Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy

CRD summary
This review found that faecal calprotectin can discriminate between patients (adults and children) with and without inflammatory bowel disease, but it is a poor test for general screening for colorectal cancer. The authors’ conclusions are supported by the data, but limitations in the review in terms of the search, quality assessment and variation between the primary studies mean that they may not be reliable.

Authors’ objectives
To determine the accuracy of faecal calprotectin (FC) for the diagnosis of inflammatory bowel disease (IBD) and colorectal cancer (CRC) in adults and children.

Searching
MEDLINE, EMBASE and the Cochrane Library were searched from inception to March 2006; the search terms were reported. No language restrictions were applied. Additional studies were identified by using the ‘related articles’ function (database not specified).

Study selection
Study designs of evaluations included in the review
Studies that did not include a control group were excluded. It appears that both diagnostic cohort and diagnostic case-control studies were included. Both prospective and retrospective studies were included.

Specific interventions included in the review
Studies of FC were eligible for inclusion. Studies that measured FC following drug, dietary or surgical intervention were excluded, as were tissue studies. Studies that assessed technical aspects of the calprotectin assay were also excluded. The included studies evaluated both the old and new FC assays. The thresholds assessed in the included studies were 18.6, 30, 31.5, 50, 90, 100, 150 or 250 microg/g.

Reference standard test against which the new test was compared
Studies that included histological diagnosis of Crohn's disease (CD), ulcerative colitis (UC) or CRC as the reference standard were eligible for inclusion. Studies that included radio-labelled white cell scanning or clinical indices of CD and UC as the reference standard were also eligible for inclusion.

Participants included in the review
Studies in both adults and children were included. Studies that assessed neonates were excluded. The studies included patients with suspected IBD, known chronic IBD, IBD in clinical remission, IBS, UC, CD, active IBD, constipation, other gastrointestinal diagnoses or symptoms, chronic diarrhoea or known colorectal malignancy, asymptomatic patients, relatives or spouses of patients with CD or CRC, patients undergoing colorectal cancer screening, and patients referred for colonoscopy, upper or lower gastrointestinal endoscopy, or radio-labelled white-cell scanning.

Outcomes assessed in the review
Inclusion criteria were not defined in terms of the outcomes. The outcomes reported in the review were the sensitivity, specificity, diagnostic odds ratio and area under the summary receiver operating characteristic (SROC) curve for the diagnosis of CD, UC and colorectal neoplasms. Weighted mean differences in FC levels between the various diagnostic groups were also reported.

How were decisions on the relevance of primary studies made?
The authors did not state how the studies were selected for the review, or how many reviewers performed the selection.
Assessment of study quality
The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to assess study quality. The studies were assigned a score of 1 if a criterion was fulfilled, 0 if it was unclear, and -1 if it was not fulfilled. The authors did not state how many reviewers performed the validity assessment.

Data extraction
Two reviewers independently extracted the data; any disagreements were resolved though referral to a third reviewer. A new assay of FC was introduced in 2000. To allow comparison of results obtained from the old assay with those obtained from the new assay, FC values obtained from studies using the original assay were multiplied by a factor of five. Data were extracted as 2x2 tables of test performance and used to estimate the sensitivity and specificity at various thresholds. The mean difference in FC levels between different diagnoses were calculated.

Methods of synthesis
How were the studies combined?
Pooled sensitivity, specificity and weighted mean differences were calculated using random-effects models. An SROC analysis was carried out to assess the correlation between sensitivity and specificity, and the area under the SROC curve was calculated. The diagnostic odds ratio was calculated from data on sensitivity and specificity. Publication bias was assessed using a funnel plot of the weighted mean difference.

How were differences between studies investigated?
Heterogeneity was assessed statistically using the Q statistic. Subgroup analysis and weighted meta-regression were used to assess the difference in accuracy of FC between adults and children and to compare two thresholds for positivity (50 and 100 microg/g). A sensitivity analysis was carried out to investigate the effects of quality (QUADAS score >11) and sample size (>100 patients) on estimates of sensitivity, specificity and diagnostic odds ratio.

Results of the review
Thirty studies were included. A total of 5,983 patients were included: 663 with CD, 361 with UC, 186 with IBD, 297 with CRC, 697 with IBS and 3,393 healthy controls.

The summary quality scores ranged from 10 to 13. The results of the individual quality assessment were not reported.

All results below refer to a FC threshold of 50 microg/g unless otherwise stated.

IBD compared with no IBD (9 studies, 1,267 patients): the pooled sensitivity and pooled specificity were 89% (95% confidence interval, CI: 86, 91) and 81% (95% CI: 78, 84), respectively. There was strong evidence of heterogeneity (p<0.001). The pooled sensitivity and specificity were both greater at a threshold of 100 microg/g but only 4 studies assessed this threshold.

CD compared with normal controls and IBS (5 studies, 733 patients): the pooled sensitivity and pooled specificity were 95% (95% CI: 92, 97) and 84% (95% CI: 80, 87), respectively. There was some evidence of statistical heterogeneity (p=0.07). Both sensitivity and specificity were greater at a threshold of 100 microg/g but only 2 studies, both restricted to children, assessed this threshold.

UC compared with normal controls and IBS (2 studies, 235 patients): the pooled sensitivity and pooled specificity were 78% (95% CI: 69, 86) and 78% (95% CI: 70, 84), respectively. There was no evidence of statistical heterogeneity (p=0.44).

CRC and adenoma compared with no neoplasia, including inflammation (7 studies, 4,112 patients): the pooled sensitivity and pooled specificity were 36% (95% CI: 34, 39) and 71% (95% CI: 70, 73), respectively. There was strong evidence of heterogeneity (p<0.001).

Adenoma versus no neoplasia, excluding inflammation (3 studies, 547 patients): the pooled sensitivity and pooled
specificity were 52% (95% CI: 42, 61) and 77% (95% CI: 73, 81), respectively. There was strong evidence of statistical heterogeneity (p<0.001). Sensitivity was lower and specificity higher in studies that used a threshold of 100 microg/g.

CRC versus no neoplasia, excluding inflammation (4 studies, 2,025 patients): the pooled sensitivity and pooled specificity were 87% (95% CI: 77, 94) and 76% (95% CI: 74, 78), respectively. There was strong evidence of statistical heterogeneity (p=0.007).

The exclusion of lower quality studies increased sensitivity and either had no effect on specificity or decreased specificity. The restriction to large studies had a similar effect. None of the variables investigated in the meta-regression analysis showed a significant association with the diagnostic odds ratio.

The funnel plot suggested a lack of publication bias.

**Authors' conclusions**
FC has the potential to discriminate between patients (both adults and children) with and without IBD. The accuracy of FC seems better at a threshold of 100 microg/g than at a threshold of 50 microg/g. FC is a poor test for general screening for CRC.

**CRD commentary**
The review addressed a focused question that was supported by defined inclusion criteria. A reasonable literature search was carried out, although attempts to identify unpublished studies were not made. The review may therefore be subject to publication bias, although this was assessed in the review and found to be absent. An appropriate quality assessment was carried out, but the results for individual quality items were not reported and a summary quality score was calculated instead. This is the only information relating to the quality assessment that was reported, therefore it is not possible to comment on the validity of the included studies. The authors reported that the data extraction was carried out independently, thus avoiding bias in this stage of the review, but there was no information on whether appropriate steps were also taken to avoid bias in other stages of the review process.

The methods used to pool the studies were not discussed in detail, so it is not possible to determine if these were appropriate. The pooled accuracy estimates presented should be interpreted with caution given the heterogeneity in the individual study estimates. In particular, pooled estimates based on small numbers of heterogeneous studies should be interpreted with extreme caution. It is also unclear whether some of the differences that the authors reported, in particular in relation to FC threshold, reflect true differences in accuracy or the result of heterogeneity between the studies. The authors' conclusions are generally supported by the data but should be interpreted with some degree of caution, especially those relating to the better accuracy of a threshold of 100 microg/g.

**Implications of the review for practice and research**
Practice: The authors stated that FC cannot be recommended as a screening tool for CRC. FC has potential as a noninvasive test for monitoring disease activity in IBD.

Research: The authors stated that further prospective studies that focus on a single clinical question and report FC accuracy at different thresholds are required.

**Bibliographic details**

**PubMedID**
17324124
DOI
10.1111/j.1572-0241.2007.01126.x

Indexing Status
Subject indexing assigned by NLM

MeSH
Biomarkers /analysis; Colorectal Neoplasms /diagnosis; Feces /chemistry; Female; Humans; Inflammatory Bowel Diseases /diagnosis; Leukocyte L1 Antigen Complex /analysis; Male; Predictive Value of Tests; Reference Values; Regression Analysis; Sensitivity and Specificity

AccessionNumber
12007001290

Date bibliographic record published
29/02/2008

Date abstract record published
29/02/2008

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.