

BMJ Open Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy

Karoline Freeman,¹ Brian H Willis,² Hannah Fraser,¹ Sian Taylor-Phillips,¹ Aileen Clarke¹

To cite: Freeman K, Willis BH, Fraser H, *et al*. Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy. *BMJ Open* 2019;**9**:e027428. doi:10.1136/bmjopen-2018-027428

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-027428>).

Received 22 October 2018
Revised 9 January 2019
Accepted 28 January 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY. Published by BMJ.

¹Warwick Medical School, University of Warwick, Coventry, UK

²Institute of Applied Health Research, University of Birmingham, Birmingham, UK

Correspondence to

Karoline Freeman;
K.Freeman@warwick.ac.uk

ABSTRACT

Objective Test accuracy of faecal calprotectin (FC) testing in primary care is inconclusive. We aimed to assess the test accuracy of FC testing in primary care and compare it to secondary care estimates for the detection of inflammatory bowel disease (IBD).

Methods Systematic review and meta-analysis of test accuracy using a bivariate random effects model. We searched MEDLINE, EMBASE, Cochrane Library and Web of Science until 31 May 2017 and included studies from auto alerts up until 31 January 2018. Eligible studies measured FC levels in stool samples to detect IBD in adult patients with chronic (at least 6–8 weeks) abdominal symptoms in primary or secondary care. Risk of bias and applicability were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 criteria. We followed the protocol registered as PROSPERO CRD 42012003287.

Results 38 out of 2168 studies were eligible including five from primary care. Comparison of test accuracy by setting was precluded by extensive heterogeneity. Overall, summary estimates of sensitivity and specificity were not recorded. At a threshold of 50 µg/g, sensitivity from separate meta-analysis of four assay types ranged from 0.85 (95% CI 0.75 to 0.92) to 0.94 (95% CI 0.75 to 0.90) and specificity from 0.67 (95% CI 0.56 to 0.76) to 0.88 (95% CI 0.77 to 0.94). Across three different definitions of disease, sensitivity ranged from 0.80 (95% CI 0.76 to 0.84) to 0.97 (95% CI 0.91 to 0.99) and specificity from 0.67 (95% CI 0.58 to 0.75) to 0.76 (95% CI 0.66 to 0.84). Sensitivity appears to be lower in primary care and is further reduced at a revised threshold of 100 µg/g.

Conclusions Conclusive estimates of sensitivity and specificity of FC testing in primary care for the detection of IBD are still missing. There is insufficient evidence in the published literature to support the decision to introduce FC testing in primary care. Studies evaluating FC testing in an appropriate primary care setting are needed.

INTRODUCTION

Inflammatory bowel disease (IBD) is an organic disease caused by inflammation of the intestine. The disease is severe and progressive,^{1,2} and over 50% of people require surgery within 10 years of diagnosis.³ Timely referral,

Strengths and limitations of this study

- This review used an innovative approach involving random combinations of test accuracy data from included studies in 25 000 meta-analyses to display the breadth of heterogeneous evidence from included studies.
- Exploratory meta-analyses investigated different test assays, clinical questions and positivity thresholds of faecal calprotectin levels.
- No overall summary estimates of sensitivity and specificity were produced due to heterogeneity.
- Comparison of test accuracy by setting was precluded by extensive heterogeneity in the small number of studies in primary care.
- The categorisation into different clinical questions was subjective because the disease categories are ill defined and studies' definitions of conditions and groups of conditions varied.

which requires filtering of patients with a high probability of IBD from a broader group of patients with mainly irritable bowel syndrome (IBS), is vital. As many as 3.1 million people in the USA⁴ and 2.5–3 million people in Europe⁵ suffer from IBD with close to 300 000 patients in the UK.⁶ An estimated 256 000 new cases of IBD are diagnosed annually throughout Europe.⁵ Northern America and Northern Europe, including the UK, have the highest incidence rates.⁷

General practitioners refer between 10% and 20% of patients presenting with abdominal symptoms if no specific test is available.^{8,9} However, only about 25% of referred patients have organic disease, of which one-third have IBD.^{10,11} Faecal calprotectin (FC), a regulator of the inflammatory response, could potentially aid more selective referral. Patients with high levels of FC have an increased probability of IBD. FC is a small calcium-binding protein. It is released into the intestinal lumen from activated neutrophils accumulating at the

site of inflammation, and levels of FC are correlated with the level of inflammation.¹² The protein is stable at room temperature and resistant to digestion, making it a relatively easy candidate for a stool test for inflammation of the intestinal tract.

The FC test is approved by the Food and Drug Administration and recommended by gastroenterological societies across the globe for its usefulness in the diagnosis of IBD.^{13–16} However, there is no clear guidance on which settings it is considered appropriate. FC testing is not supported by Medicare in the USA and Australia, but is approved for the differential diagnosis of IBD and IBS in adult patients in the UK by the National Institute for Health and Care Excellence (NICE) when referral to secondary care is being considered and cancer is not suspected (DG11).¹⁷ At the time of the recommendations,

only test accuracy studies undertaken in secondary care were available to inform the decision.¹⁸ Our knowledge on how the test performs in primary care is limited and a systematic approach to assessing the applicability of the available evidence specifically for this setting is timely. We aimed to assess the test accuracy of FC testing to detect IBD in adult patients with chronic abdominal symptoms in a primary care pathway (figure 1). In addition, we aimed to explore the differences in FC test performance between primary and secondary care.

METHODS

Review methods

We followed the protocol by Waugh *et al*¹⁸ registered as PROSPERO CRD 42012003287. We considered all studies

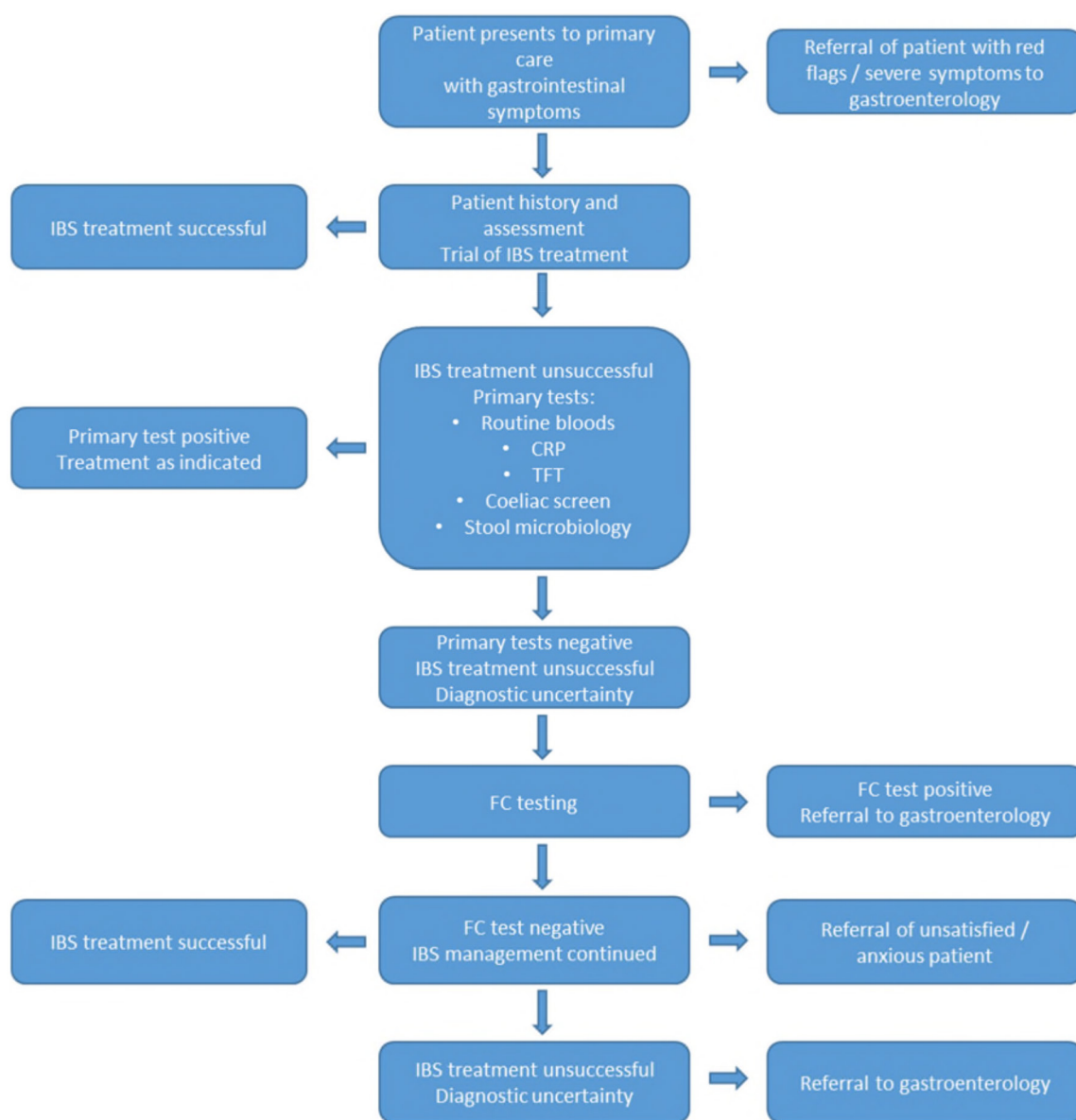


Figure 1 Indicative pathway of faecal calprotectin testing in primary care based on NICE guidelines (DG11) and expert opinion for adult patients presenting with chronic abdominal pain to primary care. CRP, C reactive protein; FC, faecal calprotectin; IBS, irritable bowel syndrome; NICE, National Institute for Health and Care Excellence; TFT, thyroid function test.

included by Waugh *et al*¹⁸ for inclusion and searched MEDLINE, EMBASE, Cochrane Library and Web of Science from 1 September 2012 to 31 May 2017 (online supplementary 1). Eligible studies from auto-alerts were included up until 31 January 2018. Reference lists of included studies were checked.

Studies were assessed for eligibility independently and in duplicate, and any disagreements were resolved by consensus. Records rejected at full-text stage and reasons for exclusion were documented.

We included studies which measured FC levels in stool samples to detect IBD in adult patients (with $\geq 80\%$ of study population 18–60 years) with chronic (at least 6–8 weeks) abdominal symptoms not yet diagnosed in primary or secondary care. The reference standard to verify FC test outcomes was colonoscopy with histology, other imaging technologies and follow-up. Our main outcomes of interest were sensitivity and specificity.

Data were extracted by KF on prespecified data extraction sheets (online supplementary 2) which were checked by HF. For studies reporting sensitivity, specificity, positive and negative predictive values and a total number of included patients, 2×2 tables of true positives, false positives (FPs), false negatives and true negatives were calculated. For studies reporting test accuracy at a lower and an upper cut-off, 2×2 data tables were extracted for the lower and the upper cut-off. Authors of relevant studies with missing 2×2 data were contacted to request data.

Quality was assessed independently and in duplicate by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 criteria,¹⁹ which included definitions for the signalling questions to match the review question.

Patient involvement

A patient advisory group was actively involved in the funding application for the project of which this study is a part. Two representatives of the patient advisory group will be involved in dissemination of findings at local and national events for patients with IBD.

Analysis

Studies were considered under three main clinical questions with different definitions of the target and non-target conditions: FC testing to differentiate: (1) IBD from IBS, (2) IBD from non-IBD conditions and (3) organic from non-organic conditions. Different FC test assays were handled independently.

We used Review Manager V.5.3 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) to produce paired forest plots of sensitivity and specificity. All analyses were performed in R V.3.4.1 (Vienna, Austria).²⁰ Meta-analyses were undertaken using a bivariate random effects model²¹ using the package *glmer*²² with a minimum of five studies required for meta-analysis. Different assays and clinical questions were considered separately. No overall summary estimates of sensitivity and specificity were produced

due to heterogeneity. Test accuracy was explored at the commonly used thresholds of 50 and 100 μg FC per g stool sample ($\mu\text{g}/\text{g}$). Outputs from meta-analyses were entered into Review Manager to produce receiver operating characteristic (ROC) plots.

Heterogeneity was explored by meta-regression analyses with assay type and clinical question added as covariates in turn. We tested the assumption of equal variances (online supplementary 3) suggesting this was reasonable for the data set. Additional models assuming unequal variances did not converge.

Each meta-analysis considered only one outcome per study. However, a number of studies reported 2×2 data for multiple test assays and clinical questions. We, therefore, undertook an exploratory sensitivity analysis. First, we produced a list of all possible combinations of test assay and clinical question for each included study at a common threshold (some studies reported outcomes for one test assay and clinical question contributing one possible combination while others reported results for up to three test assays and two clinical questions resulting in 6 possible combinations). We then ran a random sample of 25 000 meta-analyses out of a possible 10 million picking one outcome per study at random in each round of meta-analysis and considering all combinations of test assays and clinical questions reported in the studies. Pairs of sensitivity and specificity from these meta-analyses were plotted in an ROC plot and a two-dimension density plot to take account of the different definitions of disease and test assays and to allow visualisation of the variability in the data.

RESULTS

Figure 2 provides the Preferred Reporting Items for Systematic Reviews and Meta-Analyses²³ flow diagram of study selection. See online supplementary 4 for a list of records rejected at full-text stage and reasons for exclusion. Out of 2168 unique records, 38 studies were eligible for inclusion. The summary table 1 reveals extensive heterogeneity in all major aspects of test accuracy studies. Study characteristics by study can be found in online supplementary 5.

Figure 3 summarises the assessment of risk of bias and applicability (online supplementary 6 for the assessment by study). Overall, the risk of bias was high or uncertain across all four domains in at least 50% of studies. Concerns about the applicability of the patient population to a primary care setting were high in about 75% of studies.

The forest plots in online supplementary 7 describe the complete evidence on test accuracy from the 38 included studies reporting all the thresholds, assays and clinical questions explored. Overall, as might be expected, sensitivity decreases and specificity increases as the threshold increases. FC testing appears to be more accurate in the detection of the more precise clinical construct of IBD than in the detection of organic disease as a whole.

Two studies^{24 25} compared five and six different test assays at a common threshold of 50 $\mu\text{g}/\text{g}$. This allows for a test comparison in the same study population under

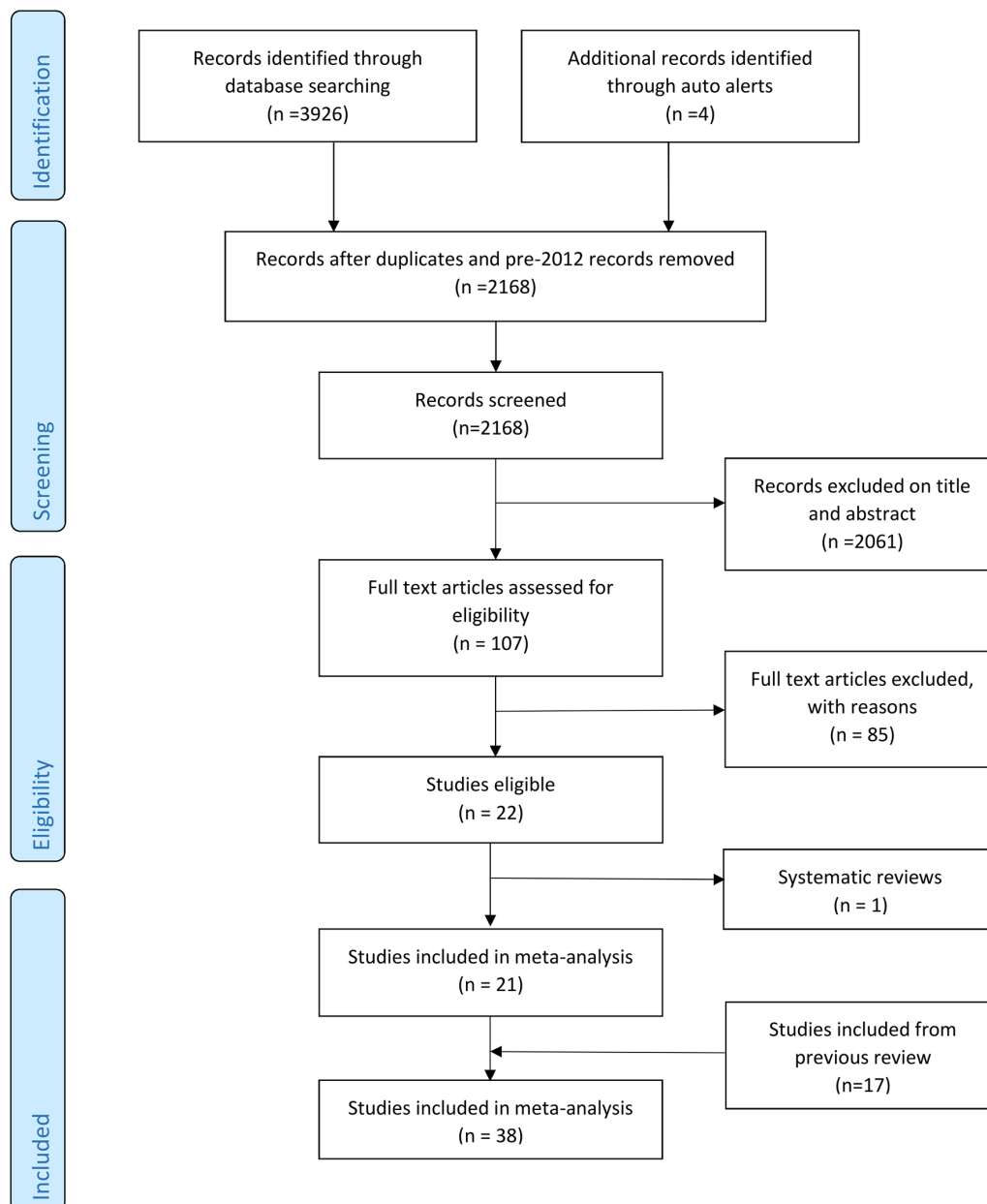


Figure 2 Flow diagram of study inclusion.

similar study conditions. It appears that the specificity varies to a greater extent than the sensitivity (figure 4). Both studies reported medium to strong correlations but low agreement between assays, meaning that analysing the same sample with different FC test assays will result in different values for FC.

Meta-analysis at 50 µg/g threshold

Out of 38, 28 studies reported test accuracy at the common 50 µg/g cut-off and could be considered for meta-analysis. Meta-analyses were undertaken separately for test assays and for clinical questions to explore heterogeneity and to allow for multiple outcomes from a number of studies.

All 18 assays were considered in the meta-analysis investigating effect of assay type on summary estimates of test accuracy. (This was based on the assumption that the clinical question is generic and one outcome was picked per

study for analysis). Four assays (PhiCal, EK-CAL, Quantum-Blue and EliA) with five or more studies each could be considered in the comparison (online supplementary 8 for 2×2 data). Nineteen studies contributed data.^{10 24–41} Five studies^{10 25 31 39 40} contributed data to two different assays. Figure 5 depicts the pairs of sensitivity and FP rates of contributing studies in the ROC space, with summary estimates by assay type. At the global 50 µg/g threshold, test performance appeared to vary slightly across assays (table 2). Quantum-Blue had the highest summary estimate for sensitivity with 0.94 (95% CI 0.75 to 0.99) but also the lowest specificity (0.67, 95% CI 0.56 to 0.76). The greatest difference in sensitivity (9%) was between EK-CAL ELISA and Quantum-Blue point-of-care test (POCT) while the greatest difference in specificity (21%) was between PhiCal ELISA and Quantum-Blue

Table 1 Summary of study characteristics of included studies addressing the test accuracy question of FC testing in the detection of IBD

Study characteristic	Outcome
Study characteristics:	
Publication type (studies)	
Full text	32
Abstract	6
Year of publication (range)	2000–2018
Population size (range)	31–1031
Geographical region (studies):	
UK	14
Rest of Europe	16
Asia	2
Middle East	2
North Africa	1
Russia	1
Canada	1
USA	1
Patient characteristics:	
Age (range)	14–97 years
IBD prevalence (range)	2.1%–76%
FC assay type (studies)*:	
Immunoassays	
ELISA	32
FEIA	5
CLIA	3
PETIA	2
POCT	10
Setting (studies)†	
Primary care	5
Secondary care	29
Outpatients and inpatients	12
Referred patients	17
Mix	2
Target condition (studies)*:	
IBD	23
Inflammatory disease	5
Organic disease	19
Non-target condition (studies)*:	
IBS	12
Functional disease	2
Non-IBD	12
Non-organic disease	16
Non-inflammatory disease	5
Other	3
FC test data collection (studies):	
Prospectively for patients with eligible symptoms in primary care	1

Continued

Table 1 Continued

Study characteristic	Outcome
Prospectively for patients at time of referral in primary care	1
Retrospectively of routine FC tests in primary or secondary care	10
Prospectively prior to a planned colonoscopy in secondary care	23
Prospectively during the assessment for the need of colonoscopy in secondary care	1
Unclear	2
Reference standard (studies):	
Colonoscopy with biopsy	13
Colonoscopy±biopsy	7
Endoscopy+other imaging tests	8
Endoscopy+follow-up	3
Endoscopy+other imaging tests+follow-up	5
Unclear	2

*Some studies evaluated multiple tests/clinical questions.

†Two studies were unclear about the setting.

CLIA, chemiluminescent immunoassay; FC, faecal calprotectin; FEIA, fluorescence enzyme immunoassay; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; PETIA, particle-enhanced turbidimetric immunoassay; POCT, point-of-care test.

POCT suggesting greater variation in specificity than in sensitivity across assays. Whether these differences are due to differing performance of test assays or due to methodological issues in study design is difficult to ascertain. The differences are large enough to suggest that test assays should not be treated as equivalent.

All reported 2×2 data (online supplementary 9) for the clinical questions IBD versus IBS, IBD versus non-IBD and organic versus non-organic disease were considered in the meta-analysis investigating the effects of the different clinical questions on summary estimates of test accuracy under the assumption that the test is generic. Twenty-eight studies were included at the common threshold of 50 µg/g.^{10 24–50} A total of 10 studies^{30 32 34 36 37 43 45–48} contributed information to more than one category. The ROC plot in figure 6 summarises the summary estimates for the comparison of the three clinical questions and the pairs of sensitivity and FP rates for the contributing studies. Table 2 shows that specificity for IBD versus non-IBD (0.67, 95% CI 0.58 to 0.75) is lower than for IBD versus IBS (0.76, 95% CI 0.66 to 0.84), acknowledging that CIs overlap, suggesting that broadening the definition of the non-target condition may produce more FPs as other non-IBD intestinal conditions are included. On the other hand, sensitivity for organic disease (0.80, 95% CI 0.76 to 0.84) is lower than for IBD (0.97, 95% CI 0.91 to 0.99) suggesting that some organic disease will be missed with FC testing as organic conditions, including adenomas and diverticulosis, are not typically associated

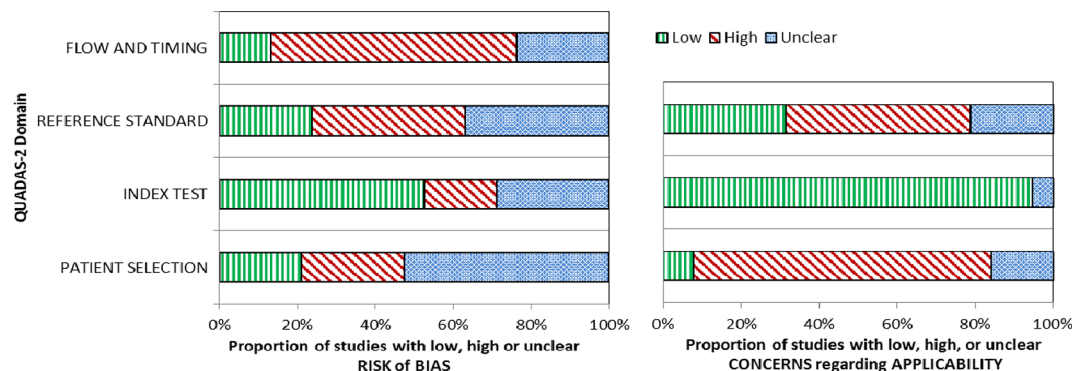


Figure 3 Overview of risk of bias and applicability concern of included studies according to QUADAS-2.¹⁹ QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2.

with inflammation. Considering different definitions of the clinical questions results in greater variation in sensitivity than specificity meaning that widening the target condition had a greater impact on test accuracy than changing the non-target condition (IBS vs non-IBD vs non-organic disease).

Meta-analysis at 100 µg/g threshold

Eleven studies^{30 32 35 43 45–48 51–53} reported test accuracy at a threshold of 100 µg/g (online supplementary 10 for 2×2 data). Of these, five contributed to more than one category of clinical questions. As expected raising the threshold to 100 µg/g reduces sensitivity and increases specificity for all clinical questions (table 2). Insufficient numbers of studies at the 100 µg/g were available to explore assay type.

Meta-regression and exploratory sensitivity analysis

Results from separate meta-analyses and from meta-regression analyses with equal variances showed similar results (table 2 and online supplementary 3).

Ten studies reported results for multiple clinical questions and eight studies compared multiple tests. Our

approach of randomly selecting one test or one clinical question per study for meta-analysis uses only one set of evidence per study. Choosing a different set might have resulted in different outcomes and conclusions as the meta-analyses suggest that it might not be appropriate to consider tests and questions alike. In an attempt to capture this variation, figure 7 displays results of 25 000 meta-analyses of 28 studies (online supplementary 11 for 2×2 data) picking one outcome per study at random for each round of meta-analysis. The results show overall high sensitivity and specificity irrespective of test assay and clinical question (Panel A) with slightly greater variation in sensitivity than specificity (Panel B). Median sensitivity and specificity of the 25 000 analyses and ranges were 0.9 (min 0.85, max 0.94) and 0.76 (min 0.73, max 0.79), respectively.

Primary versus secondary care

Five of the eligible studies were from primary care settings.^{35 44 47 48 53} The studies were too heterogeneous for meta-analysis. In an attempt to compare test performance of FC testing in primary care with secondary

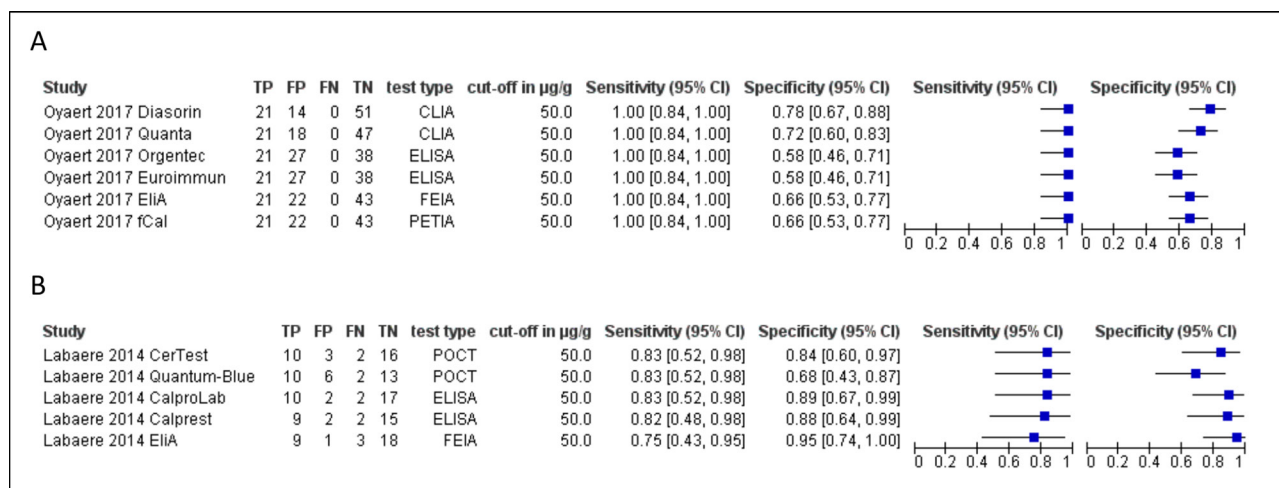


Figure 4 Forest plot of (A) Oyaert *et al*²⁴ comparing six FC tests and (B) Labaere *et al*²⁵ comparing five FC tests at cut-off 50 µg/g. CLIA, chemiluminescent immunoassay; FC, faecal calprotectin; FEIA, fluorescence enzyme immunoassay; FN, false negative; FP, false positive; PETIA, particle-enhanced turbidimetric immunoassay; POCT, point-of-care test; TP, true positive; TN, true negative.

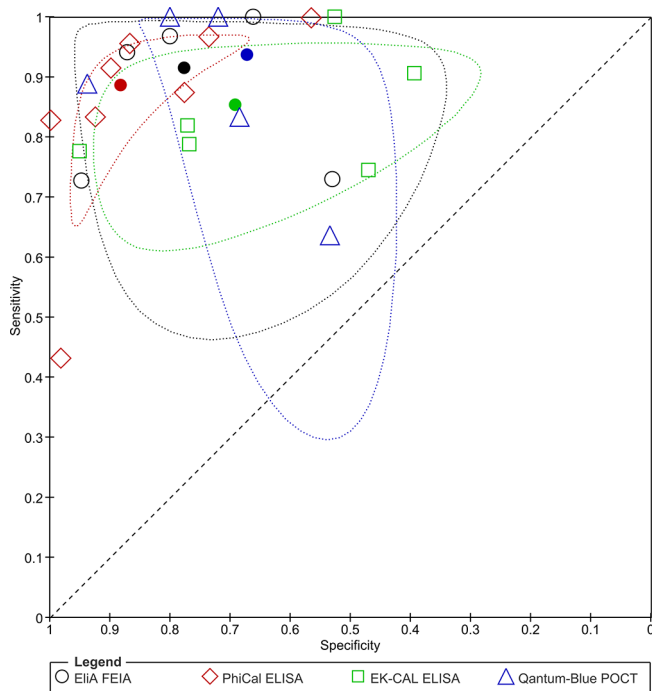


Figure 5 Receiver operating characteristic plot of sensitivity and specificity by test (filled shapes present summary estimates with 95% confidence region). FEIA, fluorescence enzyme immunoassay; POCT, point-of-care test.

care only one study qualified.⁴⁷ The study evaluated the routine use of FC testing in primary care and reported lower sensitivity of 0.72 but comparable specificity of 0.65 for differentiating IBD from non-IBD at a threshold of 50 µg/g when compared with our pooled estimates across different test assays including all settings (sensitivity 0.95, 95% CI 0.88 to 0.98 and specificity 0.67, 95% CI 0.58 to 0.75). However, indication for testing and the place of the FC test in the patient pathway were unclear in this study and the study may suffer from differential verification bias due to concerns over the reference standard. The authors explained the low sensitivity with the fact that general practitioners may have referred patients with high suspicion of IBD without testing. Another reasonable explanation could be the greater number of apparently milder IBD cases in primary care in which FC testing is falsely negative, reducing sensitivity in this setting and demonstrating the impact of the different spectrum of disease on test accuracy estimates.⁵⁴

DISCUSSION

None of the 38 included studies sufficiently addressed the review question of test accuracy of FC testing in a primary care pathway. The studies recruited patients with a different spectrum of disease from that anticipated for primary care, focused on disease groups which were broader than the intended IBD group or did not verify all patients with the preferred reference standard. Furthermore, the evidence was highly heterogeneous. Studies varied in thresholds and assays used, as well as

in target conditions and settings. Eighteen different test assays were studied, although similar they could not be considered generic for meta-analysis. Clinical questions were analysed as three distinct categories which were related but which could not be pooled for meta-analysis. A number of studies compared multiple tests and/or explored different clinical questions. Picking one test and one clinical question per study for meta-analysis was unjustifiable. We favoured an innovative approach which allowed us to explore the whole breadth of evidence and showed that irrespective of clinical question and test assay, test accuracy of FC testing was high. However, the variation in specificity could translate into considerable uncertainty of FPs when scaled up to population testing, while the variation in sensitivity reflects the limited value of FC testing for other organic conditions.

The review has a number of limitations. Due to heterogeneity, we decided not to pool across tests and clinical questions. Instead, we conducted exploratory analyses which demonstrated that the test assay and clinical question may affect the summary sensitivity and specificity by as much as 9%. We were also unable to compare primary and secondary care studies due to the small number of heterogeneous studies in primary care. The meta-regression analyses assumed equal variances between categories of tests and clinical questions and this was supported by a statistical comparison. However, the power of this test is dependent on the number of studies per subgroup and it is possible that this assumption could be challenged with a larger data set. Finally, the categorisation into different clinical questions was subjective because the disease categories are ill defined and studies' definitions of conditions and groups of conditions varied.

This review is broadly in line with the approaches and interpretation of the evidence of the previous review.¹⁸ However, Waugh *et al* reported pooled estimates of sensitivity (0.93, 95% CI 0.83 to 0.97) and specificity (0.94, 95% CI 0.73 to 0.99) across tests for the differential diagnosis of IBD versus IBS for five secondary care studies and this was the basis for national decisions to introduce FC testing in primary care. When compared with our equivalent analysis of 11 studies, we found comparable sensitivity (0.97, 95% CI 0.91 to 0.99) but considerably lower specificity (0.76, 95% CI 0.66 to 0.84) suggesting that FPs might be more of a concern than previously concluded. Thus, the predicted reduction in colonoscopies and subsequent cost savings may not be realised as a result of introducing FC testing into the primary care pathway in the UK.¹⁸ Furthermore, this review found enough disagreement between tests to caution against treating tests as equivalent. Issues in homogenisation, dilutions and extraction prior to analysis as well as lack of standardisation of different assays contribute to these differences. Recommended cut-off values would have to be determined locally until these issues are resolved.²⁴

None of the primary care studies assessed FC testing for the differential diagnosis of IBD versus IBS. The comparison of IBD versus non-IBD might reflect the

Table 2 Comparison of results of separate MAs and meta-regression with test type and clinical question as covariates at two thresholds (numbers in brackets are numbers of studies meta-analysed)

Method	PhiCal (n=8)		EK-CAL (n=6)		Quantum-Blue (n=5)		ELiA (n=5)	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Separate MA for each test	0.89 (0.76 to 0.95)	0.88 (0.77 to 0.94)	0.85 (0.75 to 0.92)	0.69 (0.47 to 0.85)	0.94 (0.75 to 0.90)	0.67 (0.56 to 0.76)	0.92 (0.78 to 0.97)	0.78 (0.60 to 0.89)
Meta-regression (test) with equal variances	0.89 (0.77 to 0.95)	0.89 (0.79 to 0.95)	0.87 (0.71 to 0.95)	0.69 (0.49 to 0.84)	0.92 (0.78 to 0.98)	0.73 (0.50 to 0.88)	0.92 (0.77 to 0.97)	0.78 (0.57 to 0.91)

Comparison of clinical questions at 50 µg/g threshold						
Method	IBD versus IBS (n=11)		IBD versus non-IBD (n=14)		Organic versus non-organic (n=15)	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Separate MA for each clinical question	0.97 (0.91 to 0.99)	0.76 (0.66 to 0.84)	0.95 (0.88 to 0.98)	0.67 (0.58 to 0.75)	0.80 (0.76 to 0.84)	0.76 (0.66 to 0.83)
Meta-regression (clinical question) with equal variances	0.96 (0.92 to 0.98)	0.77 (0.66 to 0.85)	0.94 (0.90 to 0.97)	0.67 (0.57 to 0.76)	0.81 (0.74 to 0.86)	0.75 (0.67 to 0.82)

Comparison of clinical questions at 100 µg/g threshold						
Method	IBD versus IBS (5)		IBD versus non-IBD (5)		Organic versus non-organic (7)	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Separate MA for each clinical question	0.92 (0.85 to 0.96)	0.86 (0.82 to 0.89)	0.72 (0.63 to 0.80)	0.82 (0.78 to 0.86)	0.67 (0.56 to 0.76)	0.87 (0.84 to 0.90)
Meta-regression (clinical question) with equal variances	0.92 (0.85 to 0.96)	0.86 (0.81 to 0.89)	0.73 (0.60 to 0.82)	0.82 (0.77 to 0.86)	0.67 (0.58 to 0.75)	0.87 (0.84 to 0.89)

IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; MA, meta-analysis.

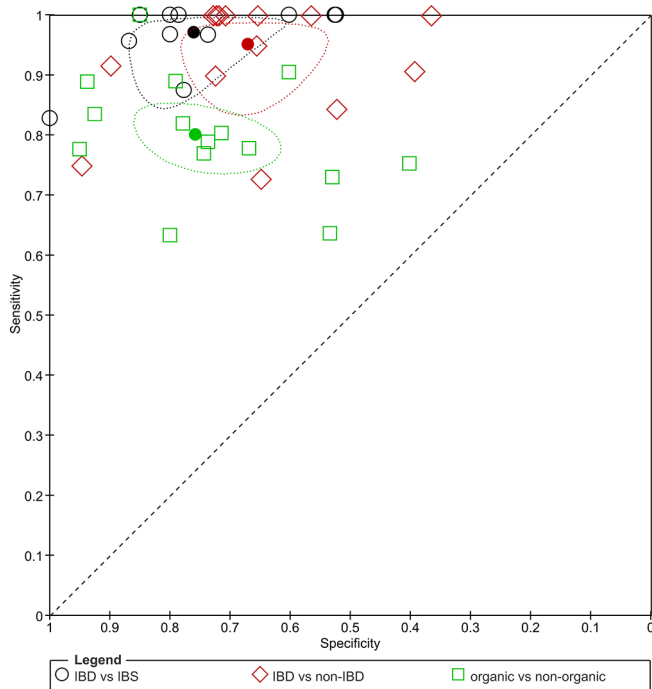


Figure 6 Receiver operating characteristic plot of sensitivity and specificity by clinical question (filled shapes present summary estimates with 95% confidence region). IBD, inflammatory bowel disease; IBS, irritable bowel syndrome.

clinical situation for FC testing more accurately, but produces more FPs as FC levels might be raised in diverticular disease, coeliac disease, rectal adenocarcinoma, non-specific inflammation, and others.^{35 53} The meaning of the additional FPs due to non-IBD inflammatory conditions is debatable in the context of clinical practice where incidental findings of true disease would not be

classified as an FP test outcome but would prompt further investigations.

NICE recently endorsed a cut-off of 100 µg/g for use in primary care in England and Wales, based on a primary care study which showed a 43% reduction in FC test positives compared with a threshold of 50 µg/g.⁴⁷ We investigated test accuracy at this cut-off including all settings and demonstrated an increase in specificity by at least 10%, while the magnitude of the resulting decrease in sensitivity was more uncertain.

Referral of FC tested patients was consistent in three UK primary care studies (41%,⁵³ 42%³⁵ and 48%⁴⁷ with over 25% of FC negative patients referred. This raises concerns about the impact of FC testing on colonoscopy rates as a considerable proportion of referred patients with negative FC levels were further investigated. Inappropriate use of FC testing as a screening test rather than to rule out IBD has been proposed as an explanation.⁵⁵ However, the number of investigations of referred FC negatives varied greatly among the three studies from 19%⁴⁷ to 46%³⁵ and even 71%⁵³ suggesting perhaps unsurprisingly, that the number of potentially unnecessary colonoscopies may be at least as dependent on secondary care decision-making as on the availability of FC testing in primary care.

The applicability of the spectrum of secondary care patients to a primary care setting is questionable even if studies investigated referred patients. First, general practitioners might have tested more patients than they referred and, second, patients referred came from different pathways including cancer pathways. Consequently, the patient population of the included studies resembled a continuum of primary to secondary care patients of which only some might plausibly resemble the primary care spectrum. Additionally, primary care populations may be different in different countries, it would,

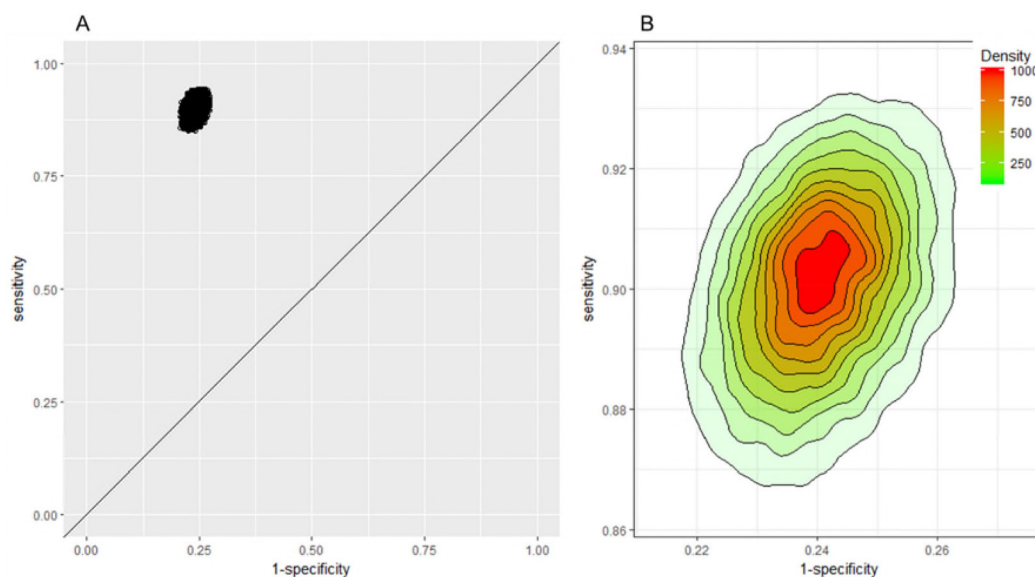


Figure 7 Pairs of sensitivity and specificity of 25 000 meta-analyses of 28 studies picking one outcome per study at random for each round of meta-analysis at a threshold of 50 µg/g. (A) Receiver operating characteristic plot, (B) 2D density plot scaled to illustrate spread and density (where the highest density represents a probability of 1000/25 000). 2D, two dimensions.

therefore, be useful to identify those studies which are relevant to primary care irrespective of the actual study setting, for instance, using tailored meta-analysis.^{56 57} Furthermore, the statistical validity of any summary estimates could be evaluated using a recently developed cross-validation technique.⁵⁸ Only five studies recruited subjects from primary care populations and analysed the general practitioners' decisions on whether to test and whether to refer.^{35 44 47 48 53} We found lower sensitivity for detection of disease in these studies compared with meta-analyses of all studies.

Despite the publication of primary care studies since the Waugh *et al* review,¹⁸ we are still lacking evidence on the defined role of FC testing in the primary care pathway for the detection of IBD. This is a concern given that it is prescribed by local and national guidance in the UK and the roll-out of a national algorithm endorsed by NICE, the National Health Service England Chief Scientific Officer and the British Society of Gastroenterologists is imminent to improve spread and adoption of FC testing nationally.⁵⁹ The decision to introduce FC testing in primary care might have been based on overoptimistic assessments of potential test accuracy in this setting. Evidence for the test accuracy of FC testing in primary care which considers the revised cut-off of 100 µg/g is needed before any recommendations for use in primary care can be made. Potentially, test accuracy studies in primary care are hampered because verification of FC test outcomes with colonoscopy may not be feasible in all tested patients. Therefore, other reference standards for assessing the performance of FC testing in primary care are also needed.

Acknowledgements We would like to thank the patient advisory group of the project 'What is the role of faecal calprotectin testing in primary care' for their input.

Contributors KF and HF undertook the review. KF and BHW carried out the analysis. KF, BHW, ST-P and AC contributed to the interpretation of the findings. KF drafted the manuscript. All authors critically revised the manuscript and approved the final version. KF takes responsibility for the integrity and accuracy of the data analysis. KF acts as guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding This study was funded by the NIHR (NIHR Doctoral Research Fellowship, KF, DRF-2016- 09-038).

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funder had no role in the study design, data collection, data analysis and interpretation, writing of the report or the decision to submit of publication.

Competing interests KF is funded by the NIHR through a doctoral research fellowship. ST-P and AC are supported by the NIHR CLAHRC West Midlands initiative. ST-P is funded by the NIHR through a career development fellowship. BHW is funded by an MRC Clinician Scientist Fellowship award (MR/N007999/1).

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data relevant to the study are included in the article or uploaded as supplementary information.

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: <https://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. Ungaro R, Mehandru S, Allen PB, *et al*. Ulcerative colitis. *Lancet* 2017;389:1756–70.
2. Torres J, Mehandru S, Colombel J-F, *et al*. Crohn's disease. *The Lancet* 2017;389:1741–55.
3. Frolkis AD, Dykeman J, Negrón ME, *et al*. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology* 2013;145:996–1006.
4. Dahlhamer JM, Zammitti EP, Ward BW, *et al*. Prevalence of inflammatory bowel disease among adults aged ≥18 years - United States, 2015. *MMWR Morb Mortal Wkly Rep* 2016;65:1166–9.
5. Burisch J, Jess T, Martinato M, *et al*. The burden of inflammatory bowel disease in Europe. *J Crohns Colitis* 2013;7:322–37.
6. Rubin GP, Hungin AP, Kelly PJ, *et al*. Inflammatory bowel disease: epidemiology and management in an English general practice population. *Aliment Pharmacol Ther* 2000;14:1553–9.
7. Molodecky NA, Soon IS, Rabi DM, *et al*. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012;142:46–54. quiz e30.
8. van den Heuvel-Janssen HA, Borghouts JA, Muris JW, *et al*. Chronic non-specific abdominal complaints in general practice: a prospective study on management, patient health status and course of complaints. *BMC Fam Pract* 2006;7:12.
9. Janssen HA, Borghouts JA, Muris JW, *et al*. Health status and management of chronic non-specific abdominal complaints in general practice. *Br J Gen Pract* 2000;50:375–9.
10. Kok L, Elias SG, Witteman BJ, *et al*. Diagnostic accuracy of point-of-care fecal calprotectin and immunochemical occult blood tests for diagnosis of organic bowel disease in primary care: the Cost-Effectiveness of a Decision Rule for Abdominal Complaints in Primary Care (CEDAR) study. *Clin Chem* 2012;58:989–98.
11. Eskeland SL, Dalén E, Sponheim J, *et al*. European panel on the appropriateness of gastrointestinal endoscopy II guidelines help in selecting and prioritizing patients referred to colonoscopy--a quality control study. *Scand J Gastroenterol* 2014;49:492–500.
12. Konikoff MR, Denson LA. Role of fecal calprotectin as a biomarker of intestinal inflammation in inflammatory bowel disease. *Inflamm Bowel Dis* 2006;12:524–34.
13. Lichtenstein GR, Loftus EV, Isaacs KL, *et al*. ACG clinical guideline: management of Crohn's Disease in Adults. *Am J Gastroenterol* 2018;113:481–517.
14. The Gastroenterological Society of Australia. Australian guidelines for general practitioners and physicians: inflammatory bowel disease, fourth edition. 2017 http://cart.gesa.org.au/membes/files/Clinical%20Guidelines%20and%20Updates/2017_IBD_guidelines_DRAFT_for_consultation_20102017.pdf (cited 08 Oct 2018).
15. Preiß J, Bokemeyer B, Buhr H, *et al*. Updated german clinical practice guideline on "Diagnosis and treatment of Crohn's disease". 2014 https://www.dgvs.de/wp-content/uploads/2016/11/DGVS_Empfehlung_fuer_Diagnostik_und_Therapie_des_Morbus_Crohn.pdf (cited 08 Oct 2018).
16. Gomollón F, Dignass A, Annesse V, *et al*. 3rd European Evidence-based Consensus on the Diagnosis and Management of Crohn's Disease 2016: Part 1: Diagnosis and Medical Management. *J Crohns Colitis* 2017;11:3–25.
17. National Institute for Health and Care Excellence (NICE). Faecal calprotectin diagnostic tests for inflammatory diseases of the bowel: NICE diagnostics guidance [DG11]. 2013 <https://www.nice.org.uk/guidance/dg11> (cited 17 Dec 2015).
18. Waugh N, Cummins E, Royle P, *et al*. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technol Assess* 2013;17:1–211.
19. Whiting PF, Rutjes AW, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [computer program]. 2017 <https://www.R-project.org/>.
21. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59:1331–2. author reply 2–3.
22. Partlett C, Takwoingi Y. Meta-analysis of test accuracy studies in R: a summary of userwritten programs and step-by-step guide to using

- glmer. Version 1.0. 2016 <http://methods.cochrane.org/sdt/> (cited 30 Jan 2018).
23. McInnes MDF, Moher D, Thombs BD, *et al.* Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018;319:388–96.
 24. Oyaert M, Boel A, Jacobs J, *et al.* Analytical performance and diagnostic accuracy of six different faecal calprotectin assays in inflammatory bowel disease. *Clin Chem Lab Med* 2017;55:21–21.
 25. Labaere D, Smismans A, Van Olmen A, *et al.* Comparison of six different calprotectin assays for the assessment of inflammatory bowel disease. *United European Gastroenterol J* 2014;2:30–7.
 26. Turvill J. High negative predictive value of a normal faecal calprotectin in patients with symptomatic intestinal disease. *Frontline Gastroenterol* 2012;3:21–8.
 27. Schoepfer AM, Trummel M, Seeholzer P, *et al.* Discriminating IBD from IBS: comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies. *Inflamm Bowel Dis* 2008;14:32–9.
 28. Otten CM, Kok L, Witteman BJ, *et al.* Diagnostic performance of rapid tests for detection of fecal calprotectin and lactoferrin and their ability to discriminate inflammatory from irritable bowel syndrome. *Clin Chem Lab Med* 2008;46:1275–80.
 29. Li XG, Lu YM, Gu F, *et al.* [Fecal calprotectin in differential diagnosis of irritable bowel syndrome]. *Beijing Da Xue Xue Bao Yi Xue Ban* 2006;38:310–3.
 30. El-Badry A, Sedrak H, Rashed L. Faecal calprotectin in differentiating between functional and organic bowel diseases. *Arab Journal of Gastroenterology* 2010;11:70–3.
 31. Burri E, Manz M, Rothen C, *et al.* Monoclonal antibody testing for fecal calprotectin is superior to polyclonal testing of fecal calprotectin and lactoferrin to identify organic intestinal disease in patients with abdominal discomfort. *Clin Chim Acta* 2013;416:41–7.
 32. Kennedy NA, Clark A, Walkden A, *et al.* Clinical utility and diagnostic accuracy of faecal calprotectin for IBD at first presentation to gastroenterology services in adults aged 16–50 years. *J Crohns Colitis* 2015;9:41–9.
 33. Dhaliwal A, Zeino Z, Tomkins C, *et al.* Utility of faecal calprotectin in inflammatory bowel disease (IBD): what cut-offs should we apply? *Frontline Gastroenterol* 2015;6:14–19.
 34. Damms A, Bischoff SC. Validation and clinical significance of a new calprotectin rapid test for the diagnosis of gastrointestinal diseases. *Int J Colorectal Dis* 2008;23:985–92.
 35. Pavlidis P, Chedgy FJ, Tibble JA. Diagnostic accuracy and clinical application of faecal calprotectin in adult patients presenting with gastrointestinal symptoms in primary care. *Scand J Gastroenterol* 2013;48:1048–54.
 36. Mowat C, Digby J, Strachan JA, *et al.* Faecal haemoglobin and faecal calprotectin as indicators of bowel disease in patients presenting to primary care with bowel symptoms. *Gut* 2016;65:1463–9.
 37. De Sloovere MMW, De Smet D, Baert FJ, *et al.* Analytical and diagnostic performance of two automated fecal calprotectin immunoassays for detection of inflammatory bowel disease. *Clin Chem Lab Med* 2017;55:11:11.
 38. Livzan MA, Lyalukova EA, Nechaeva G, *et al.* Calprotectin, transferrin, hemoglobin) in complex diagnosis of diseases of intestines]. *Eksperimental'Naia i Klinicheskaia Gastroenterologija* 2015;3:34–8.
 39. Oyaert M, Trouvé C, Baert F, *et al.* Comparison of two immunoassays for measurement of faecal calprotectin in detection of inflammatory bowel disease: (pre)-analytical and diagnostic performance characteristics. *Clin Chem Lab Med* 2014;52:391–7.
 40. Jang HW, Kim HS, Park SJ, *et al.* Accuracy of three different fecal calprotectin tests in the diagnosis of inflammatory bowel disease. *Intest Res* 2016;14:305–13.
 41. Sostres C, Lue A, Barra Pardos MV, *et al.* Combination of FOBT and fecal calprotectin may be useful for reducing unnecessary colonoscopies in symptomatic patients. *United European Gastroenterology Journal* 2017;5(Suppl 1):A779.
 42. Alrubaiy L, Malik A, Rees I, *et al.* Usefulness of fecal calprotectin in clinical practice in a district general hospital. *Inflamm Bowel Dis* 2012;18:S53–S54.
 43. Banerjee A, Srinivas M, Eyre R, *et al.* Faecal calprotectin for differentiating between irritable bowel syndrome and inflammatory bowel disease: a useful screen in daily gastroenterology practice. *Frontline Gastroenterol* 2015;6:20–6.
 44. Boyd A, Peplow C, Dhaliwal A, *et al.* Primary care faecal calprotectin pathway - A three year analysis. *Colorectal Disease* 2016;18:27.
 45. Carroccio A, Iacono G, Cottone M, *et al.* Diagnostic accuracy of fecal calprotectin assay in distinguishing organic causes of chronic diarrhea from irritable bowel syndrome: a prospective study in adults and children. *Clin Chem* 2003;49(6 Pt 1):861–7.
 46. Caviglia GP, Pantaleoni S, Touscoz GA, *et al.* Fecal calprotectin is an effective diagnostic tool that differentiates inflammatory from functional intestinal disorders. *Scand J Gastroenterol* 2014;49:1419–24.
 47. Conroy S, Hale MF, Cross SS, *et al.* Unrestricted faecal calprotectin testing performs poorly in the diagnosis of inflammatory bowel disease in patients in primary care. *J Clin Pathol* 2018;71.
 48. Högborg C, Karling P, Rutegård J, *et al.* Diagnosing colorectal cancer and inflammatory bowel disease in primary care: The usefulness of tests for faecal haemoglobin, faecal calprotectin, anaemia and iron deficiency. A prospective study. *Scand J Gastroenterol* 2017;52:69–75.
 49. Tibble JA, Sigthorsson G, Foster R, *et al.* Use of surrogate markers of inflammation and Rome criteria to distinguish organic from nonorganic intestinal disease. *Gastroenterology* 2002;123:450–60.
 50. Tan M, Subramaniam Y, Cheong C, *et al.* The role of faecal calprotectin in a primary care referral pathway in the UK. *United European Gastroenterology Journal* 2016;4(5 Suppl 1):A443.
 51. Limburg PJ, Ahlquist DA, Sandborn WJ, *et al.* Fecal calprotectin levels predict colorectal inflammation among patients with chronic diarrhea referred for colonoscopy. *Am J Gastroenterol* 2000;95:2831–7.
 52. Rosenfeld G, Greenup AJ, Round A, *et al.* FOCUS: Future of fecal calprotectin utility study in inflammatory bowel disease. *World J Gastroenterol* 2016;22:8211–8.
 53. Turvill J, O'Connell S, Brooks A, *et al.* Evaluation of a faecal calprotectin care pathway for use in primary care. *Prim Health Care Res Dev* 2016;17:428–36.
 54. Willis BH. Spectrum bias--why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract* 2008;25:390–6.
 55. Ng S, Box B, Khan K. Faecal calprotectin—a good test used badly. *Colorectal Disease* 2017;19(S4):64.
 56. Willis BH, Hyde CJ. Estimating a test's accuracy using tailored meta-analysis—How setting-specific data may aid study selection. *J Clin Epidemiol* 2014;67:538–46.
 57. Willis BH, Hyde CJ. What is the test's accuracy in my practice population? Tailored meta-analysis provides a plausible estimate. *J Clin Epidemiol* 2015;68:847–54.
 58. Willis BH, Riley RD. Measuring the statistical validity of summary meta-analysis and meta-regression results for use in clinical practice. *Stat Med* 2017;36:3283–301.
 59. Commissioning PC. Faecal calprotectin in primary care as a decision diagnostic for inflammatory bowel disease and irritable bowel syndrome. 2018 <https://www.pcc-cic.org.uk/article/updated-guidance-help-gps-easily-diagnose-serious-bowel-conditions> (cited 09 Jul 2018).