341 (1.39%)	Time to return FIT result: median 7 days (IQR 4–11 days) Diagnostic delay due to negative FIT (*n* = 15): • < 2-month delay to referral: *n* = 12/15 • < 2-month delay in diagnosis: *n* = 9/15 • Time from negative FIT to CRC diagnosis (*n* = 15): median 51 days (IQR 36.5–174.5 days)	
D'Souza 2020[18]	HM-JACKarc	Secondary care	12/298 (4.03%)	Time to analysis of FIT: averaged 10.1 days	No AEs were reported from patients undergoing FIT or colonoscopy
Georgiou Delisle 2022[49]	OC-Sensor iO	Primary care	61/4187 (1.46%)	NR	Urgent 2WW referrals: 1438/4187 FITs or 2060/5672 patients presenting to primary care
Gerrard 2023[75]	HM-JACKarc	Secondary care		Time to investigation: median 21 (IQR11–43) days	NR
Mowat 2021[61] and 2019[60]	HM JACKarc	Primary care	105/5381 (1.95%)	NR	• FIT < 10 µg/g emergency presentations: *n* = 1/5372 who had FIT • Referred to secondary care: *n* = 2848/5372 • Followed up in primary care (no immediate referral): *n* = 2521/5372 • Triaged to colonoscopy: *n* = 1381/5372 • Triaged to gastroenterology: *n* = 672/5372 • Triaged to sigmoidoscopy: *n* = 462/5372 • Triaged to colonoscopy: *n* = 83/5372 • Triaged to other assessment: *n* = 179/5372 • Routine colonoscopy: *n* = 345/1381 colonoscopy • Urgent colonoscopy: *n* = 617/1381 colonoscopy, of which *n* = 419 for suspected cancer – also reports upgrading and downgrading due to FIT result • Not referred to colonoscopy after review by gastroenterologist: *n* = 71/5660 • Referrals to colorectal services: 9.2% reduction from 4303 in previous year to 3905 • Gastroenterology outpatient referrals: 24.1% reduction from 2796 in previous years to 2121

TABLE 58 Studies reporting other outcomes listed in the NICE scope (*continued*)

Author, year	Analyser	FIT provided in	*N* with CRC/*N* analysed (%)	'Time to' outcomes	Other outcomes
Tang 2022[65]	HM-JACKarc			Time to diagnosis: median 59 days, range 8–114 days	NR
Dual FIT					
Gerrard 2023[75]	HM-JACKarc	Secondary care	88/2637 (3.34%)	Time to investigation: median 26[a] (IQR 17–45)	NR
Turvill 2018[81]	HM-JACKarc	Secondary care	27/476 (5.67%)	Time to laboratory (first sample): median 7.7hours (IQR 4.9–16.7) Time to laboratory (second sample): median 6.6 hours (IQR 4.5–14.5) Time between samples: median 6 days (IQR 5–8)	NR

N, number.
a $p < 0.050$ vs. single FIT.

TABLE 59 Study characteristics of patient acceptability studies

Author, date Study design	Study aim	Population	Sample size	Questionnaire used
Georgiou Delisle *et al.* 2022[32] Cross-sectional survey by postal questionnaire Subset of the NICE FIT study	To investigate attitudes and perception of FIT in symptomatic patients	From the NICE FIT study. UK patients referred under the 2WW pathway with suspected CRC symptoms. Patients may or may not have completed FIT or had colonic investigation	Questionnaires sent by post 3760. Questionnaires returned and analysed, *n* = 1151 (30.6% completion rate)	Developed for the study. Based on literature review of previous questionnaires, with input from study authors, experts and a patient panel. Likert scale 1–5 21 statements and four themes: feasibility of FIT; patient feelings of faecal aversion towards FIT; knowledge in relation to bowel cancer; and future test intentions
MacLean *et al.* 2022[33] Cross-sectional survey by telephone questionnaire Subset of the POC FIT study	To investigate patient opinions of FIT	From the POC FIT study. UK patients referred under the 2WW pathway with suspected CRC symptoms. All had both FIT and colonic investigation	Contacted by telephone, *n* = 117. Answered survey, *n* = 109 (93% completion rate)	Developed for the study. Design by study authors and expert academics. Likert scale 1–5 5 questions; themes: expectations; satisfaction that colonoscopy/CTC would rule out CRC; and satisfaction if their FIT results had meant avoiding colonic investigation patient experience

Georgiou Delisle *et al.*[32] summarised the results of themes by converting into binary, that is, positive (strongly agree, agree) and non-positive (neutral, disagree, strongly disagree). MacLean *et al.*[33] reported positive responses in a similar manner and also reported mode and median scores.

Study results

Georgiou Delisle *et al.*[32] sent out 3760 questionnaires, and 1151 (30.6%) were returned and analysed, whereas MacLean *et al.*[33] contacted 117 people, of whom 109 (93%) completed the survey. The difference in response rates can be explained by the fact that the Georgiou Delisle *et al.*[32] study posted questionnaires alongside a FIT kit, whereas MacLean *et al.* telephoned participants who had already engaged with services in both returning a FIT and undergoing colonic investigation, and had fewer questions.

Patient demographics were similar in the studies in terms of age, with 57.8% aged ≥ 65 years in Georgiou Delisle *et al.*[32] and 72.4% aged ≥ 60 years in MacLean *et al.*,[33] and sex, with 54.5% female in Georgiou Delisle *et al.* and 56.9% female in MacLean *et al.* (see *Table 60*).

Both studies addressed the usability of FIT (see *Tables 61* and *62*). In the Georgiou Delisle *et al.* study,[32] 95.9% patients gave a positive response (agreed or strongly agreed) that the device was easy to use, and 90.2% said that the sample was easy to collect. In the MacLean *et al.* study,[33] 88% gave a positive response to ease of use of the sampling device. Georgiou Delisle *et al.*[32] also found that 96.3% of patients found the instructions understandable.

Although there were patient feelings of faecal aversion, these were found to be able to be overcome by the patients of the Georgiou Delisle *et al.* study,[32] with 79.2% disagreeing that it was difficult to overcome embarrassment, 77.0% overcoming disgust, and 76.3% disagreeing that collecting a stool sample for FIT is unhygienic. When asked if they would prefer FIT to colonoscopy, 78.1% of patients agreed or strongly agreed, and in the Georgiou Delisle *et al.* study,[32] 95.9% agreed/strongly agreed that they would use FIT again in the future.[32]

MacLean *et al.* asked about satisfaction of clinical outcome to rule out CRC, and found that 51% of patients were satisfied/completely satisfied that if their FIT was negative they need not undergo colonic investigations;[33] 14.6% were neutral on this question, and 32.1% were unsatisfied/completely unsatisfied with not being referred for colonic investigation.[33]

Although the questions were asked in a different way, it appears that a higher proportion of patients in the Georgiou Delisle *et al.* study[32] had confidence in FIT, with 78.1% preferring it to colonoscopy, whereas only 51% of patients from the MacLean *et al.* study[33] would have been satisfied that negative FIT could rule out the need for colonic investigation. The difference in patients could explain this, as all those in the MacLean *et al.* study[33] had undergone colonic investigation already, whereas patients in the Georgiou Delisle *et al.* study[32] had not. Equally, the wording of the question may have elicited different responses.

TABLE 60 Patient characteristics of patient acceptability studies

Author, date	Population	Patient age (years)	Patient sex	Ethnicity
Georgiou Delisle *et al.* 2022[32] From the NICE FIT study	*n* = 1151 Patients completing FIT alongside survey 99.2%; patients with prior stool test 71.7% Unclear how many had prior experience of colonic investigation; survey sent with FIT prior to colonic investigation (if needed)	Mean 65 25–39 2.4%; 40–64 39.7%; ≥ 65 57.8%	Male 45.4%; female 54.6%	White 88.0%; non-White 12.0%
MacLean *et al.* 2022[33] From the POC FIT study	*n* = 109 Patients completing FIT prior to survey 100%; patients with colonic investigation prior to survey 100% (colonoscopy 46.8%; CT colonography 45.9%; flexible sigmoidoscopy 7.3%)	Age 20–39 1.8%; 40–59 25.7%; 60–79 65.1%; ≥ 80 7.3%	Male 43.1%; female 56.9%	NR

TABLE 61 Patient acceptability results, Georgiou Delisle *et al.*[32]

Author, date	Theme measured	Results, % giving positive answer
Georgiou Delisle *et al.* 2022[32] (from the NICE FIT study)	Feasibility of FIT	Instructions understandable 96.3% (95% CI 95.1% to 97.3%) Easy to use device 95.9% (95% CI 94.6% to 96.9%) Would return FIT by post (rather than via GP) 90.5% (95% CI 88.6% to 92.0%) Straightforward to collect 90.2% (95% CI 88.3% to 91.8%) Prefer FIT to colonoscopy 78.1% (95% CI 75.6% to 80.4%)
Georgiou Delisle *et al.* 2022[32] (from the NICE FIT study)	Patient feelings of faecal aversion towards FIT	Could overcome embarrassment 79.2% (95% CI 76.7% to 81.4%) Could overcome disgust 77.0% (95% CI 74.9% to 79.4%) FIT wasn't unhygienic 76.3% (95% CI 73.7% to 78.6%)
Georgiou Delisle *et al.* 2022[32] (from the NICE FIT study)	Knowledge in relation to bowel cancer	Optimistic about cure if detected early 93.0% (95% CI 91.4% to 94.4%) Worried about getting CRC 78.0% (95% CI 75.5% to 80.4%) Thought having family history of CRC carried increased risk 75.1% (95% CI 72.5% to 77.5%)
Georgiou Delisle *et al.* 2022[32] (from the NICE FIT study)	Future test intentions	Understood purpose of FIT 98.2% (95% CI 97.3% to 98.9%) FIT's ability to detect cancer important deciding factor 97.3% (95% CI 96.1% to 98.1%) Would use FIT again 95.9% (95% CI 94.9% to 96.9%) Felt my future health influences my behaviour today 93.5% (95% CI 91.9% to 94.8%)

TABLE 62 Patient acceptability results, MacLean *et al.*[33]

Author, date	Theme measured	Results, mode, median; positive response
MacLean *et al.* 2022[33] (from the POC FIT study)	Expectations	How much expected to be referred to colonic investigation (1 least expected, 5 most expected): mode 5, median 4; positive response 60%
MacLean *et al.* 2022[33] (from the POC FIT study)	Satisfaction	Satisfaction that colonic investigation could rule out CRC (1 completely unsatisfied, 5 completely satisfied): mode 5, median 5; positive response 93% If FIT negative, satisfaction to not undergo colonic investigation (1 completely unsatisfied, 5 completely satisfied): mode 5, median 4; positive response 51%
MacLean *et al.* 2022[33] (from the POC FIT study)	Patient experience	Ease of use of stool-sampling device (1 very difficult, 5 very easy): mode 5, median 5; positive response (easy) 88% Ease of colonic investigation (1 very difficult, 5 very easy): mode 5, median 4; positive response (easy) 78%

Georgiou Delisle *et al.* analysed responses in relation to covariates.[32] They found that patients were less likely to prefer to use FIT rather than undergo a colonoscopy if they were aged 40–64 years (rather than ≥ 65 years) (odds ratio 0.60, 95% CI 0.43 to 0.84) or lived in London (rather than outside London) (odds ratio 0.50, 95% CI 0.36 to 0.71). Patients were more likely to say they would use FIT in the future if they were White (odds ratio 3.20, 95% CI 1.32 to 7.75) or had prior experience of stool tests (odds ratio 2.06, 95% CI 1.03 to 4.13).[32] Patients were more likely to prefer to use FIT rather than undergo a colonoscopy if they returned a FIT that was successfully analysed to produce an f-Hb result (odds ratio 4.32, 95% CI 1.49 to 12.52), and more likely to say they would use FIT in the future if they had successfully used a FIT (odds ratio 11.08, 95% CI 2.74 to 44.75). However, only 15 patients did not complete the test successfully, so the small sample size means that the results should be interpreted with caution.[32] MacLean *et al.*[33] found that those who went on to receive CT colonography would have been less satisfied using FIT than those that went on to receive both colonoscopy (median score of 3) and sigmoidoscopy (median score of 4). Female patients would have been less satisfied using FIT alone (median score of 3) than male patients (median score of 4).[33]

In the Georgiou Delisle *et al.* study, nine patients returned the questionnaire but not the FIT kit.[32] These patients showed similar results to those returning the FIT kit (88.9% found it easy to collect a sample, 88.9% disagreed that FIT was unhygienic); however, the small sample size means that the results should be interpreted with caution.

The authors' conclusions were that most patients found FIT acceptable,[32,33] but strategies are needed to engage patients with more negative views of FIT,[32] and shared decision-making of patient and clinician should be considered for patients dissatisfied with relying on FIT results to decide on the need for further investigation.[33]

Sociodemographic factors

One conference abstract on FIT return rates across demographic subgroups in patients with suspected CRC symptoms was identified by the searches described earlier (see *Search strategy*).[138] This study was updated by an in-press article, Bailey *et al.*,[31] which was submitted by one of the authors who was a stakeholder for this assessment.

The Bailey *et al.*[31] study investigated FIT return in UK adult patients with suspected CRC symptoms, with the aim of identifying whether demographics, ethnicity or social deprivation affect FIT return rates.[31] The study was a retrospective review of records within NHS Nottingham and Nottinghamshire Clinical Commissioning Group (see *Table 63*). Data had been recorded prospectively on all adult patients presenting to primary care with suspected CRC symptoms, excluding those with rectal bleeding or mass, who were sent FIT kits by post.[31] Up to 14 days were allowed between FIT request and being defined as non-return.[31] As further FIT requests could be made to non-returners, only the first FIT request to each patient was included in the return rate analysis.[31] Exclusion criteria for the analysis were rectal bleeding or mass; duplicate request; request from out of area; sampling error; incomplete request; not indicated under 18 years old; and incomplete records.[31] Socioeconomic data were derived from six-digit postcodes using the Index of Deprivation tool.[31]

Study results

The results of the study are summarised in *Table 64*. For the overall population, the FIT return rate was 35,289 out of 38,920 (90.7%). The median age was 66 years, 70.1% were White, and more patients were from the least deprived quintile (28.4%) than any of the other socioeconomic quintiles.

The results of the multivariate analysis of non-returns showed that there were differences in return rate for sex, age, ethnicity and level of socioeconomic deprivation.[31] There was a higher return rate for female patients (91.0%) than for male patients (90.2%) (by multivariate analysis, odds ratio of non-return for male with reference female, OR 1.11, 95% CI 1.03 to 1.19).[31] There was a higher return rate for patients aged ≥ 65 years (91.9%) than for patients aged < 65 years (89.2%) (OR for non-return aged ≥ 65 years with reference aged < 65 years, OR 0.78, 95% CI 0.72 to 0.83).[31] There was a higher return rate for White patients (91.2%) than for Asian (83.8%) (OR 1.82, 95% CI 1.58 to 2.10), Black (86.6%) (OR 1.21, 95% CI 0.98 to 1.49) and mixed/other ethnic groups (87.2%) (OR 1.29, 95% CI 1.05 to 1.59).[31] There was a higher return rate for the least socioeconomically deprived quintile (93.6%) than for more socioeconomically deprived groups, with the most socioeconomically deprived quintile having a return rate of 86.3% (OR 2.20, 95% CI 1.99 to 2.43).[31]

Although not an equity study, Georgiou Delisle *et al.* reported lower rates of return for both questionnaire and FIT from sites in London than from sites outside London.[32] The questionnaire response rate was higher among older patients, but there were no significant differences for sex or deprivation; however, these data were for the questionnaire only, and demographics were not available by FIT return or non-return.[31]

Colorectal cancer was diagnosed in 599 patients in the Bailey *et al.* study, of whom 561 returned their first FIT and 38 were first FIT non-returners.[31]

TABLE 63 Study characteristics of equity study[31]

Author, date	Study design	Setting	Population	Sample size	Outcome
Bailey *et al.* 2023[31]	Observational, retrospective review of records	FIT as a triage tool in primary care, NHS Nottingham and Nottinghamshire Clinical Commissioning Group, UK, November 2017 to December 2021	Adult patients with suspected CRC symptoms (excluding rectal bleeding/mass)	First FIT requests for 38,920 individual patients	FIT return in symptomatic patients, by demographics, ethnicity and social deprivation

TABLE 64 Results of equity study by Bailey *et al.*[31]

Demographic variable	Demographic category	Population, *n* (% of participants in category)	Returned FIT, *n* (% of participants returned FIT)	Non-returned FIT, *n* (% of participants non-return)	OR of non-return (95% CI), multivariate logistic regression analysis[a]
Sex	Female	21,800 (56)	19,841 (91.0)	1959 (9.0)	Reference
Sex	Male	17,112 (44)	15,442 (90.2)	1670 (9.8)	1.11 (1.03 to 1.19)
Sex	Unknown	8 (0.0)	6 (0.0)	2 (0.0)	NA
Age	< 65 years	18,029 (46.3)	16,080 (89.2)	1949 (10.8)	Reference
Age	≥ 65 years	20,891 (53.7)	19,209 (91.9)	1682 (8.1)	0.78 (0.72 to 0.83)
Age	Unknown	0 (0.0)	0 (0.0)	0 (0.0)	NA
Ethnicity	White	27,277 (70.1)	24,864 (91.2)	2413 (8.8)	Reference
Ethnicity	Asian	1584 (4.1)	1328 (83.8)	256 (16.2)	1.82 (1.58 to 2.10)
Ethnicity	Black	801 (2.1)	694 (86.6)	107 (13.4)	1.21 (0.98 to 1.49)
Ethnicity	Mixed/other	876 (2.3)	764 (87.2)	112 (12.8)	1.29 (1.05 to 1.59)
Ethnicity	Unknown	8382 (21.5)	7639 (91.1)	743 (8.9)	0.99 (0.90 to 1.08)
Social deprivation	Fifth quintile (least deprived)	11,036 (28.4)	10,328 (93.6)	708 (6.4)	Reference
Social deprivation	Fourth quintile	6278 (16.1)	5808 (92.5)	470 (7.5)	1.18 (1.04 to 1.33)
Social deprivation	Third quintile	6454 (16.6)	5885 (91.2)	569 (8.8)	1.39 (1.24 to 1.56)
Social deprivation	Second quintile	6177 (15.9)	5521 (89.4)	656 (10.6)	1.68 (1.50 to 1.87)
Social deprivation	First quintile (most deprived)	8927 (22.9)	7703 (86.3)	1224 (13.7)	2.20 (1.99 to 2.43)
Social deprivation	Unknown	48 (0.1)	44 (91.8)	4 (8.2)	1.28 (0.46 to 3.57)

NA, not applicable.
a Variables in the multivariate logistic regression analyses were gender, age, ethnicity and socioeconomic deprivation. OR, higher numbers reflect higher *non-return rate* (i.e. return rate), Numbers in bold indicate that CI does not cross 1.

The authors conclusion was that there is a need to find strategies to mitigate the lower FIT return rates in patients with suspected CRC symptoms who are male; aged < 65 years; from Asian, Black or mixed/other ethnic groups; or socioeconomically deprived.[31] Strategies may involve following up after FIT non-return, providing information in a range of languages, and offering counselling regarding the perceived risk of disease and success of treatment.[31]

Appendix 7 Review of existing published health economic analyses

Cost-effectiveness and health-related quality of life review: methods

Systematic searches were undertaken to identify existing economic evaluations of the use of FIT in people presenting to primary care with symptoms of CRC. As a systematic review of the literature on this topic had been performed for the previous appraisal of FIT for patients with suspected CRC (NICE DG30),[11] the EAG's searches included only studies relevant to the decision problem that had been published since the previous assessment. The main focus of this review was to explore methodological choices made in previous economic evaluations and their potential relevance to the current decision problem and the model being developed by the EAG, rather than to assess the individual results of published economic evaluations.

Search strategy

A comprehensive search was undertaken to systematically identify economic evaluations of FIT in people with symptoms of CRC. A combined search was also performed using similar search strategies to identify HRQoL studies in the relevant population.

Literature searches were undertaken to identify economic evaluations and studies reporting utility estimates in people with symptoms of CRC were undertaken in February 2023 in the following electronic databases:

- MEDLINE(R) and Epub ahead of print, In-Process Citations & Daily Update (via Ovid), 1946 to 22 February 2023
- EMBASE (via Ovid), 1974 to week 7 2023
- EconLit (via Ovid), 1886 to 9 February 2023
- The Cochrane Library (via Wiley), December 2022 to 23 February 2023
- Tufts' CEA Registry (https://research.tufts-nemc.org/cear4/Home.aspx), from 2016 to 23 February 2023

Searches on the Research Papers in Economics (RePEc) (http://repec.org/) database were not carried out due to time and technical constraints; however, the EAG believes that this is likely to have had only a minor impact on the final results of the review.

The search strategies comprised MeSH and Emtree terms and free-text synonyms for FIT (including terms for each FIT brands), colorectal cancer, and (1) economic or (2) HRQoL with free-text synonyms for 'EQ-5D'. Searches were translated across databases and were not limited by language. Searches were limited to results since 2016, considering the date of the last systematic review of this topic undertaken to inform NICE DG30.[11] As the Cochrane Library had already been searched in December 2022 for the clinical review, an update search was run to identify any new studies added between December 2022 and February 2023.

Methodological study type search filters to identify economic evaluations were applied in MEDLINE and other databases where appropriate, and were based on the NHS EED filter and economic filter by the McMaster University HEDGES team (https://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx). The search strategies are presented in *Appendix 1*.

All references obtained were imported into reference management software [EndNote® version 20, Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA], with their respective bibliographic data and abstracts, where duplicate references were subsequently excluded.

Inclusion and exclusion criteria

Study selection was carried out in two stages, based on titles and abstracts, and full texts. Studies were required to meet the following criteria to be considered relevant for inclusion in the review:

- Full economic evaluations comparing interventions for CRC that included FIT.
- The population of the study should include the relevant population included in the final NICE scope (patients presenting to primary care with symptoms suggestive of CRC).
- Published in English.
- Available in full-text format (studies available in abstract form only were excluded from the review).

Other types of studies or publications (primary studies, in animal, in vitro or genetically based studies, letters to the editor or comments) and duplicated studies on the same model were excluded.

Data extraction and quality assessment

For economic studies, data extraction focused on (1) the indicated population, main results in terms of costs, consequences and the incremental cost-effectiveness of the alternatives compared; and (2) the modelling methods used, the sources of input parameters, the key modelling assumptions and the robustness of the study results. For HRQoL studies, data extraction focused on (1) the indicated population, the location of study and (2) the main results in reported EQ-5D valuations in patients with different CRC stages, and potentially for other events related to the diagnosis process.

The methodological quality of the included economic studies was assessed using published checklists for economic evaluations and modelling studies.[240]

Cost-effectiveness and health-related quality of life review results: summary of studies identified

The results of the searches and selection process for economic evaluations and HRQoL studies are presented as a PRISMA flow chart in *Figure 21*.

For economic evaluations, a total of 820 citations were initially identified after the exclusion of duplicates, with 792 being excluded at the title and abstract phase of the selection process. Most of the exclusions were of non-economic evaluation studies or economic evaluations undertaken in populations that differ from that described in the NICE scope (e.g. in asymptomatic patients). Twenty-three studies were reviewed at the full-text stage; however, none of these included FIT as part of the diagnostic options for symptomatic patients and they were deemed not relevant to the decision problem. A list of excluded studies and comments on each exclusion for both reviews are provided in *Report Supplementary Material 4*. These consisted of papers with only the abstract provided, editorial papers or comments, study types other than economic evaluations, studies in a different population from patients with symptoms suggestive of CRC presenting to primary care and studies that did not include FIT. Some studies were excluded for more than one reason (the most outstanding were considered for counting purposes). Two additional studies were retrieved from the HRQoL studies' review and included in the final review.[90,91]

For HRQoL studies, a total of 264 citations were initially identified after the exclusion of duplicates. At the title and abstract selection phase, 246 papers were excluded, while 18 were reviewed in full text. One study could not be retrieved by the EAG and was therefore excluded. None of the studies met the inclusion criteria (see *Figure 21*). The main reasons for excluding studies were that they did not report EQ-5D estimates, they were reported only as abstracts, or they reflected a different population from that listed in the NICE scope. Two studies were economic evaluations that were reviewed and included in the review of economic evaluations.

Tables 65 and *66* summarise the two included economic evaluations. Both studies were model-based cost–utility analyses that report the incremental cost per QALY gained for FIT compared with a variety of comparators as part of the diagnostic pathway for people with symptoms of CRC. Both studies were undertaken from the perspective of the NHS and PSS. The models included populations with initial ages ranging from 40 to 70 years.

Both included studies adopted similar general modelling approaches and structures. Both models combined a decision tree containing the diagnostic decision nodes with Markov models that estimate the lifetime costs and QALYs for patients with CRC based on states using CRC Duke's staging system, and a two-state model (alive–dead) for patients without underlying CRC. Westwood *et al.*[90] reported keeping similarities in their structure to the previously published model in NICE NG12,[10] while Medina-Lara *et al.*[91] provide a comparison of key model characteristics from previous models retrieved from their reviews.

Both models include only CRC patients presenting to primary care who are classified as DG30 low risk based on NG12/DG30 criteria, and do not include any other underlying lower bowel conditions. The diagnostic component of the model in both studies is based on the prevalence of disease (both assume that the prevalence of CRC in this population is 1.5%) and on the accuracy estimates of the tests used for detection of CRC. Both models assume that colonoscopy is a perfect diagnostic test, assuming its sensitivity and specificity to be 100%. Both models adopt a lifetime horizon, with cycles ranging from 28 days to 1 year. A list of assumptions adopted by the models and the sources of their key parameters is presented in *Tables 65* and *66*.

The quality assessment of the included studies is presented in *Report Supplementary Material 4*. Considering that the models identified by the review adopted a similar modelling approach, included FIT as the intervention evaluated, and included part of the population considered relevant to this appraisal (they have included only DG30 low-risk symptomatic patients), both existing models informed the development of the EAG's model.

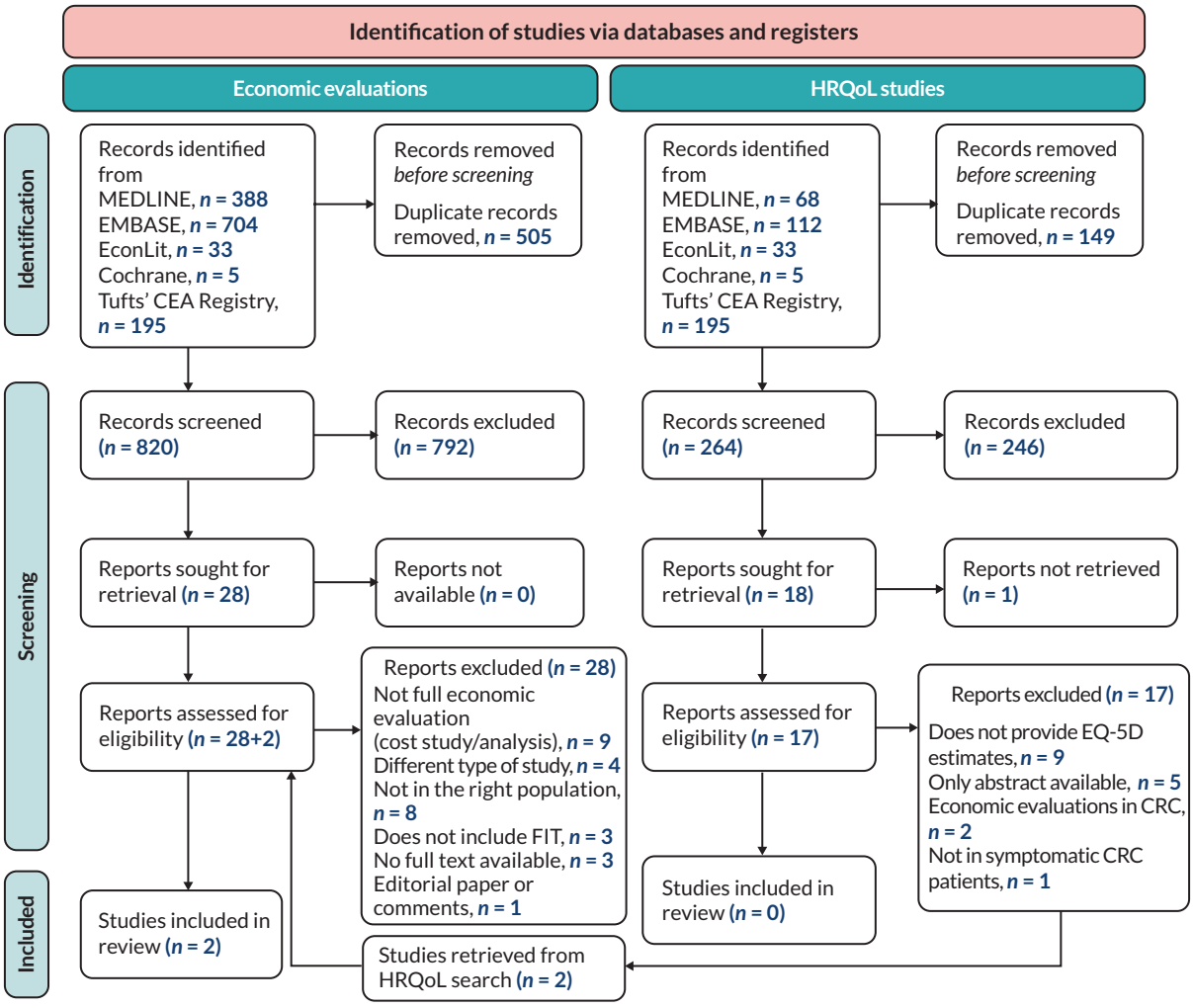

FIGURE 21 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram,[85] review for economic evaluations and HRQoL studies. EQ-5D, EuroQol-5 Dimensions.

TABLE 65 Existing economic evaluations: analytic scope

Author, year	Country	Population	Intervention	Comparator	Population characteristics	Underlying conditions included	Perspective of analysis	Time horizon	Discount rate
Westwood *et al.* 2017[90]	UK	Symptomatic people who are at low risk of CRC (as per NG12 definition) presenting to primary care	FIT (10 µg/g threshold chosen based on optimal threshold for each assay method)	• gFOBTs • no triage (all referred to colonoscopy)	Base-case: initial age 40 years; proportion female 65%; CRC prevalence 1.5%	CRC only	NHS and PSS	'Lifetime'	3.5% for QALYs and costs
Medina-Lara *et al.* 2020[91]	UK	Symptomatic patients at low risk for CRC (do not fulfil NICE's NG12 2WW referral criteria) but for whom GP has concerns	Use of diagnostic tools (RAT and QCancer) in combination with FIT	• FIT given to all • Send home/wait • Refer all	Initial age 70 years; CRC prevalence 1.5%	CRC only	NHS	Lifetime (30 years)	3.5% for QALYs and costs

CRC, colorectal cancer; FIT, quantitative faecal immunochemical test; NHS, National Health Service; PSS, Personal Social Services; QALY, quality-adjusted life year.

TABLE 66 Existing economic evaluations: modelling approach, main assumptions, definition of health states and summary of HRQoL included

Author, year	Model approach	Cycle length	Model type and states included	Sensitivity and specificity for intervention	Sensitivity and specificity for comparator	Sensitivity and specificity for other tests	Model main assumptions	Sources for survival	Costs included	EQ-5D valuation for health states
Westwood *et al.* 2017[90]	Combined decision tree and Markov STM	Assumed 1-year time frame for diagnostic model; 1-year cycle length for Markov STM	Based on sensitivity and specificity of tests (FIT, FOBT and COL/CTC) and symptoms persistence; two Markov STMs to estimate long-term effects, based on CRC Duke's stages (A to D) where patients may stay in current health state, progress to next worst stage or die (from CRC or another cause); and alive and dead states for people without underlying CRC	Base-case scenario (FIT 10 µg/g faeces threshold, single sample): OC Sensor: sensitivity 92.1% (95% CI 86.9% to 95.3%); specificity 85.8% (95% CI 78.3% to 91.0%) HM-JACKarc: sensitivity 100% (95% CI 71.5% to 100%); specificity 76.6% (95% CI 72.6% to 80.3%)	FOBT: sensitivity 50% (95% CI 15.0% to 85.0%); specificity 88% (95% CI 85.0% to 89.0%)	COL used as reference standard and assumed sensitivity and specificity for detection of CRC to be 100%	Diagnostic model: Patients whose symptoms do not persist assumed not to have CRC False-negative gFOBT or FIT patients whose symptoms persisted were assumed to receive COL and be diagnosed within 1 year and higher probability of progressing to a worse cancer state due to the delay in diagnosis Only those patients with a negative test result who symptoms do not persist do not receive COL/CTC • Patients with CRC: 1-year cycle length assumed to capture the probability of progression to next worst stage or die for treated and untreated patients Patients without CRC: difference in costs only due to tests and COL/CTC; difference in survival due to COL/CTC	Patients with CRC: 15-year predicted survival data from NG12; CRC-related mortality assumed constant after year 15 Patients without CRC: UK life tables	Initial and follow-up investigations Staging Lifetime treatment for CRC Drug costs Clinical visits and other resources required Costs taken from NG12 CRC treatment lifetime costs from Tappenden *et al.* 2007, inflated to 2015 prices HCHS index	Utilities for CRC stages based on Ness *et al.* 1999 Values used for Dukes' stages: A = 0.74; B = 0.70; C = 0.50; D = 0.25 Population without CRC: sex- and age-related utilities for every cycle from Kind *et al.* (1999)

TABLE 66 Existing economic evaluations: modelling approach, main assumptions, definition of health states and summary of HRQoL included (*continued*)

Author, year	Model approach	Cycle length	Model type and states included	Sensitivity and specificity for intervention	Sensitivity and specificity for comparator	Sensitivity and specificity for other tests	Model main assumptions	Sources for survival	Costs included	EQ-5D valuation for health states
Medina-Lara *et al.* 2020[91]	Combined decision tree and Markov STM	28-day cycles	Based on prevalence of CRC, sensitivity and specificity of strategies (diagnostic tool, FIT, send home/wait) Markov STM based on diagnosis status (diagnosed or undiagnosed) and CRC Duke's stages (A to D) where undiagnosed patients may stay in current health state, be diagnosed (via the diagnostic decision model) at their current stage, progress to next worst stage or die (from CRC or another cause); and alive and dead states for people without underlying CRC	**QCancer:** (Hippsley-Cox 2012) sensitivity 0.610; specificity 0.910 **RAT:** sensitivity 0.69; specificity 0.77 (Hamilton *et al.* 2005)	**FIT (threshold of 20 µg/g,** Murphy *et al.* 2017): Sensitivity 0.526; specificity for 50–69 years old: 0.988 Specificity for ≥ 70 years old: 0.963	COL sensitivity and specificity 1.0	In the intervention, patients with threshold score above 35 would be directly referred, while those with lower values receive FIT (threshold of 20 µg/g); Sens and spec of tests are assumed to be independent; accuracy of diagnostic tests is assumed independent of disease stage at presentation Model allows for partial adherence to the diagnostic protocol; COL sensitivity and specificity assumed to be 1 CRC patients who remain undiagnosed after first presentation will have repeated GP visits until diagnosis or death; strategies' sensitivity determines number of visits before referral; impact of delays in referral and diagnosis adapted from Whyte *et al.* using data from Tappenden *et al.*; disease progression rates from Tappenden *et al.*, estimated for asymptomatic patients is assumed to apply to symptomatic population	CRC mortality: exponential function fitted to digitised KM curves by stage at diagnosis from NCRAS HR for untreated CRC from Liu *et al.* 2014	FIT, GP visits, COL, COL AEs, heath-stage lifetime treatment costs	Age and sex-matched utilities from Ara and Brazier 2010; CRC Dukes' stages from Ness *et al.*1999: A = 0.74; B = 0.70; C = 0.50; D = 0.25

CTC, computed tomography colonography; COL, colonoscopy; CRC, colorectal cancer; FIT, quantitative faecal immunochemical test; NCRAS, National Cancer Registration and Analysis Service; sens, sensitivity; spec, specificity; STM, state transition model.

Appendix 8 Model estimation of the impact of additional time to diagnosis on colorectal cancer outcomes

April 2023

Dr Sophie Whyte

Laura Heathcote

Dr Chloe Thomas

Dr Olena Mandrik

Introduction

As time goes by from the onset of symptoms to their presentation in primary care, the disease may progress for individuals with symptomatic CRC. Additionally, the disease may also progress as time passes from the primary care presentation to receiving a diagnosis.

Patients presenting to their GP with suspected CRC may have various underlying conditions that could explain their symptoms, including non-cancerous conditions such as IBD, diverticulitis, irritable bowel syndrome or haemorrhoids. Symptoms of these conditions can overlap with CRC symptoms and include abdominal pain, rectal bleeding, changes in bowel habits, and weight loss. Moreover, patients may have adenomas, which are generally asymptomatic but can be diagnosed incidentally during investigations for suspected CRC and are clinically important because adenomas (particularly AAs) have the potential to develop into CRC.

The NHS 2WW system aims to expedite the diagnosis and treatment of patients with suspected cancer, including CRC, by ensuring that they are seen by a specialist within 2 weeks of referral by their GP. According to the latest annual data from NHS England (2020–1), 88.9% of patients with suspected CRC referred through the 2WW system were seen by a specialist within 2 weeks of referral, indicating that the vast majority of patients with suspected CRC are able to access specialist care within the recommended time frame.[241]

The 2WW system is important because it facilitates early diagnosis and treatment of cancer, which is crucial for improving survival rates and reducing the need for costly and invasive treatments. The same data[241] show that only 50.6% of patients with suspected CRC referred through the 2WW system started their first treatment within 62 days of referral, well below the target of 85%. The data for 2019–20[242] show that 66.7% of patients started their first treatment within 62 days of referral, suggesting that the failure to meet the target is not wholly explained by the extraordinary circumstances of the COVID-19 pandemic.

Patients who do not meet the criteria for an urgent referral may be referred via the non-urgent cancer referral pathway, which aims to ensure that patients with suspected cancer receive a diagnosis or are given the all-clear within 18 weeks from referral by their GP. The 18-week target applies to all non-urgent referrals, including those for suspected cancer.

Evidence on the association between time to diagnosis and CRC outcomes is heterogeneous. A systematic review explored the association between shorter times to diagnosis and more favourable outcome and found that although many studies reported no associations, more studies reported a positive, rather than a negative, association.[92]

The objective of this study was to develop a health economic model of CRC progression for symptomatic individuals associated with additional time to diagnosis, and to use the model to estimate the impact of additional time to diagnosis in terms of healthcare costs and health outcomes.

Methods

Model perspective

A lifetime horizon was adopted in this analysis to evaluate the long-term impact of additional time to diagnosis on healthcare costs and health outcomes for CRC and high-risk adenomas (HRA). A discount rate of 3.5% was used to adjust for time preferences, in line with the NICE reference case. The analysis was conducted from the perspective of the UK NHS and PSS, which included all costs and benefits associated with healthcare services and social care interventions. Both direct and indirect costs associated with CRC diagnosis and treatment, including costs of diagnostic tests, healthcare contacts, hospitalisations, medications, and palliative care, were considered. Health benefits or disbenefits were measured in terms of life-years gained (or lost), and QALYs.

Model structure

A health economic model was used to estimate the impact on patient outcomes of additional time to diagnosis. The model structure is illustrated in *Figure 22*. For patients with CRC, the impact of additional time to diagnosis is estimated by comparing the stage distribution of CRC at diagnosis without delay with the expected stage distribution of CRC at diagnosis with the delay. The change in stage distribution during the additional time to diagnosis represents disease progression during this time. For patients with HRA, disease progression is represented by the proportion of individuals who develop CRC during the additional time to diagnosis. These estimates of disease progression during the additional time to diagnosis are combined with estimates of the differential outcomes by disease stage to produce an overall estimate of the impact of additional time to diagnosis.

Population

The model population reflects the population of patients in the 2WW system for suspected CRC in England.[243] All individuals in the model have either CRC or HRA.

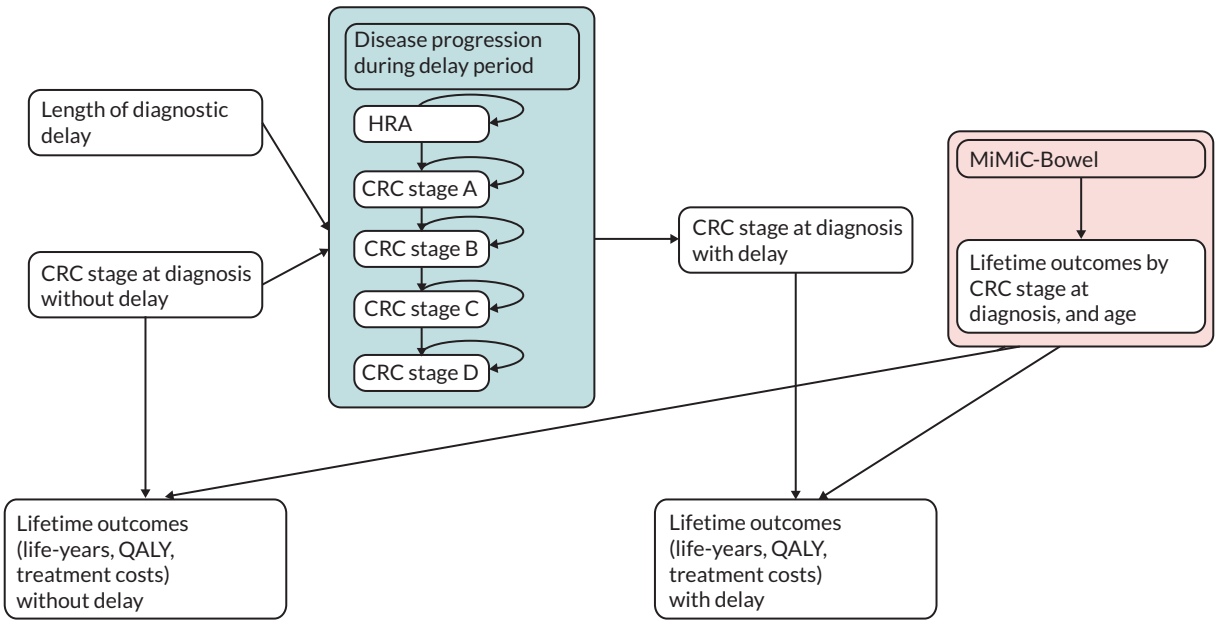

FIGURE 22 Model diagram.

The age distributions applied for both CRC and HRA are shown in *Table 67*.

Estimates for different age bands were generated and combined to produce estimates for a cohort.

The stage distribution without additional time to diagnosis (i.e. at the start of the model) was assumed to correspond to the stage distribution for patients diagnosed via symptomatic or chance detection (i.e. not via screening or surveillance), as shown in *Figure 23*. This stage distribution is a snapshot of the disease present at the point of data collection. Therefore, it reflects the stage distribution at the average time to diagnosis.

We note that chance detection may be associated with an earlier average stage at diagnosis than symptomatic presentation and so these data may show an earlier stage distribution than is appropriate for a symptomatic population, which is a minor limitation.

Modelling disease progression during delay period

Patients start the model in one of five health states: HRA, or CRC stage A, B, C or D. During the additional time to diagnosis, a proportion of patients will experience a stage shift; that is, a proportion of patients with HRA will develop CRC stage A, and a proportion of patients in each CRC stage will advance to the next stage. The probability of transitioning depends on the length of additional time to diagnosis. It is assumed that patients can only make up to one transition within 1 year. It is assumed that all patients survive the delay period; that is, there is no transition to 'dead' in the model. To include deaths in the diagnosis delay period would necessitate updating the delay progression component to depend on age, which would add a fair amount of complexity. This is a minor limitation of the methods but is not expected to have a significant impact on the results.

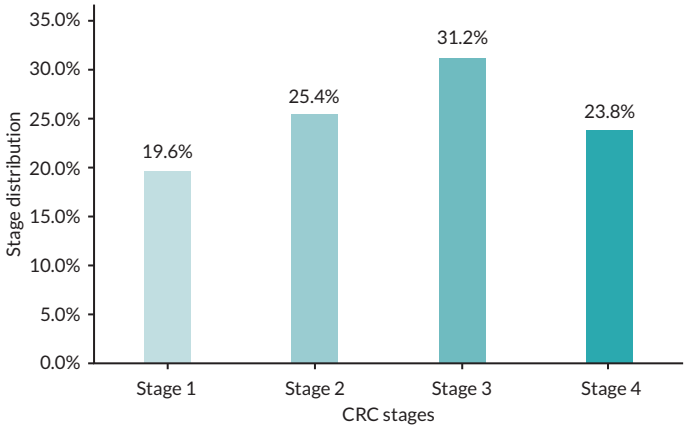

FIGURE 23 Stage distribution of CRC in the 2WW population in England.[244]

TABLE 67 Age distributions

Age distribution used for CRC		Age distribution used for HRA	
CRC diagnosed via 2WW[243]		2WW referrals population[243]	
Frequency, *n*	%	Frequency, *n*	%
734	5%	49,251	13
1814	13%	63,396	17
2841	21%	85,690	23
4274	32%	104,062	28
3789	28%	73,564	20
13,452	100%	375,963	100

2WW, 2 week wait; CRC, colorectal cancer; HRA - high-risk adenomas.

The transition probabilities in the model were taken from MiMiC-Bowel (*Table 68*).[93] MiMiC-Bowel is a microsimulation model of CRC written in R, which includes a natural history model.

The calculations assume that persons can make only one state transition within a 1-year period. This assumption is consistent with the assumptions made within MiMiC-Bowel model. For predictions related to delays of > 1 year, multiple transitions are included.

MiMiC-Bowel reports annual transition probabilities; however, to estimate transitions for shorter periods of time, it is necessary to first convert transition probabilities into rates. The formula used was rate, $r = -\ln(1 - \text{annual\_trans\_prob})$, then to estimate the transition probabilities relating to shorter time period the formula $p(t) = 1 - e - rt$, where r is the rate and t is the time period used. Note that this conversion formula has weaknesses and is most reliable for a model in which a person can experience only one type of event in a single cycle.[245]

In MiMiC-Bowel, the preclinical patient population includes both asymptomatic and symptomatic patients, hence the preclinical disease progression probabilities relate to both asymptomatic and symptomatic individuals. It is plausible that a wholly symptomatic population may experience faster disease progression, and this has been explored within a sensitivity analysis.

Lifetime outcomes for colorectal cancer and high-risk adenomas

Colorectal cancer

Lifetime outcomes for CRC without additional time to diagnosis were estimated using MiMiC-Bowel.

The model was set up to best reflect current practice in CRC screening and diagnosis, that is, individuals in the model were eligible for screening by FIT at the age of 56 years. The model records diagnoses and outcomes separately for individuals diagnosed via screening or via symptomatic presentation. Only outcomes for individuals diagnosed symptomatically were used, as these best represent individuals in the NHS 2WW pathway.

The model was run for a population of 169,975 individuals. For each individual diagnosed symptomatically, the life-years, QALYs and healthcare costs from the point of diagnosis until death were recorded. These outcomes were then subdivided according to age group and stage at diagnosis, and the mean outcomes per age and stage at diagnosis were calculated. Details on how these outcomes are estimated by MiMiC-Bowel are reported in full in the relevant published model documentation.[93]

As the costs in MiMiC-Bowel correspond to 2018 prices, aggregate costs were inflated to the latest possible price year (2021) using NHSCII from *Unit Costs of Health and Social Care 2021*.[114]

High-risk adenomas

It is implicitly assumed that individuals diagnosed with high-risk adenomas (HRA) have these removed via polypectomy. It was assumed that such individuals have the same lifetime cost and health outcomes as the general population. It is possible that individuals with HRA would be expected to have higher lifetime costs and less favourable lifetime health

TABLE 68 Transition probabilities used in the model[93]

Transition	MiMiC-Bowel, transition probability (1 year)	MiMiC-Bowel, transition rate (1 year)
CRC A -> CRC B	0.293	0.347
CRC B -> CRC C	0.554	0.807
CRC C -> CRC D	0.350	0.431
HRA -> CRC A[a]	0.027	0.028

CRC, colorectal cancer; HRA - high-risk adenomas.
a Risk of progression is age dependent for high-risk adenomas ->CRC but average transition rate for age of 62 years has been used currently.

outcomes; however, a simplifying assumption was made as this was not anticipated to have a significant impact on model outcomes.

Life expectancy was taken from life tables[246] and age- and sex-adjusted HRQoL was estimated using methods published by Ara *et al.*[247]

Results

Model estimates of disease progression

Figure 24 shows the change in CRC stage distribution with increasing additional time to diagnosis. With longer additional time to diagnosis, more individuals progress to late-stage (C or D) CRC, and fewer are diagnosed in early stages (A or B).

Costs and health outcomes by age and stage

Table 69 shows the expected outcomes by age and stage, as estimated by MiMiC-Bowel. There is an inconsistency in the results in that outcomes are less favourable for HRA than for CRC stage A; this is likely due to uncertainty in the model outcomes and the overwhelmingly positive outcomes associated with an early diagnosis of CRC. Within CRC stages, later stage is associated with fewer life-years, as is older age at diagnosis.

Fewer lifetime QALYs are accrued by individuals with CRC than with HRA, and within CRC fewer QALYs are accrued by individuals diagnosed at later stages than at early stages. Within each stage, individuals diagnosed in older age groups accrue fewer lifetime QALYs than those diagnosed in younger age groups. Expected QALY estimates are lower than the corresponding life-year estimates, reflecting the impact of CRC and CRC treatment on HRQoL.

Lifetime treatment costs show a more complex pattern. Treatment costs for individuals with CRC are much higher than for individuals with HRA. Individuals diagnosed with stage D cancer have the lowest treatment costs (likely because such individuals have a much shorter life expectancy and are more likely to be offered only palliative treatment). The pattern across the other age groups and stages is influenced by the interactions between life expectancy and treatment options.

Impact of additional time to diagnosis

Table 70 shows the estimated impact of additional time to diagnosis for individuals with CRC. Note that *additional* refers to beyond the current time to diagnosis on the 2WW pathway; time zero is current time to diagnosis. All results are incremental compared with this. With increasing additional time to diagnosis, health outcomes (life-years and QALYs) are worse. Treatment costs are also lower (due to more individuals being diagnosed at stage D, which has lower treatment costs). However, at the willingness-to-pay thresholds of £20,000 or £30,000 per QALY, the QALY loss outweighs the treatment cost savings, resulting in a lower NMB with increasing additional time to diagnosis.

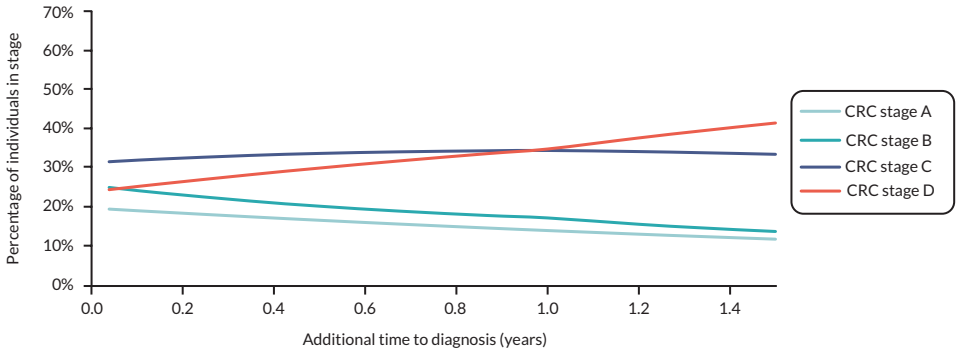

FIGURE 24 Change in CRC stage distribution by additional time to diagnosis.

TABLE 69 Expected discounted life-years, QALYs, and inflated treatment costs by age and stage at diagnosis

Age	Diagnosed with HR adenomas	Diagnosed with CRC stage A	Diagnosed with CRC stage B	Diagnosed with CRC stage C	Diagnosed with CRC stage D
Expected discounted lifetime LYs					
< 50	22.5	22.7	21.4	20.0	6.6
50–59	18.2	18.6	18.2	16.3	6.0
60–69	14.5	15.7	14.6	13.2	5.3
70–79	10.2	11.4	10.6	9.1	4.3
80+	5.8	7.3	7.0	6.1	4.0
Expected discounted lifetime QALYs					
< 50	19.3	13.7	11.3	10.1	1.5
50–59	14.7	8.5	8.2	6.7	1.3
60–69	11.5	6.0	5.5	4.5	1.1
70–79	7.4	3.6	3.1	2.4	0.8
80+	3.4	1.9	1.6	1.3	0.7
Expected discounted lifetime treatment costs (INFLATED TO 2021)					
< 50	£530	£14,621	£13,930	£19,762	£7166
50–59	£537	£14,602	£14,800	£19,186	£5536
60–69	£481	£15,972	£15,521	£16,662	£4533
70–79	£355	£14,646	£13,303	£13,486	£3215
80+	£87	£11,297	£10,317	£10,402	£2091

CRC, colorectal cancer; HR , high-risk; LY, life-years; QALY, quality-adjusted life year.

TABLE 70 Estimated outcomes by additional time to diagnosis for CRC

Additional time to diagnosis									
Months	**0.0**	**0.5**	**1.0**	**2.0**	**4.0**	**6.0**	**8.0**	**10.0**	**12.0**
Weeks	**0.0**	**2.2**	**4.3**	**8.7**	**17.3**	**26.0**	**34.7**	**43.3**	**52.0**
Incremental values vs. time zero									
NMB (WTP = £30,000/QALY)	£0	−£573	−£1133	−£2220	−£4748	−£6591	−£8494	−£10,414	−£11,649
NMB (WTP = £20,000/QALY)	£0	−£366	−£724	−£1418	−£3031	−£4205	−£5415	−£6636	−£7419
LYs	0.00	−0.04	−0.08	−0.17	−0.35	−0.49	−0.64	−0.78	−0.88
QALYs	0.00	−0.02	−0.04	−0.08	−0.17	−0.24	−0.31	−0.38	−0.42
Treatment costs	£0	−£47	−£94	−£186	−£404	−£568	−£742	−£922	−£1041
Absolute values									
LYs	10.03	9.98	9.94	9.86	9.67	9.53	9.39	9.24	9.15
QALYs	3.37	3.35	3.33	3.29	3.19	3.13	3.06	2.99	2.94
Treatment costs	£11,458	£11,410	£11,364	£11,272	£11,054	£10,890	£10,716	£10,536	£10,417

CRC, colorectal cancer; LY, life-years; NMB, net monetary benefit; QALY, quality-adjusted life year; WTP, willingness to pay.

Note
All outcomes are discounted at 3.5% per annum.

High-risk adenomas

Table 71 shows the impact of additional time to diagnosis of HRA. With increasing time to diagnosis, more life-years are accrued. This is likely due to the inconsistencies described previously in the expected life-years between HRA and CRC stage A. However, with increasing additional time to diagnosis, fewer QALYs are accrued (reflecting the lower HRQoL with stage A CRC vs HRA). Treatment costs are also higher, reflecting the higher treatment costs for CRC versus HRA.

TABLE 71 Impact of additional time to diagnosis of HRA

Additional time to diagnosis									
Months	**0.0**	**0.5**	**1.0**	**2.0**	**4.0**	**6.0**	**8.0**	**10.0**	**12.0**
Weeks	**0.0**	**2.2**	**4.3**	**8.7**	**17.3**	**26.0**	**34.7**	**43.3**	**52.0**
Incremental values vs. time zero									
NMB (WTP = £30,000/QALY)	£0	−£156	−£312	−£623	−£1397	−£2014	−£2704	−£3468	−£4000
NMB (WTP = £20,000/QALY)	£0	−£109	−£218	−£435	−£976	−£1406	−£1889	−£2422	−£2793
LYs	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03
QALYs	0.00	−0.00	−0.01	−0.02	−0.04	−0.06	−0.08	−0.10	−0.12
Treatment costs	£0	£15	£30	£59	£133	£192	£257	£330	£381
Absolute values									
LYs	13.27	13.28	13.28	13.28	13.28	13.29	13.29	13.30	13.30
QALYs	10.35	10.35	10.34	10.33	10.31	10.29	10.27	10.25	10.23
Treatment costs	£385	£400	£415	£444	£518	£577	£642	£715	£766

HRA, high-risk adenoma; LY, life-years; NMB, net monetary benefit; QALY, quality-adjusted life year; WTP, willingness to pay.

Note
All outcomes are discounted at 3.5% per annum.

Appendix 9 Methods for model evaluation: table with distributions used for each group of parameters in Evidence Assessment Group probabilistic analyses

TABLE 72 Distributions used in EAG probabilistic analyses

Model parameter group	Model parameter	Distribution	EAG comments
Patient characteristics	Age	Fixed	
	Proportion of female	Beta	
Settings	Discount rates (QALYs and costs)	Fixed	
Disease prevalence (overall population and high- and low-risk groups)	CRC prevalence AAs prevalence IBD prevalence	CODA samples	CODA sampling, for overall population an estimate based on high- and low-risk groups was estimated. Samples for proportion of high/low risk patients based on beta distribution
	Proportion of high-risk patients	Beta	
Disease stage/severity distribution	CRC stage distribution	See *Appendix 8*	
	UC/CD severity	Fixed	Assumption
Tests' accuracy	FIT	CODA samples/ beta	CODA samples when point estimates from EAG's clinical review and analysis; beta when data from unique study
	COL	Beta/fixed	Sensitivities were samples, while specificity were assumed to be 1.0
	CTC	Beta	
	Other non-invasive interventions	Fixed	Assumed to be 1.0
Safety netting and 'intermediate group' pathways	Probability of receiving each of the pathways following a FIT result < t or < t_{low}	Fixed	
	Proportion of patients receiving each of the pathways following a FIT result > t_{low} and < t_{high}	Fixed	
Interventions received in 2WW and 18WW referrals	Proportion of patients of receiving each imaging test in 2WW and 18WW referrals	Fixed	
	Complications after colonoscopies	Beta	

continued

TABLE 72 Distributions used in EAG probabilistic analyses (*continued*)

Model parameter group	Model parameter	Distribution	EAG comments
Time to diagnosis and length of delay	Time to diagnosis	Fixed	Varied in scenario analysis
	Proportion of change in total referrals (2WW + 18WW) in comparison to intervention 3 (applied to time to diagnosis in interventions 1 and 2)	Normal	SD assumed to be 0.1
Costs	FIT total costs (completed tests)	Normal	SD assumed to be 0.1
	FIT costs (non-completed tests)	Fixed	
	Repeat FIT total costs (completed tests)	Normal	SD assumed to be 0.1
	Repeat FIT costs (non-completed tests)	Fixed	
	FIT return rate	Beta	
	Referral – initial appointment	Normal	SD assumed to be 0.1
	Colonoscopy	Normal	SD assumed to be 0.1
	CTC	Normal	SD assumed to be 0.1
	Other non-invasive interventions	Normal	SD assumed to be 0.1
	Watch and wait	Normal	SD assumed to be 0.1
	COL complications	Normal	SD assumed to be 0.1, exception for death after perforation which is assumed fixed (£0.0)
	Annual cost treatment IBD	Normal	
	Increased treatment cost for IBD due to delay in diagnosis	Normal	
Long term STM model	Lifetime survival, QALYs and costs (CRC and AAs)	See *Appendix 8*	
	IBD and NSBP survival	Fixed	Based on general population's life tables
HRQoL	General population utility values (by age and sex)	Fixed	
	IBD utility value	Beta	
	Utility multiplier for IBD delayed diagnosis	Normal	
	Utility loss due to COL complications	Normal	

2WW, 2 week wait; 18WW, 18 week wait; AAs, advanced adenomas; CD, Crohn's disease; CODA, compositional data analysis; COL, colonoscopy; CRC, colorectal cancer; CTC, computed tomography colonography; EAG, Evidence Assessment Group; FIT, Faecal Immunochemical Test; HRQoL, health-related quality of life; IBD, inflammatory bowel disease; NSBP, no significant bowel pathology; QALYs, quality-adjusted life years; STM, state transition model; t, threshold; thigh, higher threshold; tlow, lower threshold; UC, ulcerative colitis.

Appendix 10 Results for the deterministic scenario analyses ran by the Evidence Assessment Group

The EAG has run 13 scenario analyses. For illustrative purposes, the sensitivity analyses have all been conducted on the comparison between HM JACKarc using one threshold of 10 µg/g (intervention 1), in comparison with current recommendations (intervention 3), using the lower intensity option for safety netting. The summary of results is presented in *Table 73*, while full tables are presented in *Tables 74–86*.

TABLE 73 Deterministic sensitivity analyses results for HM JACKarc using one threshold (10 µg/g)

	Intervention 1 (FIT using threshold of 10) vs intervention 3 (DG30/NG12)			
Scenario	Including QALYs	Including costs	ICER[a]	iNMB (20,000)
Base case (deterministic)	−0.0023	−£321	£141,344	£276
Scenario 1: shorter time to diagnosis (best-case)	−0.0012	−£321	£271,243	£297
Scenario 2: longer time to diagnosis (worst-case)	−0.0041	−£322	£77,671	£239
Scenario 3: QALY loss due to receiving a colonoscopy	−0.0015	−£321	£208,442	£291
Scenario 4: QALY loss for each month of diagnostic delay	−0.0027	−£321	£118,455	£267
Scenario 5: Dual FIT	−0.0015	−£224	£148,086	£194
Scenario 6: removing IBD and AAs from the model	−0.0012	−£359	£301,267	£335
Scenario 7: Using alternative source for FIT return rate from Moss *et al.*[122]	−0.0043	−£357	£83,129	£271
Scenario 8: Use of accuracy data for DG30 low-risk group (intervention 3) from EAG's clinical review analysis for this group	−0.0023	−£288	£124,880	£242
Scenario 9: Increased resource use of GP appointments for patients with NSBP following watch and wait or repeat FIT	−0.0023	−£309	£135,974	£264
Scenario 10: Alternative method to estimate unit costs for FIT in intervention 3 (weighted mean)	−0.0023	−£321	£141,222	£276
Scenario 11: FIT has perfect accuracy (sensitivity and specificity = 1.0) and return rate = 1.0	0.0007	−£435	Dominates	£448
Scenario 12: Reduction in prevalence for CRC, AAs and IBD by 50%	−0.0011	−£350	£326,506	£329
Scenario 13: Increase in prevalence for CRC, AAs and IBD by 50%	−0.0035	−£294	£84,249	£224

AAs, advanced adenomas; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; EAG, Evidence Assessment Group; FIT, faecal immunochemical test; GP, general practitioner; IBD, inflammatory bowel disease; ICER, incremental cost-effectiveness ratio; iNMB, incremental net monetary benefit; NG12, National Guideline 12; NSBP, no significant bowel pathology; QALY, quality-adjusted life year.
a South-west quadrant ICER.

Scenario 1: Shorter time to diagnosis (best case)

TABLE 74 Tabulated results for HM JACKarc using one threshold (Scenario 1)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.168	14.168	14.168	14.168	14.168	14.167	14.166	14.166	14.165	14.169
QALYs	10.895	10.894	10.894	10.894	10.894	10.894	10.893	10.893	10.893	10.895
Costs (£)	2956	2891	2871	2844	2818	2775	2732	2725	2708	3139
ICER (pairwise, vs. intervention 3)[a] (£)	256,849	290,663	291,210	285,102	271,243	239,634	193,441	184,837	163,058	–
iNMB λ = 20,000 (vs. intervention 3) (£)	169	231	249	274	297	333	365	369	378	–
iNMB λ = 30,000 (vs. intervention 3) (£)	161	223	240	264	285	318	344	347	352	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	1.521	1.593	1.628	1.693	1.782	2.019	2.482	2.596	2.948	1.045
Mean time to diagnosis – AAs	2.823	3.196	3.344	3.587	3.882	4.375	5.017	5.142	5.461	1.444
Mean time to diagnosis – IBD	1.865	2.089	2.177	2.325	2.505	2.910	3.503	3.623	3.952	1.396

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 2: Longer time to diagnosis (worst case)

TABLE 75 Tabulated results for HM JACKarc using one threshold (Scenario 2)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.163	14.163	14.162	14.162	14.161	14.160	14.157	14.157	14.155	14.166
QALYs	10.891	10.890	10.890	10.890	10.889	10.888	10.886	10.886	10.885	10.893
Costs (£)	2953	2887	2868	2841	2815	2771	2726	2719	2700	3137
ICER (pairwise, vs. intervention 3)[a] (£)	72,311	81,352	81,765	80,693	77,671	70,326	58,679	56,415	50,569	–
iNMB λ = 20,000 (vs. intervention 3) (£)	133	188	203	222	239	262	270	270	264	–
iNMB λ = 30,000 (vs. intervention 3) (£)	108	157	170	186	198	210	201	196	178	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	4.173	4.452	4.575	4.798	5.089	5.844	7.283	7.632	8.711	2.489
Mean time to diagnosis – AAs	8.155	9.344	9.807	10.568	11.483	13.005	14.973	15.354	16.327	3.702
Mean time to diagnosis – IBD	5.236	5.973	6.256	6.731	7.296	8.558	10.382	10.751	11.756	3.585

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 3: Quality-adjusted life-year loss due to receiving a colonoscopy

TABLE 76 Tabulated results for HM JACKarc using one threshold (Scenario 3)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.162	14.168
QALYs	10.893	10.892	10.892	10.892	10.892	10.891	10.890	10.890	10.890	10.893
Costs (£)	2955	2890	2870	2843	2817	2774	2730	2723	2705	3138
ICER (pairwise, vs. intervention 3)[a] (£)	198,131	230,347	230,264	223,102	208,442	177,605	136,678	129,500	111,895	–
NMB λ = 20,000 (vs. intervention 3) (£)	165	227	245	269	291	323	348	351	356	–
NMB λ = 30,000 (vs. intervention 3) (£)	155	216	233	255	275	303	318	319	317	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	2.310	2.461	2.529	2.653	2.815	3.237	4.045	4.242	4.849	1.388
Mean time to diagnosis – AAs	4.456	5.128	5.390	5.822	6.341	7.207	8.329	8.546	9.102	1.956
Mean time to diagnosis – IBD	2.946	3.355	3.513	3.777	4.092	4.798	5.821	6.027	6.592	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 4: Quality-adjusted life-year loss for each month of diagnostic delay

TABLE 77 Tabulated results for HM JACKarc using one threshold (Scenario 4)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.162	14.168
QALYs	10.887	10.886	10.886	10.886	10.886	10.885	10.884	10.883	10.883	10.888
Costs (£)	2955	2890	2870	2843	2817	2774	2730	2723	2705	3138
ICER (pairwise, vs. intervention 3)[a] (£)	112,711	125,642	125,872	123,636	118,455	106,583	88,619	85,182	76,362	–
NMB λ = 20,000 (vs. intervention 3) (£)	151	209	225	247	267	296	316	318	320	–
NMB λ = 30,000 (vs. intervention 3) (£)	134	189	204	224	240	262	270	269	263	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	2.310	2.461	2.529	2.653	2.815	3.237	4.045	4.242	4.849	1.388
Mean time to diagnosis – AAs	4.456	5.128	5.390	5.822	6.341	7.207	8.329	8.546	9.102	1.956
Mean time to diagnosis – IBD	2.946	3.355	3.513	3.777	4.092	4.798	5.821	6.027	6.592	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 5: Dual FIT

TABLE 78 Tabulated results for HM JACKarc using one threshold (Scenario 5)

	Intervention 1: FIT 1 threshold	Intervention 3: DG30 and NG12
t (μg/g)	10	10
LYs	14.167	14.168
QALYs	10.893	10.895
Costs (£)	2914	3138
ICER (pairwise, vs. intervention 3)[a] (£)	148,086	–
NMB λ = 20,000 (vs. intervention 3) (£)	194	–
NMB λ = 30,000 (vs. intervention 3) (£)	179	–
Number of 2WW referrals (total)	0.336	0.639
Number of 18WW referrals (total)	0.070	0.038
Number of repeat FIT (total)	0.070	0.038
Number of watch and wait (total) (total)	0.524	0.285
Number of COLs (total)	0.373	0.620
Reduction in number of referrals (total – 2WW + 18WW)	40.0%	–
Reduction in number of referrals (2WW only)	47.4%	–
Increase in number of referrals (18WW only)[b]	83.9%	–
Reduction in number of COLs	39.8%	–
Mean time to diagnosis – CRC	2.206	1.388
Mean time to diagnosis – AAs	5.514	1.956
Mean time to diagnosis – IBD	2.398	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 6: Removing inflammatory bowel disease and advanced adenomas from the model

TABLE 79 Tabulated results for HM JACKarc using one threshold (Scenario 6)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
***t* (µg/g)**	**2**	**4**	**5**	**7**	**10**	**20**	**50**	**60**	**100**	**10**
LYs	14.193	14.193	14.192	14.192	14.192	14.191	14.190	14.189	14.188	14.195
QALYs	10.944	10.944	10.944	10.944	10.944	10.943	10.943	10.942	10.942	10.945
Costs (£)	932	861	841	812	784	739	695	688	670	1143
ICER (pairwise, vs. intervention 3)[a] (£)	274,897	318,019	320,713	315,859	301,267	257,369	194,474	183,268	155,479	–
NMB λ = 20,000 (vs. intervention 3) (£)	196	264	284	310	335	373	402	406	412	–
NMB λ = 30,000 (vs. intervention 3) (£)	188	255	274	300	323	357	379	381	382	–
Number of 2WW referrals (total)	0.365	0.293	0.272	0.243	0.216	0.172	0.130	0.124	0.109	0.635
Number of 18WW referrals (total)	0.067	0.074	0.077	0.080	0.083	0.087	0.092	0.092	0.094	0.038
Number of repeat FIT (total)	0.067	0.074	0.077	0.080	0.083	0.087	0.092	0.092	0.094	0.038
Number of watch and wait (total)	0.501	0.558	0.575	0.598	0.619	0.654	0.686	0.692	0.704	0.289
Number of COLs (total)	0.394	0.335	0.318	0.295	0.272	0.237	0.203	0.198	0.185	0.613
Reduction in number of referrals (total – 2WW + 18WW)	35.8%	45.5%	48.2%	52.1%	55.7%	61.5%	67.0%	67.9%	69.9%	–
Reduction in number of referrals (2WW only)	42.4%	53.9%	57.2%	61.7%	66.0%	72.9%	79.4%	80.5%	82.9%	–
Increase in number of referrals (18WW only)[b]	73.6%	93.5%	99.2%	107.1%	114.6%	126.6%	137.9%	139.7%	143.9%	–
Reduction in number of COLs	35.7%	45.4%	48.1%	52.0%	55.6%	61.4%	66.9%	67.8%	69.8%	–
Mean time to diagnosis – CRC	2.298	2.449	2.516	2.640	2.802	3.225	4.035	4.231	4.840	1.388
Mean time to diagnosis – AAs	–	–	–	–	–	–	–	–	–	–
Mean time to diagnosis – IBD	–	–	–	–	–	–	–	–	–	–

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 7: Using alternative source for faecal immunochemical test return rate from Moss *et al.*[122]

TABLE 80 Tabulated results for HM JACKarc using one threshold (Scenario 7)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.162	14.162	14.162	14.162	14.161	14.161	14.160	14.159	14.159	14.167
QALYs	10.891	10.891	10.891	10.890	10.890	10.890	10.889	10.889	10.888	10.894
Costs (£)	2858	2809	2795	2775	2755	2723	2691	2686	2673	3113
ICER (pairwise, vs. intervention 3)[a] (£)	69,416	78,489	80,384	82,326	83,129	82,333	77,093	75,697	71,575	–
NMB λ = 20,000 (vs. intervention 3) (£)	181	226	239	256	271	295	312	314	317	–
NMB λ = 30,000 (vs. intervention 3) (£)	144	188	199	215	228	248	258	258	256	–
Number of 2WW referrals (total)	0.296	0.244	0.229	0.207	0.187	0.154	0.121	0.116	0.103	0.618
Number of 18WW referrals (total)	0.074	0.080	0.081	0.083	0.086	0.089	0.093	0.093	0.094	0.040
Number of repeat FIT (total)	0.074	0.080	0.081	0.083	0.086	0.089	0.093	0.093	0.094	0.040
Number of watch and wait (total)	0.555	0.597	0.609	0.626	0.642	0.668	0.694	0.698	0.708	0.301
Number of COLs (total)	0.340	0.297	0.284	0.267	0.250	0.223	0.196	0.192	0.181	0.602
Reduction in number of referrals (total – 2WW + 18WW)	43.7%	50.9%	52.9%	55.8%	58.6%	63.1%	67.6%	68.3%	70.0%	–
Reduction in number of referrals (2WW only)	52.0%	60.5%	63.0%	66.5%	69.8%	75.1%	80.4%	81.3%	83.4%	–
Increase in number of referrals (18WW only)[b]	84.2%	98.0%	102.0%	107.6%	112.9%	121.6%	130.2%	131.6%	134.9%	–
Reduction in number of COLs	43.6%	50.7%	52.8%	55.7%	58.5%	63.0%	67.4%	68.2%	69.9%	–
Mean time to diagnosis – CRC	4.684	4.788	4.836	4.923	5.039	5.345	5.933	6.076	6.520	1.762
Mean time to diagnosis – AAs	6.287	6.772	6.963	7.276	7.654	8.285	9.105	9.263	9.669	1.956
Mean time to diagnosis – IBD	5.133	5.425	5.539	5.730	5.958	6.472	7.217	7.368	7.780	2.757

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 8: Use of accuracy data for DG30 low-risk group (intervention 3) from Evidence Assessment Group's clinical review analysis for this group

TABLE 81 Tabulated results for HM JACKarc using one threshold (Scenario 8)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (μg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.162	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.892	10.891	10.891	10.890	10.895
Costs (£)	2955	2890	2870	2843	2817	2774	2730	2723	2705	3105
ICER (pairwise, vs. intervention 3)[a] (£)	105,436	126,338	128,481	128,488	124,880	114,277	95,823	92,151	82,602	–
NMB λ = 20,000 (vs. intervention 3) (£)	122	182	198	221	242	273	297	299	303	–
NMB λ = 30,000 (vs. intervention 3) (£)	107	165	180	201	219	244	258	258	255	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.604
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.042
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.042
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.312
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.591
Reduction in number of referrals (total – 2WW + 18WW)	30.2%	40.1%	43.0%	47.1%	50.9%	57.2%	63.4%	64.4%	66.8%	–
Reduction in number of referrals (2WW only)	36.1%	48.0%	51.4%	56.2%	60.8%	68.3%	75.7%	76.9%	79.8%	–
Increase in number of referrals (18WW only)[b]	55.1%	73.3%	78.5%	85.9%	92.9%	104.4%	115.7%	117.5%	121.9%	–
Reduction in number of COLs	30.1%	40.0%	42.8%	46.9%	50.7%	57.0%	63.2%	64.2%	66.6%	-
Mean time to diagnosis – CRC	2.325	2.475	2.542	2.665	2.826	3.248	4.055	4.251	4.858	1.375
Mean time to diagnosis – AAs	4.472	5.143	5.405	5.836	6.355	7.220	8.341	8.557	9.112	1.956
Mean time to diagnosis – IBD	2.964	3.370	3.527	3.791	4.105	4.810	5.831	6.038	6.602	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 9: Increased resource use of general practitioner appointments for patients with no significant bowel pathology following watch and wait or repeat faecal immunochemical test

TABLE 82 Tabulated results for HM JACKarc using one threshold (Scenario 9)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.162	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.892	10.891	10.891	10.890	10.895
Costs (£)	2974	2911	2892	2866	2841	2799	2756	2749	2732	3150
ICER (pairwise, vs. intervention 3)[a] (£)	126,754	143,019	143,686	141,594	135,974	122,447	101,363	97,301	86,850	–
NMB λ = 20,000 (vs. intervention 3) (£)	148	206	222	244	264	294	316	318	322	–
NMB λ = 30,000 (vs. intervention 3) (£)	134	189	204	224	241	265	277	277	273	–
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	2.310	2.461	2.529	2.653	2.815	3.237	4.045	4.242	4.849	1.388
Mean time to diagnosis – AAs	4.456	5.128	5.390	5.822	6.341	7.207	8.329	8.546	9.102	1.956
Mean time to diagnosis – IBD	2.946	3.355	3.513	3.777	4.092	4.798	5.821	6.027	6.592	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 10: Alternative method to estimate unit costs for faecal immunochemical test in intervention 3 (lower value)

TABLE 83 Tabulated results for HM JACKarc using one threshold (Scenario 10)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.166	14.166	14.166	14.166	14.166	14.165	14.163	14.163	14.162	14.168
QALYs	10.893	10.893	10.893	10.893	10.893	10.892	10.891	10.891	10.890	10.895
Costs (£)	2955	2890	2870	2843	2817	2774	2730	2723	2705	3138
ICER (pairwise, vs. intervention 3)[a] (£)	132,129	148,785	149,408	147,143	141,222	127,059	105,076	100,847	89,970	-
NMB λ = 20,000 (vs. intervention 3) (£)	155	215	232	255	276	307	330	333	337	-
NMB λ = 30,000 (vs. intervention 3) (£)	141	198	214	235	253	278	291	292	288	-
Number of 2WW referrals (total)	0.386	0.314	0.294	0.265	0.237	0.191	0.147	0.140	0.122	0.639
Number of 18WW referrals (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of repeat FIT (total)	0.065	0.072	0.074	0.077	0.080	0.085	0.090	0.091	0.092	0.038
Number of watch and wait (total)	0.485	0.541	0.558	0.581	0.603	0.638	0.674	0.679	0.693	0.285
Number of COLs (total)	0.413	0.355	0.338	0.314	0.291	0.254	0.218	0.212	0.197	0.620
Reduction in number of referrals (total – 2WW + 18WW)	33.4%	42.9%	45.6%	49.5%	53.2%	59.2%	65.1%	66.0%	68.3%	–
Reduction in number of referrals (2WW only)	39.6%	50.8%	54.0%	58.6%	63.0%	70.1%	77.0%	78.2%	80.9%	–
Increase in number of referrals (18WW only)[b]	70.1%	90.0%	95.8%	103.9%	111.6%	124.2%	136.5%	138.5%	143.4%	–
Reduction in number of COLs	33.3%	42.7%	45.5%	49.3%	53.0%	59.0%	64.9%	65.8%	68.2%	–
Mean time to diagnosis – CRC	2.310	2.461	2.529	2.653	2.815	3.237	4.045	4.242	4.849	1.388
Mean time to diagnosis – AAs	4.456	5.128	5.390	5.822	6.341	7.207	8.329	8.546	9.102	1.956
Mean time to diagnosis – IBD	2.946	3.355	3.513	3.777	4.092	4.798	5.821	6.027	6.592	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 11: Faecal immunochemical test with perfect accuracy and return rate = 1.0

TABLE 84 Tabulated results for HM JACKarc using one threshold (Scenario 11)

	Intervention 1: FIT 1 threshold	Intervention 3: DG30 and NG12
t (µg/g)	–	10
LYs	14.169	14.168
QALYs	10.896	10.895
Costs (£)	2713	3148
ICER (pairwise, vs. intervention 3) (£)	Dominates	–
NMB λ = 20,000 (vs. intervention 3) (£)	448	–
NMB λ = 30,000 (vs. intervention 3) (£)	455	–
Number of 2WW referrals (total)	0.124	0.647
Number of 18WW referrals (total)	0.092	0.037
Number of repeat FIT (total)	0.092	0.037
Number of watch and wait (total)	0.691	0.278
Number of COLs (total)	0.201	0.626
Reduction in number of referrals (total – 2WW + 18WW)	68.4%	–
Reduction in number of referrals (2WW only)	80.8%	–
Increase in number of referrals (18WW only)[a]	148.4%	–
Reduction in number of COLs	67.9%	–
Mean time to diagnosis – CRC	0.771	1.242
Mean time to diagnosis – AAs	1.668	1.956
Mean time to diagnosis – IBD	0.356	1.766

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 12: Reduction in prevalence of colorectal cancer, advanced adenomas and inflammatory bowel disease by 50%

TABLE 85 Tabulated results for HM JACKarc using one threshold (Scenario 12)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.244	14.244	14.244	14.244	14.244	14.243	14.243	14.242	14.242	14.245
QALYs	11.030	11.030	11.030	11.030	11.030	11.029	11.029	11.029	11.028	11.031
Costs (£)	1773	1703	1682	1654	1627	1582	1539	1533	1516	1977
ICER (pairwise, vs. intervention 3)[a] (£)	314,248	350,557	350,626	342,978	326,506	289,325	234,456	224,135	197,872	-
NMB λ = 20,000 (vs. intervention 3) (£)	191	258	277	304	329	367	400	404	414	-
NMB λ = 30,000 (vs. intervention 3) (£)	184	250	269	294	318	353	381	385	391	-
Number of 2WW referrals (total)	0.368	0.295	0.274	0.245	0.217	0.173	0.130	0.124	0.108	0.635
Number of 18WW referrals (total)	0.066	0.074	0.076	0.079	0.082	0.087	0.092	0.092	0.094	0.038
Number of repeat FIT (total)	0.066	0.074	0.076	0.079	0.082	0.087	0.092	0.092	0.094	0.038
Number of watch and wait (total)	0.499	0.556	0.573	0.596	0.618	0.653	0.687	0.692	0.704	0.288
Number of COLs (total)	0.397	0.338	0.321	0.297	0.274	0.238	0.203	0.198	0.185	0.615
Reduction in number of referrals (total – 2WW + 18WW)	35.5%	45.2%	47.9%	51.8%	55.5%	61.4%	67.1%	67.9%	70.1%	-
Reduction in number of referrals (2WW only)	42.0%	53.5%	56.8%	61.4%	65.8%	72.8%	79.5%	80.5%	83.0%	-
Increase in number of referrals (18WW only)[b]	73.2%	93.3%	99.0%	107.1%	114.7%	126.9%	138.6%	140.4%	144.7%	-
Reduction in number of COLs	35.4%	45.1%	47.8%	51.7%	55.4%	61.3%	67.0%	67.9%	70.0%	-
Mean time to diagnosis – CRC	2.300	2.450	2.517	2.641	2.803	3.225	4.034	4.231	4.839	1.388
Mean time to diagnosis – AAs	4.445	5.115	5.378	5.809	6.328	7.194	8.316	8.534	9.090	1.956
Mean time to diagnosis – IBD	2.936	3.343	3.500	3.764	4.079	4.785	5.808	6.015	6.580	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

Scenario 13: Increase in prevalence of colorectal cancer, advanced adenomas and inflammatory bowel disease by 50%

TABLE 86 Tabulated results for HM JACKarc using one threshold (Scenario 13)

	Intervention 1: FIT 1 threshold									Intervention 3: DG30 and NG12
t (µg/g)	2	4	5	7	10	20	50	60	100	10
LYs	14.089	14.088	14.088	14.088	14.087	14.086	14.084	14.083	14.082	14.091
QALYs	10.757	10.756	10.756	10.756	10.756	10.755	10.753	10.753	10.752	10.759
Costs (£)	4136	4075	4057	4031	4006	3964	3920	3912	3893	4300
ICER (pairwise, vs. intervention 3)[a] (£)	76,723	87,437	88,120	87,258	84,249	76,624	64,287	61,877	55,638	–
NMB λ = 20,000 (vs. intervention 3) (£)	121	174	188	207	224	248	262	262	260	–
NMB λ = 30,000 (vs. intervention 3) (£)	100	148	160	176	189	204	203	200	187	–
Number of 2WW referrals (total)	0.404	0.334	0.313	0.284	0.256	0.210	0.163	0.155	0.136	0.643
Number of 18WW referrals (total)	0.063	0.070	0.072	0.075	0.078	0.083	0.088	0.089	0.091	0.038
Number of repeat FIT (total)	0.063	0.070	0.072	0.075	0.078	0.083	0.088	0.089	0.091	0.038
Number of watch and wait (total)	0.470	0.526	0.542	0.565	0.587	0.624	0.661	0.667	0.682	0.282
Number of COLs (total)	0.430	0.372	0.355	0.332	0.309	0.271	0.232	0.226	0.210	0.625
Reduction in number of referrals (total – 2WW + 18WW)	31.4%	40.7%	43.4%	47.2%	50.9%	57.0%	63.1%	64.1%	66.6%	–
Reduction in number of referrals (2WW only)	37.2%	48.1%	51.3%	55.8%	60.2%	67.4%	74.6%	75.8%	78.8%	–
Increase in number of referrals (18WW only)[b]	67.0%	86.7%	92.4%	100.6%	108.5%	121.4%	134.4%	136.6%	141.9%	–
Reduction in number of COLs	31.2%	40.4%	43.1%	46.9%	50.6%	56.7%	62.8%	63.9%	66.4%	–
Mean time to diagnosis – CRC	2.319	2.472	2.540	2.664	2.827	3.249	4.056	4.253	4.859	1.388
Mean time to diagnosis – AAs	4.466	5.140	5.403	5.835	6.355	7.221	8.342	8.559	9.114	1.956
Mean time to diagnosis – IBD	2.957	3.367	3.525	3.790	4.106	4.812	5.833	6.040	6.603	2.044

2WW, 2 week wait; 18WW, 18 week wait; AAs, Advanced adenomas; COL, colonoscopy; CRC, colorectal cancer; DG30, Diagnostics Guidance 30; FIT, faecal immunochemical test; IBD, inflammatory bowel disease; iNMB, incremental net monetary benefit; ICER, incremental cost-effectiveness ratio; LY, life-years; QALY, quality-adjusted life year.
a South-west quadrant ICER.
b Also the value for increased repeat FIT and increased number of watch and waits.

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library