Faecal immunochemical tests (FIT) versus colonoscopy for surveillance after screening and polypectomy: a diagnostic accuracy and cost-effectiveness study


ABSTRACT

Objective The English Bowel Cancer Screening Programme (BCSP) recommends 3 yearly colonoscopy surveillance for patients at intermediate risk of colorectal cancer (CRC) post-polypectomy (those with three to four small adenomas or one ≥10 mm). We investigated whether faecal immunochemical tests (FITs) could reduce surveillance burden on patients and endoscopy services.

Design Intermediate-risk patients (60–72 years) recommended 3 yearly surveillance were recruited within the BCSP (January 2012–December 2013). FITs were offered at 1, 2 and 3 years post-polypectomy. Invitees consenting and returning a year 1 FIT were included. Participants testing positive (haemoglobin ≥40 µg/g) at years one or two were offered colonoscopy early; all others were offered colonoscopy at 3 years. Diagnostic accuracy for CRC and advanced adenomas (AAs) was estimated considering multiple tests and thresholds. We calculated incremental costs per additional AA and CRC detected by colonoscopy versus FIT surveillance.

Results 74% (5938/8009) of invitees were included in our study having participated at year 1. Of these, 97% returned FITs at years 2 and 3. Three-year cumulative positivity was 13% at the 40 µg/g haemoglobin threshold and 29% at 10 µg/g. 29 participants were diagnosed with CRC and 446 with AAs. Three-year programme sensitivities for CRC and AAs were, respectively, 59% and 33% at 40 µg/g, and 72% and 57% at 10 µg/g. Incremental costs per additional AA and CRC detected by colonoscopy versus FIT (40 µg/g) surveillance were £7354 and £180 778, respectively.

Conclusions Replacing 3 yearly colonoscopy surveillance in intermediate-risk patients with annual FIT could reduce colonoscopies by 71%, significantly cut costs but could miss 30%–40% of CRCs and 40%–70% of AAs.

Trial registration number ISRCTN18040196; Results.

INTRODUCTION

Most colorectal cancers (CRCs) arise from adenomas, and adenoma removal reduces CRC incidence and mortality. Some patients remain at increased risk of CRC post-polypectomy, and national guidelines recommend surveillance by colonoscopy. The UK surveillance guidelines recommend different strategies for patients at low risk,
intermediate risk and high risk. Surveillance is not recommended for low-risk patients (with one to two adenomas <10 mm), while 3-yearly surveillance is recommended for intermediate-risk patients (with three to four adenomas <10 mm or at least one ≥10 mm). High-risk patients (with five or more adenomas <10 mm, or three or more adenomas with at least one ≥10 mm) are recommended colonoscopy at 1 year before commencing 3-yearly surveillance.

Colonoscopy surveillance is associated with reduced CRC incidence; however, this procedure can have serious complications and be uncomfortable and anxiety inducing. Colonoscopy surveillance places great demand on endoscopy services, accounting for 20% of colonoscopies in the UK. Further increases in endoscopy demand due to widespread implementation of CRC screening and higher rates of primary care referrals for suspected CRC give grounds for finding alternative surveillance methods.

The faecal immunochemical test (FIT) may be an alternative. Like the guaiac faecal occult blood test (gFOBT), FIT detects occult blood in faeces, although it detects globin rather than haem, making it more specific for human blood. FIT is replacing gFOBT in CRC screening programmes because of its ease of use, increased uptake, quantitative analysis and greater sensitivity for advanced colorectal neoplasia (ACN). The English Bowel Cancer Screening Programme (BCSP) is planning to introduce FIT in 2018.

FIT has been extensively evaluated for screening, with sensitivity estimates of a single low-threshold FIT for CRC approaching 90%. There are few studies on FIT for surveillance, and these reported sensitivities for CRC of 70%–100%, although they included patients attending surveillance for personal or family histories of CRC. We therefore developed the ‘FIT for Follow-Up’ study to investigate the potential utility of FIT for postpolypectomy surveillance, looking specifically at patients deemed at intermediate-risk following screening within the BCSP. We focused on intermediate-risk patients as they account for most surveillance colonoscopies in the UK. Here we report the diagnostic accuracy of annual FIT for CRC and advanced adenomas (AAs) over 3 years versus 3 yearly colonoscopy, alongside estimates of cost-effectiveness.

METHODS

Study design and participants

Individuals aged 60–72 years were considered eligible if deemed at intermediate risk of CRC following colonoscopy performed <1 year previously in the BCSP for a positive gFOBT and scheduled to undergo 3 yearly colonoscopy surveillance. Individuals with ≥1 baseline colonic examination were excluded to prevent overinvestigation from examinations performed early due to positive FIT.

Potentially eligible individuals were identified by National Health Service (NHS) Digital using the Bowel Cancer Screening System (BCSS). Information on these individuals was sent in encrypted form by NHS Digital to the BCSP Southern Hub, one of five BCSP hubs across England. The hubs send invitations, mail and process faecal occult blood tests and record test results for their regional population. Although all five hubs were involved in this study, the Southern Hub coordinated all of these functions using a bespoke electronic patient management system (PMS). Each hub is associated with screening centres; 64 screening centres were involved in this study.

From January 2012 to December 2013, the Southern Hub sent invitations to consecutive individuals meeting the eligibility criteria, together with a participant information sheet, consent form and FIT kit. The kit contained a FIT device (OC-AUTO Sampling Bottle 3, Eiken Chemical Co Ltd, Japan), instructions, a plastic zip-lock bag and a prepaid envelope to return the kit to the Southern Hub.

Individuals who consented and returned an analysable FIT were included. Kits received >10 days after sample collection were not accepted. Samples were refrigerated on receipt and analysed within 1 week using the OC-Sensor DIANA analyser and new formulation sample buffer (Eiken Chemical Co Ltd, supplied by MAST Group Ltd, UK), according to manufacturer instructions. Results were entered on to the PMS. Repeat kits were dispatched to individuals who had lost theirs or had inadequate or spoil samples.

We initially used a positivity threshold of 20 µg haemoglobin per gram (µg/g) faeces. At this threshold, the positivity rate of the first 65 kits was 9.2%, higher than expected from the literature. We were concerned that if this rate was sustained, too many participants would be offered early colonic examinations and no further FITs, putting pressure on endoscopy services and affecting our ability to model FIT performance at different thresholds. The threshold was therefore raised to 40 µg/g. Of the six kits classed as positive using the 20 µg/g threshold, three had readings below 40 µg/g. These were not retrospectively reclassified as negative following the threshold change.

Participants were offered FITs at 1, 2 and 3 years postpolypectomy. Participants testing positive at years 1 or 2 were offered early colonic examination and no further FITs. Participants not attending the early examination were invited to the 3-year examination, as per UK guidelines. Participants testing negative at years 1 or 2 were offered further FITs, as were those who did not complete their year 2 FIT. All participants offered a year 3 FIT were invited to the routine 3-year examination. Findings at colonic examination (early or 3-year) were the reference standard against which we measured the diagnostic accuracy of FIT.

The default colonic examination was colonoscopy performed by accredited endoscopists, the ‘gold-standard’ for detection of colorectal lesions. CT colonography was an alternative for participants unfit for colonoscopy. Endoscopy findings were not known by those interpreting FIT results; however, endoscopists may have been aware when examinations were for positive FIT results.

Data collection and management

FIT results were reported as positive or negative to participants and their general practitioners. For participants testing positive at years 1 or 2, affiliated screening centres were informed of their participation in the study, positive FIT result and that they should be offered an early colonic examination. Early
examinations were organised by Specialist Screening Practitioners and administrators at the screening centres.

Colonic examination and pathology reports from early and 3-year examinations were obtained from screening centres. Patient, procedural and polyp data from these reports were entered on to the PMS using standard operating procedures. For some participants (n=28), we could not obtain all reports from screening centres, and the BCSS was used as a supplementary data source. Some participants received examinations at hospitals outside the BCSP. When made aware of this, usually by participants contacting the BCSP, we attempted to retrieve the reports; reports for 42 participants were retrieved in this way.

Colonic examinations occurring between consent date and 18 months after scheduled 3-year examination were included. Participants with examinations known to be incomplete or have poor visualisation (n=32) were not included as having examinations performed, as we could not determine accurately whether colorectal lesions were present. Participants may have undergone a combination of colonic examinations or surgery; all procedures were considered when defining outcomes of CRC and AAs. Cancer diagnoses to the end of 2014 were obtained from the English cancer registry. Participants with no record of a colonic examination occurring during the study but found to have CRC in registry data (n=2) were included as having examinations with CRC detected.

Polyp size was defined by the largest diameter reported on colonic examination, pathology or surgical reports. Adenomas ≥10 mm, with high-grade dysplasia, villous or tubulovillous histology were defined as AAs. We defined CRC sites by the International Classification of Diseases 10th revision, using codes C18–C20. We coded CRC morphologies using International Classification of Diseases for Oncology 2nd edition codes, including 8070/3, 8140/3, 8210/3, 8240/3, 8244/3, 8263/3, 8480/3 and 8481/3. We included all histological subtypes of CRC that arose in our dataset in our main analysis as surveillance programmes would ideally detect all cancers in the lower GI tract.

Statistical analysis
Our sample size calculation was based on an estimated relative sensitivity for ACN of three annual FITs versus 3 yearly colonoscopy. Assuming an ACN prevalence of 2.5%, relative FIT sensitivity of 75% and 40% compliance with all tests, 72 cases, 2881 adherent participants and 8000 invitees were required to provide sensitivity estimates with 95% CIs within ±10% and margin of error of ±10%.

We calculated rates of FIT return and positivity and diagnostic yields of CRC and AAs in participants who underwent colonic examination following positive FIT. We analysed the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of FIT for CRC and AAs at the 40 µg/g threshold. For the analyses of diagnostic accuracy for AAs, we excluded participants with CRC as we expected their inclusion would lead to biased estimates because FIT is more sensitive for CRC than AAs.

We estimated diagnostic accuracy at lower thresholds (10 µg/g, 20 µg/g and 30 µg/g) and with multiple tests, assuming that any ACN detected was present from year 1, remaining unchanged over successive years in the absence of colonic examination. Considering thresholds below 40 µg/g, we assumed that participants with faecal haemoglobin levels above the particular threshold underwent hypothetical early examinations, which detected what was actually found later on. Results from subsequent FITs were excluded from these analyses.

For analyses with multiple tests, we conducted cumulative test and programme analyses. For cumulative test analyses, participants were included if they completed the designated number of FITs (two for two-test analyses; three for three-test analyses) or tested positive with a previous FIT. All study participants were included in programme analyses; in two-test analyses, participants were ‘positive’ if they tested positive with either of their first two FITs, while in three-test analyses, participants were ‘positive’ if they tested positive with any FIT.

We conducted subgroup analyses by sex and age (≤65 years and >65 years) at invitation date. In sensitivity analyses, we examined effects on diagnostic accuracy estimates of excluding non-adenocarcinomas (squamous cell carcinomas and neuroendocrine tumours); these are rare cancers that account for <5% of CRCs. Further sensitivity analyses involved excluding participants with colonic examinations of unknown quality and those with IBT.

We assessed 3-year surveillance costs of annual FIT with colonoscopy in positive cases versus 3 yearly colonoscopy, calculating total costs and mean costs per participant. Cost-effectiveness was presented as incremental costs per additional AA and CRC detected by colonoscopy versus FIT surveillance. We estimated the budget impact of replacing colonoscopy with FIT surveillance nationally over a screening cycle (supplementary material: Economic analysis of faecal immunochemical test versus colonoscopy surveillance).

Analyses were performed in Stata/IC V.13. The protocol is online.

RESULTS
There were 9851 potentially eligible individuals. Of these, we excluded 186 with multiple baseline examinations and 109 due to informed dissent, clinical reasons, death or emigration. An additional 1547 individuals were not invited as the sample size of 8000 had been met. In total, 8009 individuals were invited (figure 1). The proportions of invitees aged ≤65 years and >65 years were, respectively, 49.3% and 50.7%; 34.7% were women (table 1).

Of those invited, 2060 were not recruited: 2055 did not consent and five consented but did not return an analysable FIT. A further 11 were excluded due to withdrawal of consent, baseline CRC, baseline high-risk adenomas, prolonged baseline episode or symptomatic exam before year 1 (figure 1). Therefore, 5938 of 8009 (74.1%) invitees were recruited, having consented and returned an analysable FIT (table 1). Return of FIT was 97% at years 2 and 3 (table 2). Participation was similar among men and women and across age groups (online supplementary tables 1 and 2).

FIT positivity decreased from 5.8% to 4.1% over years 1–3. Cumulative positivity over 3 years was 13.2% (table 2) and greater in men than women (14.7% vs 10.4%) and in participants aged >65 years than those ≤65 years (14.3% vs 12.1%) (online supplementary tables 1 and 2).

Colonic examinations were performed at some time during the study in 5225 of 5938 (88.0%) participants, yielding CRC in 29 (0.6%) and AAs in 446 (8.5%) (table 2). There were six participants with CRC and AAs. Of the 29 CRCs, 15 were proximal and 14 distal to the splenic flexure, and 26 were adenocarcinomas. Of the 26 CRCs with stage information, 10 were stage III or IV (online supplementary table 3).
Figure 1  Participant flow diagram from invitation through to year 3 colonic examination. *Two hundred and ninety-five potentially eligible individuals were not invited as they were excluded after the eligibility assessment (186 in order to prevent over-investigation as they had already undergone more than one colonoscopy and 109 due to informed dissent, clinical reasons, death or emigration) and a further 1547 individuals were not invited as the sample size of 8000 had already been met. †Two thousand and fifty-five individuals were not recruited due to lack of consent; one consented but did not return their FIT; and four consented but returned a FIT that could not be analysed. AAs, advanced adenomas; CRC, colorectal cancer; FIT, faecal immunochemical test; gFOBT, guaiac faecal occult blood test.
Table 1  Baseline characteristics of invited individuals by participation

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age at invitation date (years)</th>
<th>Invited*</th>
<th>Participants*</th>
<th>Non-participants*</th>
<th>P value†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>All ages</td>
<td>8009</td>
<td>5938 (74.1)</td>
<td>2071 (25.9)</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>≤65</td>
<td>3950</td>
<td>2877 (72.8)</td>
<td>1073 (27.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;65</td>
<td>4059</td>
<td>3061 (75.4)</td>
<td>998 (24.6)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>All ages</td>
<td>5228</td>
<td>3892 (74.4)</td>
<td>1336 (25.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≤65</td>
<td>2634</td>
<td>1901 (72.2)</td>
<td>733 (27.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>&gt;65</td>
<td>2594</td>
<td>1991 (76.8)</td>
<td>603 (23.2)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>All ages</td>
<td>2781</td>
<td>2046 (73.6)</td>
<td>735 (26.4)</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>≤65</td>
<td>1316</td>
<td>976 (74.2)</td>
<td>340 (25.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;65</td>
<td>1465</td>
<td>1070 (73.0)</td>
<td>395 (27.0)</td>
<td></td>
</tr>
</tbody>
</table>

* Participants were individuals who gave consent, returned an analysable FIT at year 1 and did not subsequently withdraw from the study.
† Participants who gave consent, returned an analysable FIT at year 1 and did not subsequently withdraw from the study.
‡ Percentages calculated using the number of participants who completed FIT as the denominator. In the pilot study, a threshold of 20 µg haemoglobin (Hb)/g faeces was used to denote test positivity. The positivity threshold used in the rest of the study was 40 µg Hb/g faeces.
§ Participants who underwent colonoscopy or CT colonography. Percentages calculated using the number of FIT positive participants as the denominator.
¶ Participants who had a colonoscopy. Percentages calculated using the number of participants who underwent colonoscopy as the denominator.

Over 3 years, 725 (12.2%) participants underwent colonic examination following positive FIT, finding CRC in 17 (2.3%) and AAs in 151 (20.8%) (table 2). FIT therefore identified 58.6% of participants with CRC (17/29) and 33.9% of participants with AAs (151/446) over 3 years at a threshold of 40 µg/g. Diagnostic yields at years 1, 2 and 3 were, respectively, 2.5%, 3.2% and 1.1% for CRC, and 24.4%, 17.1% and 19.0% for AAs (table 2).

FIT positivity was higher at lower thresholds. In programme analysis, 3-year positivity was 13.2% at 40 µg/g, and 28.8% at 10 µg/g (table 3). Sensitivities for CRC increased while specificities decreased at lower thresholds and with multiple tests. Sensitivities and specificities of the first FIT were, respectively, 27.6% and 94.1% at 40 µg/g, and 51.7% and 86.2% at 10 µg/g (table 4). Over 3 years, programme sensitivities were 58.6% at 40 µg/g and 72.4% at 10 µg/g, and programme specificities were 86.4% at 40 µg/g and 71.1% at 10 µg/g. Similar patterns were observed in cumulative test analyses (table 4). Diagnostic accuracy estimates at 30 µg/g and 20 µg/g are shown in online supplementary table 4.

At lower thresholds, greater proportions of FIT-detected CRCs were found at year 1 (71.4% at 10 µg/g versus 47.1% at 40 µg/g) (data not shown). PPVs for CRC were smaller at lower thresholds and with multiple tests. In programme analysis after 3 years, PPVs were 2.4% at 40 µg/g and 1.4% at 10 µg/g. NPVs fell between 99.6% and 99.9% for all thresholds and numbers of tests (table 4).

Subgroup analyses by sex and age were underpowered; however, sensitivities were generally higher in older (>65 years) participants, while specificities were generally higher in younger (≤65 years) participants and women (online supplementary tables 5 and 6).

FIT was less sensitive but more specific for AAs than CRC. Sensitivities for AAs increased and specificities decreased at lower thresholds and with multiple tests. Three-year programme sensitivities for AAs were 33.4% at 40 µg/g and 56.6% at 10 µg/g, and programme specificities were 88.3% at 40 µg/g and 73.7% at 10 µg/g (table 5). PPVs were higher and NPVs lower for AAs than for CRC (table 5).

Although lacking in power, subgroup analyses revealed that sensitivities for AAs were slightly higher and specificities lower in men and older (≥65 years) participants (online supplementary tables 7 and 8). Excluding participants with

Table 2  Uptake, positivity rate and diagnostic yield of the faecal immunochemical test (FIT) at years 1, 2 and 3 and over all 3 years

<table>
<thead>
<tr>
<th>Uptake</th>
<th>Positivity rate</th>
<th>Colonic exam performed</th>
<th>Diagnostic yield*</th>
<th>Advanced adenomas**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invited Completed FIT test</td>
<td>Tested positive</td>
<td>Any exam§</td>
<td>Colonoscopy¶</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>Year</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>1</td>
<td>8009</td>
<td>5938 (74.1)</td>
<td>346†</td>
<td>317 (99.1)</td>
</tr>
<tr>
<td>2</td>
<td>5479</td>
<td>5329 (97.3)</td>
<td>236 (4.4)</td>
<td>216 (91.5)</td>
</tr>
<tr>
<td>3</td>
<td>5179</td>
<td>5022 (97.0)</td>
<td>204 (4.1)</td>
<td>189 (92.6)</td>
</tr>
<tr>
<td>Cumulative</td>
<td>8009</td>
<td>5938 (74.1)</td>
<td>786†</td>
<td>713 (98.3)</td>
</tr>
<tr>
<td>Routine year 3 colonic exam</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire study findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Diagnostic yield in participants who underwent colonic examination.
† Participants who gave consent, returned an analysable FIT at year 1 and did not subsequently withdraw from the study.
§ Percentages calculated using the number of participants who completed FIT as the denominator. In the pilot study, a threshold of 20 µg haemoglobin (Hb)/g faeces was used to denote test positivity. The positivity threshold used in the rest of the study was 40 µg Hb/g faeces.
¶ Participants who underwent colonoscopy or CT colonography. Percentages calculated using the number of FIT positive participants as the denominator.
** Advanced adenomas were defined as adenomas ≥10 mm, with villous or tubulovillous histology, or high-grade dysplasia.
†† Three participants tested positive at year 1 during the pilot study based on a threshold of 20 µg Hb/g. They are included as FIT positive in this table even though their faecal haemoglobin levels were lower than the 40 µg Hb/g threshold used in the rest of the study.
‡‡ Participants who were FIT positive with any FIT, regardless of whether they had completed all FITs that they were offered.


non-adenocarcinomas (squamous cell carcinoma: n=1; neuro-endocrine tumours: n=2) saw 3-year programme sensitivities for CRC increase from 72.4% to 80.8% at 10\(\mu\)g/g (data not shown).

Excluding participants with colonic examinations of unknown quality (n=587) or with confirmed IBD (n=19) had negligible effects on diagnostic accuracy estimates, although excluding participants with confirmed or possible IBD (n=68) saw the 3-year programme sensitivity for CRC increase to 75.0% at 10\(\mu\)g/g (data not shown).

The mean incremental cost per participant for 3 yearly colonoscopy surveillance versus 3 years of annual FIT (40\(\mu\)g/g) was £365, and the total cost difference was £2 169 341. Incremental costs per additional AA and CRC detected by colonoscopy surveillance were £7354 and £180 778, respectively. Replacing yearly colonoscopy with FIT surveillance nationally over a screening cycle would have a significant impact on budget savings: the total saving was estimated to be £4.7 million (online supplementary figure 1 and supplementary tables 9–12).

**DISCUSSION**

In this study, we evaluated the potential of annual FIT for post-polypectomy surveillance of intermediate-risk patients instead of 3 yearly colonoscopy. Annual FIT was well accepted with initial consent and FIT return of 74% and subsequent FIT return of 97% and less costly than colonoscopy surveillance. Sensitivities increased at lower FIT positivity thresholds; at a threshold of 10\(\mu\)g/g, 3-year programme sensitivities were 72% for CRC and 57% for AAs.

Increased FIT sensitivity was accompanied by increased positivity and reduced specificity. Compared with one FIT at 40\(\mu\)g/g, 3 years of low-threshold (10\(\mu\)g/g) FIT was associated with five times as many positive and false positive results. Nevertheless, even with the latter regimen, cumulative positivity was only 29%. Substituting yearly colonoscopy in intermediate-risk patients with low-threshold annual FIT could therefore reduce numbers of colonoscopies by 71%. However, this strategy would increase...
Table 4  Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the faecal immunochemical test (FIT) at thresholds of 40 µg/g and 10 µg/g for colorectal cancer in participants who completed one, two, or three tests and underwent colonoscopic examination

<table>
<thead>
<tr>
<th>FIT threshold (µg/g)</th>
<th>Test</th>
<th>Completed test*</th>
<th>Participants with colorectal cancer</th>
<th>Participants without colorectal cancer</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>TP FN</td>
<td>FF TN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>First</td>
<td>5225</td>
<td>8 21</td>
<td>309 487</td>
<td>27.6 (12.2 to 47.2)</td>
<td>94.1 (93.4 to 94.7)</td>
<td>2.5 (1.1 to 4.0)</td>
<td>99.6 (99.3 to 99.7)</td>
</tr>
<tr>
<td></td>
<td>Second†</td>
<td>4664</td>
<td>7 9</td>
<td>218 4630</td>
<td>40.8 (19.8 to 70.1)</td>
<td>95.5 (94.9 to 96.1)</td>
<td>3.1 (1.3 to 6.0)</td>
<td>99.8 (99.6 to 99.9)</td>
</tr>
<tr>
<td></td>
<td>Third‡</td>
<td>4468</td>
<td>2 5</td>
<td>178 4631</td>
<td>26.6 (12.7 to 47.0)</td>
<td>96.0 (95.4 to 96.5)</td>
<td>1.1 (0.1 to 4.0)</td>
<td>99.9 (99.7 to 100.0)</td>
</tr>
<tr>
<td>Over two tests</td>
<td>CTA</td>
<td>5191</td>
<td>15 9</td>
<td>527 4630</td>
<td>62.5 (46.6 to 81.2)</td>
<td>89.9 (88.9 to 90.6)</td>
<td>2.8 (1.6 to 4.5)</td>
<td>99.8 (99.6 to 99.9)</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>5225</td>
<td>15 14</td>
<td>527 4630</td>
<td>57.7 (32.5 to 79.6)</td>
<td>89.9 (89.0 to 90.7)</td>
<td>2.8 (1.6 to 4.0)</td>
<td>99.7 (99.5 to 99.8)</td>
</tr>
<tr>
<td>Over three tests</td>
<td>CTA</td>
<td></td>
<td>4990</td>
<td>17 5</td>
<td>705 4630</td>
<td>77.3 (54.6 to 92.2)</td>
<td>85.8 (84.8 to 86.8)</td>
<td>2.4 (1.4 to 3.7)</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>5225</td>
<td>17 12</td>
<td>705 4491</td>
<td>58.6 (38.9 to 76.5)</td>
<td>86.4 (85.5 to 87.4)</td>
<td>2.4 (1.4 to 3.7)</td>
<td>99.7 (99.5 to 99.9)</td>
</tr>
<tr>
<td>10</td>
<td>First</td>
<td>5225</td>
<td>15 14</td>
<td>718 4478</td>
<td>51.7 (25.2 to 74.6)</td>
<td>86.2 (85.2 to 87.1)</td>
<td>2.0 (1.1 to 3.4)</td>
<td>99.7 (99.5 to 99.8)</td>
</tr>
<tr>
<td></td>
<td>Second†</td>
<td>4458</td>
<td>5 6</td>
<td>437 4010</td>
<td>45.5 (16.7 to 76.6)</td>
<td>90.2 (89.3 to 91.0)</td>
<td>1.1 (0.4 to 2.6)</td>
<td>99.9 (99.7 to 99.9)</td>
</tr>
<tr>
<td></td>
<td>Third‡</td>
<td>3853</td>
<td>1 3</td>
<td>3 46</td>
<td>25.0 (0.6 to 80.6)</td>
<td>90.0 (90.1 to 91.9)</td>
<td>0.3 (0.0 to 1.6)</td>
<td>99.9 (99.8 to 100.0)</td>
</tr>
<tr>
<td>Over two tests</td>
<td>CTA</td>
<td>5191</td>
<td>20 6</td>
<td>1155 4010</td>
<td>76.9 (56.4 to 91.9)</td>
<td>77.6 (76.5 to 78.8)</td>
<td>1.7 (1.0 to 2.6)</td>
<td>99.9 (99.7 to 99.9)</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>5225</td>
<td>20 9</td>
<td>1155 4041</td>
<td>69.0 (40.2 to 84.7)</td>
<td>77.8 (76.6 to 79.9)</td>
<td>1.7 (1.0 to 2.6)</td>
<td>99.8 (99.6 to 99.9)</td>
</tr>
<tr>
<td>Over three tests</td>
<td>CTA</td>
<td>5028</td>
<td>21 3</td>
<td>1500 3504</td>
<td>87.5 (67.6 to 97.3)</td>
<td>70.0 (60.7 to 77.3)</td>
<td>1.4 (0.9 to 2.1)</td>
<td>99.9 (99.8 to 100.0)</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>5225</td>
<td>21 8</td>
<td>1500 3696</td>
<td>72.4 (52.8 to 87.1)</td>
<td>71.1 (60.9 to 72.4)</td>
<td>1.4 (0.9 to 2.1)</td>
<td>99.8 (99.6 to 99.9)</td>
</tr>
</tbody>
</table>

*Participants who tested positive at a given threshold at year 1 or 2 were excluded from subsequent analyses.
†Includes participants who completed their second FIT either at year 2 or 3.
‡Includes participants who completed their third FIT.
§Includes participants who completed at least two FITs or who tested positive at year 1. Participants were classed as positive if they tested positive with either of their first two FITs.
¶Includes participants who completed at least one FIT. Participants were classed as positive if they tested positive with any FIT.
**Includes participants who completed all three FITs or who tested positive with any FIT. Participants were classed as positive if they tested positive with any FIT.
††Includes participants who completed at least one FIT. Participants were classed as positive if they tested positive with any FIT.

CTA, cumulative test analysis; FN, false negative; FP, false positive; PA, programme analysis; TN, true negative; TP, true positive.
Table 5  Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the faecal immunochemical test (FIT) at thresholds of 40 µg/g and 10 µg/g for advanced adenomas in participants who completed one, two or three tests and underwent colonic examination and did not have colorectal cancer diagnosed

<table>
<thead>
<tr>
<th>FIT threshold (µg/g)</th>
<th>Test</th>
<th>Completed test*</th>
<th>Participants with advanced adenomas</th>
<th>Participants without advanced adenomas</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>TP</td>
<td>FN</td>
<td>PP</td>
<td>TN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>First</td>
<td>5196</td>
<td>75</td>
<td>365</td>
<td>234</td>
<td>4522</td>
<td>17.0 (3.6 to 20.9)</td>
<td>95.1 (94.4 to 95.7)</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>4498</td>
<td>37</td>
<td>325</td>
<td>181</td>
<td>4305</td>
<td>10.2 (3.3 to 13.8)</td>
<td>96.0 (95.3 to 96.7)</td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>4441</td>
<td>35</td>
<td>273</td>
<td>143</td>
<td>3990</td>
<td>11.4 (8.0 to 15.4)</td>
<td>96.5 (95.9 to 97.1)</td>
</tr>
<tr>
<td></td>
<td>Over two tests CTA</td>
<td>5196</td>
<td>112</td>
<td>325</td>
<td>415</td>
<td>4305</td>
<td>25.6 (21.4 to 30.0)</td>
<td>91.2 (90.4 to 92.0)</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>5196</td>
<td>112</td>
<td>328</td>
<td>415</td>
<td>4341</td>
<td>25.6 (21.4 to 30.8)</td>
<td>91.3 (90.4 to 92.2)</td>
</tr>
<tr>
<td></td>
<td>Over three tests CTA**</td>
<td>4968</td>
<td>147</td>
<td>273</td>
<td>598</td>
<td>3990</td>
<td>35.0 (30.4 to 39.6)</td>
<td>97.7 (96.7 to 98.7)</td>
</tr>
<tr>
<td></td>
<td>PA††</td>
<td>5196</td>
<td>147</td>
<td>293</td>
<td>558</td>
<td>4198</td>
<td>33.4 (29.0 to 38.0)</td>
<td>98.3 (97.3 to 99.2)</td>
</tr>
<tr>
<td>10</td>
<td>First</td>
<td>5196</td>
<td>145</td>
<td>295</td>
<td>573</td>
<td>4183</td>
<td>33.0 (28.6 to 37.6)</td>
<td>98.0 (97.0 to 98.9)</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>4447</td>
<td>61</td>
<td>231</td>
<td>376</td>
<td>3779</td>
<td>20.9 (16.4 to 26.0)</td>
<td>91.0 (90.0 to 91.8)</td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>3849</td>
<td>43</td>
<td>175</td>
<td>302</td>
<td>3329</td>
<td>19.7 (14.7 to 25.6)</td>
<td>91.7 (90.7 to 92.6)</td>
</tr>
<tr>
<td></td>
<td>Over two tests CTA§</td>
<td>5165</td>
<td>206</td>
<td>231</td>
<td>949</td>
<td>3779</td>
<td>47.1 (42.4 to 51.9)</td>
<td>79.3 (78.8 to 81.1)</td>
</tr>
<tr>
<td></td>
<td>PA§</td>
<td>5196</td>
<td>206</td>
<td>234</td>
<td>949</td>
<td>3807</td>
<td>48.8 (42.1 to 55.6)</td>
<td>80.0 (79.9 to 81.2)</td>
</tr>
<tr>
<td></td>
<td>Over three tests CTA**</td>
<td>5004</td>
<td>249</td>
<td>175</td>
<td>1251</td>
<td>3329</td>
<td>58.7 (53.9 to 63.5)</td>
<td>72.7 (71.4 to 74.0)</td>
</tr>
<tr>
<td></td>
<td>PA††</td>
<td>5196</td>
<td>249</td>
<td>191</td>
<td>1251</td>
<td>3505</td>
<td>56.6 (51.8 to 61.3)</td>
<td>73.7 (72.4 to 74.9)</td>
</tr>
</tbody>
</table>

*Participants who tested positive at a given threshold at year 1 or 2 were excluded from subsequent analyses.†Includes participants who completed their second FIT either at year 2 or 3.‡Includes participants who completed their third FIT.¶Includes participants who completed at least two FITs or who tested positive at year 1. Participants were deemed as positive if they tested positive with either of their first two FITs.¶¶Includes participants who completed at least one FIT. Participants were deemed as positive if they tested positive with either of their first two FITs.**Includes participants who completed all three FITs or who tested positive with any FIT. Participants were deemed as positive if they tested positive with any FIT.††Includes participants who completed at least one FIT. Participants were deemed as positive if they tested positive with any FIT.

CTA, cumulative test analysis; FN, false negative; FP, false positive; PA, programme analysis; TN, true negative; TP, true positive
the risk of missing ACN; in our data, 28% of CRCS would have been missed. Although this percentage has limited precision owing to few CRC outcomes (n=29), it may cause concern as diagnostic delay can result in disease progression. Additional concern may arise from the fact that 30% of FIT-detected CRCS were not found until years 2 and 3, even with the low-threshold. If in practice FIT return following a negative result is depressed by false reassurance, the proportion of cancers missed by FIT could be even higher.

This discussion requires consideration of what is an 'acceptable' proportion of missed cancers, accounting for the possibility that missed cancers could be detected at treatable stages with continued FIT surveillance. This comes with the caveat that FIT return may fail following repeat negative testing. Delayed detection of AAs is less serious given the slower rate of transformation from adenoma to cancer, and a 3-year programme sensitivity of 57% may be considered acceptable.

It is interesting to consider these results in light of the findings of our 'Intermediate Adenoma' study. Analysing long-term CRC incidence in intermediate-risk patients post-polypectomy, we identified two risk subgroups. In the higher risk subgroup, comprising patients with suboptimal quality colonoscopies, adenomas ≥20 mm, high-grade adenomas or proximal polyps at baseline, colonoscopy surveillance was associated with a halving in incidence. In lower risk patients without these features, the value of surveillance was unclear. Therefore, while higher risk patients significantly benefit from 3 yearly colonoscopy, annual FIT may suffice for lower risk patients. FIT may also be useful in the surveillance of high-risk patients; for example, as a substitute for the 1-year colonoscopy. Future studies should investigate FIT performance in different risk subgroups.

Further research is needed to understand the influence of sex and age on FIT performance. Consistent with the literature, higher positivity was observed in men and older participants, reflecting the increased prevalence of ACN. Sensitivities were generally higher and specificities lower in men and older participants, although subgroup analyses lacked power. Other studies have reported variability in diagnostic accuracy of FIT by sex and age, leading to recommendations for individualisation of positivity thresholds. PPVs are particularly important as they indicate the probability of a colonoscopy being unnecessary following positive FIT. Here we did not detect a consistent pattern due to insufficient power, but a study of the Basque Country CRC Screening Programme found PPVs of FIT for ACN to be significantly higher in men than women at thresholds between 20 µg/g and 60 µg/g. Well-powered studies should investigate whether this is also the case in individuals undergoing surveillance.

FIT sensitivity varies according to lesion type and location. Here we reported higher sensitivities when the three non-adenocarcinomas were excluded; these were not detected by FIT, even at 10 µg/g. Poor sensitivity has also been reported for proximal and serrated lesions. Sensitivity may be improved by combining FIT with other faecal biomarker tests. A multtarget molecular stool test has shown greater sensitivity than FIT for ACN and for proximal and large serrated lesions, although has lower specificity. An evaluation of FIT and molecular stool testing as alternatives to colonoscopy surveillance is underway, although further research should examine how these perform in combination.

Intermediate-risk patients are currently recommended to undergo surveillance until they have two consecutive negative colonoscopies. Stopping rules for FIT-based surveillance would have to consider the susceptibility of some patients to repeatedly test false positive; something we observed in our analysis of BCSP participants who underwent gFOBT screening.

To our knowledge, this study is the largest to have evaluated FIT for surveillance and the only to have evaluated FIT specifically in patients undergoing postpolypectomy surveillance. Previous studies included patients attending surveillance for personal or family histories of CRC. In four of these, sensitivities were calculated for a single FIT performed prior to surveillance colonoscopy. One study of 808 patients reported sensitivities of 70% for CRC and 44% for AAs for Hemeselect, a qualitative FIT that was deemed positive when the concentration of faecal haemoglobin exceeded 200–300 µg/g. The OC-Sensor (10 µg/g) achieved sensitivities for CRC and AAs, respectively, of 80% and 28% in a study of 1041 patients, and 100% and 44% in a study of 1000. A fourth study stratified sensitivity estimates by indication for surveillance, showing that in 348 patients undergoing surveillance for colorectal neoplasia, sensitivities of OC-Sensor (15 µg/g) were 100% for CRC and 68% for AAs.

These studies reported higher sensitivities than we estimated for one FIT at 10 µg/g (52% for CRC; 33% for AAs), although small numbers of CRCS meant estimates were imprecise. Differences in patient mix, number of tests, scheduling of FIT relative to colonoscopy and the high quality of colonoscopy in England may have contributed to the discrepancy. There is one other study that evaluated FIT for surveillance with repeated testing; among 1071 patients who completed at least one qualitative FIT between surveillance colonoscopies, sensitivities were 86% for CRC and 63% for AAs. In our study, sensitivities in patients completing at least one FIT was lower at 72% and 57% for CRC and AAs, respectively, over 3 years at 10 µg/g.

Strengths of our study include the relatively large sample of clearly defined patients undergoing postpolypectomy surveillance and evaluation of FIT at multiple intervals with multiple tests. We tested the robustness of our results in sensitivity analyses, observing little change when participants with unknown quality examinations and IB were excluded. The high FIT return rates are a further strength, although it is likely that our inclusion criteria inflated rates at years 2 and 3 because only individuals consenting and returning an analysable year 1 FIT were offered further FITs, limiting the generalisability of our findings. This may have also biased our diagnostic accuracy estimates because individuals who provided consent but not an analysable FIT may have participated at subsequent years.

Limitations of this study include that it examined FIT performance only in relation to the first surveillance colonoscopy. We do not know the effect of continuing with FIT beyond 3 years. Furthermore, we assumed that colonic examinations detected all ACN; given that 98% of participants who underwent examination had colonoscopy, the ‘gold-standard’ investigation, this assumption is reasonable. That said, colonoscopy is not perfectly sensitive and occasionally misses adenomas and CRCS. A small percentage of participants received CT colonography only, which shows comparable sensitivity to colonoscopy for CRC and large polyps.

Another limitation is that endoscopists may have been aware when examinations were for positive FIT results, which may have influenced the thoroughness of examination, potentially leading to inflated sensitivity estimates of FIT-positive colonoscopies. Additionally, we assumed that the population of ACN was static. Adenoma progression rates are thought to be low; however, if neoplastic progression did occur, our reported sensitivities may be biased. The bias may be in either direction depending on the FIT threshold being considered, leading to both underestimation...
and overestimation of sensitivity. Biases may have resulted from missing information, as we did not know why some individuals had multiple baseline examinations and were therefore excluded or why some did not attend any colonic examination.

Interpretation of our results should take into account the specific nature of the study population, namely that it was drawn from BCSP participants assigned as intermediate-risk at colonoscopy performed following positive gFOBT. Generalisability is not limited by the use of a single FIT brand, as different brands can achieve almost equal sensitivity when thresholds are set to yield defined positivity rates.33

CONCLUSION
Our results suggest that FIT could perform a role in postpolypectomy surveillance of intermediate-risk patients. If low-threshold annual FIT was implemented instead of 3 yearly colonoscopy, numbers of colonoscopies could be reduced by more than 70% with significant cost savings. However, this would come at the cost of missed ACN; depending on the threshold, annual FIT could miss 30%–40% of CRCs and 40%–70% of AAs. Further research is warranted before decisions are made about whether it is reasonable to adopt FIT for surveillance. Future studies should investigate FIT performance in longer term surveillance and within subgroups of intermediate-risk patients. Further economic analyses should examine the cost implications of CRCs and AAs missed by FIT and the impact on quality-adjusted life years.

Author affiliations
1Cancer Screening and Prevention Research Group (CSPRG), Department of Surgery and Cancer, Imperial College London, London, UK
2Bowel Cancer Screening Programme Southern Hub, Guildford, UK
3Department of Applied Health Research, University College London, London, UK
4Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University, London, UK
5Research Department of Behavioural Science and Health, University College London, London, UK
6Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

Acknowledgements
The investigators are grateful to the people listed below for their involvement in this study. In addition to those named, we would like to thank the personnel within the 64 Bowel Cancer Screening Centres we liaised with to follow the investigations and subsequent clinician management for the participants who tested FIT positive and for all their invaluable support with retrieving the relevant colonic examination and histology reports for all participants in the study. A special thank you to all of the patients who contributed data to the study.

Collaborators
Trial steering committee: Dr Andrew Veitch (Chairperson), Ms Helen Watson and Ms Lynn Faulds-Wood (patient representatives), Professor Allan Hackshaw, Professor Sue Moss, Professor Marco Novelli, Professor Matt Rutter and Professor Christopher Todd. CSPRG study staff: Mrs Elizabeth Coles, Dr Paula Kirby, Ms Amy Brenner, Ms Jessica Martin, Mrs Unil Shah, Ms Irene Simmonds, Dr Laura Turner, Mr Jeremy Brown and Dr Fiona Lucas. Collaborators: FIT kit provision: Mr Iain McClarney, MAST Group Ltd. FIT kits were generously provided by Eiken Chemical Co Ltd (through MAST Group Ltd). BCSP Southern Hub: Mrs Sheena Pearson, Mrs Sally Benton, Mrs Helen Bruce, Ms Carole Burtonwood, Dr Magdalene Carroll, Ms Cerin John, Ms Kate Randall, Mrs Katy Allen, Dr Helen Seaman and Mr Neil Stubbs. Clinical Informatics research unit at the University of Southampton: Professor James Batchelor and Mr Wz Morris. Health Behaviour Research Centre: Ms Bernardette Bonello, Ms Harriet Bowyer, Dr Gemma Vart and Professor Jane Wardle (deceased). Public Health England (PHE): Mr Tariq Malik, Ms Claire Nickerson and Dr Suzanne Wright. Clinical advice: Professor Brian Saunders, Professor Robert Steele and Dr Roland Valori.

Contributors
WSA, IK-H, CW, SM, SWD, SPH, CP, JS and JS were responsible for obtaining funding and research approvals, and study design. EM, BP, KP and RH were responsible for trial management. WSA, AIC, AP, KW and SWD were responsible for data analysis and interpretation. SM and NS undertook the economic analysis. WSA, AIC, FCR and KW drafted the manuscript. All authors critically appraised the final manuscript. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. AIC and KW had final responsibility for the decision to submit for publication.

Funding
This is a summary of independent research funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme and the Bobby Moore Fund for Cancer Research UK.

Disclaimer
The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report, or in the decision to submit the manuscript for publication. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or Cancer Research UK. Infrastructure support for this work was provided by the NIHR Imperial Biomedical Research Centre.

Competing interests
WSA received non-financial support from Eiken Co. Ltd (MAST in UK distributor). SM is a member of the NIHR Health Services and Delivery Research (HS&DR) Funding Board. All other authors declare no competing interests.

Patient consent for publication
Obtained.

Ethics approval
Ethics approval was granted by London (City & East) Research Ethics Committee. All protocol amendments were approved (reference 11/LO/0326). The Ethics and Confidentiality Committee (ECC) of the Health Research Authority Confidentiality Advisory Group gave approval to access patient identifiers and contact details in June 2011, allowing identification and invitation of eligible participants (reference ECC 3 04p/2011).

Provenance and peer review
Not commissioned; externally peer reviewed.

Data sharing statement
Because of the sensitive nature of the data collected for this study, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the corresponding author.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: http://creativecommons.org/licenses/by/4.0/.

REFERENCES


