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

Richard Booth, Rachel Carten, Nigel D’Souza, Marie Westwood, Jos Kleijnen, and Muti Abulafi

Summary

Background The UK National Institute for Health and Care Excellence (NICE), recommended in 2017 the use of the faecal immunochemical test (FIT) to guide investigations in patients presenting with NICE-defined low-risk symptoms suspicious for colorectal cancer (CRC). At that time, NICE did not recommend FIT use for high-risk symptoms. This is the first systematic review to evaluate the diagnostic accuracy of FIT in NICE-defined high and low-risk symptoms and was designed to inform the joint ACPGBI/BSG guidelines.

Methods We performed a systematic literature review and meta-analysis. PROSPERO registration number CRD42021224674. Medline and EMBASE databases were searched from inception to 31st March 2022. We included studies recruiting adult patients presenting with suspected CRC symptoms in whom FIT was performed and diagnostic accuracy data for CRC detection could be derived at a limit of detection (LoD) and/or 10 µg haemoglobin/gram faeces threshold in four commonly used analysers. FIT performance was assessed for high-risk, low-risk and individual symptoms where possible. Bivariate meta-analysis was performed where study numbers allowed.

Findings Thirty-one studies (79,566 patients) met inclusion criteria. At 10 µg/g, for “all symptoms” (n = 35,945) sensitivity and specificity were 91.0% (95% CI: 88.9, 92.7) and 75.2% (95% CI: 69.6, 80.1); for “high-risk” symptoms (n = 18,264), 88.7% (95% CI: 84.4, 92.0) and 78.5% (95% CI: 73.0, 83.2); and for “low-risk” symptoms (n = 2161), 88.7% (95% CI: 78.1, 95.3) and 88.5% (95% CI: 87.1, 89.9), respectively. At LoD, for “all symptoms” (n = 26,056) sensitivity and specificity were 94.7% (95% CI: 90.5, 97.1) and 66.5% (95% CI: 58.7, 73.6); for “high-risk” symptoms (n = 16,768), 92.8% (95% CI: 86.4, 96.3) and 70.3% (95% CI: 66.5, 73.8); and for “low-risk” symptoms (n = 2082), 94.7% (95% CI: 85.4, 98.9) and 71.9% (95% CI: 69.9, 73.9), respectively. Summary estimates were similar across different analysers.

Interpretation FIT sensitivity for CRC detection is maximised at the LoD; its performance is similar in high and low-risk symptoms, and across different analysers where a common threshold is used. FIT performance for CRC detection is adequate and transferrable to clinical diagnostic pathways.

Funding This review was part-funded by NHS England awarded to RM Partners. RB and RC were funded by research fellowships awarded by Croydon University Hospital.

Copyright © 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Colorectal cancer; Clinical decision making; Meta-analysis; Systematic review; Stool markers; Faecal immunochemical test, FIT; NICE NG12; High-risk symptoms; NICE DG30; Low-risk symptoms; HM-JACKarc; OC-Sensor; FOB Gold; QuikRead go

Introduction

Colorectal cancer is the second most common cause of cancer death in the UK, accounting for 10% of cancer-related mortality. Despite the introduction of the national
bowel cancer screening programme in 2008, only 9.8% of cases are diagnosed via this pathway; nearly all other patients are diagnosed because of bowel symptoms.2

In England, the National Institute for Health and Care Excellence (NICE) employs symptom-based criteria to guide urgent referral for suspected colorectal cancer (CRC) (Table 1). The original 2005 NICE guidance included only high-risk symptoms for CRC.3 This was superseded in 2015 by the much-expanded NG12 guidance that also included medium and low risk symptoms.4 In 2017, NICE introduced DG30 guidance recommending use of FIT in low-risk symptoms in primary care to guide referral for further investigation.5 This guidance did not recommend the use of FIT in patients with high-risk symptoms, who continue to be referred with increasing numbers,6 creating a significant demand for diagnostic services.7 The COVID-19 pandemic further added to the backlog of patients awaiting endoscopy which forced many UK clinical services to adopt FIT into their clinical pathways to meet local needs, by reducing referrals and balancing demand for endoscopy.8

This systematic review was designed to inform the joint guidelines of The Association of Coloproctology of Great Britain and Ireland (ACPGBI) and British Society of Gastroenterology (BSG) on the role of FIT in symptomatic patients due for release in 2022. The analysis aimed to assess the diagnostic accuracy of quantitative FIT for CRC detection in purely symptomatic patient cohorts at the NICE DG30-recommended threshold of 10 µg haemoglobin per gram faeces (hereafter µg/g) and at the limit of detection (LoD) of available assays taking account of the reference standard used in the included studies; the recruitment setting (primary or secondary care); and the symptom clusters including high and low-risk symptoms.

Methods
We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to structure and report our systematic review. The review was registered with the International Prospective Register of Systematic Reviews (PROSPERO), registration number CRD42021224674.

Data sources and search strategy
We used a combination of search terms related to FIT (appendix A) and included all studies identified. Databases searched were MEDLINE and EMBASE via OVID from inception to 31 March 2022. In addition, references in included studies and systematic reviews were checked for further studies, and we also contacted several authors. Search results were combined, and duplicates removed using EndNote reference software.9

Study selection and inclusion criteria
Figure 1 illustrates the flow through the study selection process. Three authors (RB, RC and NDS)
independently screened titles with disagreement resolved via discussion or through consultation with a fourth author (MA). A similar process was followed for abstract screening. Final eligibility was determined by review of full text, with papers included if they met the following criteria:

**Population, setting and study design.** We included cohort studies performed on adult patients consulting a physician with symptoms suggestive of CRC in whom quantitative FIT was performed as part of their work-up. Studies recruiting in both primary and secondary care were included. Studies reporting on mixed cohorts including screening or follow-up populations were excluded.

**Index test.** We included data from studies that evaluated the diagnostic accuracy of quantitative FIT for CRC using the FOB Gold (Sentinel Diagnostics, Italy), HM-JACKarc (Hitachi Chemical Diagnostics Systems, Tokyo, Japan), OC-Sensor (Eiken Chemical, Japan) and QuikRead Go (Aidian Oy, Espoo, Finland) FIT assays. All included studies used single-sample FIT.

**Reference test.** We included diagnostic accuracy studies for FIT that used full colonic imaging with either colonoscopy or CT colonography (CTC) as the reference test for the diagnosis of CRC. We also included pragmatically designed studies using FIT within symptomatic pathways where registry follow-up formed part of the reference standard for some or all FIT negative patients. We grouped studies by reference standard and performed subsequent meta-analyses accordingly. The first group (hereafter referred to as tier 1) included studies where at least 90% of the cohort underwent full colonic imaging with either colonoscopy or CTC as the reference test (colonoscopy and CTC having equivalent sensitivity for CRC detection59). The second group (hereafter referred to as tier 2) included studies with mixed reference standards including plain CT and flexible sigmoidoscopy, which reflects clinical practice, as well as studies with a minimum of 3 months’ registry follow-up.

**Data extraction**

Two reviewers (RB and RC) independently extracted data which were compared and cross-checked by the other and queries clarified with the senior author (MA). The extracted data included: study design; publication year; geographical location; patient numbers; recruitment setting (primary or secondary care, or both); assay (s) used; thresholds employed (LoD or 10 µg/g); reference standard used; and presenting symptom cluster as stratified by NICE (NG12 hereafter high-risk, DG30 hereafter low-risk, or unstratified) or individual symptoms of rectal bleeding, iron deficiency anaemia and change in bowel habit.

| Symptoms                                      | 2005 NICE Guidance (CG27) | 2015 NICE Guidance (NG12) | 2017 NICE Guidance (DG30) | Risk of cancer |
|-----------------------------------------------|---------------------------|---------------------------|---------------------------|---------------|
| Rectal bleeding for 6 weeks (>60 years)       | REFER                     | REFER                     | REFER                     | High: >5%     |
| Rectal bleeding + diarrhoea for 6 weeks (>40 years) | REFER                     | REFER                     | REFER                     | Medium: 3−5% |
| Change in bowel habit for 6 weeks (>60 years) | REFER                     | Test with FOBT before referral | Test with FIT before referral | Low: 1−3%    |
| Mass (any age)                                |                           |                           |                           |               |
| Iron deficiency anaemia                       |                           |                           |                           |               |
| Abdominal pain AND weight loss (>40 years)    | REFER                     | REFER                     | REFER                     | Medium: 3−5% |
| Rectal bleeding (>50 years)                   |                           |                           |                           |               |
| Rectal bleeding + (Iron deficiency anaemia/change in bowel habit/weight loss, <50 years) | REFER                     | Refer                     |                           |               |
| Iron deficiency anaemia (>60 years)           |                           |                           |                           |               |
| Change in bowel habit (>60 years)             |                           |                           |                           |               |
| Abdominal pain OR weight loss (>50 years)     |                           |                           |                           |               |
| Change in bowel habit (<60 years)             |                           |                           |                           |               |
| Iron deficiency anaemia (<60 years)           |                           |                           |                           |               |
| Anaemia, Non-Iron deficient (>60 years)        |                           |                           |                           |               |
| Other symptoms                                |                           |                           |                           | Low: <1%     |
| FIT, if no rectal bleeding                    |                           |                           |                           |               |

Table 1: Symptoms stratified according to cancer risk according to NICE CG27, NG12 and DG30 guidelines.

NICE= National Institute for Health and Care Excellence. FOBT= Guaiac faecal occult blood test. FIT= Faecal immunochemical test.
**Figure 1.** PRISMA diagram showing selection of articles for inclusion in the review.
from reported sensitivity, specificity, positive and negative predictive value data.

Endpoints
The primary aim was to assess the diagnostic accuracy of quantitative FIT for CRC detection in purely symptomatic patients at the NICE DG30 recommended threshold of 10 µg/g and the LoD for each analyser. Secondary goals were to assess the utility of FIT for symptom clusters: high-risk; low-risk; iron-deficiency anaemia, change in bowel habit and rectal bleeding.

Risk of bias assessment
Studies were assessed for potential risks of bias and applicability independently by three authors (RB, RC and NDS) using the Quality Assessment of Diagnostic Accuracy Studies 2 tool (QUADAS-2). An example screening tool and summary table of assessments are included in appendix B.

The extent to which publication bias occurs in studies of test accuracy is uncertain, however, simulation studies have indicated that the effect of publication bias on meta-analytic estimates of test accuracy is minimal. We did not undertake a statistical assessment of publication bias; however, our search strategy included a variety of routes to identify unpublished studies and resulted in the inclusion of several conference abstracts.

Statistical analysis
Meta-analyses were performed grouping studies by reference standard (tier 1 or tier 2) and threshold for a negative FIT result (LoD or 10 µg/g). The LoD of the HM-JACKarc and FOB Gold assays is <2 µg/g, OC-Sensor <4 µg/g and QuikRead go <10 µg/g. Studies using the HM-JACKarc and OC-Sensor at these thresholds were combined within the LoD analyses. Subgroup analyses were performed to investigate the potential effects of analyser, threshold, presenting symptom cluster and recruitment location, on estimates of test accuracy.

Where four or more studies were available, bivariate random-effects analysis was performed to give summary estimates of sensitivity and specificity using STATA 13. Where this was not possible, random-effects meta-analysis was performed using Meta DiSc 1.4. Hierarchical summary receiver operating characteristic curves (HSROCs) were constructed using a bivariate model, based on Reitsma et al, and graphically depicted summary operating points and 95% confidence regions. HSROC analyses were performed using version 0.5.10 of package mada in R version 4.1.0.

Role of the funding source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results
Literature search, study characteristics and symptom clusters
The literature search yielded 10,447 discrete titles once results from Medline and EMBASE were combined with duplicates removed (Figure 1). Title screening removed 9751 irrelevant titles. For the remaining 696 titles, abstracts were reviewed and screened to 221 full-text articles. Full-text articles which were screened and excluded, with reasons for exclusion are listed in appendix C.

Thirty-one articles (n = 79,566) included inclusions criteria and had extractable data. Of these 16 (n = 35,945) were eligible for inclusion in tier 1 and 11 (n = 43,621) in tier 2. Four studies reported data on a specific symptom derived from a previously included cohort and were used for evaluation of FIT in those symptoms only, to avoid “double counting”.

Nine studies (n = 31,190) reported data on patients presenting with high-risk symptoms and 6 studies (n = 68,42) on low-risk symptoms. Four studies (n = 1050) yielded data for iron-deficiency anaemia and 3 studies each yielded data on FIT utility in change in bowel habit (n = 11,211) and rectal bleeding (n = 3665).

The median CRC prevalence within included studies was 3.7% (range 1.1−16.0%).

Table 2 shows details of geographical study setting; recruitment location (primary or secondary care); reference tests; study cancer prevalence; presenting symptoms; and individual study sensitivity, specificity, positive and negative predictive values for CRC by FIT threshold, for all included studies.

QUADAS-2 assessment
Appendix B shows the quality assessment of the 31 included studies using the QUADAS-2 instrument. Nine studies were assessed as having low risk of bias and applicability concerns across all domains. Ten studies were assessed as having high risk of bias in the patient selection domain. These concerns principally centred around either non-consecutive recruitment (for both tier 1 and tier 2 studies), or discretionary referral for investigation in the event of ongoing clinical concern for tier 2 studies. There were no significant differences seen in FIT sensitivity when comparing those studies assessed as being at high risk of bias with those at low
| Study | Setting | Region | Reference test(s) | Tier 1 → 80% Caloroscopy or CT Colonography | Study cancer prevalence (%) | Setting | Reference test(s) | Tier 2 — other reference standards |
|-------|---------|--------|------------------|---------------------------------------------|-----------------------------|--------|------------------|-----------------------------------|
| Chapman, 2021 | Primary care | UK (Eng), single centre | Colonoscopy | 52 | OC-Sensor | High-risk | Chapman, 2021 | OC-Sensor |
|            |          |        |                  | 732 | >10 | 89.5 (79.5, 95.8) | HM-JACKarc | 741 (70.7, 77.2) |
|           |          |        |                  |     | >4 U(LoD) | 97.4 (86.5, 98.9) | 4.4 | 159 (11.6, 21.4) |
|           |          |        |                  |     | >4 | 84.2 (59.2, 99.2) | 78 (74.3, 80.9) | 92 (17.2, 23.4) |
| D’Souza, 2020 | Secondary care | UK (Eng), single centre | Colonoscopy | 40 | HM-JACKarc | High-risk | D’Souza, 2020 | HM-JACKarc |
|           |          |        |                  | 160 | >10 | 91.1 (64.7, 100) | 100 (67.7, 100) | 93.3 (87.7, 96.6) |
|           |          |        |                  |     | >2 U(LoD) | 84.7 (72.9, 93.9) | 100 (51, 100) | 308 (12.7, 57.0) |
|           |          |        |                  |     | >10 | 87.5 (72, 90.9) | 81 (74, 87.7) | 100 (97.1, 100) |
|           |          |        |                  |     | >10 | 100 (67.6, 100) | 100 (51, 100) | 100 (97.1, 100) |

Table 2 (Continued)
## Table 2: Study characteristics stratified by reference standard.

*CI*: confidence interval; *CIBH*: change in bowel habit; *IDA*: iron-deficiency anaemia; *LoD*: limit of detection; *Unstratified*: Symptoms not stratified by “high-risk” or “low-risk” definitions; *NG12*: NICE guideline 12 (Suspected cancer: recognition and referral) – “high-risk symptoms”; *DG30*: NICE diagnostics guidance 30 (Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care) – “low-risk symptoms”; *republished as an abstract.*

| Study | Setting | Region | Reference test(s) | Study cancer prevalence (%) | Analyser | Presenting symptoms | n patients | Threshold (µg/g) | Sensitivity % (95% CI) | Specificity % (95% CI) | PPV % (95% CI) | NPV % (95% CI) |
|-------|---------|--------|-------------------|-----------------------------|----------|---------------------|------------|-----------------|-------------------------|--------------------------|----------------|----------------|
| Maclean, 2021 | Secondary care | UK (Eng), single centre | Colonoscopy, CT Colonoscopy, Flexible Sigmoidoscopy | 2.5 | QuickRead go | Low-risk | 553 | >10 | 92.9 (88.5, 96.7) | 70.1 (66.1, 73.8) | 7.5 (4.4, 12.4) | 99.7 (98.5, 100) |
| Mowat, 2021 | Primary care | UK (Sco), single centre | Colonoscopy, CT Colonoscopy, CT Colonoscopy, Flexible Sigmoidoscopy, Registry follow-up | 2.0 | HM-JACKarc | Unstratified | 5381 | >10 (LoD) | 97.1 (91.9, 99.9) | 89.5 (84.1, 93.8) | 7.1 (4.3, 10.9) | 99.9 (99.7, 100) |
| Nicholson, 2018 | Primary care | UK (Eng) | Colonoscopy, CT Colonoscopy, CT Colonoscopy, CT Colonoscopy, Registry follow-up | 7.0 | HM-JACKarc | Low-risk | 238 | >7 | 85.7 (48.7, 97.4) | 75.9 (74.7, 77.7) | 6.8 (3.6, 8.3) | 99.7 (99.5, 99.8) |
| Nicholson, 2020 | Primary care | UK (Eng) | Registry Fl/Up | 1.1 | HM-JACKarc | Unstratified | 9896 | >10 | 90.5 (83.4, 94.7) | 91.3 (90.8, 91.9) | 10.1 (8.3, 12.2) | 99.9 (99.6, 99.9) |
| Pin Vieto, 2020 | Primary care | Spain, multicentre | Registry F/up | 1.4 | OC-Sensor | Unstratified | 5623 | >10 | 80.2 (70.3, 87.5) | 84.1 (83.1, 85.2) | 6.9 (5.4, 8.7) | 99.7 (99.4, 99.8) |
| Widlak, 2017 | Secondary care | UK (Eng), single centre | Colonoscopy, CT Colonoscopy, CT Colonoscopy, CT Colonoscopy, Registry follow-up | 5.8 | HM-JACKarc | Unstratified | 430 | >7 | 88.0 (70.9, 98.6) | 93.1 (96.2, 95.2) | 44 (31.2, 57.7) | 99.2 (97.7, 99.7) |
| Cunin, 2020 | Secondary care | UK (Eng), single centre | Colonoscopy, CT Colonoscopy, CT Colonoscopy, CT Colonoscopy, Registry follow-up | 5.2 | HM-JACKarc | IDA | 189 | >10 | 80 (68.4, 91.9) | 81.7 (75.1, 88.8) | 54 (42.2, 68.3) | 97.2 (96.9, 98.1) |
| D Saoza, 2021 | Secondary care | UK (Eng), multicentre | Colonoscopy | 3.3 | HM-JACKarc | IDA | 479 | >10 | 100 (89.6, 100) | 81.6 (77.8, 84.9) | 28.7 (21.2, 37.5) | 100 (99, 100) |
| Digby, 2021 | Primary care | UK (Eng), single centre | Colonoscopy | 5.6 | OC-Sensor | Rectal bleeding | 464 | >10 | 96.2 (81.1, 99.3) | 38.3 (33.9, 42.9) | 8.3 (5.8, 12.8) | 99.4 (96.7, 99.9) |
| Hicks, 2021 | Secondary care | UK (Eng), multicentre | Colonoscopy | 3.3 | HM-JACKarc | Rectal bleeding | 3143 | >10 | 96.6 (92.3, 98.5) | 76.6 (75.7, 78.1) | 16.8 (14.5, 19.5) | 99.8 (99.5, 99.9) |
risk of bias, however two studies showed a moderately lower specificity at a threshold of 10 µg/g (47.0% (95% CI: 45.6, 48.5) and 57.1% (95% CI: 50.3, 60.5) respectively. Seven tier 2 studies were assessed as being at risk of bias in the reference standard domain, owing to the use of multiple reference tests or registry follow-up. Ten studies had applicability concern in the patient selection domain, often due to concerns regarding exclusions within the study population (Tiers 1 and 2).

**Tier 1 analysis**

Sixteen studies were included in tier 1: 6 using OC-Sensor; 6 using HM-JACKarc; 1 comparing both analysers; 1 using QuikRead go and 2 using FOB Gold (Table 3).

FIT performance by presenting symptom and study setting were explored by pooling all studies using a common threshold (LoD for the assay or >10 µg/g) across analysers. At >10 µg/g threshold, for the ‘all symptoms’ analysis (16 studies, n = 35,943) the summary estimates of sensitivity and specificity were 91.0% (95% CI: 88.9, 92.7) and 75.2% (95% CI: 69.6, 80.1). For studies reporting on high-risk symptoms (7 studies, n = 18,264), the summary estimates of sensitivity and specificity were 88.7% (95% CI: 84.4, 92.0) and 78.5% (95% CI: 73.0, 83.2), and for studies reporting on low-risk symptoms (3 studies, n = 2162), the summary sensitivity and specificity estimates were 94.7% (95% CI: 90.5, 97.1) and 66.5% (95% CI: 58.7, 73.6) for high-risk symptoms and specificity were 94.7% (95% CI: 90.5, 97.1) and 66.5% (95% CI: 58.7, 73.6), and for low risk symptoms (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9) respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively. Figure 2 shows HSROC curves comparing FIT performance at LoD by symptom cluster (high-risk, low-risk, or unstratified) for any analyser and at a threshold of 10 µg/g (2 studies, n = 2082). 94.7% (95% CI: 85.4, 92.8) and 71.9% (95% CI: 69.9, 73.9), respectively.
| Presenting symptoms | Analysers | Threshold (μg/g) | Setting | Number of studies (references) | n patients in analysis | Sensitivity (%) | Specificity (%) |
|---------------------|-----------|----------------|---------|-------------------------------|-----------------------|----------------|----------------|
|                     | Pooled analysers | LoD | Any | 7 (15, 16, 17, 21, 24, 33, 39) | 26056 | 94.7 (90.5, 97.1) | 66.5 (58.7, 73.6) |
| Symptom groups      | (>4 OC-Sensor) | Primary care | 3 (15, 21, 24) | 7300 | 94.9 (89.8, 97.9) | 67.5 (66.4, 68.6) |
|                     | (>2 HM-JACKarc) | Secondary care | 3 (16, 17, 39) | 15160 | 95.7 (93.5, 97.3) | 63.7 (62.9, 64.5) |
|                     | (>10) | Any | 16 (15, 16, 17, 18, 19, 20, 21, 33, 22, 23, 24, 25, 26, 27, 39, 58) | 35945 | 91.0 (88.9, 92.7) | 75.2 (69.6, 80.1) |
|                     | | | | | | 91.2 (89.2, 92.8) | 75.0 (69.4, 79.8) |
|                     | | Primary care | 4 (15, 21, 24, 27) | 12141 | 90.1 (83.9, 94.1) | 72.6 (58.6, 83.3) |
|                     | | Secondary care | 4 (15d, 21, 22, 24) | 12141 | 91.1 (85.7, 94.5) | 71.6 (57.7, 82.2) |
|                     | | Primary care | 10 (15, 17, 18, 19, 20, 25, 26, 27, 39, 58) | 19963 | 91.6 (89.2, 93.6) | 77.2 (71.1, 82.3) |
| OC-Sensor           | >4 | Any | 4 (15, 21, 24, 33) | 10896 | 95.0 (80.7, 98.9) | 65.8 (53.2, 76.5) |
|                     | | Primary care | 3 (15, 21, 24) | 7300 | 94.9 (89.8, 97.9) | 67.5 (66.4, 68.6) |
|                     | | Secondary care | 0 | | | |
|                     | >6 | Any | 1 (33d) | 3596 | 86.7 (77.9, 92.9) | 76.1 (74.7, 77.5) |
|                     | | Primary care | 0 | | | |
|                     | | Secondary care | 0 | | | |
|                     | >10 | Any | 7 (15, 20, 21, 23, 24, 26, 33) | 13716 | 90.2 (86.2, 93.1) | 74.5 (68.1, 79.9) |
|                     | | Primary care | 3 (15, 21, 24) | 7300 | 89.1 (82.7, 93.8) | 79.9 (79.0, 80.9) |
|                     | | Secondary care | 2 (20, 26) | 2575 | 93.9 (90.1, 96.5) | 70.3 (68.4, 72.1) |
| HM-JACKarc          | >2 | Any | 3 (16, 17, 39d) | 15160 | 95.7 (93.5, 97.3) | 63.7 (62.9, 64.5) |
|                     | | Primary care | 0 | | | |
|                     | | Secondary care | 3 (16, 17, 39) | 15160 | 95.7 (93.5, 97.3) | 63.7 (62.9, 64.5) |
|                     | >4 | Any | 1 (15d) | 732 | 92.1 (78.6, 98.3) | 70.0 (66.5, 73.4) |
|                     | | Primary care | 1 (15d) | 732 | 92.1 (78.6, 98.3) | 70.0 (66.5, 73.4) |
|                     | | Secondary care | 0 | | | |
|                     | >10 | Any | 7 (15, 16, 17, 18, 19, 22, 39) | 21829 | 90.6 (87.6, 92.9) | 78.2 (69.2, 85.2) |
|                     | | Primary care | 2 (15, 22) | 5573 | 93.4 (90.0, 95.9) | 51.1 (49.8, 52.5) |
|                     | | Secondary care | 5 (16, 17, 18, 19, 39) | 16256 | 89.7 (86.4, 92.3) | 82.4 (79.2, 85.2) |

Table 3 (Continued)
| Presenting symptoms | Analyser         | Threshold (µg/g) | Setting          | Number of studies (references) | n patients in analysis | Sensitivity (%) | Specificity (%) |
|---------------------|------------------|----------------|------------------|-------------------------------|------------------------|----------------|-----------------|
|                     |                  |                |                  |                               |                        |                |                 |
|                     | FOB Gold         | >10            | Any              | 2 (25, 58)                    | 890                    | 95.2 (86.5, 99.0) | 71.3 (68.0, 74.3) |
|                     |                  |                | Primary care     | 0                             |                        |                |                 |
|                     |                  |                | Secondary care   | 2 (25, 58)                    | 890                    | 95.2 (86.5, 99.0) | 71.3 (68.0, 74.3) |
|                     | QuikRead go      | >10            | Any              | 1 (27)                        | 242                    | 92.3 (64.0, 99.8) | 77.3 (71.3, 82.6) |
|                     |                  |                | Primary care     | 0                             |                        |                |                 |
|                     |                  |                | Secondary care   | 1 (27)                        | 242                    | 92.3 (64.0, 99.8) | 77.3 (71.3, 82.6) |
|                     | High-risk        |                 |                  |                               |                        |                |                 |
|                     | Pooled analysers | LoD            | Any              | 4 (16, 17, 21, 33) (†)        | 16768                  | 92.8 (86.4, 96.3) | 70.3 (66.5, 73.8) |
|                     |                  |                | Primary care     | 2 (15, 21)                    | 6550                   | 93.6 (87.3, 97.4) | 70.2 (69.1, 71.3) |
|                     |                  |                | Secondary care   | 2 (16, 17)                    | 7534                   | 97.7 (95.1, 99.2) | 63.1 (62.0, 64.3) |
|                     |                  |                |                  |                               |                        |                |                 |
|                     |                  | >10            | Any              | 7 (15, 16, 17, 18, 21, 23, 33) | 18264                  | 88.7 (84.4, 92.0) | 78.5 (73.0, 83.2) |
|                     |                  |                | Primary care     | 2 (15, 21)                    | 6550                   | 87.3 (79.6, 92.9) | 80.5 (79.5, 81.4) |
|                     |                  |                | Secondary care   | 3 (16, 17, 18)                | 7873                   | 91.0 (86.1, 94.6) | 71.2 (70.3, 72.1) |
|                     |                  |                |                  |                               |                        |                |                 |
|                     |                  | >6             | Any              | 3 (15, 21, 33)                | 10146                  | 91.0 (86.1, 94.6) | 71.2 (70.3, 72.1) |
|                     |                  |                | Primary care     | 2 (15, 21)                    | 6550                   | 93.6 (87.3, 97.4) | 70.2 (69.1, 71.3) |
|                     |                  |                | Secondary care   | 0                             |                        |                |                 |
|                     |                  | >10            | Any              | 4 (15, 21, 23, 33)            | 10391                  | 88.7 (78.8, 94.3) | 74.2 (65.0, 81.7) |
|                     |                  |                | Primary care     | 2 (15, 21)                    | 6550                   | 89.1 (81.7, 94.2) | 80.0 (79.0, 81.0) |
|                     |                  |                | Secondary care   | 0                             |                        |                |                 |
|                     |                  | >10            | Any              | 4 (15, 21, 16, 17, 18)        | 8605                   | 89.0 (82.5, 93.3) | 81.1 (79.1, 82.9) |
|                     |                  |                | Primary care     | 1 (15)                        | 732                    | 92.1 (78.6, 98.3) | 70.0 (66.5, 73.4) |
|                     |                  |                | Secondary care   | 3 (16, 17, 18)                | 7873                   | 91.3 (87.5, 94.2) | 82.3 (81.4, 83.2) |

Table 3 (Continued)
| Presenting symptoms | Analysers | Threshold (mg/g) | Setting | Number of studies (references) | n patients in analysis | Sensitivity (%) | Specificity (%) |
|---------------------|-----------|------------------|---------|--------------------------------|------------------------|----------------|----------------|
| Low-risk            | HM-JACKarc | >2<sup>a</sup>   | Any     | 2 (16, 17)<sup>b</sup>         | 2082                   | 94.7 (85.4, 98.9)<sup>c</sup> | 71.9 (69.9, 73.9)<sup>c</sup> |
|                     |           |                  | Primary care | 0                            |                        |                |                |
|                     |           |                  | Secondary care | 2 (16, 17)                  | 2082                   | 94.7 (85.4, 98.9)<sup>c</sup> | 71.9 (69.9, 73.9)<sup>c</sup> |
|                     |           | >10              | Any     | 3 (16, 17, 18)<sup>d</sup>     | 2161                   | 88.7 (78.1, 95.3)<sup>c</sup> | 88.5 (87.1, 89.9)<sup>c</sup> |
|                     |           |                  | Primary care | 0                            |                        |                |                |
|                     |           |                  | Secondary care | 3 (16, 17, 18)               | 2161                   | 88.7 (78.1, 95.3)<sup>c</sup> | 88.5 (87.1, 89.9)<sup>c</sup> |
| Individual symptoms | OC-Sensor | >4<sup>e</sup>   | Primary care | 1 (21)                        | 5818                   | 91.7 (82.7, 96.9) | 69.7 (68.5, 70.9) |
|                     |           | >10              | Any     | 1 (21)                        | 5818                   | 88.9 (79.3, 95.1) | 80.8 (79.7, 81.8) |
|                     | HM-JACKarc | >2<sup>e</sup>   | Secondary care | 1 (42)                       | 4249                   | 91.4 (83.0, 96.5) | 68.4 (67.0, 69.8) |
|                     |           | >10              | Any     | 1 (42)                        | 4249                   | 82.7 (72.7, 90.2) | 87.5 (86.5, 88.5) |
|                     | Pooled analysers | LoD (>4 OC-Sensor >2 HM-JACKarc) | Any | 2 (21, 42)                 | 10067                  | 91.5 (85.9, 95.4) | 69.1 (68.2, 70.1) |
| IDA                 | OC-Sensor | >10              | Unclear | 2 (23)                        | 245                    | 92.9 (76.5, 99.1) | 57.1 (50.3, 63.8) |
|                     | HM-JACKarc |                  | Secondary care | 1 (42)                      | 479                    | 100 (89.4, 100)  | 81.6 (77.7, 85.1) |
|                     | Pooled analysers | LoD (>4 OC-Sensor >2 HM-JACKarc) | Any | 2 (23, 42)                 | 724                    | 96.7 (88.7, 99.6) | 73.6 (70.1, 76.9) |
| Rectal bleeding     | OC-Sensor | >10              | Primary care | 1 (43)                        | 462                    | 96.2 (80.4, 99.9) | 38.3 (33.7, 43.0) |
|                     | HM-JACKarc |                  | Secondary care | 1 (44)                      | 3143                   | 96.6 (92.2, 98.9) | 76.6 (75.0, 78.1) |
|                     | QuikRead go |                  | 1 (27) | 60                            | 100 (54.1, 100)       | 74.1 (60.3, 85.0) |                |
|                     | Pooled analysers | Any          |         | 3 (27, 43, 44)              | 3665                   | 96.6 (92.8, 98.8) | 71.7 (70.2, 73.2) |

Table 3: Accuracy of FIT, tier 1 reference standard (≥90% of participants received colonoscopy or CTC), comparing individual and combined assays, symptom clusters and study setting: Summary estimates (95% CI).

<sup>a</sup> LoD for the assay.
<sup>b</sup> Bivariate Meta-analysis (STATA 13).
<sup>c</sup> Random effects meta-analysis (Meta DiSc 1.4).
<sup>d</sup> All studies conducted in patients presenting with high-risk symptoms.
<sup>e</sup> All studies conducted in secondary care settings.
<sup>f</sup> All studies conducted in primary care settings.
<sup>g</sup> Only data for the HM-JACK assay for Chapman 2021 included to avoid double counting.
<sup>h</sup> Only data for the OC-Sensor assay for Chapman 2021 included to avoid double counting.

CIBH: change in bowel habit; CI: confidence interval; DG: diagnostic guidance; IDA: iron deficiency anaemia; LoD: limit of detection.
those with mixed reference standard (tier 2), (Tables 3 and 4). The performance characteristics of FIT for CRC detection, using a LoD threshold, were similar in Tier 1 and 2 studies. The overall summary sensitivity for all analysers and all presenting symptoms, were 94.7% (95% CI: 90.5, 97.1) and 96.7% (95% CI: 94.1, 98.3) for tier 1 and tier 2, respectively and the corresponding summary specificity were 66.5% (95% CI: 58.7, 73.6) and 63.7% (95% CI: 63.0, 64.4). At a threshold of 10 µg/g, there is more variation with the overall summary sensitivity for all analysers and all presenting symptoms being 91.0% (95% CI: 88.9, 92.7) for tier 1 and 88.2% (95% CI: 84.3, 91.2) for tier 2 with the corresponding specificity being 75.2% (95% CI: 69.6, 80.1) and 85.7% (95% CI: 81.0, 89.3) respectively. Figure 4 shows HSROC curves comparing FIT performance, at LoD for tier 1 and tier 2 studies and Figure 5 provides the same comparison at 10 µg/g.

Discussion
This meta-analysis was undertaken to inform the joint ACPCBI/BSG guidelines on the use of FIT in symptomatic patients and indicates that FIT has a high sensitivity well above 90% for CRC irrespective of the presenting symptom(s), particularly when used at a threshold at or near to the LoD for the assay.

Summary of principal findings
Performance of FIT in NICE-defined and individual symptoms. This is the first review to investigate the potential effect of presenting symptoms cluster, according to NICE definitions of “high-risk” and “low-risk”, on the diagnostic accuracy of FIT. Within the limitations of the available data (in the context of the relatively recent introduction of DG30 guidelines for low-risk symptoms and hence lack of longitudinal data), the sensitivity of FIT for CRC detection is unaffected by the definition of “symptomatic patients” used. This suggests that current definitions of “high-risk” and “low-risk” symptoms for CRC are no longer required in the FIT era, and that FIT can be used for all symptomatic patients when CRC is suspected to triage their need and urgency for investigation.

Drawing conclusions regarding the diagnostic accuracy of FIT for individual symptoms is more
challenging particularly given that patients commonly present with multiple symptoms. Whilst many studies give a breakdown of numbers of patients presenting with individual symptoms, relatively few provided analysable data by symptom (change in bowel habit, rectal bleeding or iron-deficiency anaemia). Within these limitations the summary sensitivity and specificity were broadly similar in patients presenting with rectal bleeding, change in bowel habit and iron-deficiency anaemia, and were comparable to the “all symptom”, high-risk, low-risk symptoms clusters. Therefore, there is no reason to treat these symptoms differently or exclude them from FIT testing.

Comparisons between analysers. We did not find any clinically significant difference in the performance of FIT for detection of CRC between currently available analysers. One study, included in our review, directly compared the performance of HM-JACKarc and OC-Sensor in a cohort of 732 patients and reported the sensitivity of OC-Sensor was marginally higher than HM-JACKarc at low thresholds of 4 µg/g and 10 µg/g. However, this study compared FIT performance at 4 µg/g, which is the LoD of OC-Sensor, but above the LoD for HM-JACKarc.

Performance of FIT in formal diagnostic accuracy studies and clinical pathways. The performance of FIT for CRC detection at the LoD were similar in both Tier 1 and 2 studies. At a threshold of 10 µg/g, there is more variation with a lower overall summary specificity for all analysers and all presenting symptoms for Tier 1 compared to Tier 2 studies. This appears to be driven by the low specificity in the McSorley study at 47% (95% CI: 45.6, 48.5) compared with the other studies (Table 2). In this study primary care physicians were not blinded to FIT result and indeed given guidance reassuring them of the low risk of CRC with a negative FIT result. Consequently, “FIT positivity” rate amongst those referred was 55% which is more than double that seen in other diagnostic accuracy studies. This referral bias may have led to the drop in specificity observed. Within the limitations described comparisons of the
| Symptom groups | Analyser | Threshold (µg/g) | Setting | Number of studies (references) | n patients in analysis | Sensitivity (%) | Specificity (%) |
|----------------|----------|----------------|---------|--------------------------------|-----------------------|----------------|-----------------|
| All (High-risk, low-risk or unstratified) | Pooled analysers | LoD | Any | 2 (29, 35) | 18,423 | 96.7 (94.1, 98.3) | 63.7 (63.0, 64.4) |
| | | (>4 OC-Sensor) | Primary care | 2 (29, 35) | 18,423 | 96.7 (94.1, 98.3) | 63.7 (63.0, 64.4) |
| | | (>2 HM-JACKarc) | Secondary care | 0 | 43,191 | 88.2 (84.3, 91.2) | 85.7 (81.0, 89.3) |
| | | (>10) | Any | 7 (29, 30, 31, 35, 36, 37, 38) | 41,532 | 88.8 (84.6, 92.0) | 85.7 (82.3, 88.5) |
| | | | Secondary care | 3 (28, 32, 34) | 1659 | 86.4 (75.7, 93.6) | 80.3 (78.2, 82.2) |
| | | (>4) | OC-Sensor | Any | 1 (29) | 13,042 | 96.5 (93.2, 98.5) | 69.5 (68.7, 70.3) |
| | | | Primary care | 1 (29) | 13,042 | 96.5 (93.2, 98.5) | 69.5 (68.7, 70.3) |
| | | | Secondary care | 0 | 22,305 | 86.6 (73.9, 93.6) | 87.5 (80.0, 92.4) |
| | | | Primary care | 3 (29, 31, 38) | 22,127 | 89.8 (86.2, 92.7) | 82.9 (82.4, 83.4) |
| | | | Secondary care | 1 (28) | 714 (290, 96.3) | 95.9 (91.7, 98.3) |
| | | (>7) | HM-JACKarc | Any | 1 (35) | 5381 | 97.1 (91.9, 99.9) | 49.5 (48.1, 50.8) |
| | | | Primary care | 1 (35) | 5381 | 97.1 (91.9, 99.9) | 49.5 (48.1, 50.8) |
| | High-risk | >10 | QuikRead go | Any | 1 (34) | 553 | 92.9 (66.1, 99.8) | 70.1 (66.1, 74.0) |
| | | | Primary care | 0 | 553 | 92.9 (66.1, 99.8) | 70.1 (66.1, 74.0) |
| | | | Secondary care | 1 (34) | 553 | 92.9 (66.1, 99.8) | 70.1 (66.1, 74.0) |
| | | (>10) | Pooled analysers | Any | 3 (28, 29, 32) | 14,148 | 90.7 (86.6, 93.8) | 82.0 (81.3, 82.6) |
| | | | Primary care | 1 (29) | 13,042 | 92.1 (87.8, 94.9) | 81.7 (81.0, 82.4) |
| | | | Secondary care | 2 (28, 32) | 1065 | 84.3 (71.4, 93.0) | 85.0 (82.7, 87.2) |
| | | (>4) | OC-Sensor | Any | 1 (29) | 13,042 | 96.5 (93.2, 98.5) | 69.5 (68.7, 70.3) |
| | | | Primary care | 1 (29) | 13,042 | 96.5 (93.2, 98.5) | 69.5 (68.7, 70.3) |
| | | | Secondary care | 0 | 928 | 86.7 (73.2, 94.9) | 83.5 (80.8, 85.9) |
| | | (>7) | HM-JACKarc | Any | 1 (32) | 928 | 92.9 (73.2, 94.9) | 83.5 (80.8, 85.9) |

Table 4 (Continued)
| Presenting symptoms | Analyser | Threshold (µg/g) | Setting | Number of studies (references) | n patients in analysis | Sensitivity (%) | Specificity (%) |
|---------------------|----------|-----------------|---------|-------------------------------|-----------------------|----------------|-----------------|
|                     |          |                 |         |                               |                       |                |                 |
|                     |          | Secondary care  | 0       |                               | 1 (32)                | 928            | 86.7 (73.2, 94.9) | 83.5 (80.8, 85.9) |
|                     |          | Primary care    | 0       |                               | 1 (32)                | 928            | 86.7 (73.2, 94.9) | 83.5 (80.8, 85.9) |
|                     |          | Secondary care  | 0       |                               | 1 (32)                | 928            | 86.7 (73.2, 94.9) | 83.5 (80.8, 85.9) |
|                     |          | Primary care    | 0       |                               | 1 (36)                | 238            | 85.7 (48.7, 97.4) | 89.2 (84.5, 92.6) |
|                     |          | Secondary care  | 0       |                               | 2 (36)                | 4128           | 84.5 (72.6, 92.7) | 85.3 (84.2, 86.4) |
|                     |          | Primary care    | 0       |                               | 2 (36)                | 4128           | 84.5 (72.6, 92.7) | 85.3 (84.2, 86.4) |
|                     |          | Secondary care  | 0       |                               | 1 (34)                | 553            | 92.9 (66.1, 99.8) | 70.1 (66.1, 74.0) |
|                     |          | Primary care    | 0       |                               | 1 (34)                | 553            | 92.9 (66.1, 99.8) | 70.1 (66.1, 74.0) |
|                     |          | Secondary care  | 0       |                               | 1 (41)                | 189            | 80.0 (56.3, 94.3) | 81.7 (75.0, 87.2) |
|                     |          | Primary care    | 0       |                               | 1 (43)                | 462            | 96.2 (80.4, 99.9) | 38.3 (33.7, 43.0) |

**Table 4**: Accuracy of FIT, tier 2 reference standard (Mixed reference tests and registry follow up), comparing individual and combined assays, symptom clusters and study setting: Summary estimates (95% CI).

- a LoD for the assay.
- b All studies conducted in primary care settings.
- c All studies conducted in patients presenting with high-risk symptoms.
- d Bivariate Meta-analysis (STATA 13).
- e Random effects meta-analysis (Meta DiSc 1.4).
- f All studies conducted in secondary care settings.
- g All studies conducted in patients presenting with low-risk symptoms.
- h CIBH: change in bowel habit; CI: confidence interval; IDA: iron deficiency anaemia; LoD: limit of detection.
performance characteristics of FIT estimated from tier 1 studies and from tier 2 studies suggest that FIT performance for CRC detection is adequate and transferrable to clinical diagnostic pathways for CRC.

Comparison with existing literature. Earlier meta-analyses,33,45,46 were hampered by a low number of studies and heterogeneity with mixed cohorts including patients in screening populations and in some cases CRC/polyp surveillance populations. Two meta-analyses in 2021,47,48 included studies with different reference standards (ie registry or clinical follow-up cohorts and colonic investigation cohorts). This approach may introduce verification bias and prevents comparison of the performance characteristics of FIT, as has been done in our review. Despite these methodological differences and variation in the data, the results have been consistently similar which supports the robustness of use of FIT in symptomatic patients for detection of CRC.

Potential impact on referral rates. Figure 6 demonstrates the potential impact of the use of FIT applied to a hypothetical cohort of 1000 patients presenting to primary care using data from the tier 1 all analyser, all-symptoms analyses, at LoD and 10 µg/g thresholds, applied to a CRC prevalence rate of 3.3% as observed in 2019 England national data.49 Using a LoD threshold, 335 patients would return a faecal haemoglobin (f-Hb) above the LoD threshold or “FIT positive” and ideally would undergo full colonic imaging with colonoscopy or CT colonography. Of these, 31 would be correctly identified as having CRC. 2 patients with CRC would not be detected (FIT negative or undetectable f-Hb) and 643 patients would be correctly identified by FIT as not having CRC. Applying a threshold of 10 µg/g, 258 would be “FIT positive” and of these 30 CRC would be identified. 3 CRCs would not be detected and 739 would have been correctly identified by FIT as not having CRC. According to this analysis, use of a threshold of 10 µg/g rather than the LoD has the potential to reduce colonic investigations by 23%, with the caveat of a marginally higher rate of “FIT negative” or undetected cancers.

Strengths and limitations

We used the same thorough search strategy for the meta-analysis that informed NICE DG30 guidelines; the search strategy was based on terms for the test and target condition and did not include any study design filters. We excluded diagnostic case-control or ‘two-gate’
studies, as this study design has been found to produce inflated estimates of test accuracy, compared with those derived from diagnostic cohort studies. Our primary analysis considered diagnostic accuracy studies where more than 90% of patients studied had full colonic imaging, the ‘gold standard’ for diagnosis of CRC, (Tier 1). However, we also included pragmatic studies that reflect clinical practice, where not all patients were suitable to undergo full colonic imaging, but (clinically) justified other investigations such as CT scan and flexible sigmoidoscopy, and also included studies with Registry follow-up for some or all FIT negative patients (Tier 2). The minimum follow-up period was chosen following review of the Nicholson study which showed no significant difference in sensitivity after a follow-up period of 3, 6 or 12 months, with FIT negative cancers presenting within first 3 months.

The inclusion of these studies allowed comparisons between the diagnostic performance of FIT when the ‘gold standard’ method is used to determine the presence or absence of CRC, and the diagnostic performance of FIT when the presence or absence of CRC is determined as it would be in clinical practice. We did not pool data across the two categories of reference standard, whether for individual symptoms or for wider populations because, where a different reference standard is used, this essentially explores the performance of FIT to detect a different definition of the target condition.

A potential limitation of this study is the strong geographical weighting to United Kingdom-based studies (23/31 studies). The other eight included studies recruited cohorts from other European nations (5 Spanish, 1 Danish, 1 Norwegian and 1 Swedish). This geographical bias could potentially limit the generalisability of our study although we have not seen significant variation in FIT performance in these studies.

Current uncertainties
There is currently a relative paucity of diagnostic accuracy data for the QuikRead go (2 studies) and FOB Gold analysers (2 studies). Meta-analysis for these was only possible for the overall all symptoms, all analysers analysis at a threshold of 10 µg/g. No data were available for FOB Gold at the LoD threshold.

The broadening of the symptom definition within NG12 guidelines diluted their PPV and thus risks overburdening diagnostic services. Two studies demonstrated a reduction in PPV from 7.5% to 3.7% and 8.5%
Figure 6. FIT and colonoscopy outcomes comparing a LoD threshold with a threshold of 10 µg/g for a hypothetical 1000 patients with a CRC prevalence of 3.3%.
to 3.5% respectively following the introduction of NG12 with a concurrent increase in referrals. A large Danish prospective cohort study \(^5\) of 37,455 patients reported the PPV for CRC for the symptoms of abdominal pain, change in stool frequency, change in stool texture and rectal bleeding were 0.3%; 0.4%; 0.2% and 0.6% respectively. For studies included in our review, the use of FIT at a threshold of 10 µg/g increased the PPV for CRC for change in bowel habit to between 5.3% and 11.4\(^{35,37,41}\); rectal bleeding between 8.5% and 30%\(^{26,44,49}\) and iron-deficiency anaemia between 21.8% and 41.7%\(^{22,27,40,41}\). There is a current lack of data for included individual symptoms at a LoD threshold and at any threshold for other symptoms such as abdominal pain and weight loss.

Whilst we are relatively confident that FIT has sufficient operational sensitivity (irrespective of population or other variables), practical considerations about capacity for investigations as well as unnecessary alarm and investigations for false positive patients might suggest that work is still needed to refine and optimise the criteria used to select patients for testing to increase FIT specificity, at these low thresholds. Multivariable prediction modelling studies may be useful, in this context, to assess the independent predictive value of a “positive” FIT result, in the context of individual symptoms and clinical risk factors. Prediction modelling studies should consider the trade-off between the potential for improved predictive performance and ease of use (the extent to which the components of any risk score developed are readily available to and easily used by clinicians). Furthermore, despite the high sensitivity of FIT for CRC detection, there are still a small number of cancers that will not be detected, and it is therefore essential that appropriate “safety-netting” be in place to refer patients with persistent symptoms and a “negative” FIT. Research into means of optimising FIT sensitivity by repeat testing, or sampling technique may further reduce false negative results.

**Summary and conclusions**

There is evidence to suggest that FIT can be used at a threshold of 10 µg/g or the LoD as an initial test to triage patients when CRC is suspected irrespective of the presenting symptom cluster to determine the need and urgency for investigations. Within the limitations of the available data, the sensitivity of FIT assays at these thresholds appears to be unaffected by assay (OC-Sensor or HM-JACKarc), study setting or definition of “symptomatic patients”. Although the sensitivity is maximised at LoD, the specificity is relatively low. At a threshold of 10 µg/g, the specificity improves with a slightly lower sensitivity. Clinical services should consider the trade-off between the impact on diagnostic services and potential missed cancer rates when deciding on the most appropriate FIT threshold to use in their clinical setting.

**Contributors**

Muti Abulafi: Supervision, Conceptualisation, Methodology, Investigation, Writing — Original Draft, Review and Editing, Project Administration, Funding Acquisition. Richard Booth: Investigation, Writing — Original Draft, Review and Editing, Visualisation. Rachel Carten: Investigation, Writing — Review and Editing. Nigel D’Souza: Investigation, Writing — Review and Editing. Marie Westwood: Methodology, Analysis, Review and Editing. Jos Kleijnen: Methodology, Analysis, Review and Editing.

**Data sharing statement**

We obtained permission from Dr Kai Saw and Professor Ian Bissett, Faculty of Medical and Health Sciences, University of Auckland, New Zealand to use the unique R data code to generate the HSROC curves.

**Declaration of interests**

We declare no competing interests.

**Acknowledgements**

We are grateful for the advice and support of Rafay Siddiqui (RS) and Hutan Ashrafian (Imperial College, London) in the initial stages of this systematic review. RS advised authors RB and RC on use of 2 × 2 tables. We are also grateful to Dr Kai Saw and Professor Ian Bissett (Department of Surgery, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand) for their assistance in providing the relevant codes to enable us to generate the HSROC curves on R stats package. We are also grateful to Dr Robert Wolff (Kleijnen Systematic Reviews Ltd) for generating the HSROC Curves.

**Funding**

This review was part funded by NHS England awarded to RM Partners, the West London Cancer Alliance hosted by The Royal Marsden NHS Foundation Trust. Richard Booth and Rachel Carten were funded by a fellowship awarded by Croydon University Hospital. Study funders had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

**Supplementary materials**

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.lanepe.2022.100518.
Articles

4 NICE. Suspected Cancer: Recognition and Referral [NG12]. National Institute for Health and Care Excellence; 2015.
5 NICE. Diagnostics Guidance [DG15]. Quantitative Faecal Immunochemical Tests to Guide Referral for Colorectal Cancer in Primary Care. National Institute for Health and Care Excellence; 2017.
6 NHS England. Cancer Waiting Times Annual Reports. NHS England and NHS Improvement; 2019.
7 Shenbagaraj L, Thomas-Gibson S, Stebbing J, et al. Endoscopy in 2017: a national survey of practice in the UK. Front Gastroenterol. 2019;10(2):177.
8 D’Souza N, Abulafi M. The state of the faecal immunometric test in symptomatic patients in the UK. Ann R Coll Surg Engl. 2021;102:104-105.
9 Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2 group. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol. 2005;58(9):882–893.
10 StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP; 2013.
11 Rodriguez-Alonso L, Rodriguez-Moranta F, Ruiz-Cerulla A, et al. Reducing the cut-off value of the faecal immunometric test for symptomatic patients does not improve diagnostic performance. Front Med (Lausanne). 2020;7.
12 Morales Arraez DEH, Carrillo G, Adrian M, Gimeno Z, Quintero AZ. Role of faecal immunometric testing in the diagnostic workup of patients with iron deficiency anaemia. United Eur Gastroenterol J. 2018;6:493–494.
13 NICE. National Institute for Health and Care Excellence; 2017.
14 NICE. Diagnosis Guidance [DG15] in the UK. Front Gastroenterol. 2019;10(2):177.
15 D’Souza N, Georgiou Delisile T, Chen M, Benton S, Abulafi M. The diagnostic accuracy of the faecal immunometric test for colorectal cancer in risk-stratified symptomatic patients. Ann R Coll Surg Engl. 2020;102(2):174–179.
16 D’Souza N, Georgiou Delisile T, Chen M, Benton S, Abulafi M. The NICE FIT Steering Group. Faecal immunometric 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;70:1130–1138.
17 Godber IM, Todd LM, Fraser CG, MacDonald LR, Younes HB. Use of a faecal immunometric test for haemoglobin can aid in the investigation of patients with lower abdominal symptoms. Clin Chem Lab Med. 2021;59(6):721–728.
18 Farrugia A, Widlak M, Evans C, Smith SC, Arasaradnam R. Faecal immunometric testing (FIT) in symptomatic patients: what are we missing? Front Gastroenterol. 2020;10:28–31.
19 Herrero JM, Vega P, Salve M, Bujanda I, Cubilosa J. Symptom or faecal immunometric test based referral criteria for colorectal cancer detection in symptomatic patients: a diagnostic tests study. BMC Gastroenterol. 2018;18:157.
20 Khasawneh F, Osborne R, Stephenson J, et al. Faecal immunometric testing is a cost-effective way to stratify symptomatic patients for urgent straight to test investigation. Colorectal Dis. 2020;22(4(suppl. 1)):6–12.
21 McSorley ST, Digby J, Clyde D, et al. Yield of colorectal cancer at colonoscopy according to faecal haemoglobin concentration in symptomatic patients referred from primary care. Colorectal Dis. 2021;23:1615–1621.
22 Godber IM, Todd LM, Fraser CG, MacDonald LR, Younes HB. Use of a faecal immunometric test for haemoglobin can aid in the investigation of patients with lower abdominal symptoms. Clin Chem Lab Med. 2021;59(6):721–728.
23 Morales Arraez DEH, Carrillo G, Adrian M, Gimeno Z, Quintero AZ. Role of faecal immunometric testing in the diagnostic workup of patients with iron deficiency anaemia. United Eur Gastroenterol J. 2018;6:493–494.
24 D’Souza N, Georgiou Delisile T, Chen M, Benton S, Abulafi M. The NICE FIT Steering Group. Faecal immunometric 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;70:1130–1138.
25 Khasawneh F, Osborne R, Stephenson J, et al. Faecal immunometric testing is a cost-effective way to stratify symptomatic patients for urgent straight to test investigation. Colorectal Dis. 2020;22(4(suppl. 1)):6–12.
26 McSorley ST, Digby J, Clyde D, et al. Yield of colorectal cancer at colonoscopy according to faecal haemoglobin concentration in symptomatic patients referred from primary care. Colorectal Dis. 2021;23:1615–1621.
27 Morales Arraez DEH, Carrillo G, Adrian M, Gimeno Z, Quintero AZ. Role of faecal immunometric testing in the diagnostic workup of patients with iron deficiency anaemia. United Eur Gastroenterol J. 2018;6:493–494.
28 Ayling RM, Lewis SJ, Cotter F. Potential roles of artificial intelligence and faecal immunometric testing for prioritisation of colorectal cancer referral in symptomatic patients: a 2-year evaluation. BJ Innov Opm. 2021;3(4):zaaa056.
29 Bailey JA, Weller J, Chapman CJ, et al. Faecal immunometric testing and blood tests for prioritization of urgent colorectal cancer referrals in symptomatic patients: a 2-year evaluation. BJ Innov Opm. 2021;3(4):zaaa056. https://doi.org/10.1038/s41416-021-01653-x.
30 Bailey SER, Abel GA, Atkins A, et al. Diagnostic performance of a faecal immunometric 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;124:1293–1296.
31 Juul JS, Hornung N, Andersen B, Laurberg S, Olesen F, Vedsted P. The value of using the faecal immunometric test in general practice on patients with non-alarm symptoms of colorectal cancer. Br J Cancer. 2020;119:471–479.
32 Khan AA, Klimovskij M, Harshen R. Accuracy of faecal immunometric testing in patients with symptomatic colorectal cancer. BJ Innov Opm. 2020;4:1180–1183.
33 Laslo HE, Seward E, Ayling RM, et al. Faecal immunometric test for patients with ‘high-risk’ bowel symptoms: a large prospective cohort study and updated literature review. Br J Cancer. 2021. https://doi.org/10.1038/s41416-021-02159-x.
34 Maclean W, Mackenzie P, Limb C, et al. Diagnostic accuracy of point of care faecal immunometric testing using a portable high-speed quantitative analyzer for diagnosis in 2-week wait patients. Colorectal Dis. 2021;23:2376–2386.
35 Mowat C, Digby J, Strachan JA, et al. Faecal haemoglobin concentration thresholds for reasurance and urgent investigation for colorectal cancer based on a faecal immunometric test in symptomatic patients in primary care. Ann Clin Biochem. 2021;54:211–219.
36 Nicholson BD, James T, East JEE, et al. Experience of adopting faecal immunometric testing to meet the NICE colorectal cancer referral criteria for low-risk symptomatic primary care referrals in Oxfordshire, UK. Front Gastroenterol. 2019;3:137–155.
37 Nicholson BD, James T, Paddon M, et al. Faecal immunometric testing for adults with symptoms of colorectal cancer attending English primary care: a retrospective cohort study of 14487 consecutive test requests. Aliment Pharmacol Therapeut. 2020;52:1051–1054.
38 Pinheiro N, Garcia Nimo L, Bujanda I, et al. Optimal diagnostic accuracy of quantitative faecal immunometric test positivity thresholds for colorectal cancer detection in primary health care: a community-based cohort study. United Eur Gastroenterol J. 2020;9:216–226.
39 Turvill J, Tarnock D, Cottingham D, et al. The Fast Track FIT study: diagnostic accuracy of faecal immunometric test for haemoglobin in patients with suspected colorectal cancer. Br J Gen Pract. 2021;71:E643–E652.
40 Widlak MM, Thomas CI, Thomas MG, et al. Diagnostic accuracy of faecal biomarkers in detecting colorectal cancer and adenoma in symptomatic patients. Aliment Pharmacol Therapeut. 2017;45:354–365.
41 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.
42 D’Souza N, Georgiou Delisile T, Chen M, Benton S, Abulafi M. The NICE FIT Steering Group. Faecal immunometric testing in symptomatic patients to prioritize investigation: diagnostic accuracy from the NICE FIT Study. Br J Surg. 2021;108:804–810.
43 Digby J, Strachan JA, McCann R, Steele RJC, Fraser CG, Mowat C. Measurement of faecal haemoglobin with a faecal immunometric test can assist in defining which patients attending primary care with rectal bleeding require urgent referral. Ann Clin Biochem. 2020;57(4):435–437.
44 Hicks G, D’Souza N, Georgiou Delisile T, Chen M, Benton SC, Abulafi M, The NICE FIT Steering Group. Using the faecal immunometric test in patients with rectal bleeding: evidence from the NICE FIT Study. Colorectal Dis. 2021;23:1650–1658.
45 Pint Viertel, Cunin L, Nimo L, Bujanda L, et al. Optimal diagnostic accuracy of faecal immunometric test positivity thresholds for colorectal cancer detection in primary health care: a community-based cohort study. United Eur Gastroenterol J. 2020;9:216–226.
46 Westwood M, Lang S, Armstrong N, et al. Faecal immunometric 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;15:389.

47 Pin Vieto N, Tejido-Sandoval C, de Vicente-Belza N, Sánchez-Gómez 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. 2021:01–11.

48 Saw KS, Liu C, Varghese C, Parry S, Bissett I. Faecal immunochemical test to triage patients with possible colorectal cancer symptoms: meta-analysis. Br J Surg. 2022;109(3):82–100.

49 NHS England. Cancer Waiting Times Annual Reports. Cancer Waiting Times Annual Report, 2013−14; Cancer Waiting Times Annual Report, 2018−19. NHS England and NHS Improvement; 2019. https://www.england.nhs.uk/statistics/statistical-work-areas/cancer-waiting-times/cwt-annual-reports/.

50 Rutjes AWS, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PMM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005;51(8):1335–1341.

51 Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med. 2004;140(3):189–202.

52 Lijnen JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. J Am Med Assoc. 1999;282(11):1061–1066.

53 Maclean W, Singh R, Mackenzie P, et al. The two-week rule colorectal cancer pathway: an update on recent practice, the unsustainable burden on diagnostics and the role of faecal immunochemical testing. Ann R Coll Surg Engl. 2020;102(4):310–311.

54 Christopher J, Flint T, Sloan K, Hall NR, Powar M. The impact of straight-to-test for the fast-track colorectal cancer pathway under the updated NICE guidelines. Colorectal Dis. 2017;19 (S4):P006.

55 Rasmussen S, Haastrup PF, Balasubramaniam R, et al. Predictive values of colorectal cancer alarm symptoms in the general population: a nationwide cohort study. Br J Cancer. 2019;120: 595–600.

56 Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982–990.

57 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. Pract Lab Med. 2022;29(6):1026–31.

58 Pickhardt PJ, Hassan C, Halligan S, Marmo R. Colorectal cancer: CT colonography and colonoscopy for detection—systematic review and meta-analysis. Radiology. 2011;259(2):393–405.