



**Effectiveness of faecal calprotectin to identify patients with organic gastrointestinal diseases: a systematic review and diagnostic meta-analysis**

|                  |  |
|------------------|--|
| Journal:         | <i>Medical Journal of Australia</i>  |
| Manuscript ID    | mja18.01304.R2   |
| Manuscript Type: | Systematic review  |
| Keywords:        | Inflammatory bowel diseases < Digestive system diseases, Irritable bowel syndrome < Digestive system diseases, Mass screening < Environment and public health, Diagnostic tests and procedures < General medicine, Systematic review < Statistics, epidemiology and research design, Meta-analysis < Statistics, epidemiology and research design, Digestive system diseases |
|                  |  |

SCHOLARONE™  
Manuscripts

**Effectiveness of faecal calprotectin to identify patients with organic gastrointestinal diseases: *a systematic review and diagnostic meta-analysis***

**Abstract**

**Objectives:** Assessing patients with lower gastrointestinal symptoms can be challenging. Data examining the clinical effectiveness of faecal calprotectin (FC) testing to distinguish between organic gastrointestinal diseases (OGIDs), such as Inflammatory Bowel Disease (IBD), and Functional Gastrointestinal Disorders (FGIDs) were systematically reviewed.

**Study design:** Journal articles that assessed the accuracy of FC to differentiate between IBD and/or OGIDs and FGIDs were reviewed. Data on methodology and the characteristics of the diagnostic test were collected. Study quality was assessed using QUADAS-2, an evidence-based quality assessment tool for diagnostic accuracy studies.

**Data sources:** MEDLINE and EMBASE were searched for relevant literature published from 1998 to August 2018. A total of 18 studies were included.

**Data synthesis:** When distinguishing patients with OGIDs (including IBD) from FGIDs, FC had a sensitivity of 81% (95% CI=74-86%) a false positive rate of 19% (95% CI=12%-29%) (FPR= 1-Specificity), and an area under the curve (AUC) of 0.87. When distinguishing IBD from non-IBD, FC had a sensitivity of 88% (95% CI=80-93%), false positive rate of 28% (95% CI=18-41%), and an AUC of 0.89. Assuming a prevalence of 1% the positive predictive value (PPV) is 4.2% and the negative predictive value (NPV) is 100%. There was a non-significant ( $p=0.77$ ) difference in sensitivity and false-positive rates with a cut-off of 50µg/g ( $SE=3.7$ ) compared to 100µg/g ( $SE=4.0$ ).

**Conclusions:** FC is a clinically useful test to distinguish OGIDs (including IBD) from FGIDs, and its wider implementation would support more timely diagnosis, better direct colonoscopy resources (fewer patients with FGIDs undergoing colonoscopy) and yield cost savings.

**PROSPERO registration:** CRD42018105078

**TWEETS** (140 characters)

*Faecal calprotectin is highly effective at distinguishing between organic and functional gastrointestinal disease on meta-analysis.*

**Final word count:** 2924 words

## **Introduction**

Gastrointestinal symptoms account for 5-10% of primary care consultations and referral of all patients to specialist care is unnecessary.<sup>(1, 2)</sup> It can be difficult for general practitioners (GPs) and specialists to confidently distinguish between patients presenting with functional gastrointestinal disorders (FGIDs) as compared to organic gastrointestinal diseases (OGIDs) such as inflammatory bowel disease (IBD), colorectal cancer, polyps >1cm and diverticulitis, on symptoms alone.

FGIDs are chronic symptom-based disorders, with typical symptoms including abdominal pain, diarrhoea, bloating, constipation and/or alternating bowel habit. OGIDs can present with similar symptoms, so distinguishing them from FGIDs can be difficult. FGIDs are vastly more common than OGIDs with the most common FGID, Irritable Bowel Syndrome (IBS), affecting approximately 10-20% of the population, presenting between the ages of 20 and 40.<sup>(3, 4)</sup> Once a diagnosis of FGID is established further diagnostic tests are not required and the vast majority of patients can be managed successfully in the primary care setting.<sup>(5)</sup> In contrast, various OGIDs result in inflammation of the colon, and these patients are best assessed with specialist referral and further evaluation, often with colonoscopy and/or cross-sectional imaging. In the 20-40 years age group the most common OGID is IBD, comprising ulcerative colitis (UC) and Crohn's disease (CD), which affects 0.4% of those aged <50.<sup>(6)</sup>

In young adults (20-40 years of age), FC is most often used to differentiate between IBS and IBD. However, in children FC can be elevated due to a variety of conditions. Usually, it is not recommended to use FC alone as a screening tool in people older than 50, due to their higher risk of having OGIDs including neoplasms.<sup>(7-12)</sup>

The high prevalence of IBS in conjunction with the low incidence of organic disease in young adult patients have led to advocacy for a positive diagnosis of FGID based on medical history with minimal associated investigations, including faecal calprotectin (FC) testing to rule out OGIDs.<sup>(7-12)</sup> Despite primary care diagnosis and management of FGIDs being the recommendation, there are large numbers of referral of patients with likely IBS leading to unnecessary further assessment, such as colonoscopy, which are often normal.<sup>(13-15)</sup>

Inflammation is the key characteristic that differentiates OGIDs from FGIDs. FC has been identified as a non-invasive predictive test with high sensitivity and specificity for OGIDs

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

(including IBD) and IBD respectively.<sup>(16,17)</sup> Furthermore, it has been demonstrated that FC testing is superior to other biomarkers, such as CRP and ESR.<sup>(16)</sup> Despite these reported benefits, there is a scarcity of literature that provides a comprehensive evaluation of the clinical effectiveness of FC.

We have reviewed the clinical effectiveness of FC in distinguishing OGIDs and IBD from FGIDs compared to a specialist diagnosis based on clinical, laboratory, imaging and/or endoscopic findings. We present data on the effectiveness of FC in distinguishing between [1] patients with OGIDs who require further investigation and those with FGIDs, and [2] patients with IBD and non-IBD, which is a common clinical question.

**Methods**

This study has been conducted in accordance with the Preferred Items for Systematic Reviews and Meta-analysis [PRISMA] guidelines.<sup>(18)</sup>

**Search strategy**

A search of the English medical literature was conducted using MEDLINE and EMBASE from January 1980 to June 2018 that examined the clinical effectiveness of FC in distinguishing OGIDs (including IBD) and IBD alone from FGIDs. The following key words were included alone or in combination: ‘Crohn disease’, ‘ulcerative colitis’, ‘inflammatory bowel disease’, ‘irritable colon’, ‘digestive system function disorder’, ‘IBS’, ‘IBD’, ‘faecal or fecal’, ‘calprotectin’, ‘general practitioner’, ‘GP’, ‘primary care provider’ and ‘primary medical care’. In addition, we examined the references of the screened articles to identify additional studies.

**Study selection**

Retrieved studies were assessed by two reviewers [S.C. and J.B.] for eligibility for full text review. Non-relevant studies were excluded based on title, abstract and study type [i.e. review]. Subsequently, the selected full texts were assessed by two reviewers [Y.A. and D.P.]. Study inclusion criteria included: [1] study of FC as a diagnostic test in both primary care and outpatient hospital setting; [2] comparison with a reference test/standard such as colonoscopy or cross-sectional imaging; [3] paediatric and/or adult patients; [4] patients presenting with lower gastrointestinal symptoms; [5] FC using the standard ELISA-based method; [6] FC with

cut-offs of 50 or 100  $\mu\text{g/g}$  of stool; [7] articles written in English; [8] full text available [i.e. no abstracts].

We excluded studies in patients with red flag signs or symptoms, such as positive faecal occult blood test, overt rectal bleeding, iron deficiency anaemia, abdominal or rectal masses or a family history of bowel cancer that would require endoscopic evaluation.

### Data extraction

Pre-specified data were collated, including: study design, aim, number of patients, patient demographics, study setting, inclusion and exclusion criteria, reference test and index used. Additionally, the following data were collected: the FC assays used, cut-offs, true positive (TP), false positive (FP), true negative (TN), false negative (FN), sensitivity (Sn), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV) and accuracy with reference test.

### Statistical analysis and data synthesis

We followed the guidance from Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Chapter 10) in preparing this meta-analysis.<sup>(19)</sup>

Diagnostic two-by-two tables and diagnostic performance measures per research questions are presented. Forest plots were used to display sensitivity and specificity across studies, together with confidence intervals. When studies used both 50 and 100  $\mu\text{g/g}$  of stool cut-off values for the same patient, we only used the 50  $\mu\text{g/g}$  value for this analysis. Furthermore, when the included studies used the same patient results for multiple comparisons, we only used the results associated with FC used to distinguish OGIDs from FGIDs. Specifically, we only used the patient's results once per indication.

The diagnostic odds ratio (DOR) is a single measure of test accuracy that incorporates both sensitivity and specificity.<sup>(20)</sup> It is a useful summary statistic in meta-analyses because it can summarise overall test accuracy when thresholds differ across studies. Diagnostic odds ratios were summarised across studies using a random effects model following the approach of DerSimonian and Laird.<sup>(21)</sup> Summary estimates of sensitivity and specificity with 95% CI were calculated using bivariate analyses based upon the approach of Reitsma et al.<sup>(22)</sup>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The Reitsma model was also used to generate Summary Receiver Operating Characteristic Curves (SROC). The SROC summarises the overall test accuracy (true positives relative to false positives) across various thresholds based upon all relevant studies. We defined the False Positive Rate (FPR) as:  $FPR = (1 - \text{Specificity})$ . We present SROC curves per research question, together with point estimates of the true positive rates and false positive rates for each study, and the summary estimates (with 95% CI) from the bivariate meta-analysis. All analyses were performed using the mada package version 0.5.8 in the statistical software package R version 3.5.1.

**Quality Assessment**

All included studies were graded for methodological quality by two investigators [Y.A. and D.P.] with the revised tool for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2).<sup>(23)</sup> The QUADAS-2 tool is designed to assess the quality of diagnostic accuracy studies in four domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of the risk of bias and the applicability of the study results for the first three domains. A study was graded as high quality in the case of a low risk of bias in at least 6 out of the 7 subdomains. A study was graded as low quality in the case of a high risk or unclear risk in 4 or more subdomains. All other studies were graded as moderate quality. Any disagreements were resolved through discussion with a third reviewer, with the majority consensus used.

**Results**

**Study selection**

93 records were identified through database search and 22 records were identified through references. A total of 115 records were screened after removal of duplicates. After screening titles and abstracts, 30 potentially eligible studies were selected for full text review. After full text review, 18 records were eligible for inclusion in our analysis (Figure 1A).

**Grading of study quality**

The results of the QUADAS-2 assessment are shown in Figure 1B. Study quality was graded high in eleven studies, moderate in four studies, and low in three studies. Most concerns were raised in the subdomains regarding patient selection and reference standard. In some studies,

patient selection was not reflective of a true primary care population, while in others, there was lack of clarity about the reference standard test they had used to confirm the final diagnosis.

### Study characteristics

The relevant study characteristics are shown in Table 1.<sup>(24-41)</sup> Sixteen studies used a prospective and two studies a retrospective design. Seven studies were conducted in primary care and eleven studies in secondary care settings either through outpatient clinics or in endoscopy units. Most studies were performed in Europe in an adult population with a mean age range between 30 and 63; two studies also assessed a paediatric population with mean ages of 3.5 and 7. In all studies patients presented with chronic lower gastrointestinal symptoms, which were suggestive of either OGID (particularly IBD) or FGIDs. Colonoscopy was used as the reference standard in the majority of studies. The most commonly used enzyme-linked immunosorbent assay (ELISA) kit was EK-CAL ( $n=8$ ), manufactured by Buhlmann Laboratories (Schonenbuch, Switzerland). Cost-effectiveness of FC was evaluated in two studies.

The number of patients included in each study varied from 49 to 1005, with a median of 262 [IQR 399-111]. There was a total of 8150 patients included in the analyses (IBD vs non-IBD and OGIDs vs FGIDs), with 5431 associated with OGIDs vs FGIDs and 2719 associated with IBD vs non-IBD. However, results for 1887 of these patients were using a cut-off of 50  $\mu\text{g/g}$  and 100  $\mu\text{g/g}$ . When both values were reported, we only used the 50  $\mu\text{g/g}$  cut-off value. As such, our combined analysis included 6,263 patients.

There were 18 studies eligible for inclusion. However, many ( $n=7$ ) of these studies considered both diagnostic groups (OGIDs and/or IBD vs FGIDs), and both pathology reference ranges (50 $\mu\text{g/g}$  and 100 $\mu\text{g/g}$ ) ( $n=7$ ), with two studies including both an adult and paediatric population. As such, there were 26 separate comparisons. There were 16 studies that used FC as a diagnostic test for OGIDs, 10 studies that used FC to differentiate IBD from FGIDs, and 7 studies that considered both entities.

### Diagnostic performance of faecal calprotectin

We present data for two different comparisons: [1] FC used to distinguish OGIDs from FGIDs; [2] faecal calprotectin used to distinguish IBD (includes UC, CD and IBD-U) from non-IBD (includes FGIDs and other OGIDs).



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**OGIDs descriptive (forest plots) and bivariate analysis**

There were 16 studies (Supplement Table 1) included in the analysis for distinguishing OGIDs from FGIDs, with 14 including an adult population, 1 paediatric, and 1 mixed population. The adult population consisted of 5117 patients, with the remainder consisting of children ( $n=50$ ), and both adults and children ( $n=264$ ). The confidence intervals between the sensitivity and false-positive rates indicated limited variance in sensitivity (range=74-86%) and false-positive rates (range=12-29%).

When distinguishing OGIDs from FGIDs, FC had a sensitivity of 81% (95% CI=74-86%) and a false positive rate of 19% (95% CI=12%-29%), as detailed in Figure 2A. The log diagnostic odds ratio equalled 3.0 (CI=2.4, 3.6), with the area under the curve equalling 0.87, as detailed in Table 2.

**IBD descriptive (forest plot) and bivariate analysis**

There were 10 studies included in the analysis for IBD versus non-IBD (Supplement Table 2), with 9 including an adult population, and 1 mixed adult and paediatric population. The adult population consisted of 2455 patients, and the mixed study included 264 patients. The confidence intervals between the sensitivity and false-positive rates indicated limited variance in sensitivity (range=80-93%) and false-positive rates (range=18-41%), as detailed in Figure 2B.

When distinguishing IBD from non-IBD, FC had a sensitivity of 88% (95% CI=80-93%) and a false positive rate of 28% (95% CI=18-41%), as detailed in Figure 2B. The log diagnostic odds ratio equalled 3.2 (CI=2.5, 3.9), with the area under the curve equalling 0.89, as detailed in Table 2.

**Overall analysis and meta-regression to compare diagnostic accuracy cut-off points**

Since both patients with possible OGID or IBD require specialist referral and further investigation, most often a colonoscopy, to confirm the diagnosis, it is useful to consider both populations together. There were 19 studies included in the combined diagnostic analysis, with 17 including an adult population, 1 paediatric, and 1 mixed population. The adult population consisted of 5704 patients, with the remainder consisting of children ( $n=50$ ), and both adults and children ( $n=264$ ). The confidence intervals between the sensitivity and false-positive rates



indicated limited variance in sensitivity (range=73-85%) and false-positive rates (range=12-30%).

The overall sensitivity and false positive rate were 80% (95% CI=73-85%) and 19% (95% CI=14-30%) respectively, as detailed in Figure 2C. The log diagnostic odds ratio equalled 3.1 (CI=2.5, 3.6), with the area under the curve equalling 0.86, as detailed in Table 2.

Given the relatively high prevalence reported in the studies, we calculated the positive predictive value (PPV) and negative predictive value (NPV) using relevant population prevalence. The PPV and NPV at 1% prevalence equalled 4.2% and 100%, respectively. Whereas the PPV and NPV at 0.1% equalled 0.40% and 100%, respectively.

### Diagnostic accuracy between FC pathology ranges

There were 7 studies that used both FC pathology cut-off ranges. When comparing the diagnostic accuracy, there was a non-significant ( $p=0.77$ ) increase in sensitivity and decrease in false-positive rate between FC cut-offs of 50 $\mu$ g/g ( $SE=3.7$ ) and 100 $\mu$ g/g ( $SE=4.0$ ), as detailed in Figure 3A and 3B.

### Discussion

FC accurately reflects the degree of inflammation in the bowel and can be used to distinguish between OGIDs (including IBD) and FGIDs in patients presenting with lower gastrointestinal symptoms. When considering whether to perform an invasive test such as colonoscopy, the risk of harm of the procedure must be balanced against the risk of a missed diagnosis. Given the high NPV (100%) of FC testing in low prevalence populations, it is ideally suited for screening in the primary care setting, although further investigations are required for positive tests given the low PPV. The main benefit of using FC in primary care would be to confirm the clinical diagnosis of FGIDs by GPs, allowing patients to move into management phase and reduce the risk of a delayed diagnosis of OGIDs. This would greatly reduce the number of young adult patients without other alarm features being referred to specialist care, and reduce unnecessary and more expensive investigations such as cross-sectional imaging or colonoscopy.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Overall the sensitivity and false positive rate of FC testing using a cut-off of 50 µg/g in this screening patient population showed a sensitivity range of 74-86% and a false positive rate of 12-29% for distinguishing OGIDs versus FGIDs and a sensitivity range of 80-93% and a false positive rate of 18-41% for distinguishing IBD from non-IBD. Our SROC analysis showed that the overall sensitivity for distinguishing organic from functional gastrointestinal diseases was 81% with a false positive rate of 19% and for distinguishing IBD from non-IBD the sensitivity was 88% with a false positive rate of 28%. Previous studies have shown that using a cut-off of 50 µg/g results in a higher sensitivity and lower specificity, but our analysis shows that in studies examining a cut-off of 50 and 100 µg/g, there was a non-significant increase in sensitivity. Nonetheless, given the role of FC as a screening test for OGIDs we believe that a cut-off of 50 µg/g is appropriate.

Our analysis shows that FC testing is effective at identifying patients presenting with abdominal symptoms requiring further investigation. Our meta-analysis support guideline recommendations <sup>(7-11)</sup> that symptomatic patients with FC less than 50 µg/g can be confidently diagnosed with a FGID and managed appropriately and do not require further evaluation by a specialist or further testing.

The use of FC in the primary care setting allows the effective and appropriate use of resources and reduces unnecessary risk and cost to the patient and system. In patients with non-specific lower gastrointestinal symptoms, the use of FC reduces the risk of a missed diagnosis of OGIDs and ensures prompt specialist review. A recent Australian study found that FC alongside a routine panel of blood tests allowed the identification of patients who needed prompt review by a specialist, improving the safety of local triage practices.<sup>(42)</sup> FC as a negative predictor of OGIDs also enables an early, accurate and confident diagnosis of FGIDs to be made by GPs without need for gastroenterology referral or further testing. Enabling a confident FGID diagnosis to be made in primary care is likely to improve patient acceptance of the diagnosis and uptake of effective management strategies.<sup>(5)</sup> An accurate diagnosis of FGIDs reduces the number of referrals for more expensive, unnecessary and invasive investigations such as colonoscopy. An audit of colonoscopies performed in one Australian hospital found that 12% of colonoscopies were for patients with functional gastrointestinal symptoms,<sup>(43)</sup> whilst another Australian study found that 79% of patients seen in a public hospital diagnosed with IBS underwent an endoscopic procedure at an overall (includes indirect costs in addition to the

1  
2  
3 colonoscopy) cost of \$85 million AUD with an additional \$2.4 million AUD/year in  
4 Emergency Department presentations.<sup>(43, 44)</sup> The introduction of FC has consistently  
5 demonstrated significant cost savings as compared to colonoscopies,<sup>(43, 45)</sup> Reducing this  
6 referral burden of functional gastrointestinal symptoms to public health secondary care allows  
7 valuable clinical resources to be used more effectively and appropriately.  
8  
9

10  
11  
12  
13 Our analysis shows that FC is clinically effective in a primary care setting to distinguish OGIDs  
14 (including IBD) from FGIDs and would strongly encourage its wider implementation as a  
15 screening test in symptomatic patients to determine whether referral to specialist care is  
16 warranted in accordance with current Australian and international guidelines.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. Jones R. Primary care research and clinical practice: gastroenterology. *Postgrad Med J*. 2008;84(995):454-8.
2. Arasaradnam RP, Brown S, Forbes A, Fox MR, Hungin P, Kelman L, et al. Guidelines for the investigation of chronic diarrhoea in adults: British Society of Gastroenterology, 3rd edition. *Gut*. 2018;67(8):1380-99.
3. Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol*. 2016;1(2):133-46.
4. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol*. 2012;10(7):712-21 e4.
5. Linedale EC, Andrews JM. Diagnosis and management of irritable bowel syndrome: a guide for the generalist. *Med J Aust*. 2017;207(7):309-15.
6. PricewaterhouseCoopers Australia (PwC). Improving Inflammatory Bowel Disease care across Australia. 2013. <https://www.crohnsandcolitis.com.au/site/wp-content/uploads/PwC-report-2013.pdf> (accessed Sep 2018).
7. Gastroenterological Society of Australia (GESA) Digestive Health Foundation. Irritable Bowel Syndrome. 2010. <http://cart.gesa.org.au/membes/files/Consumer%20Information/IBS.pdf> (accessed Oct 2018).
8. American College of Gastroenterology Task Force on Irritable Bowel S, Brandt LJ, Chey WD, Foxx-Orenstein AE, Schiller LR, Schoenfeld PS, et al. An evidence-based position statement on the management of irritable bowel syndrome. *Am J Gastroenterol*. 2009;104 Suppl 1:S1-35.
9. NICE guidance [DG11]. Faecal calprotectin diagnostic tests for inflammatory disease of the bowel. 2013. <https://www.nice.org.uk/guidance/dg11> (accessed Oct 2018).
10. Fukudo S, Kaneko H, Akiho H, Inamori M, Endo Y, Okumura T, et al. Evidence-based clinical practice guidelines for irritable bowel syndrome. *J Gastroenterol*. 2015;50(1):11-30.
11. Shin JE, Jung HK, Lee TH, Jo Y, Lee H, Song KH, et al. Guidelines for the Diagnosis and Treatment of Chronic Functional Constipation in Korea, 2015 Revised Edition. *J Neurogastroenterol Motil*. 2016;22(3):383-411.
12. Carmona-Sanchez R, Icaza-Chavez ME, Bielsa-Fernandez MV, Gomez-Escudero O, Bosques-Padilla F, Coss-Adame E, et al. The Mexican consensus on irritable bowel syndrome. *Rev Gastroenterol Mex*. 2016;81(3):149-67.
13. Linedale EC, Shahzad MA, Kellie AR, Mikocka-Walus A, Gibson PR, Andrew JM. Referrals to a tertiary hospital: A window into clinical management issues in functional gastrointestinal disorders. *JGH Open*. 2017;1(3):84-91.
14. Chey WD, Nojkov B, Rubenstein JH, Dobhan RR, Greenston JK, Cash BD. The yield of colonoscopy in patients with non-constipated irritable bowel syndrome: results from a prospective, controlled US trial. *Am J Gastroenterol*. 2010;105(4):859-65.
15. Rizvi Q LE, Mikocka-Walus A, Gibson PR, Andrews JM. Can we better target colonoscopies using standard appropriateness guides. *J Gastroenterol Hepatol*. 2015;30(Suppl 3):58.

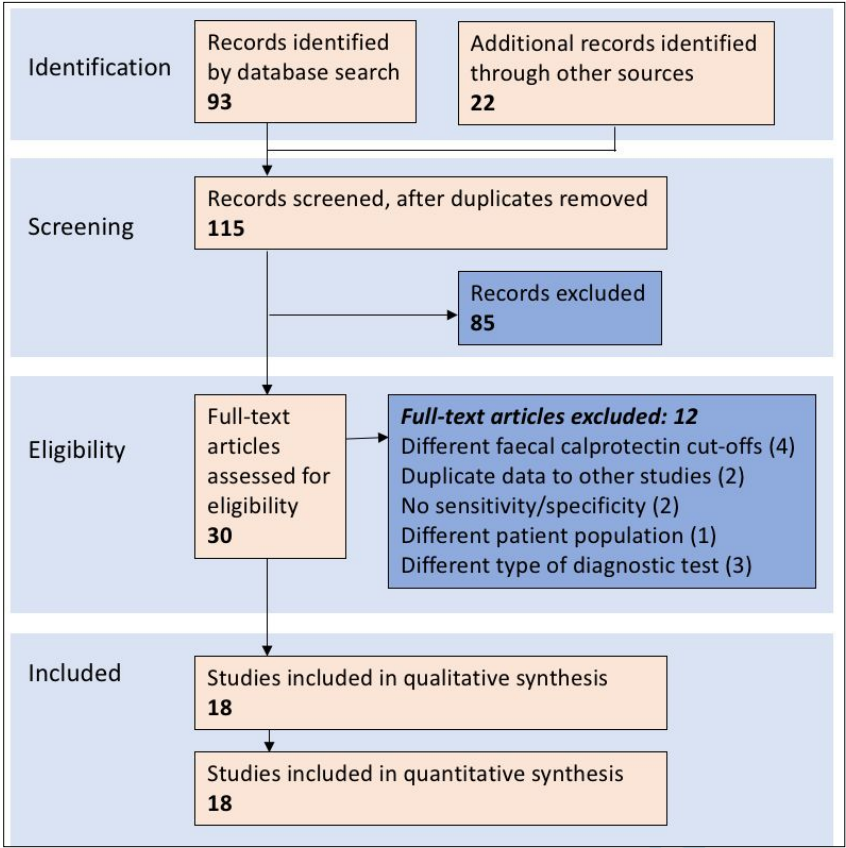
16. Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technology Assessment*. 2013;17(55):xv-xix,1-211.
17. Jellema P, van Tulder MW, van der Horst HE, Florie J, Mulder CJ, van der Windt DA. Inflammatory bowel disease: a systematic review on the value of diagnostic testing in primary care. *Colorectal Dis*. 2011;13(3):239-54.
18. Preferred Items for Systematic Reviews and Meta-analysis [PRISMA] guidelines [website]. <http://www.prisma-statement.org> (accessed Sep 2018).
19. Macaskill P, Gatsonic C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. <http://srdta.cochrane.org> (accessed Apr 2019).
20. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odd ratio: A single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129-35.
21. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88.
22. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.
23. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.
24. Limburg PJ, Ahlquist DA, Sandborn WJ, Mahoney DW, Devens ME, Harrington JJ, et al. Fecal calprotectin levels predict colorectal inflammation among patients with chronic diarrhea referred for colonoscopy. *Am J Gastroenterol*. 2000;95(10):2831-7.
25. Carroccio A. 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):861-7.
26. D'Inca R, Dal Pont E, Di Leo V, Ferronato A, Fries W, Vettorato MG, et al. Calprotectin and lactoferrin in the assessment of intestinal inflammation and organic disease. *Int J Colorectal Dis*. 2007;22(4):429-37.
27. Otten CM, Kok L, Witteman BJ, Baumgarten R, Kampman E, Moons KG, 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(9):1275-80.
28. 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(10):985-92.
29. Schoepfer AM, Trummler M, Seeholzer P, Seibold-Schmid B, Seibold F. Discriminating IBD from IBS: comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies. *Inflamm Bowel Dis*. 2008;14(1):32-9.

30. El-Badry A, Sedrak H, Rashed L. Faecal calprotectin in differentiating between functional and organic bowel diseases. *Arab Journal of Gastroenterology*. 2010;11(2):70-3.
31. Kok L, Elias SG, Witteman BJM, Goedhard JG, Muris JWM, Moons KGM, 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(6):989-98.
32. Burri E, Manz M, Rothen C, Rossi L, Beglinger C, Lehmann FS. 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.
33. 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(9):1048-54.
34. Lozoya Angulo ME, de Las Heras Gomez I, Martinez Villanueva M, Noguera Velasco JA, Aviles Plaza F. Faecal calprotectin, an useful marker in discriminating between inflammatory bowel disease and functional gastrointestinal disorders. *Gastroenterol Hepatol*. 2017;40(3):125-31.
35. Dhaliwal A, Zeino Z, Tomkins C, Cheung M, Nwokolo C, Smith S, et al. Utility of faecal calprotectin in inflammatory bowel disease (IBD): what cut-offs should we apply? *Frontline Gastroenterol*. 2015;6(1):14-9.
36. Banerjee A, Srinivas M, Eyre R, Ellis R, Waugh N, Bardhan KD, 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(1):20-6.
37. Turvill J, O'Connell S, Brooks A, Bradley-Wood K, Laing J, Thiagarajan S, et al. Evaluation of a faecal calprotectin care pathway for use in primary care. *Prim Health Care Res Dev*. 2016;17(05):428-36.
38. Hogberg C, Karling P, Rutegard J, Lilja M. 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(1):69-75.
39. Conroy S, Hale MF, Cross SS, Swallow K, Sidhu RH, Sargur R, 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(4):316-22.
40. Walker GJ, Moore L, Heerasing N, Hendy P, Perry MH, McDonald TJ, et al. Faecal calprotectin effectively excludes inflammatory bowel disease in 789 symptomatic young adults with/without alarm symptoms: a prospective UK primary care cohort study. *Aliment Pharmacol Ther*. 2018;47(8):1103-16.
41. Turvill J, Turnock D, Holmes H, Jones A, McLaughlan E, Hilton V, et al. Evaluation of the clinical and cost-effectiveness of the York Faecal Calprotectin Care Pathway. *Frontline Gastroenterol*. 2018;9(4):285-94.
42. Linedale EC, Mikocka-Walus A, Vincent AD, Gibson PR, Andrews JM. Performance of an algorithm-based approach to the diagnosis and management of functional

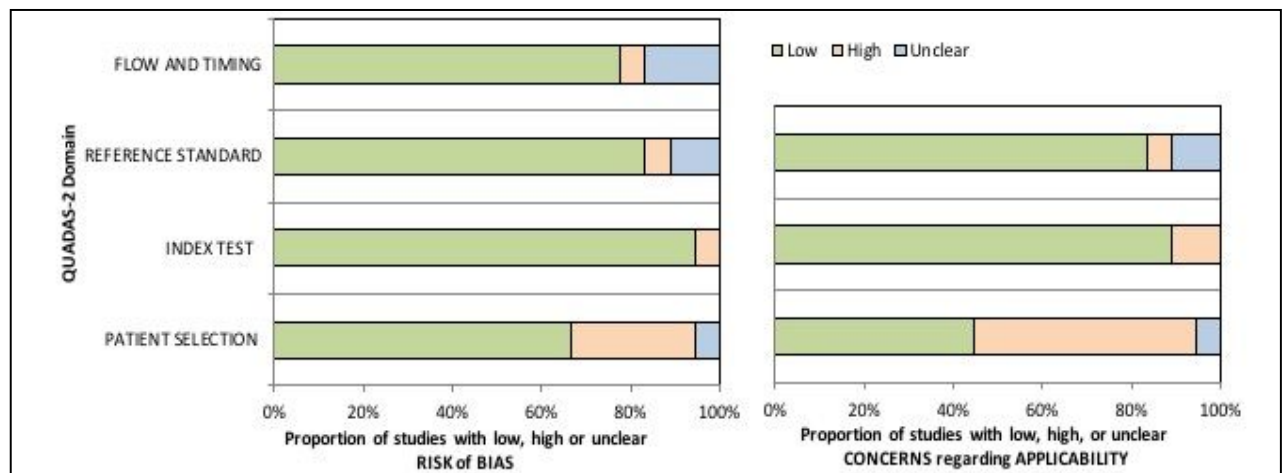
- gastrointestinal disorders: A pilot trial. *Neurogastroenterol Motil.* 2018;30(1). doi:10.1111/nmo.13243.
43. Linedale EC. Issues in the Diagnosis and Management of Functional Gastrointestinal Disorders: The Development of a Novel Clinical Pathway Adelaide The University of Adelaide; 2017.  
<https://digital.library.adelaide.edu.au/dspace/bitstream/2440/114587/2/02whole.pdf> (accessed Sep 2018).
44. Linedale EC, Chur-Hansen A, Mikocka-Walus A, Gibson PR, Andrews JM. Uncertain Diagnostic Language Affects Further Studies, Endoscopies, and Repeat Consultations for Patients With Functional Gastrointestinal Disorders. *Clin Gastroenterol Hepatol.* 2016;14(12):1735-41.e1.
45. Rizvi Q, Linedale EC, Mikocka-Walus A, Gibson PR, Andrews JM. Can we better target colonoscopies using standard "appropriateness" guides? . *J Gastroenterol Hepatol* 2015;30(Suppl 3):58.



**Figure 1:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of study selection and QUADAS-2 quality assessment of included studies [A]



[B]



Feedback: Please supply data to allow-redrawing in MJA style.

#### RISK of BIAS

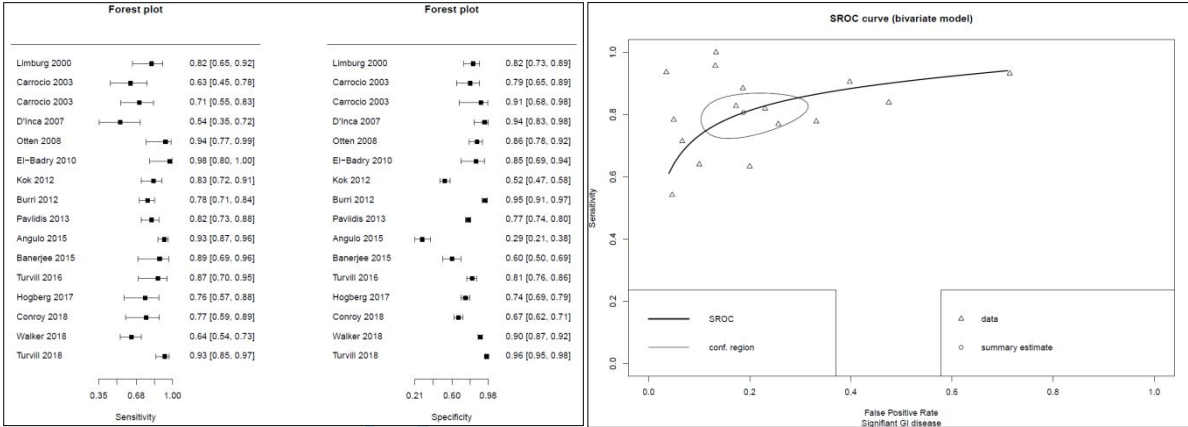
|         | PATIENT<br>SELECTION | INDEX<br>TEST | REFERENCE<br>STANDARD | FLOW<br>AND<br>TIMING |
|---------|----------------------|---------------|-----------------------|-----------------------|
| Low     | 12                   | 17            | 15                    | 14                    |
| High    | 5                    | 1             | 1                     | 1                     |
| Unclear | 1                    | 0             | 2                     | 3                     |

#### CONCERNS regarding APPLICABILITY

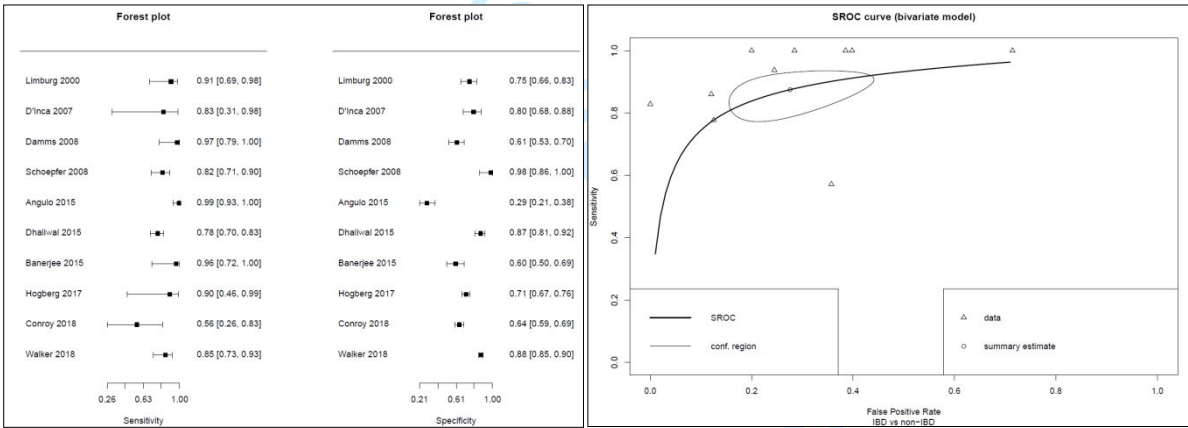
|         | PATIENT<br>SELECTION | INDEX<br>TEST | REFERENCE STANDARD |
|---------|----------------------|---------------|--------------------|
| Low     | 8                    | 16            | 15                 |
| High    | 9                    | 2             | 1                  |
| Unclear | 1                    | 0             | 2                  |

**Figure 2:** Forest plot and SROC bivariate model [A] distinguishing OGIDs from FGIDs [B] distinguishing IBD from non-IBD [C] diagnostic accuracy of FC

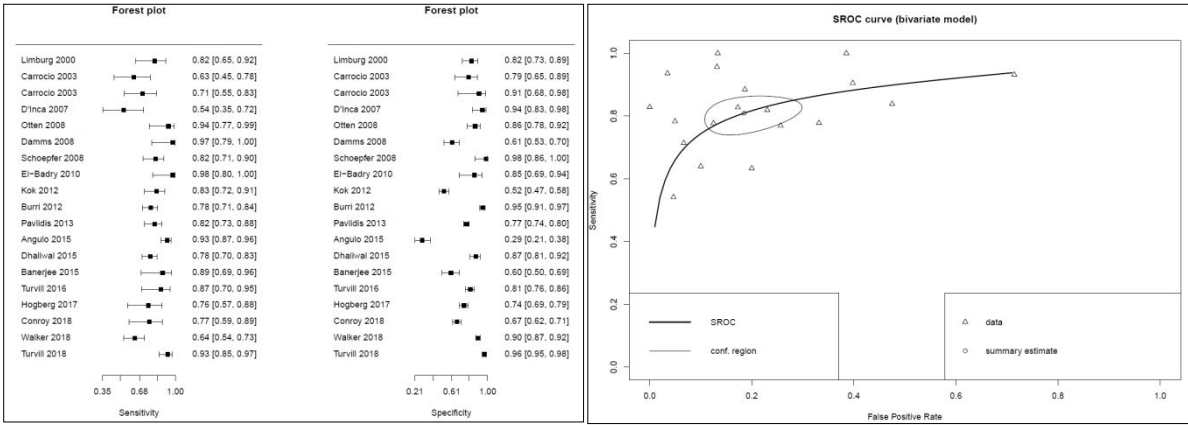
[A]



[B]

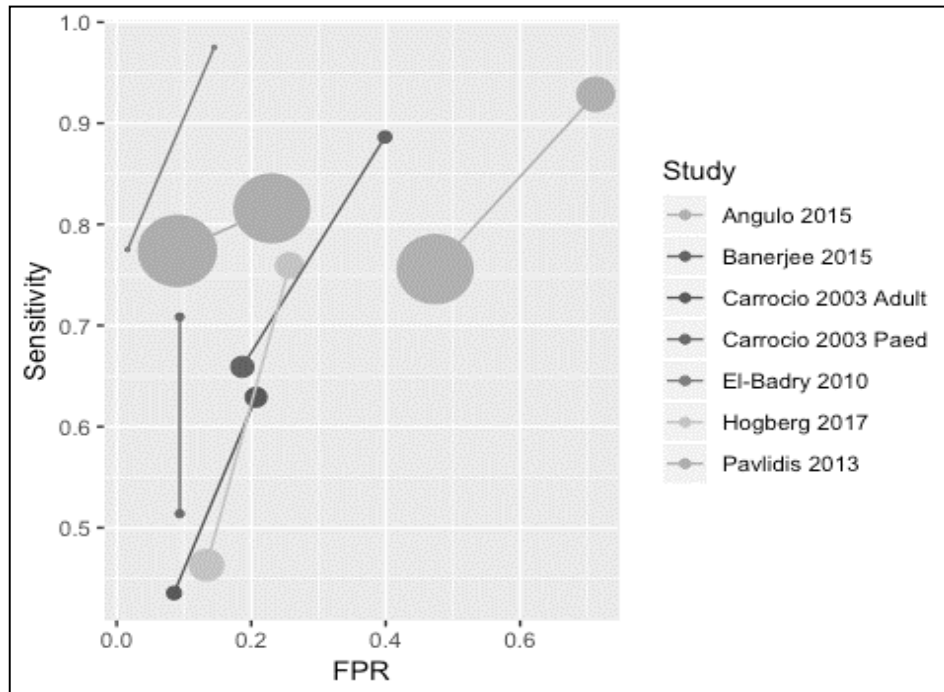


[C]



**Figure 3: [A]** Comparison of diagnostic accuracy and **[B]** diagnostic odds ratio between FC cut-offs\*

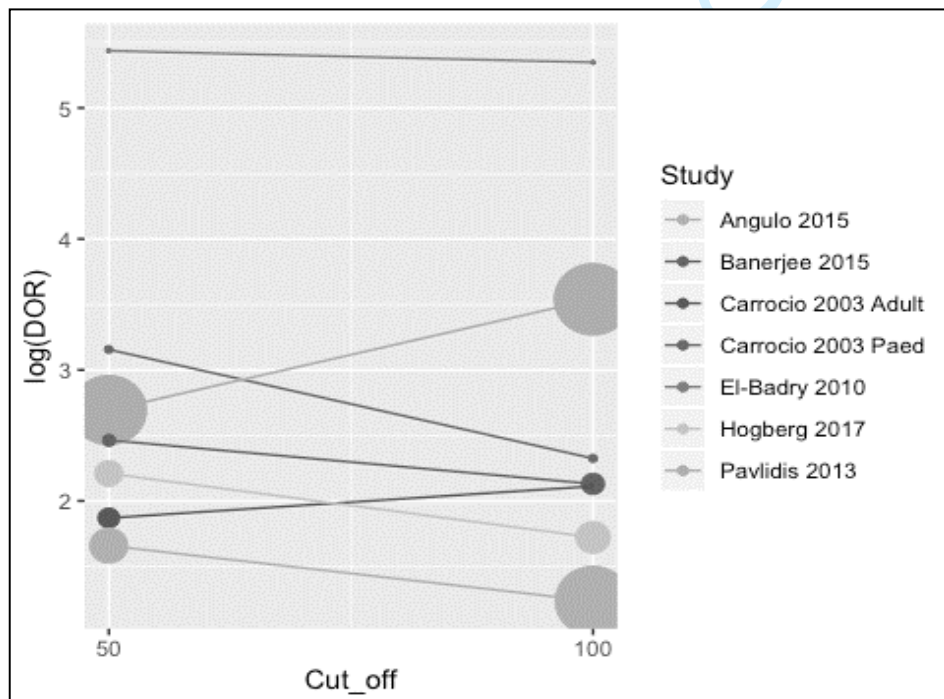
**[A]**



Nodes with lower sensitivity and FPR represent a cut-off of 100  $\mu\text{g/g}$

\* The area of the points is proportional to the inverse variance of the estimate.

**[B]**



\* The area of the points is proportional to the inverse variance of the estimate.

Figure 3 in Tabular form

| Study               | FC cut-off (µg/g) | Sn   | FPR  | DOR   | logDOR |
|---------------------|-------------------|------|------|-------|--------|
| Angulo 2015         | 50                | 0.93 | 0.71 | 5.3   | 0.40   |
| Angulo 2015         | 100               | 0.76 | 0.47 | 3.4   | 0.28   |
| Banerjee 2015       | 50                | 0.89 | 0.40 | 11.7  | 0.70   |
| Banerjee 2015       | 100               | 0.66 | 0.19 | 8.4   | 0.52   |
| Carrocio 2003 Adult | 50                | 0.63 | 0.21 | 6.5   | 0.54   |
| Carrocio 2003 Adult | 100               | 0.44 | 0.09 | 8.3   | 0.67   |
| Carrocio 2003 Paed  | 50                | 0.71 | 0.09 | 23.5  | 0.93   |
| Carrocio 2003 Paed  | 100               | 0.51 | 0.09 | 10.2  | 0.92   |
| El-Badry 2010       | 50                | 0.98 | 0.15 | 229.7 | 1.52   |
| El-Badry 2010       | 100               | 0.78 | 0.02 | 210.1 | 1.52   |
| Hogberg 2017        | 50                | 0.76 | 0.26 | 9.1   | 0.47   |
| Hogberg 2017        | 100               | 0.46 | 0.13 | 5.6   | 0.42   |
| Pavlidis 2013       | 50                | 0.82 | 0.23 | 14.8  | 0.28   |
| Pavlidis 2013       | 100               | 0.77 | 0.09 | 34.4  | 0.27   |

Table 1: Population characteristics of included studies

| Study                          | No. of patients | Mean age (yrs)              | Gender (% female) | Setting                     | Population |
|--------------------------------|-----------------|-----------------------------|-------------------|-----------------------------|------------|
| Limburg 2000 <sup>(24)</sup>   | 110             | 57                          | 0.64              | Secondary care, USA         | Adult      |
| Carrocio 2003 <sup>(25)</sup>  | 70              | 35                          | 0.57              | Secondary care, Italy       | Adult      |
|                                | 50              | 3.5                         | 0.6               | Secondary care, Italy       | Children   |
| D'Inca 2007 <sup>(26)</sup>    | 67              | 49                          | 0.58              | Secondary care, Italy       | Adults     |
| Otten 2008 <sup>(27)</sup>     | 114             | 51                          | 0.54              | Secondary care, Netherland  | Adults     |
| Damms 2008 <sup>(28)</sup>     | 140             | 58                          | 0.56              | Secondary care, Germany     | Adults     |
| Schoepfer 2008 <sup>(29)</sup> | 136             | 40                          | 0.6               | Secondary care, Bern        | Adults     |
| El-Badry 2010 <sup>(30)</sup>  | 49              | 37                          | 0.47              | Secondary care, Egypt       | Adult      |
| Kok 2012 <sup>(31)</sup>       | 382             | 60                          | 0.55              | Primary care, Netherlands   | Adults     |
| Burri 2012 <sup>(32)</sup>     | 405             | 63                          | 0.56              | Secondary care, Switzerland | Adult      |
| Pavlidis 2013 <sup>(33)</sup>  | 962             | 33                          | 0.6               | Primary care, UK            | Adult      |
| Angulo 2015 <sup>(34)</sup>    | 264             | Adults – 43<br>Children – 7 | 0.54              | Secondary care, Spain       | Mixed      |
| Dhaliwal 2015 <sup>(35)</sup>  | 311             | n/a                         | 0.67              | Secondary care, UK          | Adult      |
| Banerjee 2015 <sup>(36)</sup>  | 119             | 46                          | 0.54              | Secondary care, UK          | Adult      |
| Turvill 2016 <sup>(37)</sup>   | 262             | 37                          | 0.7               | Primary care, UK            | Adult      |
| Hogberg 2017 <sup>(38)</sup>   | 373             | 63                          | 0.65              | Primary care, Sweden        | Adult      |
| Conroy 2018 <sup>(39)</sup>    | 410             | 42                          | 0.61              | Primary care, UK            | Adults     |
| Walker 2018 <sup>(40)</sup>    | 789             | 30                          | 0.6               | Primary care, UK            | Adults     |
| Turvill 2018 <sup>(41)</sup>   | 1005            | 38                          | 0.63              | Primary care, UK            | Adults     |

Table 2: Summary statistics

| Accuracy                           | Overall        | IBD vs non-IBD | OGIDs vs FGIDs |
|------------------------------------|----------------|----------------|----------------|
| Sensitivity, % (95% CI)            | 80 (73-85%)    | 88 (80-93%)    | 81 (74-86%)    |
| False positive rate, % (95% CI)    | 19 (14-30%)    | 28 (18-41%)    | 19 (12-29%)    |
| Log Diagnostic odds ratio (95% CI) | 3.1 (2.5, 3.6) | 3.2 (2.5, 3.9) | 3.0 (2.4, 3.6) |
| Area under the curve               | 0.86           | 0.89           | 0.87           |

**Supplement Table 1:** Studies reporting test characteristics of faecal calprotectin for detecting OGIDs

|                                    | Study         | FC cut-off (µg/g) | TP  | FN | TN  | FP  | Disease prevalence (%) |
|------------------------------------|---------------|-------------------|-----|----|-----|-----|------------------------|
|                                    | Limburg 2000  | 100               | 24  | 5  | 67  | 14  | 26.36 (18.42-35.62)    |
|                                    | Carrocio 2003 | 50 (Adults)       | 19  | 11 | 32  | 8   | 42.86 (31.09-55.25)    |
|                                    | Carrocio 2003 | 50 (Children)     | 25  | 10 | 14  | 1   | 70.00 (55.39-82.14)    |
|                                    | D'Inca 2007   | 50                | 13  | 11 | 41  | 2   | 35.82 (24.47-48.47)    |
|                                    | Otten 2008    | 50                | 22  | 1  | 79  | 12  | 20.18 (13.24-28.72)    |
|                                    | El-Badry 2010 | 50                | 19  | 0  | 26  | 4   | 38.8 (25.20-53.76)     |
| Included in statistical analysis   | Kok 2012      | 50                | 52  | 10 | 168 | 152 | 16.23 (12.68-20.32)    |
|                                    | Burri 2012    | 50                | 112 | 31 | 249 | 13  | 35.31 (30.65-40.18)    |
|                                    | Pavlidis 2013 | 50                | 77  | 17 | 668 | 200 | 9.77 (7.97-11.82)      |
|                                    | Angulo 2015   | 50                | 123 | 9  | 32  | 80  | 54.10 (47.62-60.47)    |
|                                    | Banerjee 2015 | 50                | 19  | 2  | 59  | 39  | 17.65 (11.27-25.70)    |
|                                    | Turvill 2016  | 100               | 23  | 3  | 192 | 44  | 9.92 (6.59-14.20)      |
|                                    | Hogberg 2017  | 50                | 20  | 6  | 258 | 89  | 6.97 (4.60-10.05)      |
|                                    | Conroy 2018   | 50                | 21  | 6  | 256 | 127 | 6.59 (4.38-9.44)       |
|                                    | Walker 2018   | 100               | 64  | 36 | 620 | 69  | 12.67 (10.43-15.20)    |
|                                    | Turvill 2018  | 100               | 73  | 5  | 803 | 29  | 8.20 (6.54-10.13)      |
| Removed from statistical analysis* | Carrocio 2003 | 100 (Adults)      | 13  | 17 | 37  | 3   | 42.86 (31.09-55.25)    |
|                                    | Carrocio 2003 | 100 (Children)    | 18  | 17 | 14  | 1   | 70.00 (55.39-82.14)    |
|                                    | El-Badry 2010 | 100               | 15  | 4  | 30  | 0   | 38.78 (25.20-53.76)    |
|                                    | Pavlidis 2013 | 100               | 73  | 21 | 790 | 78  | 9.77 (7.97-11.82)      |
|                                    | Angulo 2015   | 100               | 100 | 32 | 59  | 53  | 54.10 (47.62-60.47)    |
|                                    | Banerjee 2015 | 100               | 14  | 7  | 80  | 18  | 17.65 (11.27-25.70)    |
|                                    | Hogberg 2017  | 100               | 12  | 14 | 301 | 46  | 6.97 (4.60-10.05)      |

\* When studies used both 50 and 100 µg/g cut-off values for the same patient, we only used the 50 µg/g value for this analysis.

**Supplement Table 2:** Studies reporting test characteristics of faecal calprotectin for detecting IBD

|                                    | Study          | FC cut-off (µg/g) | TP  | FN | TN  | FP  | Disease prevalence (%) |
|------------------------------------|----------------|-------------------|-----|----|-----|-----|------------------------|
|                                    | Limburg 2000   | 100               | 15  | 1  | 71  | 23  | 14.55 (8.55-22.54)     |
|                                    | D'Inca 2007    | 50                | 2   | 0  | 52  | 13  | 2.99 (0.36-10.37)      |
|                                    | Damms 2008     | 50                | 18  | 0  | 75  | 47  | 12.86 (7.80-19.56)     |
|                                    | Schoepfer 2008 | 50                | 53  | 11 | 30  | 0   | 68.09 (57.67-77.33)    |
| Included in statistical analysis   | Angulo 2015    | 50                | 68  | 0  | 32  | 80  | 37.78 (30.67-45.29)    |
|                                    | Dhaliwal 2015  | 50                | 115 | 33 | 126 | 18  | 50.68 (44.80-56.56)    |
|                                    | Banerjee 2015  | 50                | 12  | 0  | 59  | 39  | 10.91 (5.77-18.28)     |
|                                    | Hogberg 2017   | 50                | 4   | 0  | 264 | 105 | 1.07 (0.29-2.72)       |
|                                    | Conroy 2018    | 50                | 4   | 3  | 259 | 144 | 1.71 (0.69-3.49)       |
|                                    | Walker 2018    | 100               | 43  | 7  | 650 | 89  | 6.34 (4.74-8.27)       |
|                                    | Angulo 2015    | 100               | 60  | 8  | 63  | 49  | 37.78 (30.67-45.29)    |
| Removed from statistical analysis* | Dhaliwal 2015  | 100               | 113 | 35 | 140 | 4   | 50.68 (44.80-56.56)    |
|                                    | Banerjee 2015  | 100               | 11  | 1  | 80  | 18  | 10.91 (5.77-18.28)     |
|                                    | Hogberg 2017   | 100               | 3   | 1  | 314 | 55  | 1.07 (0.29-2.72)       |

\* When studies used both 50 and 100 µg/g cut-off values for the same patient, we only used the 50 µg/g value for this analysis



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Effectiveness of faecal calprotectin ~~in primary care~~ to identify patients with organic gastrointestinal diseases: a systematic review and diagnostic meta-analysis**

**Abstract**

**Objectives:** ~~In primary care, a~~Assessing patients with lower gastrointestinal symptoms can be challenging. Data examining the clinical effectiveness of faecal calprotectin (FC) testing ~~in primary care~~ to distinguish between organic gastrointestinal diseases (OGIDs), such as Inflammatory Bowel Disease (IBD), and Functional Gastrointestinal Disorders (FGIDs) were systematically reviewed.

**Study design:** Journal articles that assessed the accuracy of FC to differentiate between IBD and/or OGIDs and FGIDs were reviewed. Data on methodology and the characteristics of the diagnostic test were collected. Study quality was assessed using QUADAS-2, an evidence-based quality assessment tool for diagnostic accuracy studies.

**Data sources:** MEDLINE and EMBASE were searched for relevant literature published from 1998 to August 2018. A total of 18 studies were included.

**Data synthesis:** When distinguishing patients with OGIDs (including IBD) from FGIDs, FC had a sensitivity of ~~810.6%~~ 95.6% (95% CI=0.74–0.86) ~~9~~ a, false positive rate of ~~198.8%~~ 12.0% (95% CI=0.1–0.29), ~~(FPR= 1-Specificity)~~, and an area under the curve (AUC) of ~~0.8767~~. When distinguishing IBD from non-IBD, FC had a sensitivity of ~~887.5%~~ 87.5% (95% CI=80–93), ~~0.8–0.9~~, false positive rate of ~~287.6%~~ 18.4% (95% CI=0.2–0.41), and an AUC of ~~0.8986~~. Assuming a prevalence of 1% the positive predictive value (PPV) is 4.2% and the negative predictive value (NPV) is 100%. There was a non-significant ( $p=0.77$ ) ~~difference in~~ sensitivity and false-positive rates ~~with difference between using~~ a cut-off of 50µg/g ( $SE=3.7$ ) ~~versus compared to~~ 100µg/g ( $SE=4.0$ ).

**Conclusions:** FC is a clinically useful test to distinguish ~~IBD and other~~ OGIDs (including IBD) from FGIDs ~~in the primary care setting~~, and its wider implementation would ~~result in~~ support more timely diagnosis, ~~better direct~~ improve colonoscopy resources ~~(fewer utilisation by reducing the number of~~ patients with FGIDs undergoing ~~unnecessary endoscopy colonoscopy)~~, and yield cost savings.

**PROSPERO registration:** CRD42018105078

**TWEETS** (140 characters)



~~Meta-analysis shows that f~~ Faecal calprotectin is highly effective at distinguishing between organic and functional gastrointestinal disease on meta-analysis.

**Final word count: 297126 words**

For Review Only

**Introduction**

Gastrointestinal symptoms account for 5-10% of primary care consultations and referral of all patients to specialist care is ~~neither necessary nor affordable~~unnecessary.<sup>(1, 2)</sup> It can be difficult for general practitioners (GPs) and specialists to confidently distinguish between patients presenting with functional gastrointestinal disorders (FGIDs) as compared to organic gastrointestinal diseases (OGIDs) ~~which includes~~such as inflammatory bowel disease (IBD), colorectal cancer, polyps >1cm and diverticulitis, ~~based~~ on symptoms alone.

FGIDs are chronic symptom-based disorders, with typical symptoms including abdominal pain, diarrhoea, bloating, constipation and/or alternating bowel habit. OGIDs can present with similar symptoms, so distinguishing them from FGIDs can be difficult. FGIDs are vastly more common than OGIDs with the most common FGID, Irritable Bowel Syndrome (IBS), affecting approximately 10-20% of the population, presenting ~~mainly~~ between the ages of 20 and 40.<sup>(3, 4)</sup> Once a diagnosis of FGID is established further diagnostic tests are not required and the vast majority of patients can be managed successfully in the primary care setting, ~~with reassurance, education, lifestyle changes, dietary modification and psychological approaches.~~<sup>(5)</sup> In contrast, various OGIDs result in ~~degrees of~~ inflammation of the colon, and these patients are best assessed with specialist referral and further evaluation, often with colonoscopy and/or cross-sectional imaging. In ~~this~~ the young adults (20-40 years of age group) the most common OGID is IBD, comprising ulcerative colitis (UC) and Crohn's disease (CD), ~~affecting which affects~~ 0.4% of ~~the population below age those aged~~ ≤50.<sup>(6)</sup>

In young adults faecal calprotectin (~~(20-40 years of age)~~, FC) is most often used to differentiate between IBS and IBD. However, in children FC can be elevated due to a variety of conditions. Usually, it is not recommended to use FC alone as a screening tool in people older than 50, due to their higher risk of having OGIDs including neoplasms.<sup>(7-12)</sup>

The high prevalence of IBS in conjunction with the low incidence of ~~organic disease~~OGIDs in ~~young adult~~these patients have led ~~to advocacy different guideline bodies worldwide to advocate~~ for a positive diagnosis of FGID based on medical history with and minimal associated investigations, ~~including faecal calprotectin (FC) FC testing to rule out OGIDs with good negative predictive value.~~<sup>(7-12)</sup> Despite primary care diagnosis and management of FGIDs being the recommendation, there ~~continue to be~~ large numbers of ~~secondary care~~ referrals of patients with likely IBS-FGIDs for leading to unnecessary ~~and costly~~further assessment,

such as colonoscopy, which are often normal, and investigations.<sup>(13-15)</sup> Audits of colonoscopies performed for investigation of symptoms alone have found that 60-87% were normal.<sup>(14,15)</sup>

Since inflammation is the key characteristic that differentiates OGIDs from FGIDs, FC has been identified as a non-invasive predictive test with high sensitivity and specificity for OGIDs (including IBD) and IBD respectively, a biomarker of colonic inflammation such as faecal calprotectin (FC) is ideal for distinguishing the two groups of patients. The systematic review by Waugh<sup>(16,17)</sup> and Jellema<sup>(17)</sup> identified FC as a non-invasive predictive test with high sensitivity and specificity for IBD with significant potential for healthcare cost savings. However, this analysis was limited by a lack of data from studies in primary care populations. Since then multiple studies have been performed in the primary care setting to address this gap, but a comprehensive evaluation of these studies has not yet been conducted. Furthermore, it has been demonstrated that FC testing is superior to other biomarkers, such as CRP and ESR.<sup>(16)</sup> Despite these reported benefits, there is a scarcity of literature that provides a comprehensive evaluation of the clinical effectiveness of FC.

We have reviewed the clinical effectiveness of FC in distinguishing OGIDs and IBD from FGIDs compared to a specialist diagnosis based on clinical, laboratory, imaging and/or endoscopic findings gold standard endoscopy or imaging in adults who present to general practitioners (GPs) with lower gastrointestinal symptoms. We present data on the effectiveness of FC in distinguishing between [1] patients with OGIDs who require further investigation and those with FGIDs, and [2] patients with IBD and those with FGIDs and non-IBD, which is a common clinical question.

## Methods

This systematic review and meta-analysis study has been conducted in accordance with the Preferred Items for Systematic Reviews and Meta-analysis [PRISMA] guidelines.<sup>(18)</sup>

## Search strategy

A search of the English medical literature was conducted using MEDLINE and EMBASE from January 1980 to June 2018 that examined the clinical effectiveness of FC in distinguishing OGIDs (including IBD) and IBD alone from FGIDs. The following key words were included alone or in combination: 'Crohn disease', 'ulcerative colitis', 'inflammatory bowel disease',

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

‘irritable colon’, ‘digestive system function disorder’, ‘IBS’, ‘IBD’, ‘faecal or fecal’, ‘calprotectin’, ‘general practitioner’, ‘GP’, ‘primary care provider’ and ‘primary medical care’. In addition, we examined the references of the screened articles to identify additional studies.

**Study selection**

~~All~~ Retrieved studies were assessed by two reviewers [S.C. and J.B.] for eligibility for full text review. Non-relevant studies were excluded based on title, abstract and study type [i.e. review]. Subsequently, the selected full texts were assessed by two reviewers [Y.A. and D.P.]. Study inclusion criteria included: [1] study of FC as a diagnostic test in both primary care and outpatient hospital setting; [2] comparison with a reference test/standard such as colonoscopy or cross-sectional imaging; [3] paediatric and/or adult patients; [4] patients presenting with lower gastrointestinal symptoms; [5] FC using the standard enzyme-linked immunosorbent assay (ELISA)-based method; [6] FC with cut-offs of 50 or 100 µg/g ~~of stool~~; [7] articles written in English; [8] full text available [i.e. no abstracts].

We excluded studies in patients with red flag signs or symptoms, such as positive faecal occult blood test, overt rectal bleeding, iron deficiency anaemia, abdominal or rectal masses or a family history of bowel cancer that would require endoscopic evaluation.

**Data extraction**

~~The following~~ Pre-specified data were collated, including: study design, aim, number of patients, patient demographics, study setting, inclusion and exclusion criteria, reference test and index used. Additionally, the following data were collected: the FC assays used, cut-offs, true positive (TP), false positive (FP), true negative (TN), false negative (FN), sensitivity (Sn), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV) and accuracy with reference test.

**Statistical analysis and data synthesis**

We followed the guidance from Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Chapter 10) in preparing this meta-analysis.<sup>(19)</sup>

Diagnostic two-by-two tables and diagnostic performance measures per research questions are presented. Forest plots were used to display sensitivity and specificity across studies, together with confidence intervals. When studies used both 50 and 100 µg/g of stool cut-off values for the same patient, we only used the 50 µg/g value for this analysis. Furthermore, when the included studies used the same patient results for multiple comparisons, we only used the results associated with FC used to distinguish OGIDs from FGIDs. Specifically, we only used the patient's results once per indication.

The diagnostic odds ratio (DOR) is a single measure of test accuracy that incorporates both sensitivity and specificity which can summarise overall test accuracy when thresholds differ across studies.<sup>(20)</sup> ~~It is a useful summary statistic in meta-analyses because it can summarise overall test accuracy when thresholds differ across studies.~~ Diagnostic odds ratios were summarised across studies using a random effects model following the approach of DerSimonian and Laird.<sup>(21)</sup> Summary estimates of sensitivity and specificity with 95% CI were calculated using bivariate analyses based upon the approach of Reitsma et al.<sup>(22)</sup>

The Reitsma model was also used to generate Summary Receiver Operating Characteristic Curves (SROC). The SROC summarises the overall test accuracy (true positives relative to false positives) across various thresholds based upon all relevant studies. We defined the False Positive Rate (FPR) as:  $FPR = (1 - \text{Specificity})$ . We present SROC curves per research question, together with point estimates of the true positive rates and false positive rates for each study, and the summary estimates (with 95% CI) from the bivariate meta-analysis. All analyses were performed using the mada package version 0.5.8 in the statistical software package R version 3.5.1.

### Quality Assessment

All included studies were graded for methodological quality by two investigators [Y.A. and D.P.] with the revised tool for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2).<sup>(23)</sup> The QUADAS-2 tool is designed to assess the quality of diagnostic accuracy studies in four domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of the risk of bias and the applicability of the study results for the first three domains. A study was graded as high quality in the case of a low risk of bias in at least 6 out of the 7 subdomains. A study was graded as low quality in the case of a high

risk or unclear risk in 4 or more subdomains. All other studies were graded as moderate quality. Any disagreements were resolved through discussion with a third reviewer, with the majority consensus used.

**Results**

**Study selection**

93 records were identified through database search and 22 records were identified through references. A total of 115 records were screened after removal of duplicates. After screening titles and abstracts, 30 potentially eligible studies were selected for full text review. After full text review there were 10 articles examining the effectiveness of FC in distinguishing IBD from FGIDs, which was our initial predefined aim. However there were 18 publications reporting on the effectiveness of FC in distinguishing OGIDs from FGIDs. Therefore we included this test characteristic as an additional aim in our meta-analysis, 18 records were eligible for inclusion in our analysis (Figure 1A).

**Grading of study quality**

The results of the QUADAS-2 assessment are shown in Figure 1B. Study quality was graded high in eleven studies, moderate in four studies, and low in three studies. Most concerns were raised in the subdomains regarding patient selection and reference standard. In some studies, patient selection was not reflective of a true primary care population, while in others, there was lack of clarity about the reference standard test they had used to confirm the final diagnosis ~~was unclear.~~

**Study characteristics**

The relevant study characteristics are shown in Table 1. <sup>(24-41)</sup> Sixteen studies used a prospective and two studies a retrospective design. Seven studies were conducted in primary care and eleven studies in secondary care settings either through outpatient clinics or in endoscopy units. Most studies were performed in Europe in an adult population with a mean age range between 30 and 63; two studies also assessed a paediatric population with mean ages of 3.5 and 7. In all studies, patients presented with chronic lower gastrointestinal symptoms, which were suggestive of either OGID (particularly IBD) or FGIDs. Colonoscopy was used as the reference standard in the majority of studies. ~~All studies used laboratory enzyme-linked immunosorbent assay (ELISA) tests.~~ The most commonly used enzyme-linked immunosorbent assay (ELISA)

kit was EK-CAL ( $n=8$ ), manufactured by Buhlmann Laboratories (Schonenbuch, Switzerland). Cost-effectiveness of FC was evaluated in two studies.

The number of patients included in each study varied from 49 to 1005, with a median of 262 [IQR 399-111]. There was a total of 8150 patients included in the analyses (IBD vs ~~non-IBD~~ ~~FGIDs~~ and OGIDs vs FGIDs), with 5431 associated with OGIDs vs FGIDs and 2719 associated with IBD vs ~~non-IBD~~ ~~FGIDs~~. However, results for 1887 of these patients were using a cut-off of 50  $\mu\text{g/g}$  and 100  $\mu\text{g/g}$ . When both values were reported, we only used the 50  $\mu\text{g/g}$  cut-off value. As such, our combined analysis included 6,263 patients.

~~There were~~Of the 18 eligible studies, eligible for inclusion. However, many ( $n=7$ ) ~~of these studies~~ considered both diagnostic groups (OGIDs and/or IBD vs FGIDs), and both pathology reference ranges (50 $\mu\text{g/g}$  and 100 $\mu\text{g/g}$ ) ( $n=7$ ), with two studies including both an adult and paediatric population. As such, there were 26 separate comparisons. There were 16 studies that used FC as a diagnostic test for OGIDs, 10 studies that used FC to differentiate IBD from FGIDs, and 7 studies that considered both entities.

### Diagnostic performance of faecal calprotectin

We present data for two different comparisons:— [1] FC used to distinguish OGIDs from FGIDs; [2] faecal calprotectin used to distinguish IBD (includes UC, CD and IBD-U) from ~~non-IBD (includes FGIDs and other OGIDs)~~ FGIDs.

### OGIDs descriptive (forest plots) and bivariate analysis

There were 16 studies (Supplement Table 1) included in the analysis for distinguishing OGIDs from FGIDs, with 14 including an adult population, 1 paediatric, and 1 mixed population. The adult population consisted of 5117 patients, with the remainder consisting of children ( $n=50$ ), and both adults and children ( $n=264$ ). The confidence intervals between the sensitivity and false-positive rates indicated limited variance in sensitivity (range=74.1-86.5.8%) and false-positive rates (range=12.1.7-29.7%).

When distinguishing OGIDs from FGIDs, FC had a sensitivity of 81% (95% CI=74-86%) ~~80.6% (CI=0.7, 0.9)~~ and a false positive rate of 19% (95% CI=12%-29%) ~~18.8% (CI=0.1, 0.3)~~, as detailed in Figure 2A. The log diagnostic odds ratio equalled 3.0 ( $CI=2.4, 3.6$ ), with the area under the curve equalling 0.876.7, as detailed in Table 2.



**IBD descriptive (forest plot) and bivariate analysis**

There were 10 studies included in the analysis for IBD versus ~~non-IBD~~FGIDs (Supplement Table 2), with 9 including an adult population, and 1 mixed adult and paediatric population. The adult population consisted of 2455 patients, and the mixed study included 264 patients. The confidence intervals between the sensitivity and false-positive rates indicated limited variance in sensitivity (range=~~80.7-93.2~~%) and false-positive rates (range=~~18.7-41.0~~%), as detailed in Figure 2B.

When distinguishing IBD from ~~non-IBD~~FGIDs, FC had a sensitivity of ~~88% (95% CI=80-93%)~~87.5% (*CI*=0.8, 0.9) and a false positive rate of ~~28% (95% CI=18-41%)~~27.6% (*CI*=0.2, 0.4), as detailed in Figure 2B. The log diagnostic odds ratio equalled 3.2 (*CI*=2.5, 3.9), with the area under the curve equalling ~~0.89~~0.86, as detailed in Table 2.

**Overall analysis and meta-regression to compare diagnostic accuracy cut-off points**

Since both patients with possible OGID or IBD require specialist referral and further investigation, most often a colonoscopy, to confirm the diagnosis, it is useful to consider both populations together. There were 19 studies included in the combined diagnostic analysis, with 17 including an adult population, 1 paediatric, and 1 mixed population. The adult population consisted of 5704 patients, with the remainder consisting of children (*n*=50), and both adults and children (*n*=264). The confidence intervals between the sensitivity and false-positive rates indicated limited variance in sensitivity (range=~~73.3-85.3~~%) and false-positive rates (range=~~12.6-30.9~~%).

The overall sensitivity and false positive rate were ~~79.980% (95% CI=73-85%CI=0.7, 0.85)~~ and ~~19.1% (95% CI=14-30%CI=0.1, 0.3)~~ respectively, as detailed in Figure 2C. The log diagnostic odds ratio equalled 3.1 (*CI*=2.5, 3.6), with the area under the curve equalling ~~0.86~~0.86, as detailed in Table 2.

Given the relatively high prevalence reported in the studies, we calculated the positive predictive value (PPV) and negative predictive value (NPV) using relevant population prevalence. The PPV and NPV at 1% prevalence equalled 4.2% and 100%, respectively. Whereas the PPV and NPV at 0.1% equalled 0.40% and 100%, respectively.

### Diagnostic accuracy between FC pathology ranges

There were 7 studies that used both FC pathology cut-off ranges. When comparing the diagnostic accuracy, there was a non-significant ( $p=0.77$ ) increase in sensitivity and decrease in false-positive rate difference between FC cut-offs of  $50\mu\text{g/g}$  ( $SE=3.7$ ) and  $100\mu\text{g/g}$  ( $SE=4.0$ ), as detailed in Figure 3A and 3B.

### Discussion

FC is used clinically ~~Faecal-calprotectin~~FC accurately reflectsto assess the degree of inflammation in the bowel and can be used to distinguish between IBD and FGIDs in patients presenting with lower gastrointestinal symptoms. However, other OGIDs, which are only diagnosed on colonoscopy, can also lead to elevated FC. OGIDs (including IBD) and FGIDs in patients presenting in the primary care setting with lower gastrointestinal symptoms. Therefore, we performed a meta-analysis using available publications on the characteristics of FC in distinguishing between OGIDs and FGIDs in symptomatic patients without alarm features. We performed a secondary analysis on FC to distinguish patients with IBD and FGIDs. ~~From a clinical perspective in an individual patient, w~~When considering whether to perform an invasive test such as colonoscopy, the risk of harm of the procedure must be balanced against the risk of a missed diagnosis. ~~When assessing FC in this scenario a high negative predicative value is desirable, to avoid false negative results. However, given the low prevalence of IBD and OGIDs in the primary care population, specificity of the test drives relative costs in this setting. Given the high NPV (100%) of FC testing in low prevalence populations, it is ideally suited for screening in the primary care setting, although further investigations are required for positive tests given the low PPV.~~ The main benefit of using FC in primary care would be to confirm the clinical diagnosis of FGIDs by GPs, allowing patients to move into management phase and reduce the risk of a delayed diagnosis of OGIDs. This would greatly reduce the number of young adult patients without other alarm features being referred to specialist care, and reduce unnecessary and more expensive investigations such as cross-sectional imaging or colonoscopy.

~~Our systematic review identified 18 studies that examined the question of how accurately FC can distinguish IBD and other OGIDs from FGIDs in a patient population presenting with suggestive symptoms. The majority of studies were in the outpatient setting, and those performed in the gastroenterology setting were on a patient population that would be~~

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

~~appropriate for screening in primary care using FC. The studies were primarily in adult populations, but two studies included paediatric patients. Overall the quality of the studies was good with 11 of the 18 studies graded as high quality using the QUADAS 2 tool. Complete data could be extracted from all of the studies for the meta-analysis.~~

Overall the sensitivity and false positive rate of FC testing using a cut-off of 50 µg/g in this screening patient population showed a sensitivity range of 74.1-865.8% and a false positive rate of 121.7-298.7% for distinguishing OGIDs versus FGIDs and a sensitivity range of 8079.7-932.6% and a false positive rate of 187.5-410.6% for distinguishing IBD from FGIDs~~from non-IBD~~. Our SROC analysis showed that the overall sensitivity for distinguishing organic from functional gastrointestinal diseases was 810.6% with a false positive rate of 198.8% and for distinguishing IBD from non-IBD~~FGIDs~~ the sensitivity was 887.5% with a false positive rate of 287.6%. Previous studies have shown that using a cut-off of 50 µg/g results in a higher sensitivity and lower specificity, but our analysis shows that in studies examining a cut-off of 50 and 100 µg/g, there was a non-significant increase in sensitivity. Nonetheless, given the role of FC as a screening test for OGIDs we believe that a cut-off of 50 µg/g is appropriate.

Our analysis shows that FC testing is effective at identifying patients presenting with abdominal symptoms requiring further investigation~~in these populations. The widespread adoption of FC testing in the primary care setting to investigate patients presenting with abdominal symptoms has the potential to improve resource utilisation and patient outcomes.~~ Our meta-analysis ~~demonstrates support guideline recommendations~~ <sup>(7-11)</sup> that symptomatic patients with FC less than 50 µg/g ~~in a patient without any red flag symptoms is very accurate at identifying patients with~~can be confidently diagnosed with a FGID and managed appropriately and s~~that~~ do not require further evaluation by a specialist or further testing.

The use of FC in the primary care setting allows the effective and appropriate use of resources and reduces unnecessary risk and cost to the patient and system. In patients with non-specific lower gastrointestinal symptoms, the use of FC reduces the risk of a missed diagnosis of OGIDs and ensures prompt specialist review. A recent Australian study found that FC alongside a routine panel of blood tests allowed the identification of patients who needed prompt review by a specialist, improving the safety of local triage practices.<sup>(4224)</sup> ~~Almost 40% of patients referred with suspected FGIDs and triaged as ‘non-urgent’ were found to warrant earlier review.~~

1  
2  
3 ~~However,~~ FC as a negative predictor of OGIDs also enables an early, accurate and confident  
4 diagnosis of FGIDs to be made by GPs without need for gastroenterology referral or further  
5 testing. Enabling a confident FGID diagnosis to be made in primary care is likely to  
6 improve ~~Indeed, there are several important benefits to this. Firstly, the provision of an early,~~  
7 ~~accurate and confident diagnosis of a FGID by a GP is critical to~~ patient acceptance of the  
8 diagnosis and uptake of effective management strategies.<sup>(5)</sup> An accurate diagnosis of FGIDs  
9 reduces the number of referrals for more expensive, unnecessary and invasive investigations  
10 such as colonoscopy. An audit of colonoscopies performed in one Australian hospital found  
11 that 12% of colonoscopies were for patients with functional gastrointestinal symptoms,<sup>(4325)</sup>  
12 whilst another Australian study found that 79% of patients seen in a public hospital diagnosed  
13 with IBS underwent an endoscopic procedure at an overall (includes indirect costs in addition  
14 to the colonoscopy) cost of \$85 million AUD with an additional \$2.4 million AUD/year in  
15 Emergency Department presentations.<sup>(4325, 4426)</sup> ~~These costs estimates differ in the literature,~~  
16 ~~however-~~ The introduction of FC has consistently demonstrated significant cost savings as  
17 compared to colonoscopies,<sup>(4325, 4527)</sup> Reducing this referral burden of functional  
18 gastrointestinal symptoms to public health secondary care allows valuable clinical resources to  
19 be used more effectively and appropriately.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34 Our analysis shows that FC is clinically effective in a primary care setting to distinguish ~~IBD~~  
35 ~~and organic bowel disease as a whole~~ OGIDs (including IBD) from FGIDs and would strongly  
36 encourage its wider implementation as a screening test in symptomatic patients to determine  
37 whether referral to specialist care is warranted in accordance with current Australian and  
38 international guidelines.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

References

1. Jones R. Primary care research and clinical practice: gastroenterology. *Postgrad Med J.* 2008;84(995):454-8.

2. Arasardnam RP, Brown S, Forbes A, Fox MR, Hungin P, Kelman L, et al. Guidelines for the investigation of chronic diarrhoea in adults: British Society of Gastroenterology, 3<sup>rd</sup> edition. *Gut.* 2018;67(8):1380-99.

3. Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol.* 2016;1(2):133-46.

4. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol.* 2012;10(7):712-21 e4.

5. Linedale EC, Andrews JM. Diagnosis and management of irritable bowel syndrome: a guideline for the generalist. *Med J Aust.* 2017;207(7):309-15.

6. PricewaterhouseCoopers Australia (PwC). Improving Inflammatory Bowel Disease care across Australia. 2013. <https://www.crohnscolitis.com.au/site/wp-content/uploads/PwC-report-2013.pdf> (accessed Sep 2018).

7. Gastroenterological Society of Australia (GESA) Digestive Health Foundation. Irritable Bowel Syndrome 2010. <http://cart.gesa.org.au/members/files/Consumer%20Information/IBS.pdf> (accessed Oct 2018).

8. American College of Gastroenterology Task Force on Irritable Bowel Syndrome, Brandt LJ, Chey WD, Foxx-Orenstein AE, Schiller LR, Schoenfeld PS, et al. An evidence-based position statement on the management of irritable bowel syndrome. *Am J Gastroenterol.* 2009;104 Suppl 1:S1-35.

9. NICE guidance [DG11]. Faecal calprotectin diagnostic tests for inflammatory disease of the bowel. 2013. <https://www.nice.org.uk/guidance/dg11> (accessed Oct 2018).

10. Fukudo S, Kaneko H, Akiho H, Inamori M, Endo Y, Okumura T, et al. Evidence-based clinical practice guidelines for irritable bowel syndrome. *J Gastroenterol.* 2015;50(1):11-30.

11. Shin JE, Jung HK, Lee TH, Jo Y, Lee H, Song KH, et al. Guidelines for the Diagnosis and Treatment of Chronic Functional Constipation in Korea, 2015 Revised Edition. *J Neurogastroenterol Motil.* 2016;22(3):383-411.

12. Carmona-Sanchez R, Icaza-Chavez ME, Bielsa-Fernandez MV, Gomez-Escudero O, Bosques-Padilla F, Coss-Adame E, et al. The Mexican consensus on irritable bowel syndrome. *Rev Gastroenterol Mex.* 2016;81(3):149-67.

13. Linedale EC, Shahzad MA, Kellie AR, Mikocka-Walus A, Gibson PR, Andrew JM. Referrals to a tertiary hospital: A window into clinical management issues in functional gastrointestinal disorders. JGH Open. 2017;1(3):84-91.
14. Chey WD, Nojkov B, Rubenstein JH, Dobhan RR, Greenson JK, Cash BD. The yield of colonoscopy in patients with non-constipated irritable bowel syndrome: results from a prospective, controlled US trial. Am J Gastroenterol. 2010; 105(4):859-65.
15. Rizvi Q LE, Mikocka-Walus A, Gibson PR, Andrews JM. Can we better target colonoscopies using standard appropriateness guides. J Gastroenterol Hepatol. 2015;30(Suppl 3):58.
16. Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel disease: systematic review and economic evaluation. Health Technology Assessment. 2013;17(55):xv-xix,1-211.
17. Jellema P, van Tulder MW, van der Horst HE, Florie J, Mulder CJ, vander Windt DA. Inflammatory bowel disease: a systematic review on the value of diagnostic testing in primary care. Colorectal Dis. 2011;13(3):239-54.
18. Preferred Items for Systematic Reviews and Meta-analysis [PRISMA] guidelines [website]. <http://www.prisma-statement.org> (accessed Sep 2018).
19. Macaskill P, Gatsonic C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. [https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/Chapter %2010%20-%20Version%201.0.pdf](https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/Chapter%2010%20-%20Version%201.0.pdf) (accessed Apr 2019).
20. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odd ratio: A single indicator of test performance. J Clin Epidemiol. 2003;56(11):1129-35.
21. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.
22. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982-90.
23. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
24. Limburg PJ, Ahlquist DA, Sandborn WJ, Mahoney DW, Devens ME, Harrington JJ, et al. Fecal calprotectin levels predict colorectal inflammation among patients with chronic diarrhea referred for colonoscopy. Am J Gastroenterol. 2000;95(10):2831-7.



25. [Carrocio A. 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\(6\):861-7.](#)
26. [D'Inca R, Dal Pont E, Di Leo V, Ferronato A, Fries W, Vettorato MG, et al. Calprotectin and lactoferrin in the assessment of intestinal inflammation and organic disease. Int J Colorectal Dis. 2007;22\(4\):429-37.](#)
27. [Otten CM, Kok L, Witterman BJ, Baumgarten R, Kampman E, Moons KG, 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\(9\):1275-80.](#)
28. [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\(10\):985-92.](#)
29. [Schoepfer AM, Trummel M, Seeholzer P, Seibold-Schmid B, Seibold F. Discriminating IBD from IBS: comparison of the test performance of fecal markers, blood leukocytes, CRP and IBD antibodies. Inflamm Bowel Dis. 2008;14\(1\):32-9.](#)
30. [El-Badry A, Sedrak H, Rashed L. Faecal calprotectin in differentiating between functional and organic bowel diseases. Arab Journal of Gastroenterology. 2010;11\(2\):70-3.](#)
31. [Kok L, Elias SG, Witteman BJM, Goedhard JG, Muris JWM, Moons KGM, et al. Diagnostic Accuracy of Point-of-Care Fecal Calprotectin and Immunochemical Occult Blood Test 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\(6\):989-98.](#)
32. [Burri E, Manz M, Rothen C, Rossi L, Beglinger C, Lehmann FS. 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.](#)
33. [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\(9\):1048-54.](#)
34. [Lozoya Angulo ME, de Las Heras Gomez I, Martinez Villanueva M, Noguera Velasco JA, Aviles Plaza F. Faecal calprotectin, an useful marker in discriminating between inflammatory bowel disease and functional gastrointestinal disorders. Gastroenterol Hepatol. 2017;40\(3\):125-31.](#)
35. [Dhaliwal A, Zeino Z, Tomkins C, Cheung M, Nwokolo C, Smith S, et al. Utility of faecal calprotectin in inflammatory bowel disease \(IBD\): what cut-offs should we apply? Frontline Gastroenterol. 2015;6\(1\):14-9.](#)



36. Banerjee A, Srinivas M, Eyre R, Ellis R, Waugh N, Bardhan KD, 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(1):20-6.
37. Turvill J, O'Connell S, Brooks A, Bradley-Wood K, Laing J, Thiagarajan S, et al. Evaluation of a faecal calprotectin care pathway for use in primary care: *Prim Health Care Res Dev.* 2016;17(05):428-36.
38. Hogberg C, Karling P, Rutegard J, Lilja M. 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(1):69-75.
39. Conroy S, Hale MF, Cross SS, Swallow K, Sidhu RH, Sargur R, 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(4):316-22.
40. Walker GJ, Moore L, Heerasing N, Hendy P, Perry MH, McDonald TJ, et al. Faecal calprotectin effectively excludes inflammatory bowel disease in 789 symptomatic young adults with/without alarm symptoms: a prospective UK primary care cohort study. *Aliment Pharmacol Ther.* 2018;47(8):1103-16.
41. Turvill J, Turnock D, Holmes H, Jones A, McLaughlan E, Hilton V, et al. Evaluation of the clinical and cost-effectiveness of the York Faecal Calprotectin Care Pathway. *Frontline Gastroenterol.* 2018;9(4):285-94.
42. Linedale EC, Mikocka-Walus A, Vincent AD, Gibson PR, Andrews JM. Performance of an algorithm-based approach to the diagnosis and management of functional gastrointestinal disorders: A pilot trial. *Neurogastroenterol Motil.* 2018;30(1).
43. Linedale EC. Issues in the Diagnosis and Management of Functional Gastrointestinal Disorders: The Development of a Novel Clinical Pathway. The University of Adelaide;2017. <https://digital.library.adelaide.edu.au/dspace/bitstream/2440/114587/2/02whole.pdf> (accessed Sep 2018).
44. Linedale EC, Chur-Hansen A, Mikocka-Walus A, Gibson PR, Andrews JM. Uncertain Diagnostic Language Affects Further Studies, Endoscopies, and Repeat Consultations for Patients With Functional Gastrointestinal Disorders. *Clin Gastroenterol Hepatol.* 2016;14(12):1735-41.e1.
45. Rizvi Q, Linedale EC, Mikocka-Walus A, Gibson PR, Andrews JM. Can we better target colonoscopies using standard "appropriateness" guides?. *J Gastroenterol Hepatol* 2015;30(Suppl 3):58.
1. Jones R. Primary care research and clinical practice: gastroenterology. *Postgrad Med J.* 2008;84(995):454-8.
2. Arasaradnam RP, Brown S, Forbes A, Fox MR, Hungin P, Kelman L, et al. Guidelines for the investigation of chronic diarrhoea in adults: British Society of Gastroenterology. 3rd edition. *Gut.* 2018;67(8):1380-99.

3. — Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol*. 2016;1(2):133–46.
4. — Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol*. 2012;10(7):712–21 e4.
5. — Linedale EC, Andrews JM. Diagnosis and management of irritable bowel syndrome: a guide for the generalist. *Med J Aust*. 2017;207(7):309–15.
6. — PricewaterhouseCoopers Australia (PwC). Improving Inflammatory Bowel Disease care across Australia. 2013. <https://www.crohnsandcolitis.com.au/site/wp-content/uploads/PwC-report-2013.pdf> (accessed Sep 2018).
7. — Gastroenterological Society of Australia (GESA) Digestive Health Foundation. Irritable Bowel Syndrome. 2010. <http://cart.gesa.org.au/membes/files/Consumer%20Information/IBS.pdf> (accessed Oct 2018).
8. — American College of Gastroenterology Task Force on Irritable Bowel S, Brandt LJ, Chey WD, Foxx-Orenstein AE, Schiller LR, Schoenfeld PS, et al. An evidence-based position statement on the management of irritable bowel syndrome. *Am J Gastroenterol*. 2009;104 Suppl 1:S1–35.
9. — NICE guidance [DG11]. Faecal calprotectin diagnostic tests for inflammatory disease of the bowel. 2013. <https://www.nice.org.uk/guidance/dg11> (accessed Oct 2018).
10. — Fukudo S, Kaneko H, Akiho H, Inamori M, Endo Y, Okumura T, et al. Evidence-based clinical practice guidelines for irritable bowel syndrome. *J Gastroenterol*. 2015;50(1):11–30.
11. — Shin JE, Jung HK, Lee TH, Jo Y, Lee H, Song KH, et al. Guidelines for the Diagnosis and Treatment of Chronic Functional Constipation in Korea, 2015 Revised Edition. *J Neurogastroenterol Motil*. 2016;22(3):383–411.
12. — Carmona-Sanchez R, Icaza-Chavez ME, Bielsa-Fernandez MV, Gomez-Escudero O, Bosques-Padilla F, Coss-Adame E, et al. The Mexican consensus on irritable bowel syndrome. *Rev Gastroenterol Mex*. 2016;81(3):149–67.
13. — Linedale EC, Shahzad MA, Kellie AR, Mikočka-Walus A, Gibson PR, Andrew JM. Referrals to a tertiary hospital: A window into clinical management issues in functional gastrointestinal disorders. *JGH Open*. 2017;1(3):84–91.
14. — Chey WD, Nojkov B, Rubenstein JH, Dobhan RR, Greenston JK, Cash BD. The yield of colonoscopy in patients with non-constipated irritable bowel syndrome: results from a prospective, controlled US trial. *Am J Gastroenterol*. 2010;105(4):859–65.
15. — Rizvi Q LE, Mikočka-Walus A, Gibson PR, Andrews JM. Can we better target colonoscopies using standard appropriateness guides. *J Gastroenterol Hepatol*. 2015;30(Suppl 3):58.
16. — Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technology Assessment*. 2013;17(55):xv–xix,1–211.
17. — Jellema P, van Tulder MW, van der Horst HE, Florie J, Mulder CJ, van der Windt DA. Inflammatory bowel disease: a systematic review on the value of diagnostic testing in primary care. *Colorectal Dis*. 2011;13(3):239–54.

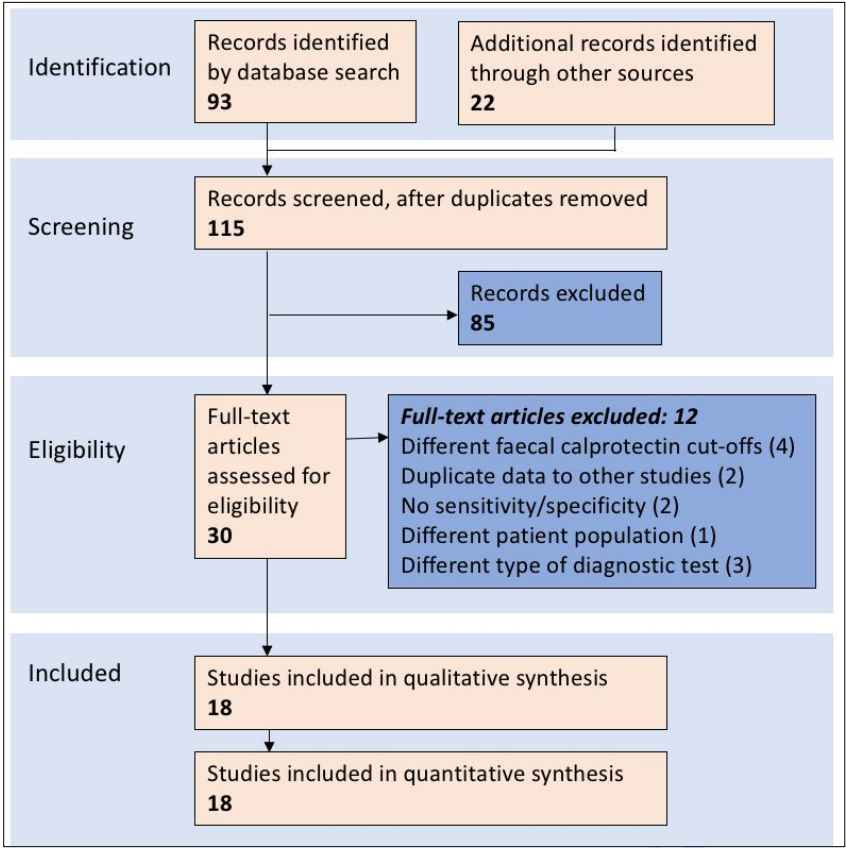
18. — Preferred Items for Systematic Reviews and Meta-analysis [PRISMA] guidelines [website]. <http://www.prisma-statement.org> (accessed Sep 2018).
19. — Macaskill P, Gatsonic C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonic C (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. <http://srdta.cochrane.org> (accessed Apr 2019).
20. — Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odd ratio: A single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129-35.
21. — DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88.
22. — Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.
23. — Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.
24. — Limburg PJ, Ahlquist DA, Sandborn WJ, Mahoney DW, Devens ME, Harrington JJ, et al. Fecal calprotectin levels predict colorectal inflammation among patients with chronic diarrhea referred for colonoscopy. *Am J Gastroenterol*. 2000;95(10):2831-7.
25. — Carroccio A. 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):861-7.
26. — D'Inca R, Dal Pont E, Di Leo V, Ferronato A, Fries W, Vettorato MG, et al. Calprotectin and lactoferrin in the assessment of intestinal inflammation and organic disease. *Int J Colorectal Dis*. 2007;22(4):429-37.
27. — Otten CM, Kok L, Witteman BJ, Baumgarten R, Kampman E, Moons KG, 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(9):1275-80.
28. — 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(10):985-92.
29. — Schoepfer AM, Trummler M, Seeholzer P, Seibold-Schmid B, Seibold F. Discriminating IBD from IBS: comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies. *Inflamm Bowel Dis*. 2008;14(1):32-9.
30. — El Badry A, Sedrak H, Rashed L. Faecal calprotectin in differentiating between functional and organic bowel diseases. *Arab Journal of Gastroenterology*. 2010;11(2):70-3.
31. — Kok L, Elias SG, Witteman BJM, Goedhard JG, Muris JWM, Moons KGM, 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(6):989-98.

- 32.— Burri E, Manz M, Rothen C, Rossi L, Beglinger C, Lehmann FS. 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.
- 33.— 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(9):1048-54.
- 34.— Lozoya Angulo ME, de Las Heras Gomez I, Martinez Villanueva M, Noguera Velasco JA, Aviles Plaza F. Faecal calprotectin, an useful marker in discriminating between inflammatory bowel disease and functional gastrointestinal disorders. *Gastroenterol Hepatol*. 2017;40(3):125-31.
- 35.— Dhaliwal A, Zeino Z, Tomkins C, Cheung M, Nwokolo C, Smith S, et al. Utility of faecal calprotectin in inflammatory bowel disease (IBD): what cut-offs should we apply? *Frontline Gastroenterol*. 2015;6(1):14-9.
- 36.— Banerjee A, Srinivas M, Eyre R, Ellis R, Waugh N, Bardhan KD, 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(1):20-6.
- 37.— Turvill J, O'Connell S, Brooks A, Bradley-Wood K, Laing J, Thiagarajan S, et al. Evaluation of a faecal calprotectin care pathway for use in primary care. *Prim Health Care Res Dev*. 2016;17(05):428-36.
- 38.— Hogberg C, Karling P, Rutegard J, Lilja M. 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(1):69-75.
- 39.— Conroy S, Hale MF, Cross SS, Swallow K, Sidhu RH, Sargur R, 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(4):316-22.
- 40.— Walker GJ, Moore L, Heerasing N, Hendy P, Perry MH, McDonald TJ, et al. Faecal calprotectin effectively excludes inflammatory bowel disease in 789 symptomatic young adults with/without alarm symptoms: a prospective UK primary care cohort study. *Aliment Pharmacol Ther*. 2018;47(8):1103-16.
- 41.— Turvill J, Turnock D, Holmes H, Jones A, McLaughlan E, Hilton V, et al. Evaluation of the clinical and cost-effectiveness of the York Faecal Calprotectin Care Pathway. *Frontline Gastroenterol*. 2018;9(4):285-94.
- 42.— Linedale EC, Mikočka-Walus A, Vincent AD, Gibson PR, Andrews JM. Performance of an algorithm-based approach to the diagnosis and management of functional gastrointestinal disorders: A pilot trial. *Neurogastroenterol Motil*. 2018;30(1). doi:10.1111/nmo.13243.
- 43.— Linedale EC. Issues in the Diagnosis and Management of Functional Gastrointestinal Disorders: The Development of a Novel Clinical Pathway Adelaide The University of Adelaide; 2017. <https://digital.library.adelaide.edu.au/dspace/bitstream/2440/114587/2/02whole.pdf> (accessed Sep 2018).

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
44. Linedale EC, Chur Hansen A, Mikočka-Walus A, Gibson PR, Andrews JM. Uncertain Diagnostic Language Affects Further Studies, Endoscopies, and Repeat Consultations for Patients With Functional Gastrointestinal Disorders. Clin Gastroenterol Hepatol. 2016;14(12):1735-41.e1.
45. Rizvi Q, Linedale EC, Mikočka-Walus A, Gibson PR, Andrews JM. Can we better target colonoscopies using standard "appropriateness" guides? . J Gastroenterol Hepatol. 2015;30(Suppl 3):58.

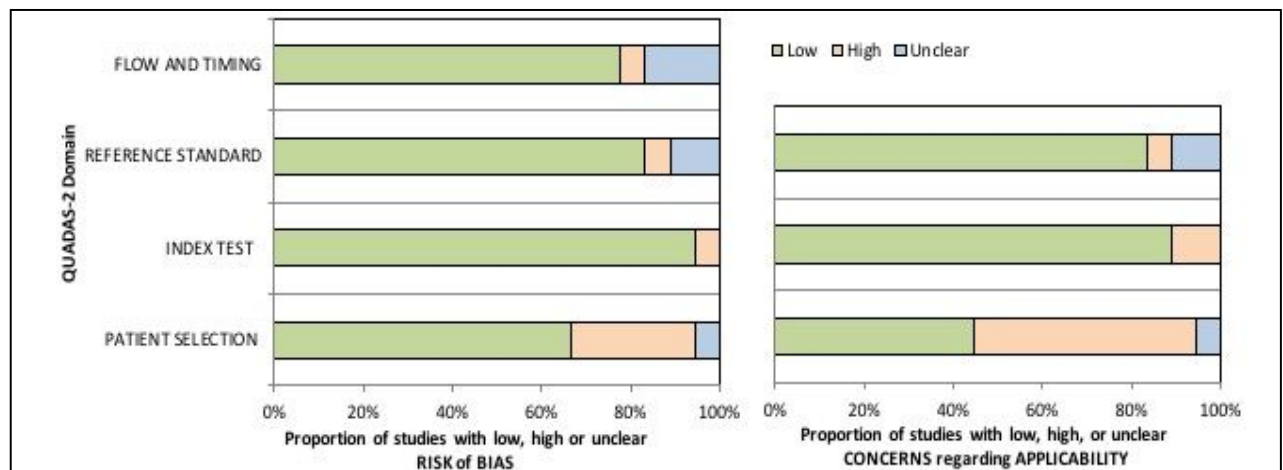
For Review Only

**Figure 1:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of study selection and QUADAS-2 quality assessment of included studies [A]





[B]



Feedback: Please supply data to allow-redrawing in MJA style.

#### RISK of BIAS

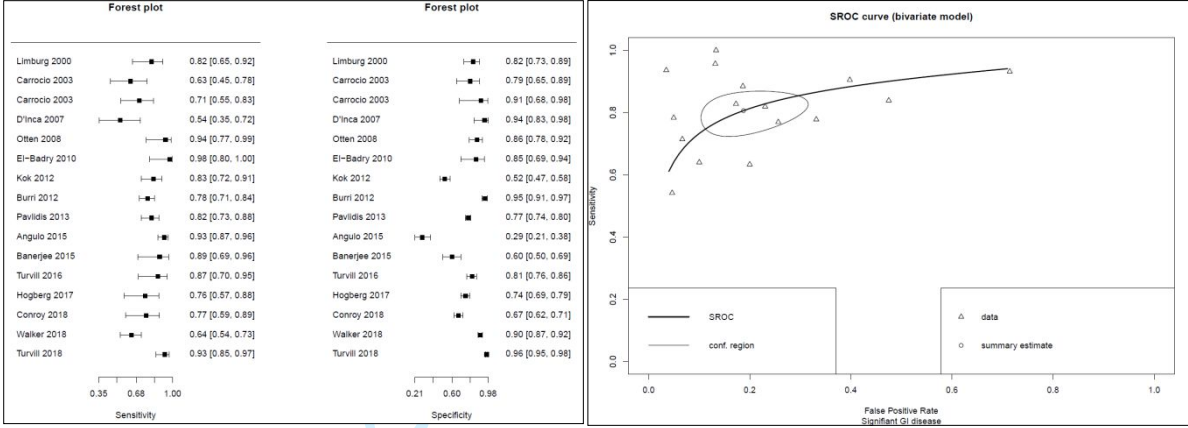
|         | PATIENT SELECTION | INDEX TEST | REFERENCE STANDARD | FLOW AND TIMING |
|---------|-------------------|------------|--------------------|-----------------|
| Low     | 12                | 17         | 15                 | 14              |
| High    | 5                 | 1          | 1                  | 1               |
| Unclear | 1                 | 0          | 2                  | 3               |

#### CONCERNS regarding APPLICABILITY

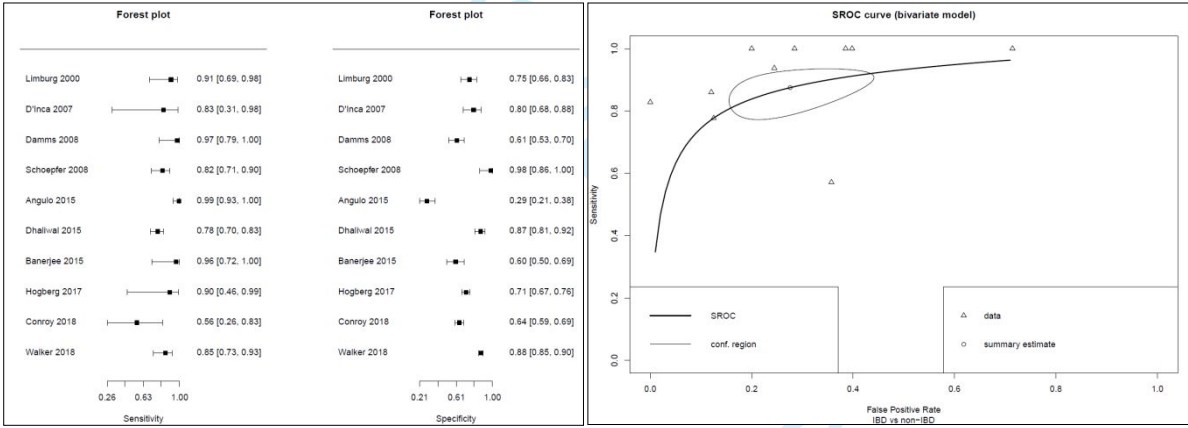
|         | PATIENT SELECTION | INDEX TEST | REFERENCE STANDARD |
|---------|-------------------|------------|--------------------|
| Low     | 8                 | 16         | 15                 |
| High    | 9                 | 2          | 1                  |
| Unclear | 1                 | 0          | 2                  |

**Figure 2:** Forest plot and SROC bivariate model [A] distinguishing OGIDs from FGIDs [B] distinguishing IBD from non-IBD [C] diagnostic accuracy of FC

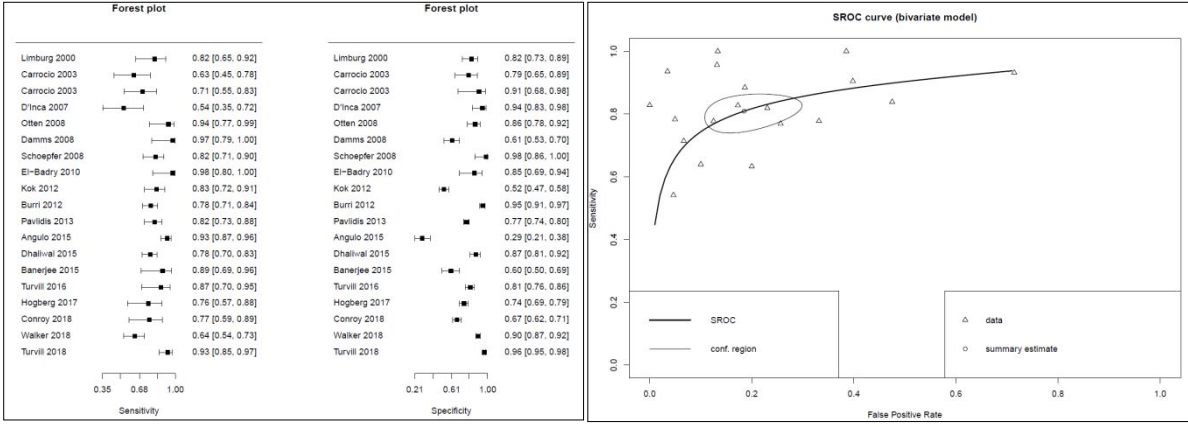
[A]



[B]

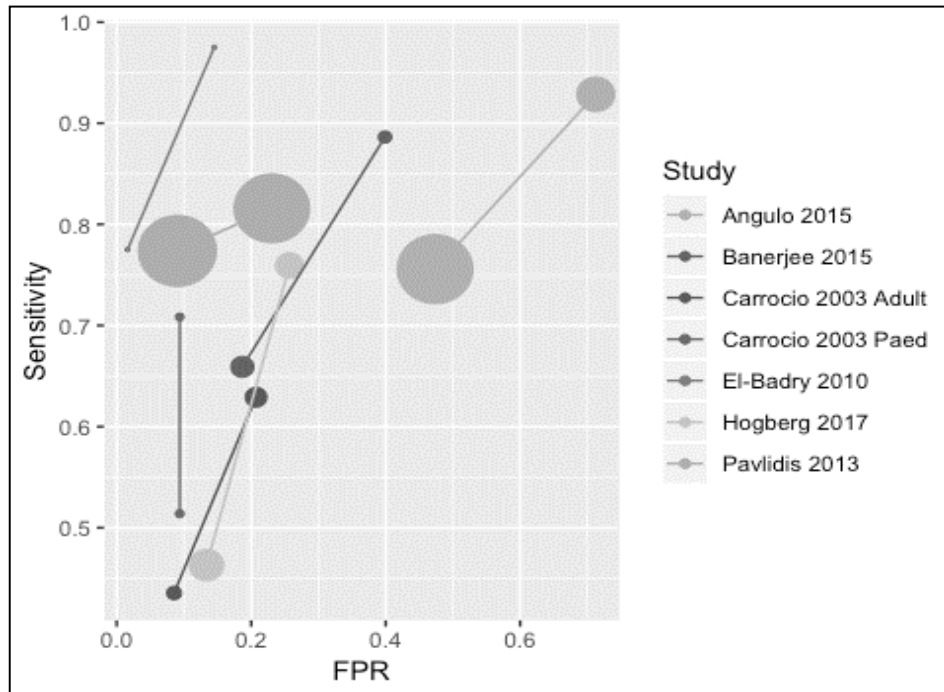


[C]



**Figure 3: [A]** Comparison of diagnostic accuracy and **[B]** diagnostic odds ratio between FC cut-offs\*

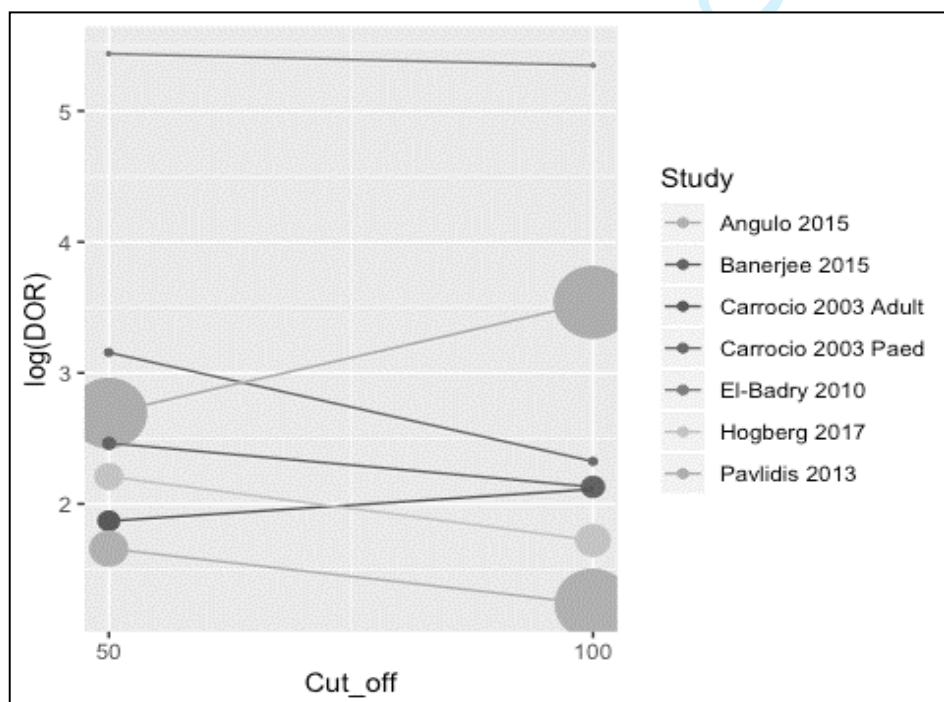
**[A]**



Nodes with lower sensitivity and FPR represent a cut-off of 100  $\mu\text{g/g}$

\* The area of the points is proportional to the inverse variance of the estimate.

**[B]**



\* The area of the points is proportional to the inverse variance of the estimate.

Figure 3 in Tabular form

| Study               | FC cut-off (µg/g) | Sn   | FPR  | DOR   | logDOR |
|---------------------|-------------------|------|------|-------|--------|
| Angulo 2015         | 50                | 0.93 | 0.71 | 5.3   | 0.40   |
| Angulo 2015         | 100               | 0.76 | 0.47 | 3.4   | 0.28   |
| Banerjee 2015       | 50                | 0.89 | 0.40 | 11.7  | 0.70   |
| Banerjee 2015       | 100               | 0.66 | 0.19 | 8.4   | 0.52   |
| Carrocio 2003 Adult | 50                | 0.63 | 0.21 | 6.5   | 0.54   |
| Carrocio 2003 Adult | 100               | 0.44 | 0.09 | 8.3   | 0.67   |
| Carrocio 2003 Paed  | 50                | 0.71 | 0.09 | 23.5  | 0.93   |
| Carrocio 2003 Paed  | 100               | 0.51 | 0.09 | 10.2  | 0.92   |
| El-Badry 2010       | 50                | 0.98 | 0.15 | 229.7 | 1.52   |
| El-Badry 2010       | 100               | 0.78 | 0.02 | 210.1 | 1.52   |
| Hogberg 2017        | 50                | 0.76 | 0.26 | 9.1   | 0.47   |
| Hogberg 2017        | 100               | 0.46 | 0.13 | 5.6   | 0.42   |
| Pavlidis 2013       | 50                | 0.82 | 0.23 | 14.8  | 0.28   |
| Pavlidis 2013       | 100               | 0.77 | 0.09 | 34.4  | 0.27   |

Table 1: Population characteristics of included studies

| Study                          | No. of patients | Mean age (yrs)              | Gender (% female) | Setting                     | Population |
|--------------------------------|-----------------|-----------------------------|-------------------|-----------------------------|------------|
| Limburg 2000 <sup>(24)</sup>   | 110             | 57                          | 0.64              | Secondary care, USA         | Adult      |
| Carrocio 2003 <sup>(25)</sup>  | 70              | 35                          | 0.57              | Secondary care, Italy       | Adult      |
|                                | 50              | 3.5                         | 0.6               | Secondary care, Italy       | Children   |
| D'Inca 2007 <sup>(26)</sup>    | 67              | 49                          | 0.58              | Secondary care, Italy       | Adults     |
| Otten 2008 <sup>(27)</sup>     | 114             | 51                          | 0.54              | Secondary care, Netherland  | Adults     |
| Damms 2008 <sup>(28)</sup>     | 140             | 58                          | 0.56              | Secondary care, Germany     | Adults     |
| Schoepfer 2008 <sup>(29)</sup> | 136             | 40                          | 0.6               | Secondary care, Bern        | Adults     |
| El-Badry 2010 <sup>(30)</sup>  | 49              | 37                          | 0.47              | Secondary care, Egypt       | Adult      |
| Kok 2012 <sup>(31)</sup>       | 382             | 60                          | 0.55              | Primary care, Netherlands   | Adults     |
| Burri 2012 <sup>(32)</sup>     | 405             | 63                          | 0.56              | Secondary care, Switzerland | Adult      |
| Pavlidis 2013 <sup>(33)</sup>  | 962             | 33                          | 0.6               | Primary care, UK            | Adult      |
| Angulo 2015 <sup>(34)</sup>    | 264             | Adults – 43<br>Children – 7 | 0.54              | Secondary care, Spain       | Mixed      |
| Dhaliwal 2015 <sup>(35)</sup>  | 311             | n/a                         | 0.67              | Secondary care, UK          | Adult      |
| Banerjee 2015 <sup>(36)</sup>  | 119             | 46                          | 0.54              | Secondary care, UK          | Adult      |
| Turvill 2016 <sup>(37)</sup>   | 262             | 37                          | 0.7               | Primary care, UK            | Adult      |
| Hogberg 2017 <sup>(38)</sup>   | 373             | 63                          | 0.65              | Primary care, Sweden        | Adult      |
| Conroy 2018 <sup>(39)</sup>    | 410             | 42                          | 0.61              | Primary care, UK            | Adults     |
| Walker 2018 <sup>(40)</sup>    | 789             | 30                          | 0.6               | Primary care, UK            | Adults     |
| Turvill 2018 <sup>(41)</sup>   | 1005            | 38                          | 0.63              | Primary care, UK            | Adults     |

Table 2: Summary statistics

| Accuracy                           | Overall        | IBD vs <del>non-IBD</del> FGIDs | OGIDs vs FGIDs |
|------------------------------------|----------------|---------------------------------|----------------|
| Sensitivity, % (95% CI)            | 80 (73-85%)    | 88 (80-93%)                     | 81 (74-86%)    |
| False positive rate, % (95% CI)    | 19 (14-30%)    | 28 (18-41%)                     | 19 (12-29%)    |
| Log Diagnostic odds ratio (95% CI) | 3.1 (2.5, 3.6) | 3.2 (2.5, 3.9)                  | 3.0 (2.4, 3.6) |
| Area under the curve               | 0.86           | 0.89                            | 0.87           |

**Supplement Table 1:** Studies reporting test characteristics of faecal calprotectin for detecting OGIDs

|                                    | Study         | FC cut-off (µg/g) | TP  | FN | TN  | FP  | Disease prevalence (%) |
|------------------------------------|---------------|-------------------|-----|----|-----|-----|------------------------|
|                                    | Limburg 2000  | 100               | 24  | 5  | 67  | 14  | 26.36 (18.42-35.62)    |
|                                    | Carrocio 2003 | 50 (Adults)       | 19  | 11 | 32  | 8   | 42.86 (31.09-55.25)    |
|                                    | Carrocio 2003 | 50 (Children)     | 25  | 10 | 14  | 1   | 70.00 (55.39-82.14)    |
|                                    | D'Inca 2007   | 50                | 13  | 11 | 41  | 2   | 35.82 (24.47-48.47)    |
|                                    | Otten 2008    | 50                | 22  | 1  | 79  | 12  | 20.18 (13.24-28.72)    |
|                                    | El-Badry 2010 | 50                | 19  | 0  | 26  | 4   | 38.8 (25.20-53.76)     |
| Included in statistical analysis   | Kok 2012      | 50                | 52  | 10 | 168 | 152 | 16.23 (12.68-20.32)    |
|                                    | Burri 2012    | 50                | 112 | 31 | 249 | 13  | 35.31 (30.65-40.18)    |
|                                    | Pavlidis 2013 | 50                | 77  | 17 | 668 | 200 | 9.77 (7.97-11.82)      |
|                                    | Angulo 2015   | 50                | 123 | 9  | 32  | 80  | 54.10 (47.62-60.47)    |
|                                    | Banerjee 2015 | 50                | 19  | 2  | 59  | 39  | 17.65 (11.27-25.70)    |
|                                    | Turvill 2016  | 100               | 23  | 3  | 192 | 44  | 9.92 (6.59-14.20)      |
|                                    | Hogberg 2017  | 50                | 20  | 6  | 258 | 89  | 6.97 (4.60-10.05)      |
|                                    | Conroy 2018   | 50                | 21  | 6  | 256 | 127 | 6.59 (4.38-9.44)       |
|                                    | Walker 2018   | 100               | 64  | 36 | 620 | 69  | 12.67 (10.43-15.20)    |
|                                    | Turvill 2018  | 100               | 73  | 5  | 803 | 29  | 8.20 (6.54-10.13)      |
| Removed from statistical analysis* | Carrocio 2003 | 100 (Adults)      | 13  | 17 | 37  | 3   | 42.86 (31.09-55.25)    |
|                                    | Carrocio 2003 | 100 (Children)    | 18  | 17 | 14  | 1   | 70.00 (55.39-82.14)    |
|                                    | El-Badry 2010 | 100               | 15  | 4  | 30  | 0   | 38.78 (25.20-53.76)    |
|                                    | Pavlidis 2013 | 100               | 73  | 21 | 790 | 78  | 9.77 (7.97-11.82)      |
|                                    | Angulo 2015   | 100               | 100 | 32 | 59  | 53  | 54.10 (47.62-60.47)    |
|                                    | Banerjee 2015 | 100               | 14  | 7  | 80  | 18  | 17.65 (11.27-25.70)    |
|                                    | Hogberg 2017  | 100               | 12  | 14 | 301 | 46  | 6.97 (4.60-10.05)      |

\* When studies used both 50 and 100 µg/g cut-off values for the same patient, we only used the 50 µg/g value for this analysis.

**Supplement Table 2:** Studies reporting test characteristics of faecal calprotectin for detecting IBD

|                                    | Study          | FC cut-off (µg/g) | TP  | FN | TN  | FP  | Disease prevalence (%) |
|------------------------------------|----------------|-------------------|-----|----|-----|-----|------------------------|
|                                    | Limburg 2000   | 100               | 15  | 1  | 71  | 23  | 14.55 (8.55-22.54)     |
|                                    | D'Inca 2007    | 50                | 2   | 0  | 52  | 13  | 2.99 (0.36-10.37)      |
|                                    | Damms 2008     | 50                | 18  | 0  | 75  | 47  | 12.86 (7.80-19.56)     |
|                                    | Schoepfer 2008 | 50                | 53  | 11 | 30  | 0   | 68.09 (57.67-77.33)    |
| Included in statistical analysis   | Angulo 2015    | 50                | 68  | 0  | 32  | 80  | 37.78 (30.67-45.29)    |
|                                    | Dhaliwal 2015  | 50                | 115 | 33 | 126 | 18  | 50.68 (44.80-56.56)    |
|                                    | Banerjee 2015  | 50                | 12  | 0  | 59  | 39  | 10.91 (5.77-18.28)     |
|                                    | Hogberg 2017   | 50                | 4   | 0  | 264 | 105 | 1.07 (0.29-2.72)       |
|                                    | Conroy 2018    | 50                | 4   | 3  | 259 | 144 | 1.71 (0.69-3.49)       |
|                                    | Walker 2018    | 100               | 43  | 7  | 650 | 89  | 6.34 (4.74-8.27)       |
|                                    | Angulo 2015    | 100               | 60  | 8  | 63  | 49  | 37.78 (30.67-45.29)    |
| Removed from statistical analysis* | Dhaliwal 2015  | 100               | 113 | 35 | 140 | 4   | 50.68 (44.80-56.56)    |
|                                    | Banerjee 2015  | 100               | 11  | 1  | 80  | 18  | 10.91 (5.77-18.28)     |
|                                    | Hogberg 2017   | 100               | 3   | 1  | 314 | 55  | 1.07 (0.29-2.72)       |

\* When studies used both 50 and 100 µg/g cut-off values for the same patient, we only used the 50 µg/g value for this analysis

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1:** Population characteristics of included studies

| Study                          | No. of patients | Mean age (yrs)              | Gender (% female) | Setting                     | Population |
|--------------------------------|-----------------|-----------------------------|-------------------|-----------------------------|------------|
| Limburg 2000 <sup>(24)</sup>   | 110             | 57                          | 0.64              | Secondary care, USA         | Adult      |
| Carrocio 2003 <sup>(25)</sup>  | 70              | 35                          | 0.57              | Secondary care, Italy       | Adult      |
|                                | 50              | 3.5                         | 0.6               | Secondary care, Italy       | Children   |
| D'Inca 2007 <sup>(26)</sup>    | 67              | 49                          | 0.58              | Secondary care, Italy       | Adults     |
| Otten 2008 <sup>(27)</sup>     | 114             | 51                          | 0.54              | Secondary care, Netherland  | Adults     |
| Damms 2008 <sup>(28)</sup>     | 140             | 58                          | 0.56              | Secondary care, Germany     | Adults     |
| Schoepfer 2008 <sup>(29)</sup> | 136             | 40                          | 0.6               | Secondary care, Bern        | Adults     |
| El-Badry 2010 <sup>(30)</sup>  | 49              | 37                          | 0.47              | Secondary care, Egypt       | Adult      |
| Kok 2012 <sup>(31)</sup>       | 382             | 60                          | 0.55              | Primary care, Netherlands   | Adults     |
| Burri 2012 <sup>(32)</sup>     | 405             | 63                          | 0.56              | Secondary care, Switzerland | Adult      |
| Pavlidis 2013 <sup>(33)</sup>  | 962             | 33                          | 0.6               | Primary care, UK            | Adult      |
| Angulo 2015 <sup>(34)</sup>    | 264             | Adults – 43<br>Children – 7 | 0.54              | Secondary care, Spain       | Mixed      |
| Dhaliwal 2015 <sup>(35)</sup>  | 311             | n/a                         | 0.67              | Secondary care, UK          | Adult      |
| Banerjee 2015 <sup>(36)</sup>  | 119             | 46                          | 0.54              | Secondary care, UK          | Adult      |
| Turvill 2016 <sup>(37)</sup>   | 262             | 37                          | 0.7               | Primary care, UK            | Adult      |
| Hogberg 2017 <sup>(38)</sup>   | 373             | 63                          | 0.65              | Primary care, Sweden        | Adult      |
| Conroy 2018 <sup>(39)</sup>    | 410             | 42                          | 0.61              | Primary care, UK            | Adults     |
| Walker 2018 <sup>(40)</sup>    | 789             | 30                          | 0.6               | Primary care, UK            | Adults     |
| Turvill 2018 <sup>(41)</sup>   | 1005            | 38                          | 0.63              | Primary care, UK            | Adults     |

Table 1. Population characteristics of included studies

333x188mm (144 x 144 DPI)



Table 2: Summary statistics

| Accuracy                           | Overall        | IBD vs FGIDs   | OGIDs vs FGIDs |
|------------------------------------|----------------|----------------|----------------|
| Sensitivity, % (95% CI)            | 80 (73-85%)    | 88 (80-93%)    | 81 (74-86%)    |
| False positive rate, % (95% CI)    | 19 (14-30%)    | 28 (18-41%)    | 19 (12-29%)    |
| Log Diagnostic odds ratio (95% CI) | 3.1 (2.5, 3.6) | 3.2 (2.5, 3.9) | 3.0 (2.4, 3.6) |
| Area under the curve               | 0.86           | 0.89           | 0.87           |

Table 2. Summary statistics

401x105mm (144 x 144 DPI)

Figure 3 in Tabular form

| Study               | FC cut-off (µg/g) | Sn   | FPR  | DOR   | logDOR |
|---------------------|-------------------|------|------|-------|--------|
| Angulo 2015         | 50                | 0.93 | 0.71 | 5.3   | 0.40   |
| Angulo 2015         | 100               | 0.76 | 0.47 | 3.4   | 0.28   |
| Banerjee 2015       | 50                | 0.89 | 0.40 | 11.7  | 0.70   |
| Banerjee 2015       | 100               | 0.66 | 0.19 | 8.4   | 0.52   |
| Carrocio 2003 Adult | 50                | 0.63 | 0.21 | 6.5   | 0.54   |
| Carrocio 2003 Adult | 100               | 0.44 | 0.09 | 8.3   | 0.67   |
| Carrocio 2003 Paed  | 50                | 0.71 | 0.09 | 23.5  | 0.93   |
| Carrocio 2003 Paed  | 100               | 0.51 | 0.09 | 10.2  | 0.92   |
| El-Badry 2010       | 50                | 0.98 | 0.15 | 229.7 | 1.52   |
| El-Badry 2010       | 100               | 0.78 | 0.02 | 210.1 | 1.52   |
| Hogberg 2017        | 50                | 0.76 | 0.26 | 9.1   | 0.47   |
| Hogberg 2017        | 100               | 0.46 | 0.13 | 5.6   | 0.42   |
| Pavlidis 2013       | 50                | 0.82 | 0.23 | 14.8  | 0.28   |
| Pavlidis 2013       | 100               | 0.77 | 0.09 | 34.4  | 0.27   |

Table 3. Diagnostic odds ratio between FC cut-offs

294x139mm (144 x 144 DPI)