Use of ColonFlag score for prioritisation of endoscopy in colorectal cancer

Ruth M Ayling, A Wong, Finbarr Cotter

ABSTRACT

Objective Colorectal cancer (CRC) is the fourth most common cancer in UK. Symptomatic patients are referred via an urgent pathway and although most are investigated with colonoscopy <4% are diagnosed with cancer. There is therefore a need for a suitable triage tool to prioritise investigations. This study retrospectively examined performance of various triage tools in patients awaiting investigation on the urgent lower gastrointestinal cancer pathway.

Design All patients over 40 years of age on the urgent pathway awaiting investigation for suspected CRC on 1 May were included. After 6 months, outcomes were evaluated and the performance of the faecal immunochemical test (FIT), faecal haemoglobin concentration, age and sex test (FAST) and the artificial intelligence algorithm ColonFlag were examined.

Results 532 completed investigations and received a diagnosis; 15 had CRC. 388 had a valid FIT result, of whom 11 had CRC; FAST Score ≥4.5 had sensitivity of 72.7%, specificity of 80.6% and would have failed to detect three tumours. Faecal haemoglobin (f-Hb) at cut-off of 10 µg/g and ColonFlag had equal sensitivity of 81.82%, ColonFlag had greater specificity 73.47%, compared with 64.99%. Both tests would have failed to detect two tumours but not in the same patients; when used in combination, sensitivity and specificity were 100% and 49.4%. When ColonFlag was applied to the cohort of 532, an additional four tumours would have been detected in patients without a valid FIT.

Conclusion This study showed ColonFlag to have equal sensitivity and greater specificity than f-Hb at a cut-off of 10 µg/g as a triage tool for CRC.

INTRODUCTION

Colorectal cancer (CRC) is the fourth most common cancer in the UK, with approximately 43,000 patients diagnosed each year and the second largest cause of cancer death.1 To assist early detection, dedicated urgent pathways were established in England in 2000 to facilitate symptomatic patients receiving specialist assessment within 2 weeks if cancer was suspected. This pathway was revised with the provision of the NG12 guidance by the National Institute for Health and Care Excellence (NICE); this suggests a positive predictive value (PPV) of 3% as a threshold for urgent referral.2 In 2017, NICE issued diagnostics guidance (DG30)3 on measurement of faecal haemoglobin (f-Hb) using faecal immunochemical testing (FIT) and updated NG12 to incorporate a recommendation for its measurement in patients with unexplained abdominal symptoms, without rectal bleeding, who do not meet criteria for urgent referral for suspected CRC. A f-Hb ≥10 µg haemoglobin/g faeces (µg/g) was recommended as the cut-off to prompt urgent referral. The use of FIT in line with DG30 was incorporated into our local urgent referral pathway from April 2019.4

In addition to the investigation of symptomatic patients, bowel cancer screening exists in many countries and a programme has been in place in England since 2006. Screening has been shown to lower CRC mortality and reduce the incidence of CRC, due mainly to detection and removal of adenomatous polyps.5 However, engagement in screening...
is not universal and, of relevance to our practice, has been shown to be lower in inner London than elsewhere in the country.6

While FIT is well established on the pathways in use for both symptomatic and asymptomatic patients, there is scope for additional triage tools and in this context, a number of CRC prediction models have been designed. Some of these incorporate clinical variables which could potentially be captured in a single consultation, but further work is needed to ensure robust validation.7

The faecal haemoglobin concentration, age and sex test (FAST) is a three-variable model for CRC based on f-Hb concentration, age and sex. Two thresholds of 2.12 and 4.5 have been identified with 99% and 90% sensitivity, respectively, for CRC.8

Association between components of the full blood count (FBC) and the detection of CRC has been reported in the literature for many years.9 NICE guidance recommends urgent referral to a gastroenterologist of unexplained iron deficiency anaemia in men with Hb <120 g/L and postmenopausal women with Hb <100 g/L.10

Triage tools exist which incorporate FBC changes into risk scores, one such being ColonFlag (Medial EarlySign, Kfar Malal, Israel) which identifies patients of 40 years or older at risk of CRC using artificial intelligence (AI) learning using their age, sex and FBC parameters based on an ensemble of decision trees. The rationale of the tool is that patients develop subtle changes in multiple indices of the FBC before becoming symptomatic from CRC. Although the parameters themselves may remain within the reference range, if more than one FBC is available, such changes can be detected by the AI algorithm and flagged as a potential indicator of CRC. The model was developed using data from healthy Israelis and CRC patients and trained using Israeli databases; validation was performed using additional cohorts within the USA and UK.11 Any number of FBCs can be used to derive the score, but a minimum of three over a period of 3–5 years is enough to reach almost optimal performance. We have previously reported its potential for use in triage of patients with anaemia.12

In the COVID-19 pandemic, England was placed into lockdown for the first time in March 2020 and non-emergency colonoscopy was temporarily suspended. During recovery from the first peak of infection, resumption of service was associated with reduced capacity because of the necessity for new health and safety protocols and with an accumulated backlog of patients requiring investigation. At this point, FIT was mandated as a criterion for referral. Some patients who had already been referred on our urgent pathway had not had the test performed in line with previous guidance and we co-ordinated a central process to offer them the test.

This study retrospectively investigated test performance of ColonFlag for prioritisation in patients who had already been referred and were awaiting colonoscopy on 1 May in our hospital and compared it with the performance of FIT and FAST score.

METHODS

Patients

All adult patients over 40 years of age, referred to Barts Health NHS Trust on an urgent pathway with suspected CRC and awaiting investigation on 1 May 2020 were included in the study.

Outcome definition

After 6 months, clinical outcomes were collected and diagnoses of CRC, high risk adenomas (HRA) and inflammatory bowel disease (IBD) were confirmed from clinical notes, radiological reports and endoscopy and histological findings. HRA were defined using British Society of Gastroenterology guidance 2020.13

Sample analysis

Faecal samples were taken into a specimen collection device and returned to the Clinical Biochemistry department at Barts Health NHS Trust. They were stored at 4°C before analysis, which took place within 1 week of receipt and 2 weeks of sampling. The laboratory is accredited by the UK Accreditation Service to ISO 15189 standards. Analysis was performed using a single OG-SensoriO (Eiken Chemical Co., Tokyo, Japan). Inter-run imprecision was assessed with quality control materials (Eiken) in each run. Coefficients of variation were 2.8% at 14 µg/g and 3.0% at 91 µg/g. External quality assurance was achieved via satisfactory performance in the relevant National External Quality Assurance Scheme. The lower limit of quantification was 4 µg/g. The upper analytical limit was 200 µg/g and samples with a concentration above this were not diluted and reassayed but reported as >200 µg/g. If a patient returned more than one FIT sample, only the first test result was selected for inclusion in the analysis. FBCs were measured on a Sysmex XE 2100 (Sysmex, Milton Keynes, UK).

Calculation of risk scores

FAST score was calculated using the equation f-Hb score + (0.031x age in years)+0.479 if male. The f-Hb score is 0 if the concentration is 0 µg/g, 0.6841 if 1–19 µg/g, 2.824 if 20–199 µg/g and ≥200 µg/g 4.184. 4 µg/g was taken to be the lower limit of quantification.

ColonFlag was calculated by Medial EarlySign. The parameters from all available FBCs from 2015 onwards were used, as these were easily accessible from current laboratory records. FBC parameters, age and sex were used to assign an individual risk score for each patient and to place them in one of four bands (high to low: bands 3 to 0) indicating their likelihood of CRC. The highest scoring band equated to a theoretical PPV of about 10%.

Statistical considerations

Data were summarised and tabulated; population characteristics were summarised by appropriate descriptive statistics by data type.

Continuance measures (eg, sensitivity, specificity, PPV and NPV) were described by average and the exact 95% CI for each measure.
Comparison for categorical parameters was tested by \( \chi^2 \) or Fisher’s exact test (in case of low frequency). A calculated \( p \)-value of 5% or less was considered as statistically significant.

**Ethical considerations**
Data were gathered during routine patient care, therefore, ethical approval was not required. In order to mitigate risks associated with privacy, the data were provided to Medial EarlySign in de-identified format.

**RESULTS**
Data were obtained from 617 patients, 314 (50.81%) male. Their median age was 63 years (range 40–98). The study flow chart is shown in figure 1.

**Further investigation**
Investigations were performed on clinical grounds and a final diagnosis was obtained in 532 patients: 316 (59.4%), underwent colonoscopy, 153 (28.8%) abdominopelvic CT (with flexible sigmoidoscopy in addition in 15 and proctoscopy in six), 54 (10.0%) CT colonography (with flexible sigmoidoscopy in addition in five) and six (1.1%) flexible sigmoidoscopy alone. The patients who underwent colonoscopy were younger than those who underwent other definitive investigations (median 60.94 years, vs 67.21 years, \( p<0.001 \)). The others were not specifically investigated on the urgent pathway as they were under the care of other teams—one because he underwent emergency surgery for intussusception before investigations could be performed, one with known metastatic carcinoma of the prostate and one woman with iron deficiency anaemia and menorrhagia. 74 patients declined investigations, one was overseas and unable to attend and seven were unable to be contacted, despite multiple attempts.

**Final diagnoses**
On the basis of these investigations, 17 of the 532 patients (3.2%) were found to have CRC, 28 (5.2%) patients had HRA and 10 (1.9%) had IBD. Low risk adenomas were detected in 85 patients. Malignancy was diagnosed in an additional nine patients (lung, breast, cervix (2), prostate, kidney (2), unknown primary and caecal neuroendocrine tumour). Of the 388 patients in whom a FIT test was performed and a final diagnosis available, 11 (2.9%) were found to have CRC, 24 (6.3%) HRA and 8 (2.1%) IBD.

**Haemoglobin concentration**
The median blood haemoglobin (Hb) concentration in men was 134 g/L (range 68–182) and in women 122 g/L (range 73–167). 92 men had a Hb <120 g/L, and 30 women older than 50 years had a Hb <100 g/L. In the group with a valid FIT, the 11 patients with CRC four men were anaemic (Hb <130 g/L) and three women (haemoglobin <120 g/L).

**Faecal immunochemical test results**
Of all the patients on the waiting list on 1 May, FIT had only been requested from 440 (71.3%) and, of these samples, only 427 met preanalytical criteria for analysis. There was no significant difference in age and sex.
between those who had and had not provided a sample (p: 0.82 and 0.10, respectively).

Of the 532 patients in whom investigation was performed and a final diagnosis available, 388 had performed a FIT and the f-Hb concentrations were <4 µg/g in 134 (34.5%), <10 µg/g in 247 (63.7%), ≥10–99 µg/g in 94 (24.2%) and ≥100 µg/g in 47 (12.1%).

FAST score
Of the 388 patients with a final diagnosis who provided a FIT sample, 312 had a FAST score above 2.12 and 81 had a score above 4.5.

ColonFlag
Using ColonFlag, patients were divided into one of four bands indicating their likelihood of CRC. Of those in whom a final diagnosis was made 165 patients (31.3%) were in band 3 (highest risk), 111 (21.0%) in band 2, 128 (24.2%) in band 1 and 128 (24.2%) in band 0. The distribution of banding in the subset of 388 patients who provided a FIT sample was as follows band 3 109 (28.1%), band 2 85 (21.9%), band 1 94 (24.2%) and band 0 100 (25.8%) patients. The median number of FBC available was eight (range 1–85). In the group of 388 patients, 34 had less than three FBC and of these, 17 had only one. Of the 11 patients in this group with CRC, the scoring was on 5–57 measurements.

Table 1 shows the performance of FIT, ColonFlag and FAST score in diagnosis of CRC, HRA and IBD in the 388 patients in whom a viable FIT was available. FAST score at a cut-off of 4.15 was the least sensitive test for detection of CRC but had the highest specificity and would have detected eight tumours with only 81 patients (20.9%) requiring colonoscopy. Using a cut-off of 2.12 would have detected all tumours but reduced specificity to 20.16% requiring colonoscopy of 2.12 would have detected eight tumours with only 81 patients (20.9%) requiring colonoscopy.

DISCUSSION
In this study, we retrospectively investigated the potential use of ColonFlag to prioritise patients awaiting colonoscopy during the COVID-19 pandemic. In our cohort of 532 patients, triage according to the highest risk band 3 would have led to 165 (31%) colonoscopies being prioritised with detection of 15 of 17 tumours. ColonFlag differs from other risk-prediction models in that it has the ability to test people automatically using routinely available laboratory data and we would have enabled us to detect four tumours in patients who did not have a f-Hb result available.

Measurement of f-Hb using FIT was recommended by NICE to assist with triage of patients deemed to be at low risk of CRC, using a cut-off of 10 µg/g (NG12). However, the test has been widely investigated and is now used in patients at both low14 and high14 risk of CRC. Previous studies investigating the test performance of f-Hb in CRC have found sensitivities between 83.3% and 90.9% at a cut-off of 10 µg/g15–17 comparable to our findings. Specificities in these studies were slightly higher at 79.1%–83.5%, but this may reflect a difference in population as our patients were referred as the start of the pandemic which may not represent a typical sample. In a study of 755 patients, f-Hb at a cut-off of ≥10 µg/g was shown to be an objective predictor of underlying pathology when used at point of referral with sensitivity of 68.6%, specificity of 89.6%, PPV of 39.8% and NPV of 94.4%.16 Our results add further to weight to these findings and show potential advantage of FIT in combination with ColonFlag. Both tests are reliant on blood loss from colonic lesions to provide parameters for indication of risk. For FIT, this is direct measurement of haemoglobin in faeces but with ColonFlag, it is the result of subtle changes in FBC parameters.

FAST score has been proposed as a tool to assist with prioritisation of referral of patients from primary care and for segregation into a high risk group in which 90% of the CRC would be expected, an intermediate risk group where an additional 9% of CRC might be found and a third group hypothetically containing the remaining 1% of CRC.8 Our results show sensitivities of 71.07%–95% CI: 66.94 to 74.94), PPV 9.15% (95% CI: 7.47 to 11.15) and NPV 99.45% (95% CI: 98.03 to 99.85).
We accept that our study is small, the cohort containing only 11 patients with CRC and that further work is required, but hypothesise that ColonFlag is potentially of value as a triage tool. It differs from other risk-prediction models in that it has the ability to test people automatically using routinely available laboratory data. Reluctance of our patients to complete FIT and to undergo invasive investigations and a reduction of colonoscopy provision in the current climate resulting from COVID-19 are drivers for consideration of an AI triage tool based on the parameters of the FBC. ColonFlag has the potential to be embedded into hospital computer systems in order to generate a prediction score associated with existing FBC analyses and patients’ scores could be automatically updated with each FBC and to assist case finding in various contexts. Approaching the test from the angle of high specificity, patients with a raised or increasing score could be highlighted in order to be considered for investigation. Alternatively, at lower cut offs, where sensitivity is greater, the algorithm might be of benefit to target patients who have not engaged with bowel screening services. Uptake of faecal testing is not universal, in screening programmes or patients with symptoms and 20% of CRC still present through Emergency pathways.19

<table>
<thead>
<tr>
<th>CRC</th>
<th>FAST score ≥4.5</th>
<th>f-Hb ≥10 µg/g</th>
<th>Colon flag band 3</th>
<th>ColonFlag band 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No of cases</strong></td>
<td>11</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>True positives</strong></td>
<td>303</td>
<td>245</td>
<td>277</td>
<td>187</td>
</tr>
<tr>
<td><strong>False positives</strong></td>
<td>73</td>
<td>132</td>
<td>100</td>
<td>190</td>
</tr>
<tr>
<td><strong>False negatives</strong></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>72.7 (39.0–94.0)</td>
<td>81.8 (48.2–97.7)</td>
<td>81.8 (48.2–97.7)</td>
<td>100 (71.5–100)</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>80.6 (76.2–84.5)</td>
<td>64.9 (59.9–69.8)</td>
<td>73.5 (68.7–77.9)</td>
<td>49.6 (44.4–54.8)</td>
</tr>
<tr>
<td><strong>PPV</strong></td>
<td>9.9 (6.7–14.3)</td>
<td>6.4 (4.8–8.5)</td>
<td>8.3 (6.1–11.1)</td>
<td>5.5 (4.9–6.0)</td>
</tr>
<tr>
<td><strong>NPV</strong></td>
<td>99.0 (97.5–99.6)</td>
<td>99.2 (97.2–99.8)</td>
<td>99.3 (97.6–99.8)</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRC+HRA</th>
<th>FAST score ≥4.5</th>
<th>f-Hb ≥10 µg/g</th>
<th>Colon flag band 3</th>
<th>ColonFlag band 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No of cases</strong></td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td><strong>True positives</strong></td>
<td>21</td>
<td>25</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td><strong>True positives</strong></td>
<td>293</td>
<td>237</td>
<td>259</td>
<td>182</td>
</tr>
<tr>
<td><strong>False positives</strong></td>
<td>60</td>
<td>116</td>
<td>94</td>
<td>171</td>
</tr>
<tr>
<td><strong>False negatives</strong></td>
<td>14</td>
<td>10</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>60.0 (42.1–76.1)</td>
<td>71.4 (53.7–85.4)</td>
<td>42.9 (26.3–60.7)</td>
<td>85.7 (69.7–95.2)</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>83.0 (78.7–86.8)</td>
<td>67.3 (62.2–72.2)</td>
<td>73.4 (68.4–77.9)</td>
<td>51.6 (46.2–56.9)</td>
</tr>
<tr>
<td><strong>PPV</strong></td>
<td>25.9 (19.70–33.3)</td>
<td>17.9 (14.4–21.0)</td>
<td>13.7 (9.5–19.5)</td>
<td>14.9 (12.9–17.3)</td>
</tr>
<tr>
<td><strong>NPV</strong></td>
<td>95.4 (93.3–96.9)</td>
<td>95.6 (93.3–97.6)</td>
<td>92.8 (90.6–94.6)</td>
<td>97.3 (94.1–98.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRC+HRA+IBD</th>
<th>FAST score ≥4.5</th>
<th>f-Hb ≥10 µg/g</th>
<th>Colon flag band 3</th>
<th>ColonFlag band 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No of cases</strong></td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td><strong>True positives</strong></td>
<td>29</td>
<td>33</td>
<td>17</td>
<td>38</td>
</tr>
<tr>
<td><strong>True positives</strong></td>
<td>293</td>
<td>237</td>
<td>253</td>
<td>182</td>
</tr>
<tr>
<td><strong>False positives</strong></td>
<td>52</td>
<td>108</td>
<td>92</td>
<td>163</td>
</tr>
<tr>
<td><strong>False negatives</strong></td>
<td>14</td>
<td>10</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>67.4 (51.5–80.9)</td>
<td>76.7 (61.4–88.2)</td>
<td>39.5 (25.0–55.6)</td>
<td>88.4 (74.9–96.1)</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>84.9 (80.7–88.5)</td>
<td>68.7 (63.5–73.6)</td>
<td>73.3 (68.3–78.0)</td>
<td>52.8 (47.3–58.1)</td>
</tr>
<tr>
<td><strong>PPV</strong></td>
<td>35.8 (28.7–43.6)</td>
<td>23.4 (19.6–27.7)</td>
<td>15.6 (10.9–21.8)</td>
<td>18.9 (16.64–21.4)</td>
</tr>
<tr>
<td><strong>NPV</strong></td>
<td>95.4 (93.1–97.0)</td>
<td>96.0 (93.2–97.6)</td>
<td>69.6 (64.7–74.1)</td>
<td>97.3 (94.1–98.8)</td>
</tr>
</tbody>
</table>

CRC, colorectal cancer; f-Hb, faecal haemoglobin; HRA, high risk adenomas; IBD, inflammatory bowel disease; NPV, negative predictive value; PPV, positive predictive value.
This has multiple causes including cultural and health inequalities related to deprivation.\textsuperscript{20} Optimal performance of ColonFlag requires prior engagement by the population to have had blood tests and serial results may be least accessible in those people who decline faecal testing. However, potentially the use of ColonFlag as a test in iron deficiency\textsuperscript{12} or as part of routine hospital admissions in the CRC screening age group could assist both with early detection and reduction of the mortality associated with late presentation.

In our cohort, both f-Hb at a cut-off of 10\textmu{g}/g and ColonFlag at band 3 failed to detect two patients, but these were not the same patients. Both the patients whose CRC was not detected by f-Hb using a cut-off of 10\textmu{g}/g had tumours located in the right side of the colon. It has been reported that f-Hb is less accurate in detecting right-sided than left-sided CRC.\textsuperscript{21} Possible reasons suggested include right-sided lesions having potential to grow more rapidly and bleed less because of their specific phenotypic characteristics, less susceptibility to mechanical triggers exacerbating bleeding and a longer transit with more opportunity for degradation of haemoglobin.

In conclusion, we evaluated the performance of various tools to assist prioritisation for colonoscopy in an urgent lower gastrointestinal cancer pathway. We propose that AI based on FBC parameters using ColonFlag has potential both for prioritisation of symptomatic patients and to target patients who have not engaged with bowel screening services.

**Contributors** RMA and FC designed the study, AW and RMA collected the data, all authors were involved in writing of and approved the final version of the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information. Deidentified participant data are stored securely by the authors and would potentially be available upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is made available under the terms of this licence.

**ORCID iD** Ruth M Ayling http://orcid.org/0000-0003-4662-2502

**REFERENCES**