Diagnosis of Crohn's disease in ambulatory care: a systematic review.

Dr. Camille Hoefnagels, KU Leuven

Dr. Laura Verhaegen, KU Leuven

Dr. Jarek Wauters, KU Leuven

Promotor: prof. dr. Jan Verbakel, Epi-Centre, Department of Public Health and Primary Care, KU Leuven

Co-promotor: dr. Thomas Struyf, Epi-Centre, Department of Public Health and Primary Care, KU Leuven

Master of Family Medicine

Masterproef Huisartsgeneeskunde
Academiejaar: 2025-2026

©Copyright by KU Leuven

Deze masterproef is een examendocument dat niet werd gecorrigeerd voor eventueel vastgestelde fouten. Met uitzondering van het door de wet toegestaan gebruik, is zonder voorafgaande schriftelijke toestemming van de auteur(s) het overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie in verband met het overnemen en/of gebruik van gedeelten uit deze publicatie, wendt u zich tot de KU Leuven, Faculteit Geneeskunde, Campus Gasthuisberg O&N2, Herestraat 49 – bus 400, 3000 Leuven.

Voorafgaande schriftelijke toestemming van de promotor(en) is eveneens vereist voor het aanwenden van de in dit afstudeerwerk beschreven (originele) methoden, producten en programma's voor commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

COVER LETTER

Leuven, 1 May 2026

Dear members of the editorial board,
Dear madam, sir,

We would hereby like to submit our article, entitled: “Diagnosis of Crohn’s disease in ambulatory care: a systematic review.”, for consideration to be published.

Crohn’s disease remains an important diagnosis to catch early on, to prevent further complications. With its multifactorial etiology and diverse presentations in ambulatory care, it is hard to get a firm grip on its precise diagnosis. The conservative diagnostic tools in ambulatory care have been investigated. We offer an overview of the current literature concerning Crohn’s disease and the scientific evidence gathered on its diagnosis with means of non-invasive biomarkers. We contribute to the clarification of the value of these biomarkers in ambulatory care, to narrow down the diagnostic delay and accelerate the start of therapeutic intervention, but also to identify the need for further research.

All three authors have read and approved this article. We declare that this article is original and has not been published or submitted for publication in any other journal. We have no competing interests, nor are there any ethical conflicts.

Thank you for considering this literature review for publication. We hope you find our work of interest, and we are available for any further questions you may have.

Yours Sincerely,

Dr. Camille Hoefnagels, Dr. Laura Verhaegen, and Dr. Jarek Wauters

Faculty of Medicine, KU Leuven
Herestraat 49, 3000 Leuven

Camille.hoefnagels@student.kuleuven.be
Laura.verhaegen@student.kuleuven.be
Jarek.wauters@student.kuleuven.be

TABLE OF CONTENT

ABSTRACT **1**
NEDERLANDSE SAMENVATTING **2**
METHODS **4**
Search strategy 4
Searching other resources *4*
Selection criteria 4
Types of studies *4*
Participants *4*
Index tests *4*
Target conditions *4*
Reference standards *5*
Data collection and analysis 5
Selection of studies *5*
Data extraction and management *5*
Assessment of methodological quality *5*
Statistical analysis and data synthesis *5*
Investigations of hterogeneity *6*
Sensitivity analyses *6*
RESULTS **6**
Study selection 6
Study characteristics 6
Risk of bias 9
Diagnostic test characteristics 9
DISCUSSION **13**
Principal Findings 13
Clinical Implications 14
Strengths and limitations of our systematic review 15
CONCLUSION **16**
Future 17
BIBLIOGRAPHY ..... **A**
CONFLICTS OF INTEREST ..... **C**
FINANCIAL DISCLOSURE ..... **C**
FUNDING ..... **C**
ACKNOWLEDGMENT ..... **C**
APPENDIX ..... **D**
Appendix I. Search Strategy ..... D
Appendix II. Red Flags Index ..... E
Appendix III. Forest plot and ROC plot for CRP ..... F
Appendix IV. Forest plot and ROC plot for ESR ..... G
Appendix V. Forest plot and ROC plot for FC ..... H
Appendix VI. Forest plot and ROC plot for RFI I

LIST OF ABBREVIATIONS

AUROC	Area Under The Receiver Operating Characteristic Curve
BRMA	Bivariate random-effects model
BSE	Bezinkingssnelheid van erytrocyten
CD	Crohn Disease
CH	Camille Hoefnagels
CRP	C-Reactive Protein
CT	Computed Tomography
CU	Colitis Ulcerosa
ESR	Erythrocyte Sedimentation Rate
FC	Fecal Calprotectin
FN	False Negative
FOB	Fecal Occult Blood
FP	False Positive
GI	Gastrointestinal
GP	General Practice
IBD	Inflammatory Bowel Disease
JW	Jarek Wauters
Lf	Fecal Lactoferrin
LR + / -	Likelihood Ratios Positive / Negative
LV	Laura Verhaegen
m	Months
M	Males
Max	Maximum
MRI	Magnetic Resonance Imaging
n	number of patients
NSAID	Non-Steroidal Anti-Inflammatory Drugs
OPC	Outpatient Clinic
PIRTS	Population, Index test, Reference standard, Target disease, Setting
Pt	Patients
QoL	Quality of Life
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies
RBC	Red Blood Cells
RFI	Red Flags Index
ROC	Receiver Operating Characteristic

SR	Systematic Review
TN	True Negative
TP	True Positive
UK	United Kingdom
W	Weeks
WBC	White Blood Cells
y	Years

ABSTRACT

Objective: we aimed to evaluate the diagnostic accuracy of biomarkers for diagnosing Crohn's Disease (CD) in adults and adolescents presenting with abdominal pain complaints in ambulatory care settings.

Methods: we systematically searched the databases MEDLINE, EMBASE, and Cochrane Library from February to June 2024, with predetermined inclusion and exclusion criteria, looking for studies evaluating biomarkers in patients ≥14 years (y) with suspected CD in ambulatory care. The index tests included C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), fecal calprotectin (FC), fecal lactoferrin (Lf), fecal occult blood (FOB), and a red flags index (RFI). The gold standard was defined as a colonoscopy or sigmoidoscopy with histological confirmation, with or without additional imaging. In pairs of two reviewers, studies were independently screened, data extracted, and risk of bias assessed using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2). The analysis was performed using forest plots, a Receiver Operating Characteristic (ROC) and dumbbell plot, to visualize the data.

Results: six studies were included, with a total of 1,241 patients. The analyses show that FC is the most clinically useful biomarker in ambulatory care. At a value of <50 μg/g, FC is an excellent rule-out marker, with high sensitivity but lower specificity. However, these results vary greatly depending on the clinical setting and waiting room probability. In a more clearly defined ambulatory care setting, FC of <50 μg/g has minimal diagnostic value, while in the described outpatient clinics (OPC) with the same cutoff value, it performs considerably better. A threshold of FC ≥250 μg/g demonstrates strong rule-in value, particularly when combined with clinical disease activity (RFI ≥8), thereby substantially increasing diagnostic confidence. The combination of ESR ≥15 mm/h and CRP ≥6 mg/L also demonstrates potential rule-in value. For exclusion of CD, the combination of RFI <8 and FC <250 μg/g, CRP <6 mg/L, and ESR <15 mm/h provide reliable rule-out capacity. Lf and FOB showed minimal diagnostic value. Methodological limitations include this spectrum bias, where only two studies included true ambulatory care populations; verification bias, meaning inappropriate intervals between index and reference tests in four studies; and inconsistent reference standards.

Conclusion: FC is the most valuable biomarker for excluding and confirming CD in ambulatory care, although highly dependent on the setting. CD can be reliably excluded in patients with low inflammatory markers and low symptom burden. The results are limited by the small number of heterogeneous studies, most of which were conducted in OPC, which cannot be considered purely ambulatory care. These biomarkers serve their purpose in triage settings, where they can help in decision-making regarding whether to refer or not. As definitive diagnostic tests, endoscopic evaluation with histology remains the gold standard. Prospective diagnostic studies in ambulatory care populations are needed to further define optimal screening strategies.

NEDERLANDSE SAMENVATTING

Doelstelling: we trachten de diagnostische nauwkeurigheid van biomarkers te evalueren, voor het diagnosticeren van de ziekte van Crohn bij volwassenen en adolescenten die zich presenteren met abdominale pijnklachten in eerstelijnszorg setting.

Methode: we doorzochten systematisch de databases MEDLINE, EMBASE, en Cochrane Library van februari tot juni 2024, met vooropgestelde in- en exclusiecriteria, op zoek naar studies die biomarkers evalueren bij patiënten van ≥14 jaar met vermoedelijke ziekte van Crohn in eerstelijnszorg. De index testen omvatten C-reactief proteïne (CRP), bezinkingssnelheid van erytrocyten (BSE), fecaal calprotectine (FC), fecaal lactoferrine (Lf), fecaal occult bloed (FOB) en een rode vlaggen index (RFI). De gouden standaard werd bepaald als de coloscopie of sigmoïdoscopie met histologische bevestiging, al dan niet met extra beeldvorming. In sets van twee reviewers werden de studies onafhankelijk gescreend, de data geëxtraheerd en het risico op bias geëvalueerd d.m.v. Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2). De analyse werd uitgevoerd d.m.v. forest plots, een Receiver Operating Characteristic (ROC) en dumbbell plot, ter visualisatie van de gegevens.

Resultaten: er werden zes studies geïncludeerd, met een totaal van 1,241 patiënten. Uit de analyses blijkt dat FC de meest klinisch bruikbare biomarker in de eerstelijns zorg. Bij een waarde van <50 µg/g is FC een uitstekende uitsluiter, met een hoge sensitiviteit, maar een lagere specificiteit. Deze resultaten variëren echter sterk naargelang de klinische setting en wachtkamerkans. In een meer duidelijke eerstelijnssetting heeft een FC van <50 µg/g een minimale diagnostische waarde, terwijl deze in de beschreven poliklinieken met dezelfde afkapwaarde aanzienlijk beter presteert. Een drempelwaarde van FC ≥250 µg/g duidt op een sterke aantonende waarde, met name in combinatie met klinische ziekteactiviteit (RFI ≥8), waardoor de diagnostische zekerheid aanzienlijk toeneemt. De combinatie van CRP ≥6 mg/L en BSE ≥15 mm/u toont ook een potentiële aantonende waarde. De combinatie van RFI <8 en FC <250 µg/g, CRP <6 mg/L en BSE <15 mm/u toonden een betrouwbare uitsluitingscapaciteit. Lf en FOB toonden zich van minimale diagnostische waarde. Methodologische beperkingen omvatten deze spectrum bias, waar slechts twee studies echte eerstelijns populaties includeerden; verificatiebias, met ongepaste intervallen tussen index- en referentietesten in vier studies; en inconsistente referentiestandaarden.

Conclusie: FC is de meest waardevolle biomarker voor het uitsluiten en aantonen van de ziekte van Crohn in ambulante zorg, hoewel sterk setting-afhankelijk. De ziekte van Crohn kan betrouwbaar worden uitgesloten bij patiënten met lage ontstekingsmarkers en een lage symptoomlast. De resultaten zijn beperkt door het kleine aantal heterogene studies, waarvan de meeste werden uitgevoerd in poliklinieken, dewelke niet zuiver als eerstelijnszorg te beschouwen zijn. Deze biomarkers dienen hun nut in een triage setting, waarbij ze kunnen helpen in de beslissing om al dan niet te verwijzen. Als definitieve diagnostische test blijft de endoscopische evaluatie met histologie de gouden standaard. Prospectieve diagnostische studies in eerstelijnspopulaties zijn nodig om optimale screening strategieën verder te definiëren.

INTRODUCTION

Worldwide, 189 patients annually consult their general practitioner for abdominal pain, an often harmless and self-limiting complaint (1). Sometimes, however, this might point to a more severe underlying condition such as chronic inflammatory bowel disease (IBD)(2). IBD includes ulcerative colitis (CU) and Crohn's disease (CD). CD presents as inflammation that can involve any segment of the gastrointestinal (GI) tract though it most commonly occurs in the terminal ileum, colon, and rectum (3). Pathophysiologically, if left unrecognized and thus untreated, recurrent flare ups of transmural granulomatous inflammation (4) can lead to complications such as abscesses, adhesions, strictures, perforations, obstructions and fistulas, resulting in a significant reduction in an individual's Quality of Life (QoL) (5). Recognizing the disease early allows timely treatment, reducing the risk of these complications and improving long-term prognosis (6), (7).

The global incidence and prevalence of CD are steadily increasing, particularly in newly industrialized countries (8). IBD is affecting an estimated 3.8 million people globally (9), and it is theorized that 20-60% of the IBD population comprises CD patients, percentages that strongly vary by region and research methods (10), (11). The etiology of CD is unclear due to its postulated multifactorial origin. It is considered to involve complex interactions between genetic susceptibility, environmental exposures, and an inadequate immune response (7). The diagnosis remains challenging, particularly in ambulatory care settings, as patients often present with non-specific symptoms such as abdominal pain, diarrhea, fatigue, and weight loss. These are atypical symptoms that overlap with functional GI disorders such as irritable bowel syndrome, and other inflammatory diseases (12).

Currently, the gold standard for diagnosis consists of a colonoscopy or sigmoidoscopy with biopsy and histological confirmation from the terminal ileum and colon (13), while Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) enterography is used to evaluate disease extent and complications (14). However, these investigations are invasive, expensive, and not readily available in primary or outpatient care. Given the complex and uncertain etiology, unspecific symptomatology and inaccessibility of these diagnostic tools in primary care, CD remains a difficult condition to identify, which results in a possible diagnostic delay ranging from two months (m) up to eight years (y) from the initial onset of symptoms (15). How can we select patients in the general practice (GP), who might benefit from a referral?

Non-invasive biomarkers are promising tools to support the diagnostic work-up of suspected CD. Systemic inflammation and disease activity can be reflected in C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and fecal calprotectin (FC) —yet these markers alone lack sufficient specificity for reliable diagnosis (16). They might, however, be useful as a valuable screening tool in primary care to select a subpopulation to refer for further investigations such as a colonoscopy with biopsies (17)

Our review question is as follows: **what is the diagnostic accuracy of blood and fecal biomarkers for the diagnosis of Crohn's disease in adults visiting ambulatory care?** This

review provides a critical synthesis of existing literature on the use of blood and fecal biomarkers in ambulatory care, assessing their accuracy in distinguishing benign abdominal complaints from Crohn's disease.

METHODS

Search strategy

Electronic searches

We systematically searched MEDLINE, EMBASE and Cochrane Library from February to June 2024. Search strategies were developed with support from an experienced postdoctoral researcher and included MeSH/Emtree terms and free-text terms related to diagnosis, biomarkers (e.g. CRP, FC, blood proteins, iron, ESR), CD, and ambulatory care settings. The full search strategy is available in Appendix 1.

Searching other resources

We screened the reference lists of included studies (snowballing), relevant reviews, and clinical guidelines to identify additional eligible studies.

Selection criteria

Types of studies

We included cross-sectional diagnostic accuracy studies with prospective or retrospective data collection that assessed the performance of blood, urine, and/or fecal biomarkers for the diagnosis of CD in ambulatory care. Only studies published in Dutch, English, or French were considered. We excluded animal or cadaver studies, single case studies, case reports, case series, case-control studies, ongoing studies, conference abstracts, reviews, editorials, lectures, questionnaires, self-reports, and study protocols.

Participants

Eligible studies included adolescents and adults (aged ≥14y) presenting to ambulatory care settings (including GP, primary care centers, OPC, and emergency departments) with symptoms suggestive of CD.

Index tests

Index tests of interest included a broad range of blood, fecal and urine biomarkers, assessed either individually or in combination. These comprised blood biomarkers such as CRP, white blood cell count (WBC), red blood cell count (RBC), ferritin, iron, vitamin B12, folic acid, liver function tests, and ESR, as well as urine and fecal biomarkers, including FC.

We included studies evaluating threshold-based tests, continuous biomarker values, or diagnostic prediction models based on these biomarkers.

Target conditions

The target condition was a diagnosis of CD, including regional ileitis, terminal ileitis, ileocolitis, granulomatous colitis, or IBD type 1, as defined by the concerned paper.

Reference standards

The preferred reference standard was colonoscopy or sigmoidoscopy with biopsies and histological confirmation. In addition, imaging techniques such as MRI or CT enterography were accepted to assess disease extent. Where alternative or less invasive diagnostic pathways were used, their clinical validity was judged by the review team and, if necessary, discussed with a clinical expert.

Data collection and analysis

Selection of studies

Three reviewers, Jarek Wauters (JW), Camille Hoefnagels (CH) and Laura Verhaegen (LV), independently screened titles and abstracts for eligibility. Full-text review was performed for all potentially eligible articles. The selection criteria were defined a priori. Disagreements were resolved through discussion or consultation with a fourth reviewer (Thomas Struyf). For the deduplication, we used Covidence online software.

Data extraction and management

Data were extracted independently by three reviewers (JW, CH, LV) using a pre-specified data extraction form. Extracted data included study characteristics (author, year of publication, country, clinical setting, and year of data collection), participant characteristics (age, sex, inclusion, and exclusion criteria), details of the index tests (test type, thresholds, and timing), reference standards (description, thresholds, and timing), and outcome data for diagnostic accuracy, including sensitivity, specificity, predictive values, likelihood ratios (LR), and the area under the receiver operating characteristic curve (AUROC).

We contacted authors for missing or unclear data when needed.

Assessment of methodological quality

Three reviewers (JW, CH, LV) independently assessed risk of bias using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. Disagreements were resolved through discussion. Studies were not excluded based on quality assessment, but results were considered in the interpretation and sensitivity analyses. Studies with retrospective data collection and non-consecutive inclusion were noted and analyzed separately.

Statistical analysis and data synthesis

We planned to summarize diagnostic accuracy in 2×2 tables where data were available. If ≥4 studies per biomarker were available without significant clinical heterogeneity, pooled sensitivity and specificity were to be estimated using a bivariate random-effects model (BRMA). The pooled results were to be visualized in summary Receiver Operating Characteristic (ROC) plots. Descriptive statistics were used where pooling was not feasible.

To visually compare the diagnostic performance of biomarkers, we constructed dumbbell plots displaying paired positive and negative LR with corresponding confidence intervals. We considered tests with a negative LR ≤0.20 as potentially useful to rule out CD and those with a positive LR ≥5.0 as potentially useful to confirm the diagnosis. In addition, LR were categorized as

minimal (LR+ 1-2, LR- 0.5-1), small (LR+ 2-5, LR- 0.2-0.5), moderate (LR+ 5-10, LR- 0.1-0.2), and significant (LR+ >10, LR- <0.1).

Investigations of heterogeneity

We explored clinical and statistical heterogeneity, including heterogeneity in test thresholds. Subgroup analyses were planned according to age group (adults versus adolescents aged 14–18y), clinical setting (GP versus primary care center versus OPC versus emergency department), data collection method (prospective versus retrospective), and risk of bias domains as assessed using the QUADAS-2 tool.

Sensitivity analyses

If possible, we aimed to perform sensitivity analyses excluding studies at high risk of bias or those with retrospective designs or non-consecutive sampling.

RESULTS

Study selection

The initial search identified 8376 unique studies of which 63 were selected for full-text screening. Ultimately, six studies were included in this review (Figure 1). The additional screening of the references of the six included studies identified four studies for full-text screening. None of the latter were included in this review.

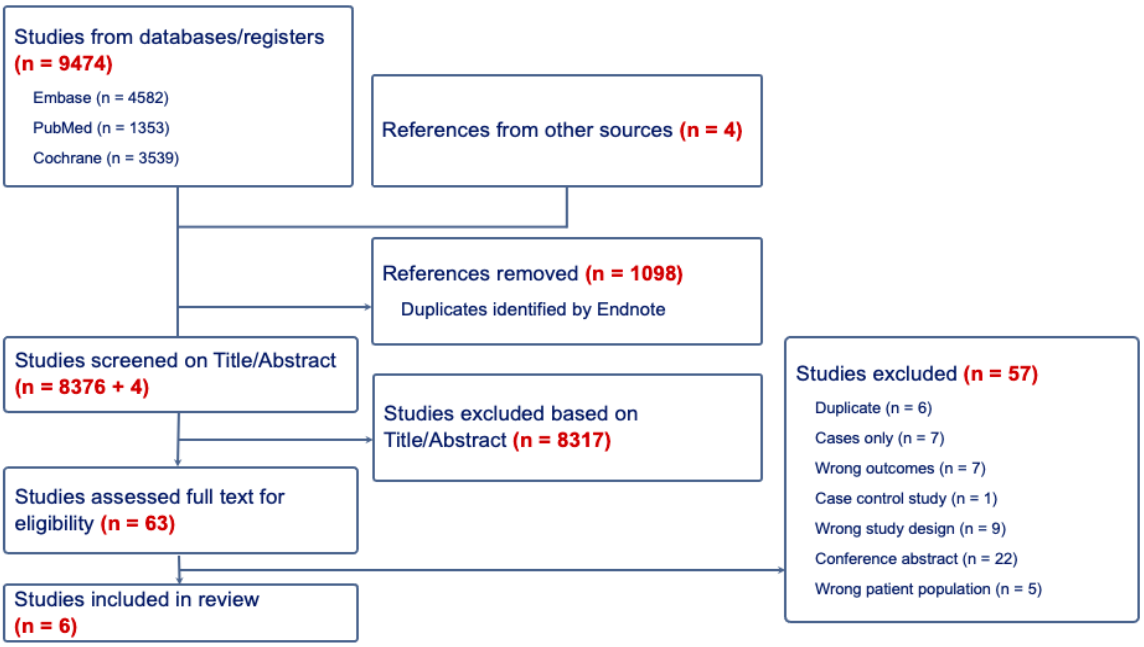


Figure 1. Flowchart of the study selection. n, number.

Study characteristics

Study characteristics of the six included studies are shown in Table 1. The studies were published between 1985 (18) and 2021 (19). Five studies were prospective in nature (19), (20), (21), (18), (22), while one employed a retrospective design (23). The total number of included patients was

1,241. The number of patients (n) per study ranging from n=82 (18) to n=410 (23), with CD prevalence ranging between 0.07 (23) and 48% (21). Reported median ages varied between 30 (21) and 42y (23). One study reported the mean age of 43.5y (20). Two studies did not mention any median or mean age (18), (22). The proportion of male participants ranged from 29 (21) to 40% (23) and was not reported in three studies (20), (18), (22). Index tests included FC (23), (19), (21), CRP (20), (18), ESR (18), red flags index (RFI) (19), fecal lactoferrin (Lf) (22), and fecal occult blood (FOB) (22). The RFI is described in Appendix II. Reference standards were based on histological confirmation from endoscopic biopsies, radiological or capsule endoscopy findings (23), (20), (21), (22), or a combination of clinical, endoscopic, and histopathological criteria (19), (18). The time interval between index and reference testing was variable, ranging from same-day assessment (18) to a maximum of 18m (23). While two studies involved patients from a GP setting (23), (19), three studies were undertaken in OPC (20), (21), (18). The setting was unclear in one study (22). Common exclusion criteria included pre-existing IBD (23), (21), (18), recent Non-Steroidal Anti-Inflammatory Drugs (NSAID) use (21), other GI or systemic diseases (19), (20), (21), (18), (22), and incomplete diagnostic data (23), (19).

Study	Design Population (CD/N; median/mean age; %M; prevalence) Setting; country	Index tests	Reference standard	Time interval between index and reference test	Inclusion criteria	Exclusion criteria
Conroy et al. (2018)	Retrospective 3/410; median 42y; 40%; 0.07% GP; UK	FC	Histology in biopsies taken at lower GI endoscopy, typical radiological features or accepted diagnostic appearances at small bowel video capsule endoscopy	within 18m	Pt underwent FC testing between October 1, 2013, and February 28, 2014	Pt <6y, with known IBD, where FC testing was initiated in secondary care, for whom GPs did not respond regarding further investigations
Fiorino et al. (2021)	Prospective 4/112; median 35y; 37%; 4% GP; Italy	RFI, FC	Blood cell counts, CRP, FC, abdominal ultrasound, and, if necessary, colonoscopy or imaging	max of 3w	Adult pt with intestinal symptoms (chronic abdominal pain, chronic diarrhea, nocturnal diarrhea, unexpected weight loss, or perianal lesions)	Pt with GI diseases. Pt lost to follow-up were considered as not having CD
Poullis et al. (2002)	Prospective 7/203; mean 43.5y; unknown; 3% OPC; UK	CRP	Sigmoidoscopy and rectal biopsy. All pt with abnormal rectal biopsy had full colonoscopy	unclear	New pt referred to a gastroenterology OPC with symptoms suggestive of a lower-bowel disorder, including disturbed bowel habits	Pt with abnormalities detected on routine screening bloods or diagnoses made on other investigations
Jensen et al. (2011)	Prospective 40/83; median 30y; 29%; 48% OPC; Denmark	FC	Ileo-colonoscopy and biopsy	unclear	New pt referred to the gastroenterology OPC, with suspected CD (being clinical, endoscopic and or/ histological) who provided informed consent	Growth of enteropathogenic bacteria in stool cultures, positive serological markers for celiac disease, acute bowel obstruction, elevated serum-creatinine, use of NSAID less than 4w prior to diagnostic procedures, severe claustrophobia, cardiac pacemaker, implanted magnetic foreign bodies, pregnancy and lactation. Small bowel stenosis contraindicated capsule endoscopy. Pt suspected of CU
Shine et al. (1985)	Prospective 19/82; unknown; unknown; 23% OPC; UK	CRP, ESR	Clinical examination, rectal biopsy, full blood count and liver function tests. Further investigations if necessary	same day	Pt referred for evaluation of abdominal symptoms	Pt already treated for IBD or already with non-IBD
Saitoh et al. (2000)	Prospective 13/351; unknown; unknown; 4% unclear; Japan	Lf, FOB	Colonoscopy with biopsy	Pt obtained fecal samples the day before colonoscopy	Pt undergoing colonoscopy at the hospital who provided informed consent	Pt with severe liver, gallbladder, or pancreatic diseases

Table 1. Characteristics of included studies. CD, Crohn's disease; CRP, C-reactive protein; CU, colitis ulcerosa; ESR, erythrocyte sedimentation rate; FC, fecal calprotectin; FOB, fecal occult blood; GI, gastrointestinal; GP, general practice; IBD, inflammatory bowel disease; Lf, fecal lactoferrin; M, males; m, months; Max, maximum; n, number of patients; OPC, outpatient clinic; Pt, patients; RFI, red flags index; UK, United Kingdom; w, weeks; y, years.

Risk of bias

An overview of the risk of bias assessment is presented in Figure 2. Four studies did not avoid inappropriate exclusions (19), (20), (22), (18). In four studies, it was unclear whether the reference standard results were interpreted without knowledge of the results of the index tests (23), (19), (20), (18). In another study, colonoscopy or imaging were only performed when indicated, increasing the risk of partial verification bias (19). Two studies employed limited reference standards, which may have compromised diagnostic accuracy (20), (18). In two studies, there was an inappropriate interval between index test and reference standard (23), (19). In two other studies, the interval was unclear (20), (21). In three studies, the use of multiple reference standards introduced inconsistency (23), (19), (20). In two studies, not all patients were included in the analysis (23), (19).

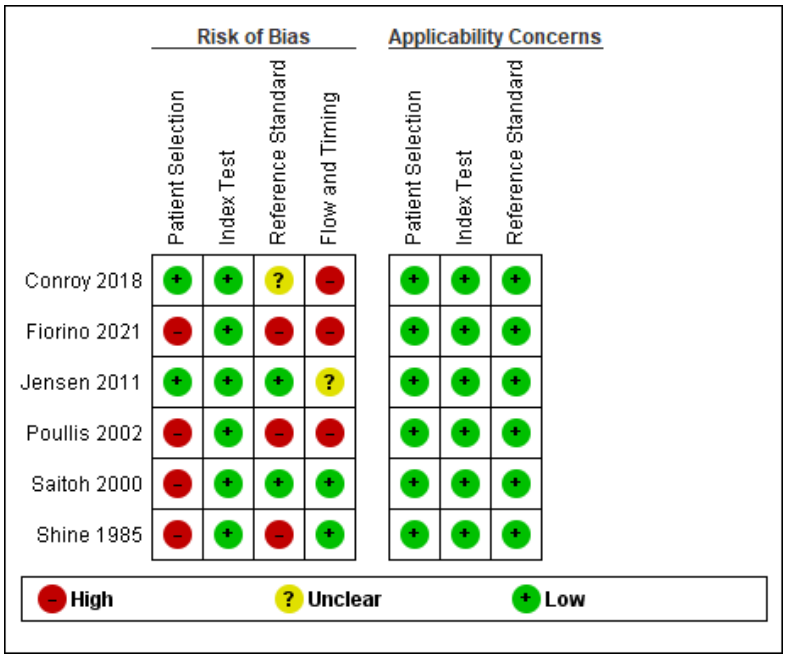


Figure 2. Risk of bias summary: QUADAS-2 risk of bias and applicability.

Diagnostic test characteristics

Diagnostic test accuracy results for the studies are presented in a forest plot (Figure 3), and are plotted in ROC space (Figure 4). The forest and ROC plot of the individual biomarkers are presented in Appendix III to VI.

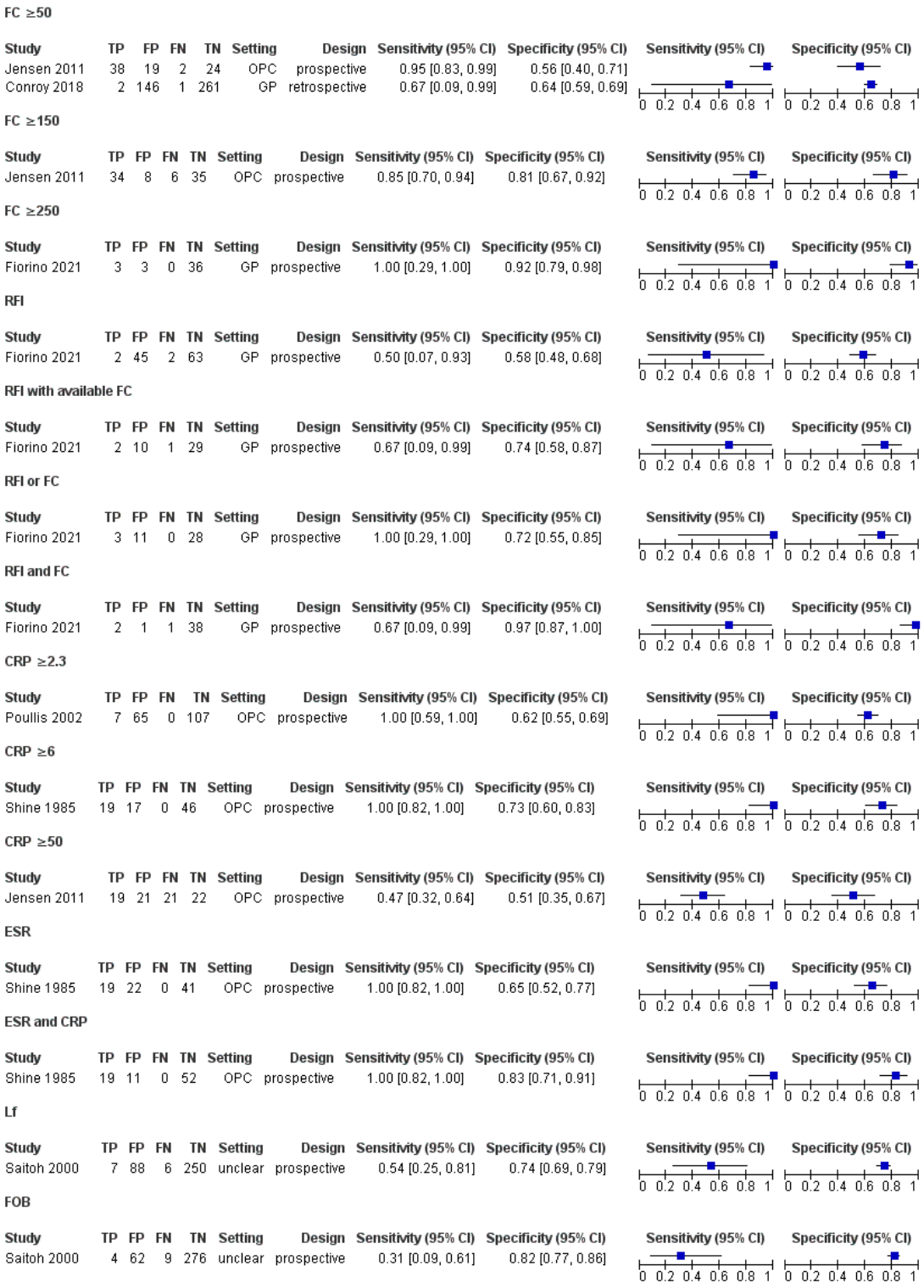


Figure 3. Forest plot: sensitivity and specificity shown for each test and threshold. 95% CI, 95% confidence interval; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FC, fecal calprotectin; FN, false negative; FOB, fecal occult blood; FP, false positive; Lf, fecal lactoferrin; RFI, red flags index; TN, true negative; TP, true positive.

The ROC plot (Figure 4) illustrates the diagnostic performance of the evaluated biomarkers by plotting sensitivity against specificity for each test and threshold. The dashed diagonal line represents the line of no discrimination. Only CRP ≥50 mg/L performed below the reference line, indicating a diagnostic ability worse than chance. FC at the higher cut-offs (≥150 and ≥250 µg/g), and the combination of ESR ≥15 mm/h and CRP ≥6 mg/L demonstrated good sensitivity and specificity, clustering in the upper-right region of the plot. FC ≥50 µg/g (21), CRP ≥2.3 mg/L, and ESR ≥15 mm/h demonstrated good sensitivity but low specificity. In contrast, CRP ≥6 mg/L and both RFI ≥8 and FC ≥250 µg/g showed good sensitivity with moderate specificity. FOB and the combined criterion of RFI ≥8 with FC ≥250 µg/g demonstrated good specificity but low sensitivity. RFI in the subgroup with available FC data (19) and Lf ≥65 ng/mL showed low sensitivity but moderate specificity. Finally, FC ≥50 µg/g (23) and RFI ≥8 demonstrated both low sensitivity and low specificity. Notably, CRP at lower cut-offs (≥2.3 and ≥6 mg/L), ESR ≥15 mm/h, the combination of ESR ≥15 mm/h and CRP ≥6 mg/L, RFI ≥8 or FC ≥250 µg/g, and FC ≥250 µg/g had a perfect sensitivity of 1.0.

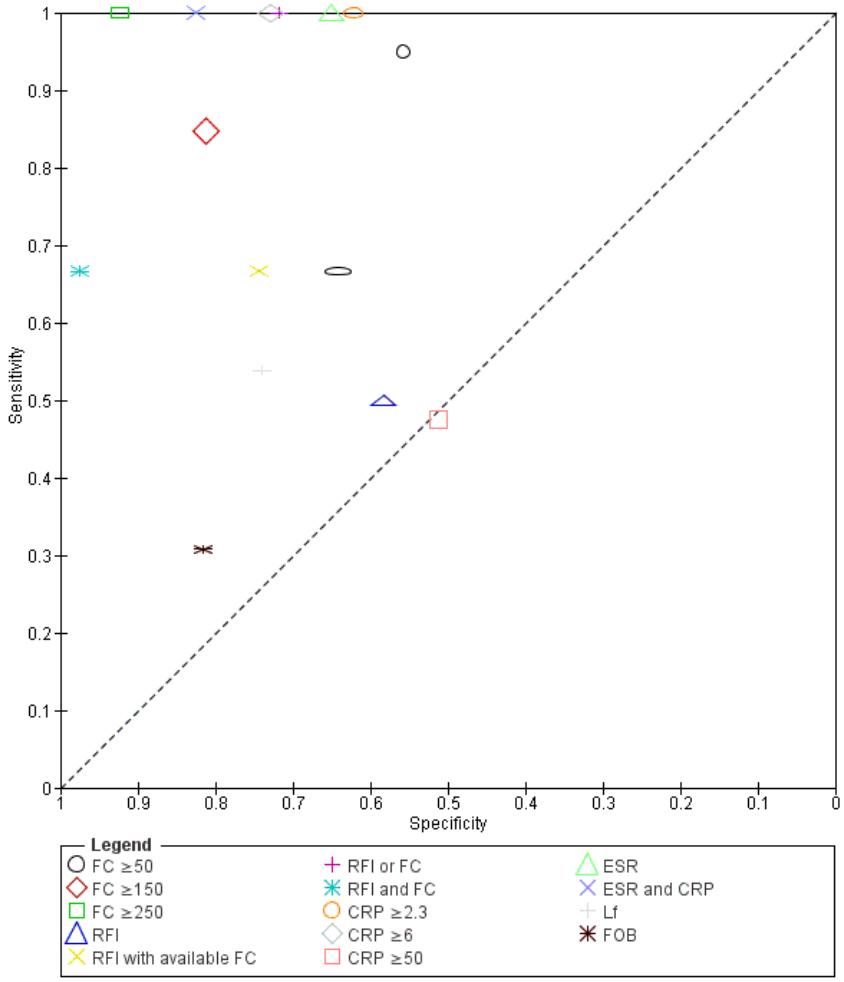


Figure 4: ROC plot: accuracy of the biomarkers by plotting sensitivity against specificity for each test and threshold. CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FC, fecal calprotectin; FOB, fecal occult blood; Lf, fecal lactoferrin; RFI, red flags index.

To further illustrate the ability of a test to either rule in or rule out CD, a dumbbell plot showing pre- and post-test probabilities was constructed (Figure 5). For each test, the pre-test probability was plotted, which is the prevalence of CD in the study (blue dot). The probability of having CD after testing (post-test probability) then changes depending on a positive rest result (red dot marked +) or a negative test result (green dot marked -). Consistent with these findings, tests with significant LR+ demonstrated the greatest ability to rule in CD. FC ≥250 µg/g and the combination of RFI ≥8 and FC ≥250 µg/g showed the strongest rule-in performance, reflected by a marked increase in post-test probability following a positive result. Moderate LR+ values were observed for the combination of ESR ≥15 mm/h and CRP ≥6 mg/L. Several tests showed only small increases in post-test probability after a positive result, including FC ≥50 µg/g (21), FC ≥150 µg/g, RFI ≥8 in the subgroup with available data on FC (19), RFI ≥8 or FC ≥250 µg/g, CRP ≥2.3 mg/L, CRP ≥6 mg/L, ESR ≥15 mm/h, and Lf ≥65 ng/mL. Minimal rule-in value was observed for FC ≥50 µg/g (23), RFI ≥8, and FOB. Notably, CRP ≥50 mg/L demonstrated a LR+ of 0.97, indicating no meaningful change in post-test probability and therefore no discriminative value for ruling in CD.

In contrast, several tests demonstrated strong performance in ruling out CD, as reflected by significant LR-. The absence of FC ≥50 µg/g in Jensen et al. (21), FC ≥250 µg/g, RFI ≥8 or FC ≥250 µg/g, CRP ≥2.3 mg/L, CRP ≥6 mg/L, ESR ≥15 mm/h, and the combination of ESR ≥15 mm/h and CRP ≥6 mg/L were associated with the largest reductions in post-test probability following a negative result. A moderate LR- was observed for FC ≥150 µg/g, while smaller reductions in disease probability were seen for the combination of RFI ≥8 and FC ≥250 µg/g, and RFI ≥8 in the subgroup with available data on FC in Fiorino et al. (19). Minimal rule-out value was observed for FC ≥50 µg/g in Conroy et al. (23), RFI ≥8, CRP ≥50 mg/L, Lf ≥65 ng/mL, and FOB, where negative results resulted in little to no change from the pre-test probability.

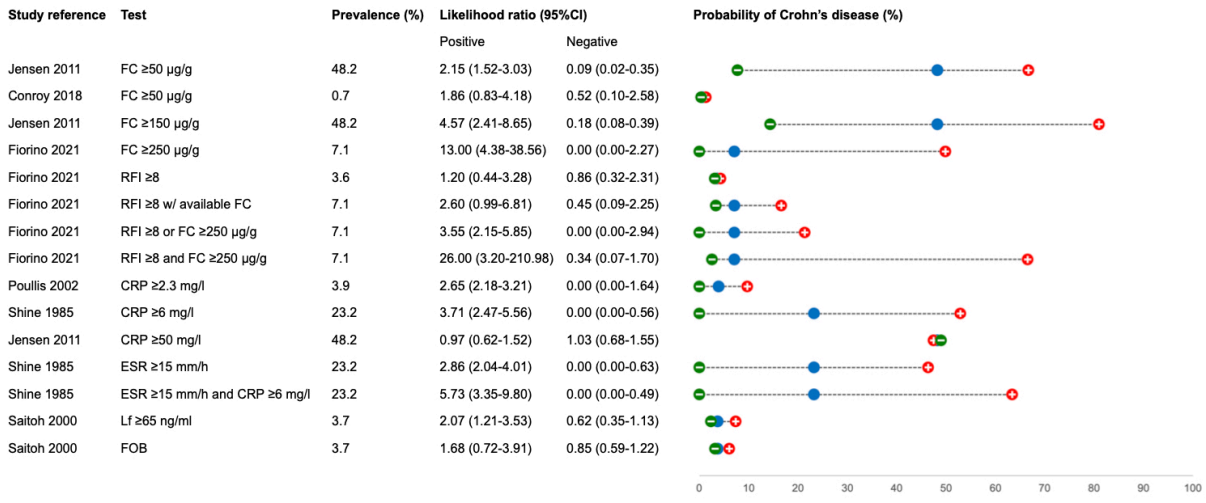


Figure 5. Likelihood ratios (LR) and pre- and post-test disease probability (%) for RFI, blood, and fecal biomarkers. 95% CI, 95% confidence interval; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FC, fecal calprotectin; FOB, fecal occult blood; Lf, fecal lactoferrin; RFI, red flags index.

DISCUSSION

Our systematic review (SR) aimed to provide a clear evaluation of the accuracy of biomarkers in diagnosing CD in adults and adolescents visiting ambulatory care. Six studies were included with a total of 1,241 patients, examining CRP, ESR, FC, LF, FOB, and RFI.

Principal Findings

CRP was evaluated in three studies (20), (18), (21). None of the cutoff values (i.e. 2.3, 6, and 50 mg/L) were useful for confirming the diagnosis of CD due to a LR+ of <5.0. As expected, the lower cutoff values of 2.3 and 6 mg/L demonstrated to be potentially useful to rule out CD due to a LR- of ≤0.20, while the cutoff value of 50 mg/L was not useful. Notably, the lower cutoff values demonstrated a perfect sensitivity (i.e. 1).

A selection bias is present regarding CRP, due to the exclusion of patients already in active treatment for IBD or with confirmed non-IBD diagnoses in Shine et al. (18) and due to strong exclusion of patients with abnormalities detected on routine screening blood or diagnoses made on other investigations in Poullis et al. (20). Moreover, the Shine et al. study (18) had a small sample size (n=82), which also contributes to the perfect sensitivity.

A higher CRP cutoff, as implemented by Jensen et al. (≥50 mg/L) (21) proved non-significant, with an impossibility to discriminate. This might be due to the lack of incidence of such high values in ambulatory care facilities. It must be noted that CRP can be falsely negative while patients may suffer from active GI inflammation (24). On the other hand, CRP is also not specific to intestinal inflammation, which can lead to false positives due to inflammation in other parts of the body (25).

ESR was only described by Shine et al. (18), showing that ESR ≥15 mm/h in combination with CRP ≥6 mg/L is potentially useful to confirm the diagnosis of CD, while relying solely on ESR or CRP is not. Like CRP, ESR is also a non-specific marker of inflammation, which can lead to false positives. Both the combination of ESR and CRP, and ESR alone were potentially useful to rule out CD, with both perfect sensitivity (i.e. 1). However, this result is believed to be due to the selection bias and small sample size of the study, as already mentioned above.

FC proved to be the most studied and clinically useful biomarker. Same as CRP, FC was evaluated in three studies with different cutoff thresholds (i.e. 50, 150, and 250 µg/g). The cutoff value as well as the difference in study design both impacted the diagnostic performance of the test. The cutoff of 250 µg/g is potentially useful to confirm the diagnosis of CD, while the lower cutoff values of 50 and 150 µg/g are not. Whereas the higher cutoff values of 150 and 250 µg/g are potentially useful to rule out CD, a discrepancy was shown between the study of Jensen et al. (21) and Conroy et al. (23), as the former demonstrated the cutoff value of 50 µg/g to be potentially useful to rule out CD, but the latter did not. This discrepancy is likely attributable to differences in study populations and methodological limitations. Conroy et al. (23) represent a true primary care population, whereas Jensen et al. (21) included a gastroenterology OPC, implying a variation in disease severity and case mix. In addition, the time interval between the index and reference test

was within 18m in Conroy et al. (23) and unclear in Jensen et al. (21), whereas it is known that inflammatory biomarkers can fluctuate over time. Moreover, more FN are expected in GP settings, as more patients with early disease and mild inflammation are presented. The higher sensitivity in Jensen et al. (21) could also be due to the strong selection bias and smaller sample size in contrast to the study of Conroy et al. (23). Also, the use of multiple reference standards and incomplete patient inclusion in Conroy et al. (23) may have introduced inconsistency, as well as the retrospective design, which is more susceptible to information bias, non-standardized timing between index test and reference standard, and selective testing, and may therefore inflate FN results and reduce sensitivity in comparison with the study of Jensen et al. (21).

It is worth mentioning that there was also an overt difference in the preanalytical handling of the stool-samples. While some examined the stool immediately, or in a narrow time window, Jensen et al. (21) opted to freeze the stool which could affect the stability of the FC values. Research suggests, however, that FC does not dramatically change after defrosting, compared to non-freezing storage for longer than 24 hours, where it does result in degradation of the FC with 12-35%. This underscores the importance of strict standardized protocols.

Lf, a neutrophil-derived protein which is theoretically useful for the detection of inflammation, has not been shown to be useful at a cutoff value of 65 ng/mL (22), for either ruling in or ruling out CD. In the same study, FOB has also not been shown to be useful for either ruling in or ruling out CD. The high LR- might have been expected in both Lf and FOB given their lack of specificity for inflammatory conditions.

RFI was presented by Fiorino et al. (19) and has shown not to be useful as a standalone test for either ruling in or ruling out CD. However, the combination of RFI ≥8 and FC ≥250 µg/g is potentially useful to confirm the diagnosis of CD, whereas the absence of both RFI ≥8 and FC ≥250 µg/g is potentially useful to rule out CD. Moreover, the combination of RFI ≥8 and FC ≥250 µg/g has a greater LR+ than FC ≥250 µg/g alone. As a standalone test, FC ≥250 µg/g is potentially useful to confirm the diagnosis of CD, whereas the combination of FC ≥250 µg/g or RFI ≥8 is not, possibly due to the low specificity of RFI.

Clinical Implications

These findings indicate that FC is the most clinically useful single index test in the diagnostic work-up of suspected CD. A cut-off value of FC ≥250 µg/g shows a high LR+ (13.00), supporting its role as a useful rule-in test, particularly when combined with clinical disease activity (RFI ≥8), which further increases diagnostic confidence (LR+ 26.00). The combination of ESR ≥15 mm/h and CRP ≥6 mg/L also demonstrates potential rule-in value; however, its moderate LR+ (5.73) limits its discriminative power. Moreover, in Belgium, combined ESR–CRP testing is discouraged, reducing its practical applicability. All other evaluated index tests appear insufficiently accurate for ruling in CD.

For ruling out CD, FC <50 µg/g shows strong potential in Jensen et al. (LR- 0.09)(21), although this finding is not consistent across studies, as Conroy et al. (23) report limited usefulness (LR- 0.52). In contrast, combinations incorporating clinical and biochemical parameters perform well: RFI <8

combined with FC <250 µg/g, CRP <6 mg/L, ESR <15 mm/h, and ESR <15 mm/h or CRP <6 mg/L all demonstrate excellent rule-out capacity. These results suggest that CD can be reliably excluded in patients with low inflammatory markers and low symptom burden, potentially reducing unnecessary endoscopic procedures. All remaining index tests showed insufficient diagnostic performance and should not be used to guide clinical decision-making.

Importantly, endoscopy with histological confirmation remains the gold standard, which cannot be replaced by these markers. Biomarkers should be viewed as triage tools rather than definitive diagnostic tests. Cost-effectiveness is an important consideration. While FC testing incurs costs, it may reduce overall healthcare expenditure by decreasing unnecessary colonoscopies in low-risk patients. However, economic analyses specific to ambulatory care settings are needed.

Strengths and limitations of our systematic review

The strength lies mainly in a rigorous and critical attitude throughout the whole process. We implemented a very broad string-search (PIRTS: Population, Index test, Reference standard, Target disease, Setting) with the guidance of an expert, resulting in a sensitive and comprehensive search strategy. This search was executed independently in three major databases (MEDLINE, EMBASE, Cochrane Library) and should have led us to find all relevant literature. During the selection process, all decisions were independently and blindly assessed by two authors, minimizing selection bias. The selection criteria were defined a priori, and any disagreements were resolved through discussion involving all three authors, and if necessary, a fourth. We also screened the reference lists of included studies, relevant reviews, and clinical guidelines to identify additional eligible studies. In October 2025, we re-searched articles that were postulated as posters or conference abstracts in the beginning of our SR (July 2025), yet we could not find any updates that might be worthy of inclusion. The qualitative assessment of bias using QUADAS-2 was conducted rigorously, and a re-evaluation was performed after we observed that one author was penalized more harshly than another, thus ensuring consistency between reviewers.

The limitations warrant discussion. First, one major limitation might lie in our previously stated strength of selection. We screened 9,474 articles and landed on merely six that were to our satisfaction, based on our predefined in- and exclusion criteria. This reflects either genuine paucity of ambulatory care diagnostic accuracy studies, or we might have been too restrictive in our selection process. The exclusion of case-control studies, while methodologically appropriate to avoid spectrum bias, may have eliminated relevant data. Similarly, only including Dutch, English, and French publications could have introduced language bias, though modern translation tools might have permitted broader inclusion.

Second, on the aspect of setting, we might have been too lenient, resulting in significant heterogeneity in our study settings, limiting its applicability to true primary care. Only Conroy et al. (23) and Fiorini et al. (19) enrolled patients from genuine GP settings. The other four were also recruiting from gastroenterology clinics where disease prevalence and severity differ substantially. This spectrum bias might inflate the diagnostic accuracy estimations compared to unselected primary care populations. We took a liberal definition to “ambulatory care” to include

OPC, which might have compromised the primary care focus. On the other hand, if we had the proposed severity in setting selection, our review would have resulted in only two studies.

Third, there are some questions around the methodological quality of the studies. Four studies failed to avoid inappropriate exclusions, introducing selection bias. Reference standard interpretation was not clearly blinded to index test results in four studies, risking incorporation bias. Two studies employed limited reference standards that may have misclassified disease status. Inappropriate or unclear intervals between index and reference tests in four studies introduce verification bias, as disease status may change over time. The use of multiple reference standards across studies creates inconsistency in the target condition definition.

Fourth, our review excluded all studies with known IBD diagnosis, focusing exclusively on CD rather than IBD. Our review could have benefited from evaluating these biomarkers in not only CD, but in IBD in general. Primarily because it is often a clinical challenge to distinguish IBD subtypes at presentation in primary care settings. A broader IBD focus might have enhanced the clinical utility of our review. We also excluded the pediatric population, younger than 14y of age, which has led to missing an entire subpopulation of literature, with distinct disease characteristics and diagnostic challenges.

Fifth, urinary biomarkers were implemented in our search strategy but excluded due to the underuse in ambulatory care. It could have highlighted the benefits of implementing this easily accessible screening tool in ambulatory care.

Finally, we could not perform meta-analyses due to considerable inter-study heterogeneity per biomarker-cutoff combination. This limited the ability to generate pooled diagnostic accuracy estimates and explore sources of heterogeneity through subgroup analyses.

CONCLUSION

FC emerges as the most clinically useful biomarker in the diagnostic evaluation of suspected CD in a patient population older than 14y, with abdominal complaints, in ambulatory care. However, it is highly dependent on disease prevalence, i.e. pre-test probability, and thus dependent on the type of care-facility in which it is conducted. A threshold of FC ≥250 μg/g demonstrates strong rule-in value, particularly when combined with clinical disease activity (RFI ≥8), thereby substantially increasing diagnostic confidence. The combination of ESR and CRP shows limited additional rule-in utility and is further constrained by restricted applicability in certain healthcare settings. For exclusion of CD, FC at a lower cutoff (<50 μg/g) provides reliable rule-out capacity, as well as RFI <8 and FC <250 μg/g, CRP <6 mg/L, and ESR <15 mm/h, indicating that patients with low inflammatory markers and minimal symptom burden can be safely spared invasive investigations. Our evidence is however limited by the small number of studies, which are heterogeneous in their setting and not merely primary care. These markers should be viewed as a supplementary triage tool to guide decision making, rather than serving as a definitive diagnostic test. The golden standard in diagnosing CD remains the use of an endoscopy with histology, but the availability of FC in primary care should be facilitated to differentiate with more

ease, to prevent unnecessary referrals, and to promote researchers to further study its benefits in primary care.

Future

The need remains for prospective, high-quality diagnostic accuracy studies in ambulatory care facilities, with populations with unknown status of IBD. This is necessary to establish optimal screening strategies for the challenging differential diagnosis in the case of abdominal pain. The study design should incorporate larger patient populations, enrolling consecutive patients with undifferentiated abdominal symptoms before specialist referral. It should start with stricter standardization of the in- and exclusion criteria and should exclude in-between settings such as private gastroenterology clinics and other OPC. The studies should report performance at multiple prespecified FC cutoffs to enable evidence-based threshold selection. Furthermore, comparisons of biomarker combinations (e.g., FC plus CRP) versus single biomarkers would clarify optimal screening strategies. Additionally, investigations should be made in the cost-effectiveness of biomarker-guided versus empiric pathways in ambulatory care. The investments in making FC available in primary care could be more beneficial for the health care system, in a way of reducing unnecessary referrals and investigations. Studies should also start including pediatric populations, and broader IBD phenotypes would enhance generalizability. We would like to see studies on a larger scale, including ambulatory care givers in a longitudinal study with a focus on FC, they truly evaluate the benefit of the marker in ambulatory care and advocate for more accessibility.

BIBLIOGRAPHY

1. Viniol A, Keunecke C, Biroga T, Stadje R, Dornieden K, Bösner S, et al. Studies of the symptom abdominal pain--a systematic review and meta-analysis. Fam Pract. 2014 Oct;31(5):517–29. doi:10.1093/fampra/cmu036 PubMed PMID: 24987023.

2. Perler B, Ungaro R, Baird G, Mallette M, Bright R, Shah S, et al. Presenting symptoms in inflammatory bowel disease: Descriptive analysis of a community-based inception cohort. BMC Gastroenterol. 2019 Apr 2;19(1). doi:10.1186/s12876-019-0963-7 PubMed PMID: 30940072.

3. Richtlijnendatabase. Inflammatoire darmziekten (IBD) bij kinderen [Internet]. 2025 [cited 2026 Mar 19]. Available from: https://richtlijnendatabase.nl/richtlijn/inflammatoire_darmziekten_ibd_bij_kinderen_en_adolescenten/diagnostiek_bij_verdenking_op_ibd.html

4. Alexander C von Roon, George E Reese, Timothy R Orchard, Paris P Tekkis. Crohn's disease. London; 2007. Report.

5. Cohen RD. The quality of life in patients with Crohn's disease. Aliment Pharmacol Ther. 2002 Sep 28;16(9):1603–9. doi:10.1046/j.1365-2036.2002.01323.x

6. Munster LJ, Pronk AJM, Mundt MW, Hompes R, Bemelman WA, van der Bilt JDW, et al. Prolonged Time to Diagnosis of Crohn's Disease in Patients With Perianal Fistulas Negatively Affects Long-Term Outcomes. J Crohns Colitis. 2025 Mar 1;19(3). doi:10.1093/ecco-jcc/jjae146 PubMed PMID: 39297524.

7. Fan YH, Wang MW, Gao YN, Li WM, Jiao Y. Genetic and environmental factors influencing Crohn's disease. World Journal of Experimental Medicine. Baishideng Publishing Group Inc; 2025. doi:10.4240/wjgs.v17.i3.98526

8. Catalan-Serra I, Sebastian S. Global Inflammatory Bowel Disease: Opportunities and Challenges for a New Era. United European Gastroenterology Journal. John Wiley and Sons Inc; 2025. doi:10.1002/ueg2.70075

9. Ruan G, Sun Y, Yu Z, Bai X, Yang H, Qian J. Global, regional, and national burden of inflammatory bowel disease from 1990 to 2021: Findings from the Global Burden of Disease 2021. Gastroenterol Rep (Oxf). 2025;13. doi:10.1093/gastro/goaf082

10. Rajbhandari R, Blakemore S, Gupta N, Mannan S, Nikolli K, Yih A, et al. Crohn's Disease Among the Poorest Billion: Burden of Crohn's Disease in Low- and Lower-Middle-Income Countries. Dig Dis Sci. 2023 Apr 1;68(4):1226–36. doi:10.1007/s10620-022-07675-6 PubMed PMID: 36044105.

11. Sarter H, Crétin T, Savoye G, Fumery M, Leroyer A, Dauchet L, et al. Incidence, prevalence and clinical presentation of inflammatory bowel diseases in Northern France: a 30-year population-based study. The Lancet Regional Health - Europe. 2024 Dec 1;47. doi:10.1016/j.lanepe.2024.101097

12. Claytor J, Kumar P, Ananthakrishnan AN, Colombel JF, Agrawal M, Ungaro RC. Mild Crohn's Disease: Definition and Management. Curr Gastroenterol Rep. 2023 Mar 8;25(3):45–51. doi:10.1007/s11894-023-00863-y

13. Magro F, Langner C, Driessen A, Ensari A, Geboes K, Mantzaris GJ, et al. European consensus on the histopathology of inflammatory bowel disease. J Crohns Colitis. 2013 Nov 1;7(10):827–51. doi:10.1016/j.crohns.2013.06.001 PubMed PMID: 23870728.

14. Biondi M, Bicci E, Danti G, Flammia F, Chiti G, Palumbo P, et al. The Role of Magnetic Resonance Enterography in Crohn's Disease: A Review of Recent Literature. Diagnostics. Multidisciplinary Digital Publishing Institute (MDPI); 2022. doi:10.3390/diagnostics12051236

15. Cross E, Saunders B, Farmer AD, Prior JA. Diagnostic delay in adult inflammatory bowel disease: A systematic review. Indian Journal of Gastroenterology. 2023 Feb 1;42(1):40–52. doi:10.1007/s12664-022-01303-x PubMed PMID: 36715839.

16. Hale MF, Drew K, McAlindon ME, Sidhu R. The diagnostic accuracy of faecal calprotectin and small bowel capsule endoscopy and their correlation in suspected isolated small bowel Crohn's disease. Eur J Gastroenterol Hepatol. 2016;28(10):1145–50. doi:10.1097/MEG.0000000000000696 PubMed PMID: 27384306.

17. Wagatsuma K, Yokoyama Y, Nakase H. Role of Biomarkers in the Diagnosis and Treatment of Inflammatory Bowel Disease. Life. MDPI; 2021. doi:10.3390/life11121375

18. Shine B, Berghouse L, Jones JEL, Landon J. C-Reactive protein as an aid in the differentiation of functional and inflammatory bowel disorders. Clinica Chimica Actu. 1985. Report.

19. Fiorino G, Bonovas S, Gilardi D, Di Sabatino A, Allocca M, Furfaro F, et al. Validation of the Red Flags Index for Early Diagnosis of Crohn's Disease: A Prospective Observational IG-IBD Study among General Practitioners. J Crohns Colitis. 2021 Dec 1;14(12):1777–9. doi:10.1093/ecco-jcc/jjaa111 PubMed PMID: 32990721.

20. Poullis AP, Zar S, Sundaram KK, Moodie SJ, Risley P, Theodossi A, et al. A new, highly sensitive assay for C-reactive protein can aid the differentiation of in¯ammatory bowel disorders from constipation-and diarrhoea-predominant functional bowel disorders. European Journal of Gastroenterology & Hepatology. Lippincott Williams & Wilkins; 2002. Report.

21. Jensen MD, Kjeldsen J, Nathan T. Fecal calprotectin is equally sensitive in Crohn's disease affecting the small bowel and colon. Scand J Gastroenterol. 2011 Jun;46(6):694–700. doi:10.3109/00365521.2011.560680 PubMed PMID: 21456899.

22. Saitoh O, Kojima K, Kayazawa M, Sugi K, Tanaka S, Nakagawa K, et al. Comparison of Tests for Fecal Lactoferrin and Fecal Occult Blood for Colorectal Diseases: A Prospective Pilot Study. 2000 Oct. Report.

23. 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 Apr 1;71(4):316–22. doi:10.1136/jclinpath-2017-204506 PubMed PMID: 28844038.

24. Chang S, Malter L, Hudesman D. Disease monitoring in inflammatory bowel disease. World J Gastroenterol. 2015 Oct 28;21(40):11246–59. doi:10.3748/wjg.v21.i40.11246 PubMed PMID: 26523100.

25. Mouliou DS. C-Reactive Protein: Pathophysiology, Diagnosis, False Test Results and a Novel Diagnostic Algorithm for Clinicians. Diseases. Multidisciplinary Digital Publishing Institute (MDPI); 2023. doi:10.3390/diseases11040132

26. Danese S, Fiorino G, Mary JY, Lakatos PL, D'Haens G, Moja L, et al. Development of Red Flags Index for Early Referral of Adults with Symptoms and Signs Suggestive of Crohn's Disease: An IOIBD Initiative. J Crohns Colitis. 2015 Aug;9(8):601–6. doi:10.1093/ecco-jcc/jjv067 PubMed PMID: 25908718.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

FINANCIAL DISCLOSURE

There has been no financial gain for either of the authors, in writing this article.

FUNDING

There is no funding involved to support the development of our systematic review.

ACKNOWLEDGMENT

First and foremost, we wish to show our gratitude to Prof. Dr. Jan Verbakel and Thomas Struyf for the necessary advice, inspiration and encouragement in completing this dissertation. It would not have been possible without their experienced guidance and efforts. Secondly, we are also grateful for the support of our families throughout this dissertation, but also throughout the years of our education. We acknowledge their sacrifice during the late hours of the many nights.

APPENDIX

Appendix I. Search Strategy

P Adults and adolescents, defined as being 14 years or older

I Blood, fecal and urine biomarkers

R Clinical presentation, colonoscopy/sigmoïdoscopy with biopsies and histology, imaging techniques such as echography, MRI/CT enterography

T Morbus Crohn

S Ambulatory care

Index tests	("Diagnosis"[Mesh] OR "diagnos*"[tiab] OR "Leukocyte L1 Antigen Complex"[Mesh] OR "Leukocyte L1 Antigen Complex"[tiab] OR "Calprotectin"[tiab] OR "C-Reactive Protein"[Mesh] OR "C-Reactive Protein"[tiab] OR "Serum Albumin"[Mesh] OR "albumin"[tiab] OR "Blood Proteins"[Mesh] OR "Blood Proteins"[tiab] OR "Blood Cell Count"[Mesh] OR "Blood Cell Count"[tiab] OR "Iron-Binding Proteins"[Mesh] OR "Iron-Binding Proteins"[tiab] OR "Antibodies"[Mesh] OR "Antibodies"[tiab] OR "Antibody"[tiab] OR "biomarkers"[Mesh] OR "biomarker*"[tiab] OR "CRP"[tiab] OR "C Reactive Protein"[tiab] OR "albumin*"[tiab] OR "RBC"[tiab] OR "WBC"[tiab] OR "Iron"[Mesh] OR "Iron"[tiab] OR "Zinc"[Mesh] OR "Zinc"[tiab] OR "Selenium"[Mesh] OR "Selenium"[tiab] OR "Avitaminosis"[Mesh] OR "Avitaminosis"[tiab] OR "Vitamins"[Mesh] OR "Vitamin*"[tiab] OR "Folic Acid"[Mesh] OR "Folic Acid"[tiab] OR "Blood Sedimentation"[Mesh] OR "Blood Sedimentation"[tiab] OR "ESR"[tiab] OR "erythrocyte sedimentation"[tiab] OR "Fatty Acids, Volatile"[Mesh] OR "Fatty Acid*"[tiab] OR "Fatty Acid-Binding Proteins"[Mesh] OR "I-FABP"[tiab] OR "Intestinal fatty acid binding protein"[tiab])
Target disease	("Crohn Disease"[Mesh] OR Crohn[tiab] OR Crohn's[tiab] OR Crohns[tiab] OR "Regional Ileitis"[tiab] OR "Terminal Ileitis"[tiab] OR "Ileocolitis"[tiab] OR "Granulomatous Colitis"[tiab] OR "regional enteritis"[tiab] OR "Inflammatory Bowel Disease 1"[tiab] OR "Granulomatous Enteritis"[tiab])
Setting	("General Practice"[Mesh] OR "General Practitioners"[Mesh] OR "general practi*"[tiab] OR "family practi*"[tiab] OR "family doctor*"[tiab] OR "family physician*"[tiab] OR "Gp"[tiab] OR "GPs"[tiab] OR "primary care centre*"[tiab] OR "primary care center*"[tiab] OR "primary care setting*"[tiab] OR "Ambulatory care"[Mesh] OR "Ambulatory care facilities"[Mesh] OR "Ambulatory"[tiab] OR "Outpatients"[Mesh] OR "Outpatient*"[tiab] OR "Out-patient*"[tiab] OR "Emergency Service, Hospital"[Mesh] OR "Emergenc*"[tiab])

Appendix II. Red Flags Index

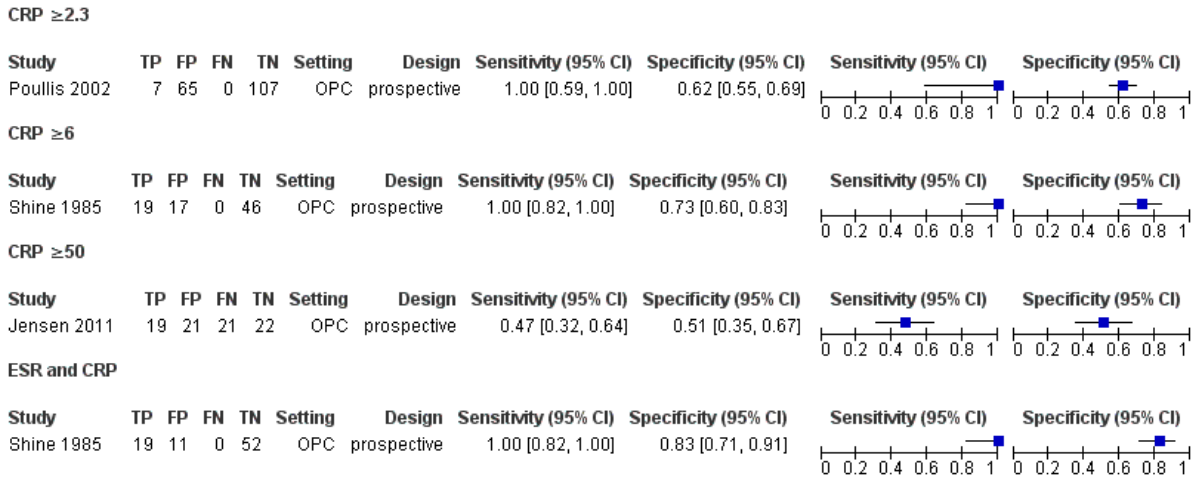
Item	Multiplier
Non-healing or complex perianal fistula or abscess or perianal lesions (apart from hemorrhoids)	5
First-degree relative with confirmed inflammatory bowel disease	4
Weight loss (5% of usual body weight) in the last 3 months	3
Chronic abdominal pain (>3 months)	3
Nocturnal diarrhea	3
Mild fiver in the last 3 months	2
No abdominal pain 30-45 minutes after meals	2
No rectal urgency	2

Every item is scored 0 if absent or 1 if present, and then multiplied by the respective multiplier. The sum results in the score. A score of ≥8 identifies a subject with a high probability of having underlying Crohn's disease (26).

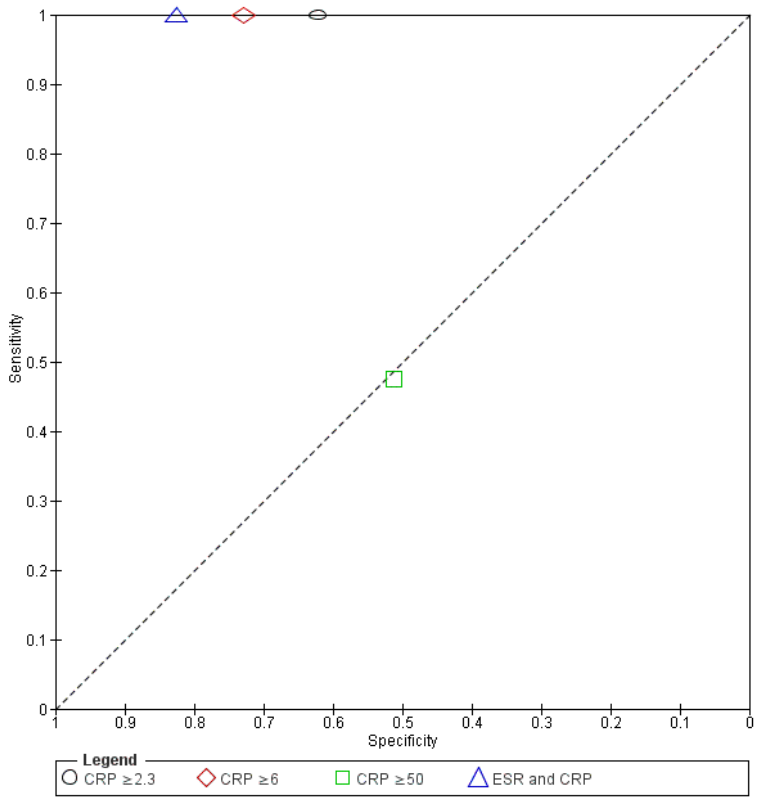
Appendix III. Forest plot and ROC plot for CRP

Sensitivity and specificity shown for CRP, in combination with ESR and the different thresholds, in forest plot (a) and ROC plot (b). 95% CI, 95% confidence interval; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FN, false negative; FP, false positive; TN, true negative; TP, true positive.

a.



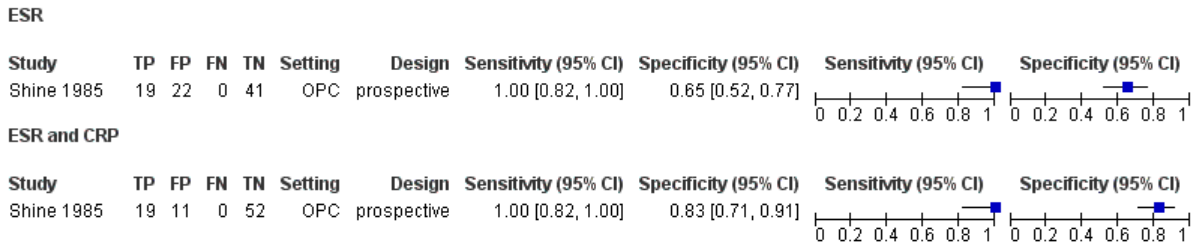
b.



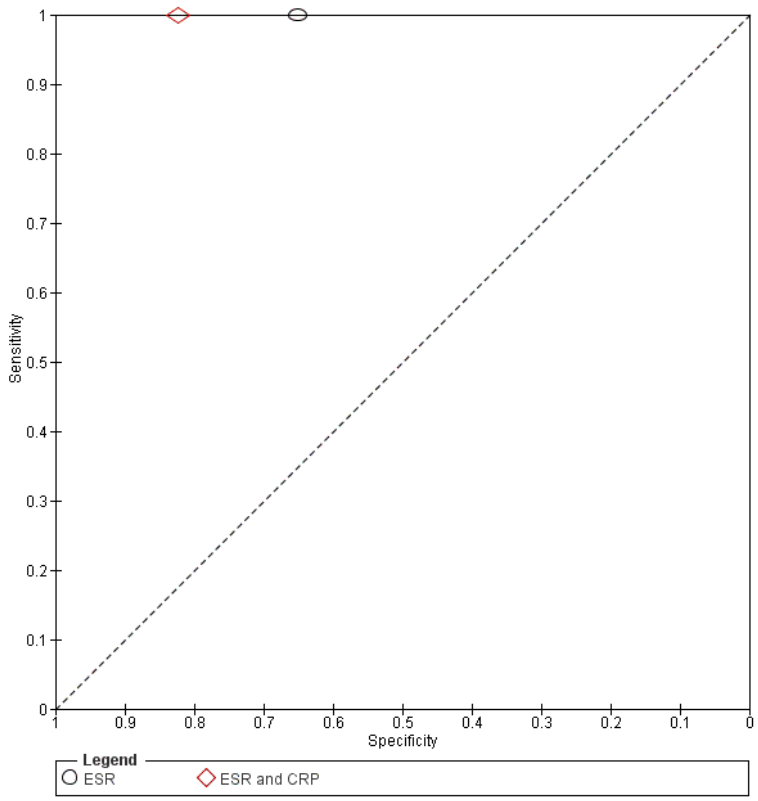
Appendix IV. Forest plot and ROC plot for ESR

Sensitivity and specificity shown for ESR, in combination with CRP, in forest plot (a) and ROC plot (b). 95% CI, 95% confidence interval; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FN, false negative; FP, false positive; OPC, outpatient clinic; TN, true negative; TP, true positive.

a.



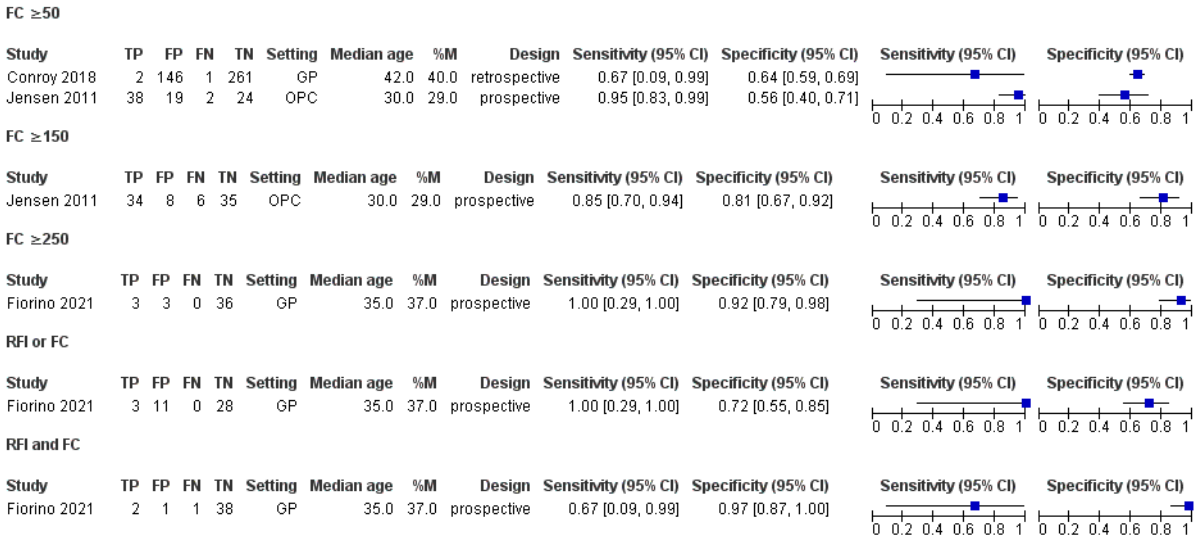
b.



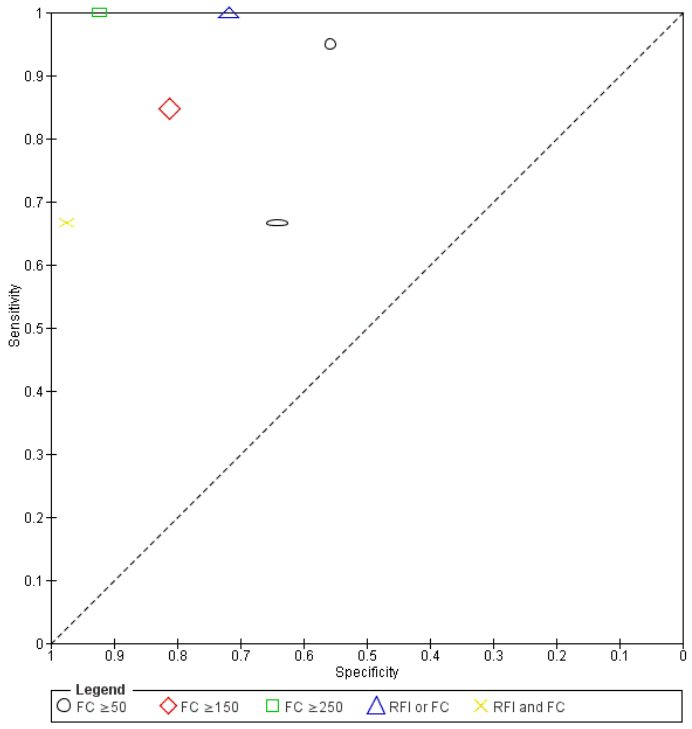
Appendix V. Forest plot and ROC plot for FC

Sensitivity and specificity shown for FC, in combination with RFI and different thresholds, in forest plot (a) and ROC plot (b). 95% CI, 95% confidence interval; CRP, C-reactive protein; FC, fecal calprotectin; FN, false negative; FP, false positive; GP, general practice; OPC, outpatient clinic; RFI, red flags index; TN, true negative; TP, true positive.

a.



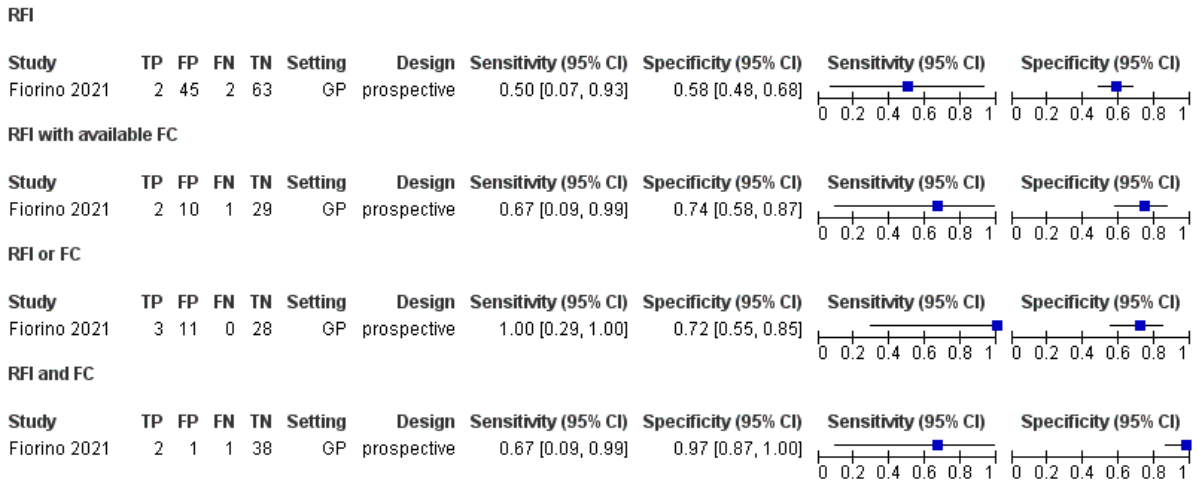
b.



Appendix VI. Forest plot and ROC plot for RFI

Sensitivity and specificity shown for RFI, in combination with FC and different thresholds, in forest plot (a) and ROC plot (b). 95% CI, 95% confidence interval; FC, fecal calprotectin; FN, false negative; FP, false positive; GP, general practice; RFI, red flags index; TN, true negative; TP, true positive.

a.



b.

