JPGN Journal of Pediatric Gastroenterology and Nutrition Publish Ahead of Print

DOI : 10.1097/MPG.0000000000000615

Faecal calprotectin in suspected paediatric inflammatory bowel disease: an individual patient data meta-analysis

Pieter LJ **Degraeuwe**[1], Monique PA **Beld**[1], Merja **Ashorn**[2], Roberto **Berni Canani**[3], Andrew S **Day**[4], Antonella **Diamanti**[5], Ulrika L **Fagerberg**[6], Paul Henderson[7], Kaija-Leena **Kolho**[8], Els **Van de Vijver**[9], Patrick F **van Rheenen**[10], David C **Wilson**[7], Alfons GH **Kessels**[11]

[1] Department of Paediatrics, Maastricht University Medical Centre, Maastricht, The Netherlands

[2] Department of Paediatrics, Tampere University Hospital, Tampere, Finland

[3] Department of Paediatrics and European Laboratory for the Investigation of Food Induced Diseases, University of Naples "Federico II", Naples, Italy

[4] Paediatric Gastroenterology, Sydney Children's Hospital, Randwick, Australia; and University of Otago (Christchurch), Christchurch, New Zealand

[5] Hepathology, Gastroenterology and Nutrition Unit, Bambino Gesù Children's Hospital, Rome, Italy

[6] Dept. of Paediatrics, Centre for Clinical Research, Västmanlands Hospital, Västerås, Karolinska Institutet, Stockholm, Sweden

[7] Department of Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Edinburgh, and Child Life and Health, University of Edinburgh, UK

[8] Hospital for Children and Adolescents, University of Helsinki, Helsinki, Finland

[9] Department of Paediatrics, University Hospital Antwerp, Edegem, Belgium

[10] Department of Paediatric Gastroenterology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

[11] Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre, Maastricht, The Netherlands

Short title: Diagnostic accuracy of faecal calprotectin

Corresponding Author:

Pieter L.J. Degraeuwe, MD, PhD,

Department of Paediatrics, Maastricht University Medical Centre,

P. Debyelaan 25, PO Box 5800, 6202 AZ Maastricht, The Netherlands

Tel: - 31 – 43 - 387 72 46

Fax: - 31 – 43 - 387 52 46

Email: pieter.degraeuwe@mumc.nl

Funding Source: This study had no explicit funding, but some of the authors were supported by research grants: Dr. P. Henderson was funded by an MRC patient cohorts research initiative grant for PICTS to Prof Wilson by the Medical Research Council, UK (No.G0800675), Dr. K-L Kolho received grants from the Finnish Pediatric Research Foundation and a Helsinki University Central Hospital Grant.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; A.S.D.: received payment for advisory board membership and educational presentations from Janssen; U.L.F.: institution received lecture fee from Tillotts AB, MSD and AbbVie and travel grant from

Tillotts AB, Otsuka Pharma Scandinavia AB; K-L.K.: received consultation fee from Tillotts Pharma, payment for advisory board membership from MSD and Abbvie, and support for travel from Biocodex; P.F.vR. received consultation fee (supply of reagents free of charge) from Buhlmann Laboratories AG and an unrestricted start-up grant from Ferring-NL B.V., D.C.W. received consultation fees or honoraria from Pfizer, MSD, a research grant from MSD, lecture honoraria from MSD and Dr. Falk, and support for travel by Ferring; the other authors have no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work**.**

Author Contributions: Dr. Degraeuwe conceived and directed the study and drafted the manuscript. He also identified, selected and appraised the studies, and carried out the statistical analyses; Dr. Beld identified, selected and appraised the studies, and reviewed and revised the manuscript; Drs. Ashorn, Berni Canani, Day, Diamanti, Fagerberg, Henderson, Kolho, Van de Vijver, van Rheenen, and Wilson acquired the individual patient data and reviewed and revised the manuscript; Dr. Kessels gave methodological and statistical advice, carried out the statistical analyses, and reviewed and revised the manuscript; all authors had full access to all of the data, approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

No writing assistance was utilized in the production of the submitted manuscript.

Word count: 3558

Number of figures: 4

Number of tables: 2

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (*www.jpgn.org*).

ABSTRACT

Objectives*:* The diagnostic accuracy of faecal calprotectin (FC) concentration for paediatric inflammatory bowel disease (IBD) is well described at the population level, but not at the individual level. We reassessed the diagnostic accuracy of FC in children with suspected IBD and developed an individual risk prediction rule using individual patient data.

Methods*:* MEDLINE, EMBASE, DARE, and MEDION databases were searched to identify cohort studies evaluating the diagnostic performance of FC in paediatric patients suspected of having IBD. A standard study level meta-analysis was performed. In an individual patient data meta-analysis we reanalysed the diagnostic accuracy on a merged patient dataset. Using logistic regression analysis we investigated whether and how the FC value and patient characteristics influence the diagnostic precision. A prediction rule was derived for use in clinical practice and implemented in a spreadsheet calculator.

Results*:* According to the study level meta-analysis (9 studies, describing 853 patients), FC has a high overall sensitivity of 0.97 (95% confidence interval 0.92 to 0.99) and a specificity of 0.70 (0.59 to 0.79) for diagnosing IBD. In the patient level pooled analysis of 742 patients from 8 diagnostic accuracy studies, we calculated that at a FC cut-off level of 50 μg/g there would be 17 % (95% confidence interval 15 to 20) false positive and 2% (1 to 3) false negative results. The final logistic regression model was based on individual data of 545 patients and included both FC level and age. The area under the ROC curve of this derived prediction model was 0.92 (95% confidence interval 0.89 to 0.94).

Conclusions*:* In high prevalence circumstances, FC can be used as a non-invasive biomarker of paediatric IBD with only a small risk of missing cases. To quantify the individual patients' risk, we developed a simple prediction model based on FC concentration and age. Although the derived prediction rule cannot substitute the clinical diagnostic process it can help in selecting patients for endoscopic evaluation.

Key words: inflammatory bowel diseases; leukocyte L1 antigen complex; meta-analysis as topic; infant; child; adolescent

Introduction

As many as 25% of all IBD cases present in childhood. Given the impact of the disease and its therapy, a reliable and timely diagnosis is mandatory (1). However, the reference standard, endoscopic evaluation with biopsies (2), is invasive, not without risks, and also costly.

Faecal calprotectin (FC) has been investigated as a surrogate marker of neutrophil influx into the bowel lumen and a non-invasive diagnostic test for IBD (3, 4).

van Rheenen and co-workers (5) published a study-level meta-analysis on the diagnostic accuracy of FC for IBD in both children and adults. Based on the calculated summary estimates for specificity, sensitivity and likelihood ratios they concluded that the test is a useful screening tool for identifying those patients most likely needing endoscopic evaluation.

A key problem is however that this meta-analysis was based on aggregate data and did not account for different cut-off levels or different patient characteristics (6).

By contrast, individual patient data (IPD) meta-analysis allows using the original test results as continuous data rather than as dichotomous classification data. In addition, the effect of patient characteristics on test accuracy can be evaluated and quantified (6, 7).

We undertook the current study to update the study-level meta-analysis on the diagnostic performance of FC for paediatric IBD and to complement it with an individual-level meta-analysis. We also sought to develop a prediction model for IBD based on calprotectin level and readily available patient variables.

Materials and Methods

The research was conducted in accordance with the Code of Conduct for Medical Research of the Dutch Federation of Biomedical Scientific Societies (8).

Identification, selection and appraisal of the relevant studies was carried out independently, by two reviewers (PLJD, MPAB). Disagreements were resolved through discussion.

Search methods for identification of studies

A systematic search was performed initially from inception to December 2010 in MEDLINE (Ovid), EMBASE (Ovid), the Database of Abstracts of Reviews of Effects (DARE) of the Centre for Reviews and Dissemination (9) and the MEDION database of the University of Maastricht (10). The electronic search was last updated on April 1, 2012. We refrained from using a diagnostic search filter (11, 12) or language restrictions. Details of the search are given in an online-only document (eText 1, *http://links.lww.com/MPG/A396*). Duplicate articles identified in both Medline and Embase were manually deleted. The reference lists of selected studies were checked for further relevant studies.

Study selection

Only cohort studies evaluating the diagnostic performance of FC concentration in paediatric patients suspected of having IBD were considered for review. Other study designs, such as case-control, are indeed prone to spectrum bias (13). In addition, the following inclusion criteria were applied: FC measurement and reference standard available for all paediatric participants (or a follow-up period long enough to exclude IBD), and sufficient data to calculate 2 x 2 tables.

Data abstraction

The following data were extracted from each study on a pre-designed form: the spectrum of the studied population (indication for testing, age and gender), details of the index and reference test, and counts in the 2 x 2 table.

Individual patient data sets

For the IPD meta-analysis, we contacted the authors of the selected studies and invited them to share the raw, de-identified study data (FC level, age, sex, and final diagnosis). The raw

data were checked for internal consistency against the summary results published in the original paper. Some small discrepancies were found, discussed with the authors and ascertainable divergences were corrected.

Quality assessment

The quality of the included studies was assessed using the revised QUADAS tool (14). Since the calprotectin assay is an objective measurement, 3 of 14 items were omitted: blinded interpretation of the index test, availability of clinical data, and reporting of uninterpretable results. Two reviewers independently answered the 11 remaining questions in the affirmative, in the negative or as being unclear.

Data synthesis and (statistical) analysis

Literature-based meta-analysis:

The aggregate data meta-analysis was performed in Stata/SE version 11.2 (Stata Corporation, College Station, TX, USA) using the Midas command (15). Accordingly, summary statistics for all diagnostic performance indices were calculated within the bivariate mixed-effects logistic regression modelling framework. Between-study heterogeneity of the results was assessed graphically by using forest plots of the diagnostic odds ratios, and statistically by using the χ^2 test of homogeneity and the inconsistency index (I^2).(16) An I^2 value greater than 50% was taken to indicate significant heterogeneity. The potential for publication bias was estimated by using a Deeks' funnel plot. As recommended, a p-value < 0.1 was considered statistically significant (17). A hierarchical summary receiver operating characteristic graph with 95% confidence region and 95% prediction region was constructed.

Individual patient-based meta-analysis:

Although we largely prefer the analytical approach using logistic regression, we also evaluated the diagnostic performance of FC based on the merged individual data. The diagnostic performance was calculated through the MedCalc software, version 11.5.01. (MedCalc Software, Mariakerke, Belgium).

Logistic regression analysis – predicted probability:

The contribution of calprotectin concentration and age, as continuous variables, and gender and study as categorical variables, to the diagnosis of IBD was explored using stepwise forward (likelihood ratio) binary logistic regression analysis. The logit and logistic command in Stata/SE 10.1 were used. An entry probability for each variable was set at 0.05. A clinical prediction rule was derived from the final regression model (see eText 1, *http://links.lww.com/MPG/A396,* for calculation details).

Results

Description of studies

Our search (see Figure 1) returned 161 citations. After removal of 36 duplicates and because 104 of the abstracts were considered not pertinent, 23 records were retrieved as full texts. An additional 4 studies were excluded because of the inappropriateness of the population studied. Insufficient information was available for constructing a 2-by-2 table from 3 studies. Eventually, eight cohort studies fulfilled the inclusion criteria (18-25). A ninth relevant population-based cohort study was discovered in a recently published diagnostic meta-analysis (5). The original study (26) does not provide data to construct a 2x2 table, but the meta-analysis does.

The authors of eight cohort studies (18-25) were willing to share a data set with individual patient data (age, gender, calprotectin concentration, and final diagnosis). Age and gender

were not available from the Diamanti et al. study due to a computer crash. The authors of the Ashorn et al. paper (Dr. Kolho) provided data on 31 additional patients recruited after publication of their study results (18). In the most recently published study (25) not all patients underwent endoscopy, but a sufficiently extended follow-up period (disease free period of 6 months) should minimize the risk of verification bias (27). To completely eliminating this type of bias, we excluded the 43 patients without histologically confirmed IBD from the IPD meta-analysis. The Norwegian group (26) justified their refusal by arguing that the data will again be used for their own follow-up study. None of the authors is aware of missed (un)published diagnostic accuracy studies fulfilling the selection criteria.

Characteristics of included and excluded studies

All nine included cohort studies (18-26) were undertaken in referral centres for paediatric gastroenterology and involved 853 patients from the toddler age group to young adults. Table 1 summarizes the characteristics of the study population, a description of the applied index and reference tests, and the findings. In one study (19) four patients were excluded because infectious gastroenteritis was the final diagnosis. Upon enquiry, none of the authors were aware of similar patients in whom readily available diagnostic investigation could have prevented unjustified study entry and unnecessary false positive or true negative cases. Another study (24) contained 16 patients whose calprotectin results were expressed as greater or less than a numerical value. These test results were substituted by a value midway between the reported numerical value and a higher or lower value present in the data set. This procedure is unlikely to influence the test accuracy parameters since 14 of the 16 adjusted values were situated in the highest or lowest quintile. The bar graph in eFigure 1 (*http://links.lww.com/MPG/A395*) is a representation of the quality of the 9 selected cohort studies (18-26).

Summary results for all included studies

Literature-based meta-analysis:

A summary of test accuracy estimates is shown in Table 2. The Cochran Q test and the I^2 values are indicative for substantial heterogeneity (28). Figure 2 shows a forest plot of the diagnostic odds ratio and the pooled estimate for the 9 cohort studies (18-26). The average prevalence of IBD in the cohort studies was 0.54. Accordingly, the post-test probability of a positive calprotectin test could be estimated to be 0.79 (95% CI 0.72-0.85). The post-test probability of a negative calprotectin test is 0.05 (0.01-0.14).

Figure 3 shows the hierarchical summary receiver operating characteristic (HSROC) graph with 95% confidence region and 95% prediction region.

Publication bias was assessed by Deeks' funnel plot asymmetry test for small study effect/publication bias. The non-significant slope (p=0.62) indicates that no significant bias was found.

Individual patient-based meta-analysis:

Individual patient data on final diagnosis and FC were collated from 742 children from 8 studies (18-25). The studies were significantly different with respect to age (p=0.004, one-way ANOVA test), mean FC concentration (p <0.001, one-way ANOVA test), and area under the ROC curve (p=0.001, one-way ANOVA). The pre-test probability ranged from 0.51 (95% CI 0.39-0.63) to as high as 0.84 (0.75-0.90).

In the pooled data set of all children the "optimal" cut-off value (the value with the highest accuracy – minimal false negative and false positive results) for FC was 212 µg/g corresponding with a sensitivity of 0.90 (95% CI 0.87-0.93), a specificity of 0.85 (0.81-0.88), a positive likelihood ratio of 5.99 (4.6-7.8), and a negative likelihood ratio of 0.11 (0.09-0.20). The area under the receiver operating characteristic (ROC) curve of FC for the diagnosis of paediatric IBD was 0.94 (95% CI 0.92-0.95).

Logistic regression analysis – predicted probability

The results of the logistic regression analysis (eText 1, *http://links.lww.com/MPG/A396*) disclosed that calprotectin concentration, age and study centre were independent predictors of IBD. The influence of study centre (likelihood ratio test significant) confirms heterogeneity across the studies.

Outside the 8 study centres a regression equation containing a term related to the study centre is not of use, therefore the impact of study centre on the diagnosis was omitted from the final regression model: $\text{logit}(p) = S = -3.294 + 0.004 * FC + 0.175 * AGE$, where the explanatory variable FC is the FC concentration in µg/g and AGE is the age in years. The predictivity of this simplified model was evaluated by ROC curve analysis. The estimated area under the ROC curve (AUC) for the final model was calculated to be 0.92 (95% CI 0.89-0.94). The logistic model using calprotectin concentration and age predicts IBD correctly in 85.5% (466/545) of children (sensitivity 0.81 (95% CI 0.76-0.85), specificity 0.92 (0.88-0.95), PPV 0.93 (0.89-0.96) and NPV 0.73 (0.72-0.83).

The probability of having IBD is determined by the equation: $p = \exp(S)/(1+\exp(S))$. We programmed an Excel spreadsheet (eFile1, *http://links.lww.com/MPG/A397*) computing the probability (with CI) of having IBD for each combination of FC, age and disease prevalence. Figure 4 illustrates that assuming a disease prevalence of 56%, a FC of 700 µg/g in a 6 year old child corresponds to an IBD probability of 64% (95% CI: 52-71). This prediction lacks precision and the post-test probability is hardly improved compared to the average pretest-probability. In a 17 year old adolescent the same test result makes the diagnosis quite probable, but ruling out IBD at this age and this pre-test probability is practically impossible.

Discussion

Principal findings

This systematic review has provided an updated aggregate data meta-analysis confirming the high diagnostic accuracy of FC for the detection of IBD in referral centres for paediatric gastroenterology. In addition, a meta-analysis using individual participant data enabled us to develop an algorithm predicting the likelihood that an *individual* child suffers from IBD. It was previously not recognized that the probability of having IBD depends both on the FC level and the age of the child.

General considerations

The literature-based meta-analysis indeed confirms that FC has an excellent overall sensitivity of 0.97 (95% CI 0.92-0.99) and a modest specificity of 0.71 (0.59-0.80) for diagnosing paediatric IBD. A first observation is that, due to the relatively small number of pooled investigated patients, the imprecision of the predictive values is still considerable.

In addition, although we used recommended, robust state of the art statistical methodology (29), the estimation of summary points does not hold true if the included studies have used different threshold values (30). For the clinician, it is not evident at which cut-off value the calculated summary estimates of the test accuracy measures apply. Therefore we recalculated the diagnostic performance of FC using merged individual patient data. The "optimal" cut-off for the whole group equals 212 µg/g. Taking into account the pretest probability and the likelihood ratios, the clinician can now better interpret a test result and discuss the predictive value with the patient and his/her parents. We should however realize that the predictive values or likelihood ratios are applicable to the whole group of patients presenting with a test result above or below the threshold and that dichotomizing carries with it a loss of information for the individual child (31). This drawback can be overcome by using logistic

regression analysis, a technique which allows taking into consideration patient-level covariates as well.

The contribution of age in the prediction of IBD turned out to be significant. FC concentrations have been shown to be higher in preschool children, especially infants, than in older children (32, 33). This age-dependency does not seem to play an important role in our logistic regression model since there is no significant correlation between age-quartile and FC in non-IBD subjects ($r = -0.06$; $p = 0.34$). In contrast, the prevalence of IBD is significantly increased with higher age-quartile (Chi-square test, $p<0.0001$), suggesting that the age factor corrects for the age dependent prevalence.

Our literature search revealed some of the well-known shortcomings in the quality and reporting of diagnostic research (27). We had to exclude nearly half of the paediatric accuracy studies because they used a case-control design known for introducing spectrum bias (13). None of the studies pre-specified a target value for sensitivity, specificity, predictive accuracy and a minimal acceptable lower confidence limit, enabling sample size calculation (34, 35). Admittedly, neither did we define a target region within which the summary estimates of our meta-analysis should fall. We suggest that the absence of a predefined target makes authors all too often undeservedly enthusiastic about the diagnostic value of an index test.

Our meta-analysis differs in some aspects from the study of van Rheenen and co-workers (5). They included one study which we excluded (36) whereas we added three new studies totalling 404 participants (20, 22, 25). Doctor Golden (personal communication – 03-18-2010) discouraged us from using her group's data (36) because Magne Fagerhol's original assay, measuring calprotectin as mg/L assay buffer, does not correlate very well with the new commercial assays expressing the results as µg/g faeces.

The major difference between van Rheenen's (5) and this paper is the IPD meta-analysis. As already known, the collection of individual patient data is time consuming and difficult. A high level of persuasiveness was needed to collect the raw study data from 8 of the 9 cohort

studies. Consequently, we were able to use easily available patient characteristics and the original continuous calprotectin data instead of the dichotomized study results in a logistic regression analysis.

It is also noteworthy that we detected and were able to correct small discrepancies between the raw data and the published data.

By publishing all the available raw data (eTable 1, *http://links.lww.com/MPG/A398*), we comply with the appropriate and growing request to share complete data, allowing re-analysis by the reader and sequential meta-analysis if new studies appear. There is indeed a growing awareness that sharing data is an ethical obligation (37-42).

Strength of the study

Compared to the systematic review of van Rheenen et al. (5), the key strength of our study is the IPD meta-analysis. Although a literature-based meta-analysis provides a good overall impression of the diagnostic accuracy of the test, pooling of studies with different diagnostic thresholds (from 50 to 160 µg/g faeces) precludes the calculation of the predictive value at a self-selected threshold or the patient's result. Only IPD permits prediction based on numerical test results and adjustment for patient characteristics. Our prediction tool calculates an individualized disease risk with accompanying confidence range taking into account the FC concentration and the patient's age. In a paediatric gastroenterology referral centre, ruling out IBD is difficult in older children, whereas in younger patients a higher FC concentration is needed to get a post-test probability larger than the prevalence.

Study weaknesses

Our search methodology aimed to avoid language and citation bias but we cannot exclude publication or reporting bias in our meta-analysis. Publication bias is probably even more of a

problem for diagnostic and prognostic than for therapeutic studies (43, 44). The effects of reporting bias in therapeutic meta-analyses have recently been shown to be substantial (45). Our IPD meta-analysis may also suffer from availability bias. Data from two large studies are incomplete or unavailable. In this context, it is noteworthy that the study of Perminow et al. (26) was not intended to be a diagnostic accuracy study and showed the lowest diagnostic odds ratio.

We also recognize that the number of patients in the meta-analysis is still limited. Therefore the precision of the diagnostic accuracy measures and our predictive algorithm leaves room for improvement. Furthermore, the final logistic regression equation contains only two independent variables. This likely represents an oversimplification of a complex clinical diagnostic process, including information collected during history taking and physical examination. The user of the algorithm should be aware of this. Other laboratory test results, such as CRP, haemoglobin, and iron indicators, could improve diagnostic precision but were not available for this study. Clinical suspicion or "gut feeling" may also be of additional diagnostic value.

Finally, the between-study heterogeneity is of concern and not fully understood. Although the selection criteria are more or less the same (suspected inflammatory bowel disease), between-study differences were shown for age and FC concentration. The prevalence of IBD varied from 51% to 84%, and the false positive rate ranged from 0 to 29% suggesting dissimilar study entry criteria. We have also noticed that there was a negative, just significant linear correlation ($r = -0.71$, $P=0.049$) between prevalence and the false positive rate.

Differences in clinical laboratory measurement procedures for FC may be another source of heterogeneity (46). We know that even small analytic biases can indeed shift the laboratory values and lead to diagnostic misclassification (47, 48).

Implications for clinical practice and future research

Despite the absence of validation, and given the mentioned concern about study heterogeneity and (reporting) bias, our predictive algorithm is currently the best available tool for predicting IBD in the individual child. As such it deserves a valid place in the decision making process. We nevertheless advocate a large prospective multi-centre study to improve and refine the IBD screening tool. More data are needed to validate the prediction algorithm on a different data set to improve its precision. Every effort should be made to standardize and harmonize calprotectin measurements and the predictive value of other clinical variables and (faecal) biomarkers (49, 50) should be investigated simultaneously.

Conclusion

Using an IPD meta-analysis and through regression modelling we identified FC concentration and age as independently associated with the diagnosis of IBD. We developed a prediction rule that enables the practicing paediatric gastroenterologist to numerically interpret the FC value together with the patient's age. As such, this rule can be a valuable adjunct to the diagnostic armamentarium making physician and patients/families better equipped to make personalized decisions.

References

1 Timmer A, Behrens R, Buderus S, et al. Childhood Onset Inflammatory Bowel Disease: Predictors of Delayed Diagnosis from the CEDATA German-Language Pediatric Inflammatory Bowel Disease Registry. J Pediatr 2011;158(3):467-73.

2 Levine A, Koletzko S, Turner D, et al. ESPGHAN Revised Porto Criteria for the Diagnosis of Inflammatory Bowel Disease in Children and Adolescents. Journal of Pediatric Gastroenterology and Nutrition 2014;58(6):795-806.

3 Konikoff MR, Denson LA. Role of fecal calprotectin as a biomarker of intestinal inflammation in inflammatory bowel disease. Inflamm Bowel Dis 2006;12(6):524-34.

4 von Roon AC, Karamountzos L, Purkayastha S, et al. Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy. Am J Gastroenterol 2007;102(4):803-13.

5 van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: diagnostic meta-analysis. BMJ 2010;341:c3369.

6 Riley RD, Dodd SR, Craig JV, et al. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med 2008;27(29):6111-36.

7 Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC Med Res Methodol 2005;5:14.

8 Code of Conduct for Medical Research of the Dutch Federation of Biomedical Scientific Societies. http://www.federa.org/sites/default/files/bijlagen/coreon/code_of_conduct_for_medical_research_1.pdf. Accessed October 10th 2012.

9 The University of York Centre for Reviews and Dissemination http://www.crd.york.ac.uk/crdweb/.

10 The Medion Database. www.mediondatabase.nl.

11 Doust JA, Pietrzak E, Sanders S, et al. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J Clin Epidemiol 2005;58(5):444-9.

12 Leeflang MM, Scholten RJ, Rutjes AW, et al. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol 2006;59(3):234-40.

13 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282(11):1061-6.

14 Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006;6:9.

15 MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies [Computer Program]. Boston College Department of Economics; 2009/02/05/.

16 Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.

17 Deeks JJ, Macaskill P, Irwig L The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58(9):882-93.

18 Ashorn S, Honkanen T, Kolho KL, et al. Fecal calprotectin levels and serological responses to microbial antigens among children and adolescents with inflammatory bowel disease. Inflamm Bowel Dis 2009;15(2):199-205.

19 Berni Canani RB, de Horatio LT, Terrin G, et al. Combined use of noninvasive tests is useful in the initial diagnostic approach to a child with suspected inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2006;42(1):9-15.

20 Diamanti A, Panetta F, Basso MS, et al. Diagnostic work-up of inflammatory bowel disease in children: The role of calprotectin assay. Inflamm Bowel Dis 2010;16(11):1926-30.

21 Fagerberg UL, Loof L, Myrdal U, et al. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. J Pediatr Gastroenterol Nutr 2005;40(4):450-5.

22 Henderson P, Casey A, Lawrence SJ, et al. The Diagnostic Accuracy of Fecal Calprotectin During the Investigation of Suspected Pediatric Inflammatory Bowel Disease. Am J Gastroenterol 2012;107(6):941-49.

23 Kolho KL, Raivio T, Lindahl H, et al. Fecal calprotectin remains high during glucocorticoid therapy in children with inflammatory bowel disease. Scand J Gastroenterol 2006;41(6):720-5.

24 Sidler MA, Leach ST, Day AS. Fecal S100A12 and fecal calprotectin as noninvasive markers for inflammatory bowel disease in children. Inflamm Bowel Dis 2008;14(3):359-66.

25 Van De Vijver E, Schreuder AB, Muller-Kobold AC, et al. Safely ruling out IBD in children and teenagers without referral for endoscopy. Archives of Disease in Childhood 2012;97(12):1014-18.

26 Perminow G, Brackmann S, Lyckander LG, et al. A characterization in childhood inflammatory bowel disease, a new population-based inception cohort from South-Eastern Norway, 2005-07, showing increased incidence in Crohn's disease. Scand J Gastroenterol 2009;44(4):446-56.

27 Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 2003;138(1):40-4.

28 Higgins JPT, Green, S. Cochrane Handbook for Systematic Reviews of Intervention. In: J. P. T. Higgins, Green, S. ed.: The Cochrane Collaboration; 2011.

29 Willis BH, Quigley M. Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. BMC Med Res Methodol 2011;11(1):27.

30 Leeflang MM, Deeks JJ, Gatsonis C, et al. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149(12):889-97.

31 Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332(7549):1080.

32 Olafsdottir E, Aksnes L, Fluge G, et al. Faecal calprotectin levels in infants with infantile colic, healthy infants, children with inflammatory bowel disease, children with recurrent abdominal pain and healthy children. Acta Paediatr 2002;91(1):45-50.

33 Rugtveit J, Fagerhol MK. Age-dependent variations in fecal calprotectin concentrations in children. J Pediatr Gastroenterol Nutr 2002;34(3):323-4; author reply 24-5.

34 Bachmann LM, Puhan MA, ter Riet G, et al. Sample sizes of studies on diagnostic accuracy: literature survey. BMJ 2006;332(7550):1127-9.

35 Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J Clin Epidemiol 2005;58(8):859-62.

36 Bunn SK, Bisset WM, Main MJ, et al. Fecal calprotectin: validation as a noninvasive measure of bowel inflammation in childhood inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2001;33(1):14-22.

37 Anderson BJ, Merry AF. Data sharing for pharmacokinetic studies. Paediatr Anaesth 2009;19(10):1005-10.

38 Fischer BA, Zigmond MJ. The essential nature of sharing in science. Sci Eng Ethics 2010;16(4):783-99.

39 Hrynaszkiewicz I, Norton ML, Vickers AJ, et al. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ 2010;340:c181.

40 Terry SF, Terry PF. Power to the people: participant ownership of clinical trial data. Sci Transl Med 2011;3(69):69cm3.

41 Vickers AJ. Making raw data more widely available. BMJ 2011;342(d2323.

42 Walport M, Brest P. Sharing research data to improve public health. Lancet 2011;377(9765):537-9.

43 Irwig L, Macaskill P, Glasziou P, et al. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995;48(1):119-30; discussion 31-2.

44 Rifai N, Altman DG, Bossuyt PM. Reporting bias in diagnostic and prognostic studies: time for action. Clin Chem 2008;54(7):1101-3.

45 Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ 2012;344:d7202.

46 Greg Miller W, Myers GL, Lou Gantzer M, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57(8):1108-17.

47 Klee GG. Establishment of outcome-related analytic performance goals. Clin Chem 2010;56(5):714-22.

48 Klee GG, Schryver PG, Kisabeth RM. Analytic bias specifications based on the analysis of effects on performance of medical guidelines. Scand J Clin Lab Invest 1999;59(7):509-12.

49 Lewis JD. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. Gastroenterology 2011;140(6):1817-26 e2.

50 van Schaik FDM, Oldenburg B, Hart AR, et al. Serological markers predict inflammatory bowel disease years before the diagnosis. Gut 2013;62(5):683-88.

Figure, Table Legends and Supplemental Data

Figure 1. Flowchart showing the search for and selection of papers evaluating faecal calprotectin in children with suspected inflammatory bowel disease. The final search was carried out on April 1st, 2012

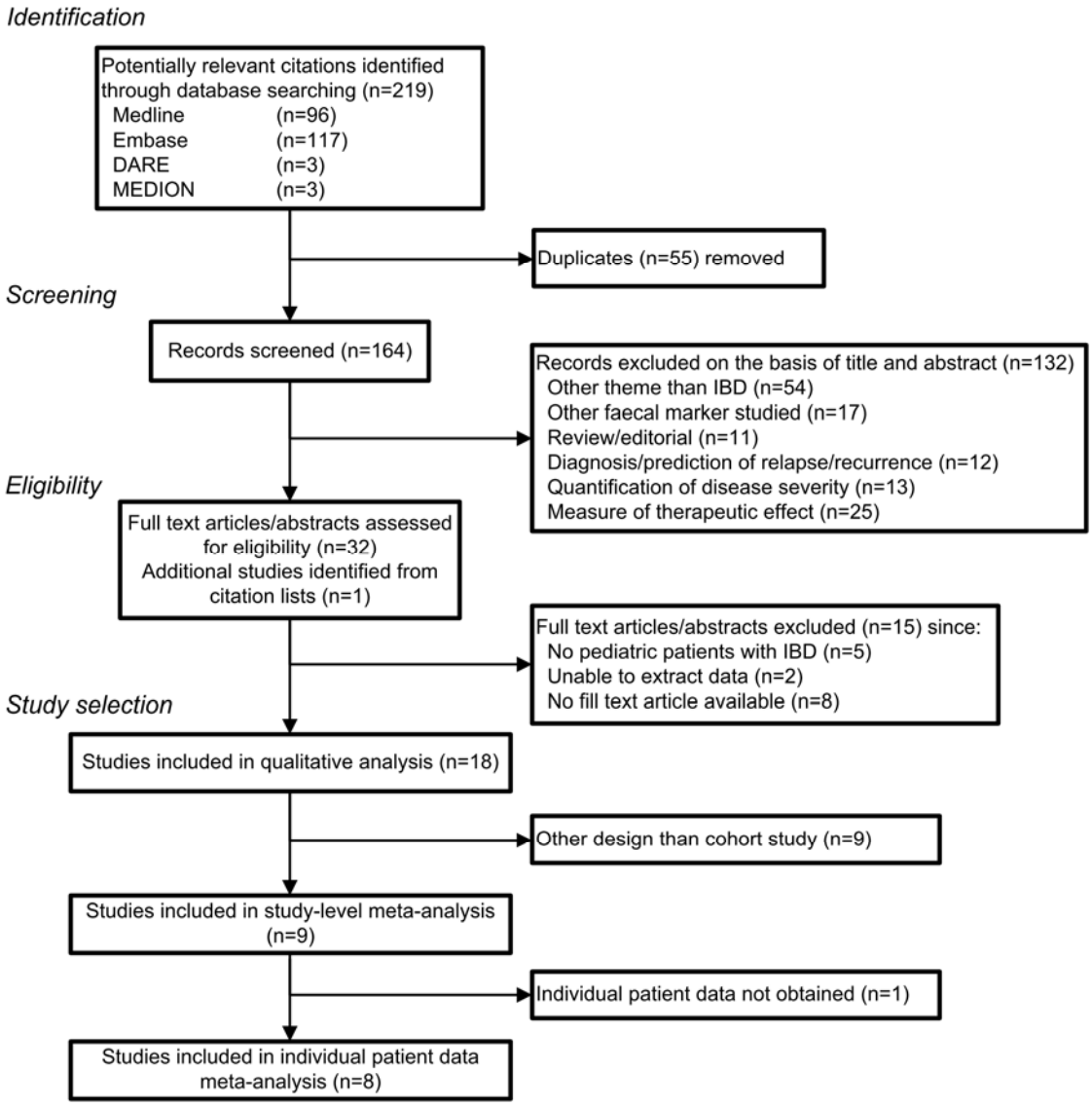
Figure 2. Forrest plot of diagnostic odds ratio of each individual cohort study, pooled odds ratio, Cochran Q test heterogeneity and I^2 statistic for inconsistency.

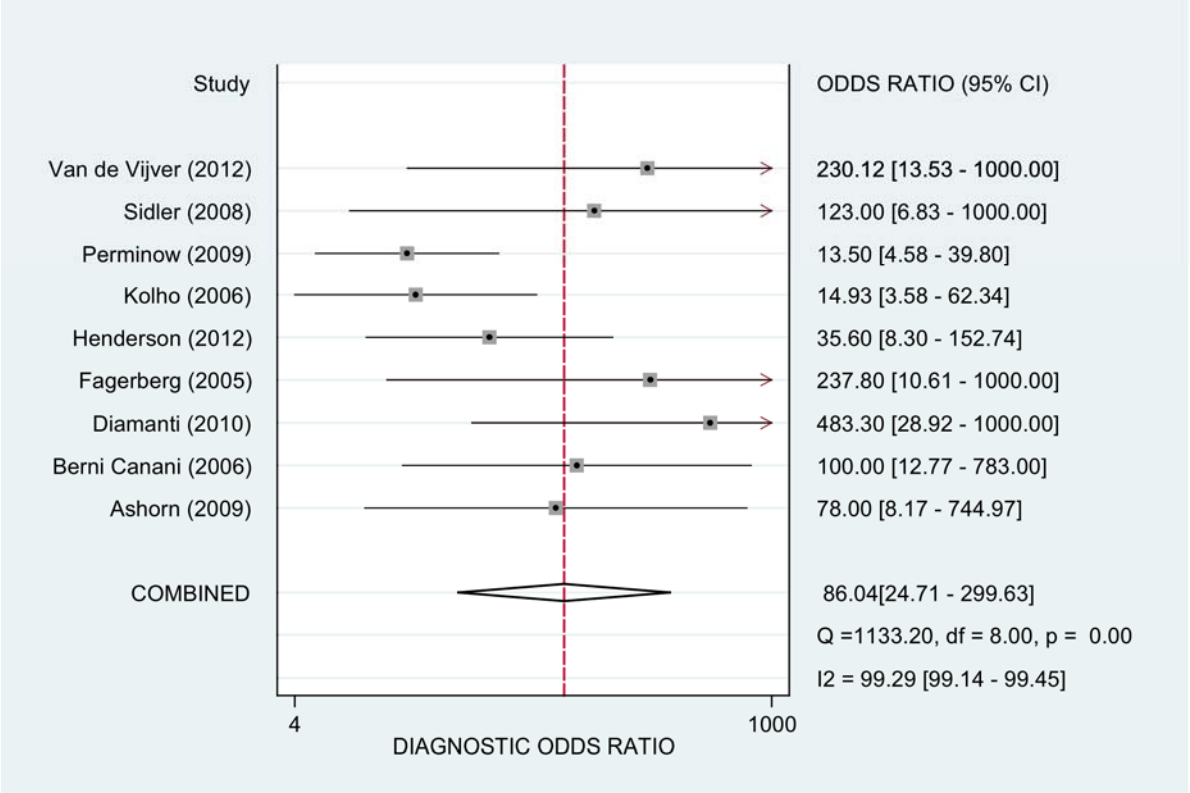
Figure 3. Summary ROC space of sensitivity and specificity for faecal calprotectin in the diagnosis of paediatric IBD. The circles depict the observed bivariate pairs of sensitivity and specificity of the 9 diagnostic cohort studies (18-26). The blue solid line is the summary ROC curve. The diamond is the bivariate summary point. The red dashed line triangle is the bivariate boundary of the 95% confidence region for the bivariate summary point and the green dotted line encloses the 95% prediction region. The study numbers correspond to those reported in Table 1.

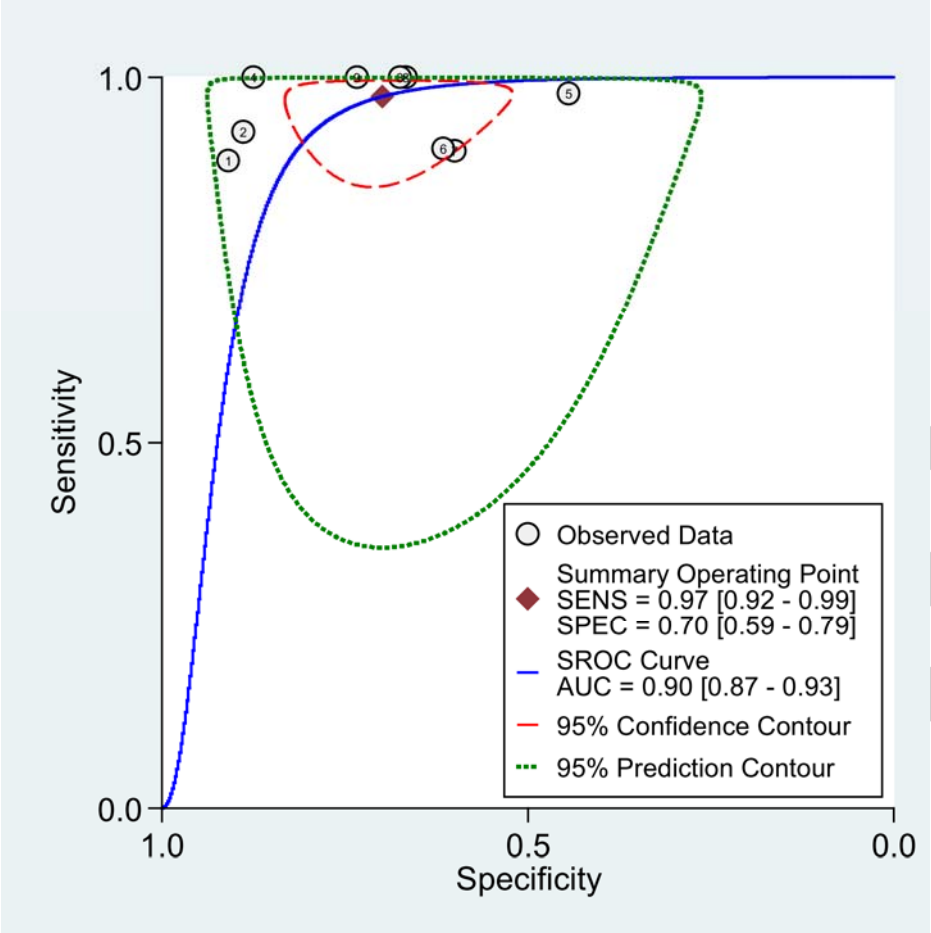
Figure 4. Plots of the predicted probabilities as a function of faecal calprotectin concentration, age and disease prevalence. This figure illustrates that age has an important additional value to calprotectin testing for the diagnosis of IBD. UL: upper limit of 95% CI, LL: lower limit of 95% CI

Table 1. Characteristics and 2 x 2 data of all included studies. TP: true positive, FP: false positive, FN: false negative, TN: true negative

Table 2. Numerical results of the literature-based meta-analysis (faecal calprotectin for the diagnosis of IBD): test performance parameters and their 95% CI. DOR: diagnostic odds ratio, LR+: likelihood ratio of a positive test, LR-: likelihood ratio of a negative test.







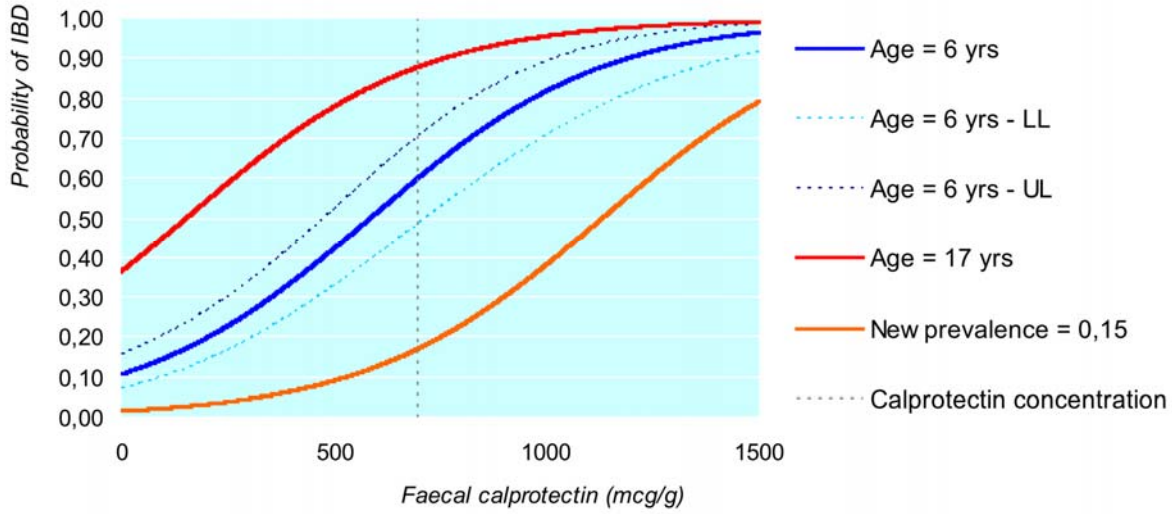


Table 1. **Characteristics and 2 x 2 data of all included studies**

Study, year (reference)		Patient characteristics				Index test	Cut-off value (µg/g)	Reference test	Prev	TP	FP	FN	TN
		Setting, location	Spectrum	Age range or mean(sd) (yrs)	Female/ male (n)								
1	Ashorn et al, 2009[20]	Children´s Hospital, University of Helsinki, Finland	Suspicion of IBD	5.8-19.9	50	Phical Test	100	Histopathology	0.80	39	1	5	10
									0.84	62	1	10	13
2	Berni Canani et al, 2006[21]	Paediatric Gastro-enterology Unit; Naples, Italy	Children referred for initial assessment of suspected IBD	Non-IBD 11.0(3.3) CD 14.5(5.1) UC 11.0(5.0)	21/24	Calprest®	95.3	Standard diagnostic criteria, including histopathology	0.60 ↓ 0.66	25	2 ↓ 0	2	16 ↓ 14
3	Diamanti et al, 2010[22]	Paediatric Gastro-enterology and Nutrition Unit, Rome, Italy	Children referred for recurrent abdominal pain and altered bowel habits	Non-IBD 1-18 IBD 1-18	88/109	Calprest®	100	Histological examination	0.59	117	26	0	54
							160		0.59	117	16	0	64
4	Fagerberg et al, 2005[23]	Department of gastroenterology, Stockholm and Paediatrics, Västerås, Sweden	Scheduled for colonoscopy to rule out IBD; no bacterial gastroenteritis	IBD 6.7-17.8 Non-IBD 6.5-17.3	19/17	Calprest®	50	Conventional histopathological criteria for diagnosis of IBD	0.56	20	2	0	14
5	Henderson et al, 2012[24]	Paediatric gastroenterology department, Edinburgh UK	Children evaluated for suspected bowel inflammation	IBD 12.6, IQR: 9.5-14.0 Non-IBD 9.3, IQR: 5.2-12.7	79/111	PhiCal Test	50	Complete endoscopic evaluation + biopsies, small bowel imaging plus follow-up of ≥ 12 months	0.48	89	55	2	44
6	Kolho et al, 2006[25]	Children´s Hospital, University of Helsinki, Finland	Children undergoing colonoscopy	0.9-18 9.8(4.7)	29/28	Phical Test	50	Colonoscopy + histology	0.54	28	10	3	16
							100		0.54	27	8	4	18
7	Perminow et al, 2009[28]	South-Eastern Norway	Children with suspected IBD (treatment naïve)	IBD 0.8-17.9 Non-IBD 1.9-18	95	?	160	Histopathological (Porto criteria)	0.63	54	14	6	21
8	Sidler et al, 2008[26]	Gastroenterology Outpatient Clinic, Randwick Australia	Symptoms suggestive of an organic gut disease	Non-IBD 2.2-15.5 IBD 2.4-16	24/37	PhilCal test	50	Standard diagnostic criteria, including histopathology	0.51	31	10	0	20
9	Van de Vijver et al, 2012[27]	Department of Paediatric Gastroenterology, Groningen, Netherlands	Clinical suspicion of IBD	IBD 6-18 Non-IBD: 6-18	61/56	Calpro® elisa test	50	Esophagogastro-duodenoscopy and ileocolonoscopy including histopathological verification or 6 months follow-up	0.36	42	20	0	55

Table 2. Numerical results of the literature-based meta-analysis (faecal calprotectin for the diagnosis of IBD): test performance parameters and their 95% CI along with Cochran-Q and inconsistency. DOR: diagnostic odds ratio, LR+: likelihood ratio of a positive test, LR-: likelihood ratio of a negative test.

Parameter	Estimate (95% CI)	Cochran-Q (P-value)		Inconsistency (I^2) (%) (95% CI)
Sensitivity	0.97 (0.92 to 0.99)	30.21	0.00	73.52 (55.84 to 91.20)
Specificity	0.70 (0.59 to 0.79)	34.33	0.00	76.70 (61.62 to 91.77)
DOR	86 (25 to 300)	1133.20	0.00	99.29 (99.14 to 99.45)
LR+	3.2 (2.3 to 4.5)	39.24	0.00	66.87 (66.87 to 92.36)
LR-	0.04 (0.01 to 0.12)	21.09	0.01	62.06 (34.57 to 89.56)