Technoleg Iechyd Cymru
Health Technology Wales

Evidence Appraisal Report

Faecal immunochemical testing (FIT)-based prediction tools as triage for referral for colorectal cancer investigations

1. Purpose of the evidence appraisal report

The Evidence Appraisal Report is a rapid systematic literature search of published evidence and websites to identify the best clinical and economic evidence on health technologies. Researchers critically evaluate this evidence. The draft Evidence Appraisal Report is reviewed by experts and by Health Technology Wales multidisciplinary advisory groups before publication.

2. Health problem

Colorectal cancer (CRC), is the 4th most common cancer in the UK, accounting for 12% of all new cancer cases (Cancer Research UK 2018). In 2015, there were 2,259 new cases of CRC in Wales, and 904 deaths (Cancer Research UK 2018). Overall, 78% of people with bowel cancer survive for more than one year and 58% survive for five years or more (Bowel Cancer UK 2018). Wales has been ranked 25 out of 29 European countries in five year survival rates; reasons for this include lifestyle choices, strength of NHS systems (e.g. GP and primary care referral) and late diagnosis (Bowel Cancer UK 2018).

People with colorectal cancer are most commonly diagnosed through primary care; the 2015 National Bowel Cancer Audit Report stated that, of all people diagnosed with CRC in 2014, 55% were diagnosed following GP referral (Westwood et al. 2017b). Determining whether someone presenting with symptoms should be referred for CRC detection (usually via colonoscopy) is critical for prompt diagnosis and treatment. NICE guidelines (NG12) recommends a referral within two weeks based on the following criteria:

- ≥ 40 years or over with unexplained weight loss and abdominal pain
- ≥ 50 years or over with unexplained rectal bleeding
- ≥ 60 years and over with iron-deficiency anaemia or changes in their bowel habit
- Tests show occult blood in their faeces (NICE 2015).

For people presenting with lower abdominal symptoms who are at low risk of CRC (i.e. do not satisfy the NG12 two-week referral criteria), NICE diagnostics guidance (DG30) recommends use of faecal immunochemical tests (FIT) to guide referral (NICE 2017).

Increased demand for endoscopy services is set to increase due to a number of factors, including an aging population, increased awareness of symptoms and increased referrals following FIT screening (Bowel Cancer UK 2018). This can result in people who do not have CRC undergoing an unnecessary, invasive diagnostic procedure with associated risks like bowel perforation, bleeding, infection and abdominal pain (Westwood et al. 2017a). On the other hand, lack of referral risks a delayed diagnosis. Accurate prediction tools are required to triage patients who are designated low-risk but present with symptoms.

3. Health technology

FIT is a type of faecal occult blood test that uses antibodies to detect the globin of human haemoglobin (Westwood et al. 2017a). FIT has the reduced potential to give false positives from upper gastrointestinal bleeding, as globin is degraded in the upper gastrointestinal tract, and is not detected in faeces. Guaiac faecal occult blood testing detects haem, which is not degraded and can therefore be detected in faeces, giving a false positive (Westwood et al. 2017a).

FIT can be either qualitative (which gives either a positive or negative visual result) or quantitative (which gives the actual concentration of faecal haemoglobin). NICE DG30 recommends use of quantitative FIT assays to triage symptomatic people who are at low-risk of CRC (NICE 2017). At the time of this report, there are four quantitative FITs available in the UK:

- HM-JACKarc system (Kyowa Medex/Alpha Laboratoried Ltd)
- FOB Gold system (Sentinel/Sysmex, Sentinel Diagnostics)
- OC-Sensor (Eiken Chemical Co./ MAST Diagnostics)
- RIDASCREEN Hb and Hb/Hp test (R-Biopharm)

Different FITs use different sampling methods and characteristics, making comparison between different FIT types difficult (Westwood et al. 2017a).

In Wales, FIT will be introduced in phases from February 2019 for the asymptomatic screening population (J Torkington 2019, expert comment). FIT is not yet used in Wales for the symptomatic population.

4. Evidence search methods

The selection criteria used to identify evidence for this appraisal were developed following comments from the Health Technology Wales (HTW) Assessment Group and topic experts (Appendix 1).

A systematic literature search to study clinical effectiveness was undertaken on 12, 19 and 24 September 2018. The search strategy is available on request. This aimed to identify the following types of evidence:

(i) systematic reviews of diagnostic studies
(ii) primary diagnostic studies
(iii) cost-effectiveness studies
(iv) ongoing clinical trials.

Background studies and other papers identified at the scoping stage were also assessed for relevance.

The searches returned 3,533 articles. Appendix 2 summarises the selection of articles for inclusion in the review.

Patient safety and organisational issues were identified from the papers included in the clinical effectiveness section, and expert advice; no specific searches were undertaken.

5. Clinical effectiveness

Searches identified limited evidence on FIT-based prediction tools for investigating CRC in low-risk people who present with lower abdominal symptoms. Two systematic reviews were identified (Niedermaier et al. 2016, Westwood et al. 2017b); both reviews were broader than the scope of this appraisal and included studies that did not fit the selection criteria.

As the secondary evidence was limited, primary evidence for FIT-based prediction tools was also considered. Three studies were included.

5.1. Guideline recommendations

No guidelines were identified that gave recommendations on FIT-based prediction tools for investigating CRC in low-risk people who present with lower abdominal symptoms.

NICE Diagnostics Guidance 30 (DG30) recommends adoption of FITs in primary care to guide referral for suspected CRC in people who present with symptoms but do not meet the criteria for suspected cancer pathway referral (as outlined in NG12, see Section 2. Health problem), using a threshold of 10 microgram of haemoglobin per 1 gram of faeces (NICE 2017). DG30 based recommendations on the four FIT tests currently available in the UK; the OC Sensor, HM-JACKarc and FOB Gold FIT tests were recommended, but there was not enough evidence to recommend adoption of RIDASCREEN.

5.2. Systematic reviews

Evidence on the clinical effectiveness of FIT-based prediction tools in symptomatic patients was identified in two systematic reviews (Niedermaier et al. 2016, Westwood et al. 2017b). The first review by Westwood et al. (2017b) was undertaken for the health technology appraisal to inform NICE DG30 (NICE 2017, Westwood et al. 2017a). The primary focus was the effectiveness of FIT assays in symptomatic patients and not FIT-based prediction tools, and does not report the full evidence on FIT-based prediction tools. Therefore the systematic review will not be included in this report; the studies including FIT-based prediction tools were identified in our review of the primary evidence and will be discussed further below.

The second systematic review by Niedermaier et al. (2016) summarises evidence of FIT combined with other markers compared to FIT alone in the detection of CRC. The review was not limited to symptomatic populations; only the studies of symptomatic patients are included in this report. The characteristics and relevant results of the review are included in Table 1.

Table 1. Systematic review: Niedermaier et al. (2016)

Included studies	PICO	Quality	Observations/notes
Total number of studies: 18 8 of the included articles were relevant to this appraisal. Kalimutho (2011), (n= 204) Harada (2014), (n = 508) Mizuno, (1995), (n = 81) Sieg (1998), (n = 739) Sheng (2009), (n = 110) Jin (2012), (n = 2,144) Parente (2012), (n = 280) Vronen (2004), (n = 199) Search period up to May 2015.	**Research objective:** to summarise the evidence on studies evaluating performance of FIT alone for the detection of CRC or AA compared with a combination of FIT and stool markers. **Population:** Any population subject to the index and reference tests. **Index test:** FIT, and FIT combined with any second diagnostic tool. **Reference test:** colonoscopy **Primary outcome measure:** sensitivity and specificity, AUC and p-value for AUC.	**Study design:** Systematic review **Risk of bias:** Assessed with QUADAS-2. All of the studies had applicability concerns regarding patients. One of the studies (Kalimutho) was reported to have high-risk of bias in the index test, and index test applicability concerns.	• The author's objective was broader than the scope of this report; the studies reported here are those relevant to the selection criteria. • This study included FIT of any type (qualitative, quantitative). In some cases the type of FIT was not clear. One study (Sieg) did not use a FIT assay, but rather polyclonal Abs within a laboratory environment.

Results

Diagnostic accuracy of FIT combined other tests

Marker		Sensitivity, % (95% CI)	Specificity, % (95% CI)	Combination
DNA- or RNA-based test				
p53	FIT	52% (32-71%)	98% (93-100%)	NR/pn
	FIT+	79% (59-92%)	91% (83-96%)	
APC	FIT	52% (32-71%)	98% (93-100%)	
	FIT+	79% (59-92%)	91% (83-96%)	
APC & p53	FIT	52% (32-71%)	98% (93-100%)	
	FIT+	81% (62-94%)	98% (93-100%)	
APC, p53, KRAS & BRAF	FIT	52% (32-71%)	98% (93-100%)	
	FIT+	89% (72-98%)	95% (88-98%)	
M-score* ≥ 1	FIT	82% (57-96%)	57% (82-63%)	pn
	FIT+	100% (80-100%)	31% (26-36%)	

Authors' observations

"Overall, improvements in FIT performance might be possible by combining FITs with other stool tests, in particular with DNA- or RNA-based tests…The combinations with stool protein-based tests so far assessed did not yield strong improvements in FIT performance."

"Overall evidence is limited by the fact that most studies were conducted in clinical settings, and most markers or marker combinations were evaluated in a single study only without external validation. Owing to relatively small sample sizes, resulting in overlapping confidence intervals in most comparisons of FITs alone and combined tests, reported changes in sensitivity and specificity are unlikely to be statistically significant."

"…comparability of studies was limited due to differences in inclusion and exclusion criteria, used FITs, cutpoints for FIT positivity, and protocols for stool sampling and processing."

"In conclusion, this systematic review suggests that improvements in performance of FITs are achievable through combination with further

M-score* ≥ 1	FIT	82% (57-96%)	57% (82-63%)	pp
	FIT+	71% (44-90%)	81% (77-85%)	
M-score* ≥ 2	FIT	82% (57-96%)	57% (82-63%)	pn
	FIT+	100% (80-100%)	47% (42-53%)	
M-score* ≥ 2	FIT	82% (57-96%)	57% (82-63%)	pp
	FIT+	41% (18-67%)	93% (89-95%)	
M-score* = 3	FIT	82% (57-96%)	57% (82-63%)	pn
	FIT+	94% (71-100%)	54% (49-60%)	
M-score* = 3	FIT	82% (57-96%)	57% (82-63%)	pp
	FIT+	12% (1-36%)	95% (92-97%)	
Stool protein-based test				
Stool DAF	FIT	73% (56-85%)	95% (83-99%)	pn
	FIT+	85% (70-94%)	85% (71-94%)	
Albumin	FIT	95% (84-99%)	97% (95-99%)	pn
	FIT+	95% (84-99%)	94% (91-96%)	
Transferrin	FIT	75% (59-87%)	88% (73-97%)	pn
	FIT+	90% (76-97%)	71% (53-85%)	
Transferrin	FIT	57% (34-78%)	64% (55-72%)	pn
	FIT+	86% (64-97%)	42% (34-51%)	
	FIT	57% (34-78%)	64% (55-72%)	pp
	FIT+	48% (26-70%)	82% (74-88%)	
Calprotectin	FIT	62% (47-74%)	89% (84-92%)	pn
	FIT+	91% (79-96%)	36% (30-43%)	
M2-PK	FIT	62% (47-74%)	89% (84-92%)	
	FIT+	92% (80-97%)	57% (51-63%)	
Calprotectin & M2-PK	FIT	62% (47-74%)	89% (84-92%)	
	FIT+	96% (86-99%)	24% (19-30%)	
Tissue tests				
PNA test applied on tissue samples	FIT	72% (55-86%)	88% (79-95%)	NR/pn
	FIT+	100% (90-100%)	58% (47-70%)	

stool tests. However, no definite conclusion could be drawn for most marker combinations, mainly because of heterogeneous cutpoints leading to different specificities across studies for both, FITs alone and the combination of FITs with other stool tests. Thus, further investigations are desirable."

*DNA methylation on bowel lavage fluid pp: both tests positive; pn: at least one test positive; NR/pn: not reported, but increasing sensitivity and decreasing specificity indicate a “pn” interpretation.	
AA: advanced adenomas; CI: confidence intervals; CRC: colorectal cancer; FIT: faecal immunochemical test.	

5.3. Additional studies

Three studies were identified evaluating FIT-based prediction tools in symptomatic patients (Cubiella et al. 2017, Cubiella et al. 2016, Rodriguez-Alonso et al. 2015). All three studies were based on prediction modelling, and the development and validation of a risk score prediction tool based on predictive variables (including FIT). The calculations for each risk score are listed in Appendix 3. Two of the studies included evaluation of stratifying patients into low, intermediate and high risk groups (Cubiella et al. 2017, Cubiella et al. 2016, Rodriguez-Alonso et al. 2015). One study included predictive modelling for CRC and advanced neoplasia (Rodriguez-Alonso et al. 2015); however, the risk score was developed for advanced neoplasia only. Study characteristics and results are summarised in Tables 2 to 4.

Table 2. Diagnostic cohort study: Cubiella et al. (2017), FAST score

Descriptive details	PICO	Quality of study	Observations/notes
Multicentre (number not described), Scotland and Spain. n = 3,976 in the validation cohort (n = 1,572 in the derivation cohort) 46.2% male; 53.8% female Median age 65 years (range 15 to 100 years) 37.6% were primary health care referral Recruitment period: not described for validation cohort studies. (March 2012 and September 2013 for derivation cohort.)	**Population:** Symptomatic patients referred for colonoscopy. **Index test:** FAST score. A risk score based on faecal haemoglobin concentration (by FIT), age and sex. **Reference test:** colonoscopy. **Primary outcome measure:** CRC detection. Other outcomes included AN and SCL detection.	**Study design:** Retrospective cohort study including combined data from five diagnostic accuracy cohort studies. **Risk of bias:** Assessed using QUADAS-2. The study reported high or unclear risk of bias regarding patent selection and reference standard. There was also patient applicability concerns.	• The derivation cohort was based on the same derivation cohort included in the COLONPREDICT study. The validation cohort included a large number of patients from studies in different locations. • The population included people presenting with rectal bleeding (38.6% in the validation cohort); rectal bleeding is included in the NG12 referral criteria. • The score thresholds 4.50 and 2.12 were established on the 90 and 99% sensitivity values, respectively.

Results

Diagnostic accuracy (validation cohort)

	Sensitivity % (95% CI)	Specificity (95% CI)
FAST score ≥ 4.5		
CRC	89.3 (84.1 to 93.0)	82.3 (81.1 to 83.5)
AN	60.7 (56.6 to 64.7)	85.4 (84.1 to 86.5)
SCL	57.8 (54.3 to 61.3)	87.4 (86.2 to 88.5)
FAST score ≥ 2.12		
CRC	100.0 (97.7 to 100)	19.8 (18.6 to 21.1)
AN	96.7 (94.9 to 98)	21.5 (20.1 to 22.9)
SCL	94.5 (92 to 6-96)	22.0 (20.6 to 23.5)

Authors' observations

"This prediction tool, termed the FAST Score, is highly accurate for CRC detection in symptomatic patients and allows establishment of three risk groups: CRC can be ruled out in those in the lowest risk group. In addition, we have confirmed that the threshold scores we have examined are equally sensitive for CRC irrespective of, country, CRC prevalence, age, sex, healthcare level and analytical system used to estimate f-Hb. In contrast, this prediction tool was less accurate for AN and SCL detection in the validation cohort."

"In the case of the FAST Score, its greatest advantage is simplicity and it could be calculated automatically by laboratory information management…or through simple calculators. We think the simplicity of this prediction model may encourage the implementation of the FAST Score in comparison with previous developed prediction models for CRC detection that are based on several clinical and analytical variables which must be collected from disparate sources before the model is applied."

"Even though we now need to evaluate the applicability of the FAST Score in a routine primary care setting, the implications for clinicians and policymakers are quite clear. Prediction tools based on f-Hb measured by high quality automated quantitative FIT should be the cornerstone of the evaluation of the symptomatic patient and used as

Diagnostic yield based on FAST score risk stratification (validation cohort)

		Low (FAST score <2.12)	Intermediate (FAST score ≥ 2.12 to <4.5)	High (FAST score ≥ 4.5)
	% patients	18.8	59.8	21.4
CRC	PPV (%)	0	0.9	21.7
AN	PPV (%)	2.6	8.9	41.7
SCL	PPV (%)	5.6	11.9	52.6

Diagnostic accuracy based on healthcare setting, post-hoc (validation cohort)

	Primary (n = 1,496)	Secondary (n = 2,480)	p-value
FAST score, AUC (95% CI)	0.90 (0.87–0.93)	0.92 (0.91–0.94)	0.2
FAST score ≥ 4.5			
Sensitivity, %	89.2	89.3	1
Specificity, %	81.8	82.6	0.5
FAST score ≥ 2.12			
Sensitivity, %	NR	NR	NR
Specificity, %	21.2	19.0	0.1

the criterion to determine the appropriateness of the referrals for colonoscopy. We consider that the FAST Score does provide an objective tool to guide who requires further investigation in secondary care"

"The diagnostic accuracy and applicability of the FAST Score tool in a primary care setting must be addressed in a prospective study and, ideally, compared with the COLONPREDICT score and current age and symptom-based referral guidelines. Second, further prediction tools based on laboratory findings other than f-Hb should be designed and evaluated in a primary care setting."

AN: advanced neoplasia (defined in the study as CRC plus advanced adenoma); AUC: area under the curve; CI: confidence interval; CRC: colorectal cancer; NR: not reported; PPV: positive predictive value; SCL: significant colonic lesions (defined in the study as CRC plus AN plus other significant pathology).

Table 3. Diagnostic cohort study: Cubiella et al. (2016), COLONPREDICT score

Descriptive details	PICO	Quality of study	Observations/notes
11 centres, Spain n = 1,481 in the validation cohort (n = 1,572 in the derivation cohort) 48.5% male; 51.5% female Median age 64 (range 19 to 101) 38.4% were primary health care referral. Recruitment period: March 2014 to March 2015 for the validation cohort. (March 2012 and September 2013 for derivation cohort).	**Population:** patients with gastrointestinal symptoms referred for colonoscopy. **Index test:** COLONPREDICT score. A risk score based on 12 variables; age, sex (male), faecal haemoglobin concentration (measured by FIT; ≥20 µg/g), blood haemoglobin (<10 g/dL), blood haemoglobin (10–12 g/dL), carcinoembryonic antigen (≥3 ng/mL), acetylsalicylic acid treatment, previous colonoscopy, rectal mass, benign anorectal lesion, rectal bleeding, and change in bowel habit. **Reference test:** colonoscopy. **Primary outcome measure:** CRC detection.	**Study design:** prospective, cross-sectional cohort study. **Risk of bias:** Assessed using QUADAS-2. The study was considered low risk of bias, but had patient applicability concerns.	• The population included people presenting with rectal bleeding (51.2% in the validation cohort); rectal bleeding is included in the NG12 referral criteria. • It is not clearly described in the study whether the post-hoc comparison of primary and secondary referrals is based on the derivation or validation cohort. • The score thresholds 5.6 and 3.5 were established on the 90 and 99% sensitivity, respectively.

Results

Diagnostic accuracy in the validation cohort

	Sensitivity % (95% CI)	Specificity (95% CI)
COLONPREDICT score ≥ 5.6		
CRC	87.1 (79.9 to 92.1)	79.3 (76.9 to 81.4)
AN	66.0 % (60.3–71.3)	83.5 % (81.2–85.7)
SCL	59.2 % (53.9–64.3)	84.2 % (81.8–86.3)
COLONPREDICT score ≥ 3.5		
CRC	100 % (96.0–100.0)	46.8 % (44.0–49.6)
AN	88.2 % (83.9–91.5)	50.7 % (47.7–53.7)
SCL	84.7 % (80.5–88.2)	51.8 % (48.7–54.9)

Authors' observations

"We have developed and externally validated a prediction model for CRC and SCL detection in symptomatic patients referred for colonoscopy. The COLONPREDICT model is based on easily obtainable variables – demographic, laboratory results, symptoms and anorectal examination findings – and is thus applicable both in primary and secondary healthcare. This prediction model is highly accurate, as the calibration plot shows, and allows for differentiation of a high-risk group and, especially, a low-risk group with a probability of CRC detection below 1 %."

"…we believe that the diagnostic accuracy of our prediction model should be externally evaluated in a population with gastrointestinal symptoms attended to in primary care before its use is recommended. Hypothetically, we believe that the diagnostic accuracy of the COLONPREDICT score may increase in CRC low prevalence populations due to an increase in specificity…we cannot answer… whether patients undergoing the diagnostic test fare better than similar untested patients."

"simpler prediction models with similar performance based on laboratory findings must be designed and evaluated in a primary care

Diagnostic yield based on COLONPREDICT score risk stratification (validation cohort)

		Low (COLONPREDICT score <3.5)	Intermediate (COLONPREDICT score ≥ 3.5 to <5.6)	High (COLONPREDICT score ≥ 5.6)
	% patients	39.5	29.6	30.9
CRC	PPV, % (95% CI)	0.2 (0.0- 1.1)	4.4 (2.8-6.8)	40.7 (36.7-45.9)
AN	PPV, % (95% CI)	7.1 (5.3- 9.6)	20.8 (17.2- 24.8)	58.1 (53.4- 62.5)
SCL	PPV, % (95% CI)	8.5 (6.4- 10.0)	24.5 (20.7-28.8)	61.4 (56.9-65.8)

Diagnostic accuracy based on healthcare setting, post-hoc analysis

	Primary	Secondary	p-value
COLONPREDICT score, AUC (95% CI)	0.91, (0.89 to 0.94)	0.93, (0.91 to 0.9)	0.3
COLONPREDICT score ≥ 5.6			
Sensitivity, %	89.0 % (81.6-93.8)	88.9 % (83.9-92.6)	1
Specificity, %	75.8 % (72.6-78.8)	80.3 % (78.4-82.1)	0.01
COLONPREDICT score ≥ 3.5			
Sensitivity, %	100 % (96.1-100.0)	99.6 % (97.2-100.0)	1
Specificity, %	44.1 % (40.6-47.7)	47.1 % (44.8-49.4)	0.1

setting. In this respect, the introduction of new CRC biomarkers may ease the CRC diagnosis process in symptomatic patients."

AN: advanced neoplasia (defined in the study as CRC plus advanced adenoma); AUC: area under the curve; CI: confidence interval; CRC: colorectal cancer; PPV: positive predictive value; SCL: significant colonic lesions (defined in the study as CRC plus AN plus polyposis, colitis (any aetiology), polyps ≥10 mm, complicated diverticular disease, colonic ulcer and/or bleeding angiodysplasia).

Table 4. Diagnostic cohort study: Rodriguez-Alonso et al. (2015)

<table>
<tr><th>Descriptive details</th><th>PICO</th><th>Quality of study</th><th>Observations/notes</th></tr>
<tr>
<td>Single centre, Spain

n = 1,003 in the main study
n = 680 in the training cohort
n = 323 in the validation cohort

46.8% male; 53.2% female

Median age not described, but the distribution is as follows: 7.1% <40 years; 13.0% 41-50 years; 22.4% 51-60 years; 29.1% 61-70 years; 28.4% >70 years.

66.3% were referred from primary care.

Recruitment period September 2011 to October 2012.</td>
<td>Population: patients with abdominal symptoms referred for colonoscopy.

Index test: risk score based on age, sex and faecal haemoglobin concentration (measured by FIT; ≥10 ug Hb/g faeces).

Reference test: colonoscopy.

Primary outcome measure: detection of CRC or AN.</td>
<td>Study design: prospective, diagnostic cohort study.

Risk of bias:
Assessed using QUADAS-2. The study was considered to have unclear risk regarding patient selection, and had concerns regarding applicability of patient selection.</td>
<td>• The primary aim of this study was the efficacy of FIT (alone) versus current referral guidelines in a larger patient cohort. Included here are the results relating to selection criteria (FIT-based prediction tool) only.
• This study included a smaller number of patients in the prediction tool cohort compared to other studies.
• The characteristics (age, sex, etc.) of the training and validation cohorts were unclear.
• People presenting with rectal bleeding were included in the study.
• The training and validation cohorts were obtained by split-sampling the original study cohort; the prediction tool analyses were therefore retrospective.</td>
</tr>
<tr><th colspan="3">Results</th><th>Authors' observations</th></tr>
<tr>
<td colspan="3">Diagnostic accuracy (validation cohort)

<table>
<tr><th></th><th>Sensitivity, % (95% CI)</th><th>Specificity, % (95% CI)</th></tr>
<tr><td>Risk score ≥ 5</td><td>88.1 (74.3 to 96.0)</td><td>63.3 (57.4 to 69.0)</td></tr>
</table>
</td>
<td>"Our study also attempted to identify risk factors for AN that could be used to derive a risk score for patients with colorectal symptoms. We found FIT positivity, age and male gender to be independent predictors of advanced adenoma or cancer."

"A quantitative FIT based strategy, performs better than current high-risk symptoms-based strategies in fast tracking suspected cancer referrals. A score that combines quantitative FIT, age and gender can accurately estimate the risk of neoplastic lesions in symptomatic patients and help the physician in the decision-making process and in the proper allocation of endoscopic resources."</td>
</tr>
<tr><td colspan="4">AUC: area under the curve; CI: confidence interval; CRC: colorectal cancer; AN: advanced neoplasia (defined in the study advanced adenoma or invasive carcinoma);</td></tr>
</table>

5.4. Ongoing trials

No ongoing trials evaluating FIT-based prediction tools for investigation CRC in symptomatic patients were identified by this appraisal.

6. Safety

None of the evidence identified by this appraisal reported on the safety of a FIT-based prediction tool for CRC.

7. Cost effectiveness

There are no direct health economic studies which compare the relative benefits of FIT-based prediction tools as triage for referral for CRC. The clinical evidence highlights two prominent tools, FAST and COLONPREDICT; these scoring systems utilise additional data alongside FIT to improve diagnostic accuracy. The brief costing approach undertaken by HTW looks to estimate the costs associated with the additional activities required for each prediction tool offset against savings gained through reduction in colonoscopies.

Compared to FIT alone, both FAST and COLONPREDICT require additional information. FAST requires information on age and sex. Data requirements for FAST is assumed, in the base case, to be collected in a standard GP consultation, FIT and a phlebotomy test. An increase in face to face GP contact time of 2 minutes is used in the sensitivity analysis for FAST.

COLONPREDICT utilises a greater level of information compared to FAST, and therefore additional tests are required. The description from Cubeilla et al. (2017) lists processes undertaken to estimate COLONPREDICT as an anorectal examination, a venous blood test, histopathology and the collection of a detailed patient history. The GP consultation duration is assumed to be sufficient to undertake the digital rectal examination, but the additional collection of the detailed patient history and subsequent calculation of the score is assumed to extend the time required above that of the traditional 9.22 minutes. The base case scenario GP consultation is assumed to be extended by 5 minutes; scenario analysis assumptions range from an additional 2 to 8 minutes. The pathology tests, such as carcinoembryonic antigen tests, are included as a single histopathology test. The venous blood test is costed as a single phlebotomy test. All costs are reported in 2017-2018 GBP.

Table 5. Unit costs

Procedure	Resource detail	Cost (£)	Range	Reference
GP consultation	9:22 minutes	37.00		PSSRU 2018-2019
Collection of details (age, sex)	0 minutes GP[1] contact	0	2 minutes (£7.90)	PSSRU 2018-2019
Extension to GP consultation (detailed patient history, DRE, calculation)	5 minutes GP contact	19.74	2 – 8 minutes GP contact (£7.90 - £31.58)	PSSRU 2018-2019
Phlebotomy		2.83		National Schedule of Reference Costs - Year 2017-18
Histopathology and histology		32.73		National Schedule of Reference Costs - Year 2017-18
FIT (OC-Sensor)		4.53		NICE Diagnostics Guidance 30 (DG30)
Colonoscopy		319.41	(45% therapeutic) 329.55	National Schedule of Reference Costs - Year 2017-18
[1]GP contact is calculated at £37 per 9:22 minutes consultation.				

The base case cost estimates of the FIT based predication tools are reported with the lower and higher bounds in parentheses. The FAST prediction tool is estimated to cost £44.36 (£44.36 to £52.26). The COLONPREDICT tool is estimated to cost £96.83 (£84.99 to £108.67). The cost of a colonoscopy reflects the weighted average of gastroenterology based colonoscopy treatments for adults. This includes the possibility of the inclusion of biopsy and also a therapeutic colonoscopy.

The cost consequence focus of this rapid review is in the use of the prediction tools as a 'rule out' test for requiring a colonoscopy. The PPV figures reported in the clinical evidence suggest that FAST would identify 18.8% of patients as true negatives. COLONPREDICT identifies 39.5% as low risk patients not requiring a colonoscopy; there are 0.2% false negatives within this group (95% CI 0.0% to 1.1%). The methodology used in NICE Diagnostics Guidance 30 (DG30) assumes that a proportion of symptomatic patients receiving a 'rule out' finding from their FIT test would remain symptomatic and either require a second test (20%) or a colonoscopy (32.5%).

A cohort of symptomatic patients is estimated from the total colonoscopy levels in Wales of around 27,000. This includes approximately 3,000 screening patients and an assumed 20% surveillance. The symptomatic cohort numbers 19,200 patients and is modelled to illustrate the relative costs and outcomes according to whether FAST or COLONPREDICT was deployed.

Table 6. Cost impact

	FAST ([A])	COLONPREDICT ([B])	Difference ([B-A])
Initial test cost	£851,712	£1,859,136	£1,007,424
Colonoscopy avoided	-£1,152,965	-£2,422,452	-£1,269,488
Repeat test	£32,024	£146,872	£114,847
Colonoscopy following rule out	£374,713	£787,297	£412,583
Total			£265,367

Estimates suggest that the higher initial cost of the COLONPREDICT tool, due to the higher data requirements, is not sufficiently offset by the greater number of colonoscopies avoided. Using FAST saves around £14 more per patients than COLONPREDICT. Low cost estimates find a reduced saving favouring FAST by around £1 whilst higher cost estimates extend the relative saving to £18. The higher reduction in colonoscopies achieved using COLONPREDICT may be associated with improved health due to the avoidance of complications, however, this may be offset by the higher level of false negatives in the 'rule out' group. The 0.2% risk associated with COLONPREDICT results in approximately 15 patients being ruled out, secondary testing or a subsequent colonoscopy may diagnose these patients.

A scenario analysis is undertaken with the assumption that all the components required to estimate COLONPREDICT are routinely collected in primary care, resulting in equality of tool cost. In this scenario, COLONPREDICT would save approximately £41 per patient. The percentage of initial consultations collecting all the components of COLONPREDICT required to break even between the tools is approximately 26%.

The sensitivity analysis suggests that changes to the time requirements do not significantly change the outcome. Findings are not sensitive to whether the base case FIT was OC-Sensor, as used in the main analysis, or HM-JACKarc. The inclusion of HM-JACKarc into the analysis includes the assumption that either FIT outcomes can be utilised within the prediction tool with no additional costs.

8. Organisational issues

Limited evidence on organisational issues was identified for FIT-based prediction tools in primary care. The authors of the FAST study noted that a more complex prediction tool (COLONPREDICT) may be limiting in practice, due to the number and type of variables included in the tool, such as anorectal examination and venous blood tests (Cubiella et al. 2017).

One study surveyed GP attitudes towards use of FIT in primary care (von Wagner et al. 2018). The survey focused on use of FIT as a rule-out test prior to a two-week referral assessment (as outlined in NG12) versus a two-week referral with no FIT test; due to the different pathway contexts, the results cannot be generalised to FIT or FIT prediction tools in the 'low risk' symptomatic population. However, the study did identify that GP awareness of FIT, as both a rule-in and rule-out test, was low.

The topic proposer highlighted that FIT to guide referral in primary care is currently being explored and planned in Wales, but the approach is inconsistent. Expert comment noted that with the introduction of FIT for asymptomatic screening, it is unlikely that a different analyser would be used for both populations.

9. Patient issues

The evidence searched did not identify any research on patients' experiences or uptake relating to FIT-based predictions tools in symptomatic patients.

10. Conclusions

The available evidence on FIT-based prediction tools in symptomatic patients who are at low risk of CRC is limited. All the studies identified for this report included symptomatic patients, but applicability of these to the selection criteria (Appendix 1) comes with uncertain/high applicability concerns, as the studies included patients who could be considered to fulfil the referral criteria outlined in NICE NG12 guidance (such as patients with rectal bleeding).

One relevant systematic review was identified, summarising studies that evaluated FIT combined with other markers, versus FIT alone. Sensitivity and specificity of FIT alone varied across the studies; sensitivity ranged from 52-95% and specificity ranged from 57-98%. Heterogeneity across the studies, including characteristics and FIT assays used, made comparative analysis difficult. Overall, the study authors reported sensitivity improved with FIT combined with DNA markers, but it did not appear to improve when combined with stool protein markers.

Three primary studies were identified and included in this report, all of which developed and validated risk scores. One study developed a simple risk score for advanced neoplasia (described as advanced adenoma or invasive carcinoma). Specificity of this score for detecting advanced neoplasia was 88.1%.

The other two primary studies described the development and validation of risk scores to detect CRC in symptomatic patients: the COLONPREDICT score and the FAST score. Both risk scores included two threshold scores, established on 95% and 99% specificity for CRC detection. Sensitivity was similar between COLONPREDICT and FAST at the higher threshold (87.1% and 89.3% sensitivity, respectively), and both reported 100% specificity when the lower cut-off values were applied, although specificity was greatly reduced. Both studies also stratified patients into low-, intermediate-, and high-risk categories based on the score thresholds. Negative predictive values (NPVs) for the stratified populations were not reported; however, positive predictive values (PPVs) for the 'low-risk' group were 0.2% with COLONPREDICT and 0.0% with FAST. Use of these lower score thresholds may therefore be useful as a 'rule out' criteria for CRC, avoiding unnecessary colonoscopies. Finally, both studies reported no difference between primary and secondary settings.

No evidence on the cost effectiveness of FIT-based prediction tools was identified from the literature. This report included a brief cost consequence analysis for FAST and COLONPREDICT, focusing on the use of the tools as a 'rule out' test for further colonoscopy, based on the 'low-risk' stratification cohort for each tool. It should be noted that the analyses were based on the proportion of patients stratified into the low-risk group (18.8% for FAST and 39.5% for COLONPREDICT) within the validation cohort, and this may not be representative of the populations in primary care settings. The higher number of tests/variables required for COLONPREDICT results in a higher initial test cost, but this is partially offset by the higher reduction in colonoscopies. Overall, the cost consequence analysis approach suggests that FAST offers greater cost savings than achieved through COLONPREDICT.

11. Further research

Further studies to develop and validate FIT-based prediction tools for people presenting with lower abdominal symptoms in primary care, but who do not meet the two-week referral criteria, are recommended. Large, prospective, comparative multicentre studies are recommended to evaluate the comparative effectiveness of FIT-based prediction tools against FIT alone, other FIT-based predictions tools, and prediction tools that do not include FIT.

12. Contributors

The HTW staff and contract researchers involved in writing this report were:

- L Elston – main author and systematic literature reviewer
- J Washington – literature search
- T Winfield - economic appraisal
- D Jarrom – quality assurance

The HTW Assessment Group advised on methodology throughout the scoping and development of the report.

A range of clinical experts from the UK provided material and commented on a draft of this report. Their views were documented and have been actioned accordingly. All contributions from reviewers were considered by HTW's Assessment Group. However, reviewers had no role in authorship or editorial control, and the views expressed are those of Health Technology Wales.

Experts who contributed to this appraisal:

Prof. Stephen P. Halloran
Dr Tom Crosby
Dr Sunil Dolwani
Prof John Geen
Ms Leanne James
Miss Gina Sanki
Dr Robert Logan
Prof Jared Torkington

Review period

Two years after the date of publication, a high-level literature search will be undertaken to determine if there is new evidence that could alter the conclusions of this report. If so, the appraisal will be updated.

13. References

Bowel Cancer UK. (2018). Spotlight on bowel cancer in Wales. Available at: www.bowelcanceruk.org.uk/news-and-blogs/news/new-report-on-bowel-cancer-in-wales-highlights-serious-delays-in-diagnosis-and-lowest-survival-rates-in-europe/ [Accessed October 2018].

Cancer Research UK. (2018). Bowel cancer incidence statistics. Available at: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence [Accessed October 2018].

Cubiella J, Digby J, Rodriguez-Alonso L, et al. (2017). The fecal hemoglobin concentration, age and sex test score: development and external validation of a simple prediction tool for colorectal cancer detection in symptomatic patients. International Journal of Cancer. 140(10): 2201-11. doi: https://dx.doi.org/10.1002/ijc.30639

Cubiella J, Vega P, Salve M, et al. (2016). Development and external validation of a faecal immunochemical test-based prediction model for colorectal cancer detection in symptomatic patients. BMC Medicine. 14: 128. doi: https://dx.doi.org/10.1186/s12916-016-0668-5

NICE. (2015). Suspected cancer: recognition and referral. Last updated 2017. NICE guideline NG12. National Institute for Health and Care Excellence. Available at: www.nice.org.uk/guidance/ng12 [Accessed October 2018].

NICE. (2017). Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care. Diagnostics guidance DG30. National Institute for Health and Care Excellence. Available at: www.nice.org.uk/guidance/dg30 [Accessed October 2018].

Niedermaier T, Weigl K, Hoffmeister M, et al. (2016). Fecal immunochemical tests combined with other stool tests for colorectal cancer and advanced adenoma detection: a systematic review. Clinical And Translational Gastroenterology. 7: e176. doi: https://dx.doi.org/10.1038/ctg.2016.29

Rodriguez-Alonso L, Rodriguez-Moranta F, Ruiz-Cerulla A, et al. (2015). An urgent referral strategy for symptomatic patients with suspected colorectal cancer based on a quantitative immunochemical faecal occult blood test. Digestive & Liver Disease. 47(9): 797-804. doi: https://dx.doi.org/10.1016/j.dld.2015.05.004

von Wagner C, Stoffel S, Freeman M, et al. (2018). Attitudes towards faecal immunochemical testing in patients at increased risk of colorectal cancer: an online survey of GPs in England. British Journal of General Practice. 68(676): e757-e64. doi: 10.3399/bjgp18X699413

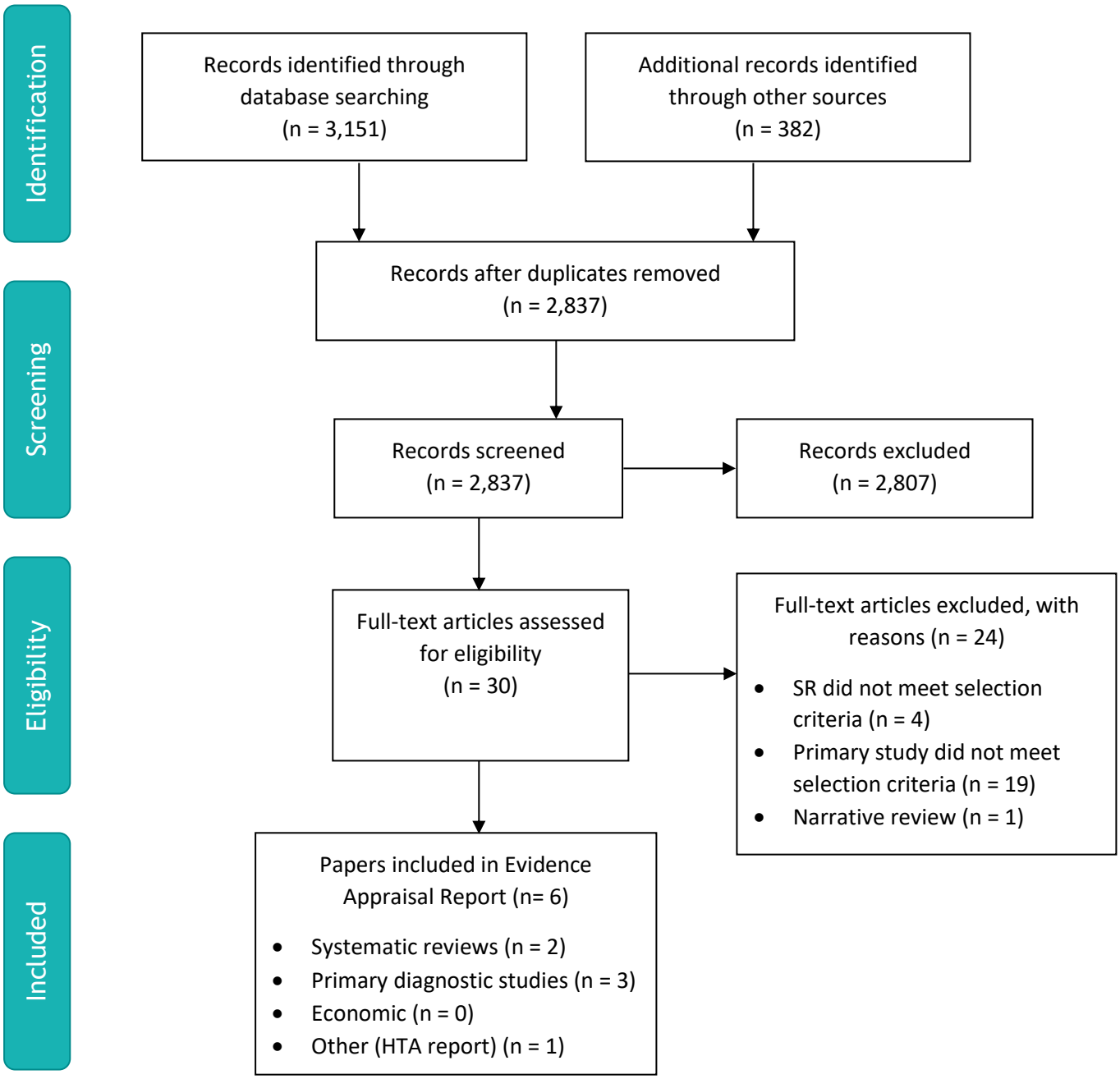
Westwood M, Corro Ramos I, Lang S, et al. (2017a). Faecal immunochemical tests to triage patients with lower abdominal symptoms for suspected colorectal cancer referrals in primary care: a systematic review and cost-effectiveness analysis. Health Technology Assessment. 21(33): 1-234. doi: https://dx.doi.org/10.3310/hta21330

Westwood M, Lang S, Armstrong N, et al. (2017b). Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. BMC Medicine. 15: 189. doi: https://dx.doi.org/10.1186/s12916-017-0944-z

Appendix 1. PICO framework

Question	*What is the most effective FIT-based prediction tool for investigating colorectal cancer in people with lower abdominal symptoms who are at low risk of colorectal cancer?*
P (population)	People presenting with lower abdominal symptoms who are being investigated for possible colorectal cancer.
I (intervention)	FIT, used as part of a prediction tool, such as FAST or COLOPREDICT
C (comparator(s))	FIT alone Other faecal blood tests No faecal blood test, ie "straight to colonoscopy"
O (outcomes)	Diagnostic accuracy Rates of referral for further investigation Appropriate referral to secondary care (proportion of people referred for further investigation in whom colorectal cancer was confirmed; proportion *not* referred in whom colorectal cancer was later diagnosed) Colorectal cancer mortality Patient acceptability/satisfaction/HRQOL (any measures)
Study design	Studies of diagnostic accuracy Randomised controlled trials (or non-randomised trials if no RCTs are identified) Where well-conducted and relevant systematic reviews exist, these will be used as the basis of the report, along with any primary studies published since the review(s) completion.
Other factors/subgroups of interest	Where available, we will focus on studies conducted in the primary care setting. Where available, we will focus on patients at low risk of colorectal cancer (defining "low risk" as patients who do not meet the criteria for a suspected cancer referral outlined in NICE Guideline NG12). Where possible, we will report results separately for different risk categories of patients. Where available, data will be collected on 'safety netting' pathways for patients who do not meet the threshold for referral according to each prediction tool, and whether this affects the rate of subsequent cancer diagnosis in such patients.

Appendix 2. PRISMA flow diagram outlining selection of papers



Appendix 3. Risk score calculations from the identified studies

Cubiella et al. (2017), FAST score

FAST Score (*b* coefficient) = 0.684 x f-Hb (0, 20) microgram Hb/g feces + 2.824 x f-Hb (20, 200) microgram Hb/g feces + 4.184 x f-Hb ≥200 microgram Hb/g feces + 0.031 x age (years) + 0.479 x sex (male).

Cubiella et al. (2016), COLONPREDICT score

COLONPREDICT score = 0.789 × rectal bleeding + 0.536 × change in bowel habit + 2.694 × rectal mass − 1.283 × benign anorectal lesions + 2.831 × f-Hb ≥20 µg/g of faeces +1.561 ×b-Hb <10 g/dL +0.588 ×b-Hb 10–12 g/dL + 1.511 × CEA ≥3 ng/mL + 0.040 × age (years) + 0.813 × sex (male) − 2.073 × previous colonoscopy (last 10 years) − 0.849 × continuous treatment with aspirin.

Rodriguez-Alonso et al. (2015)

Risk Factor	Score
Age	
<40 years	0
41–50 years	1
51–60 years	2
61–70 years	3
>70 years	4
Sex	
Female	0
Male	2
Faecal haemoglobin concentration	
<10 ug Hb/g faeces	0
≥10 ug Hb/g faeces	5
Total score	0-11