Patient-relevant outcomes of stool testing in pediatric inflammatory bowel disease.
Heida, Anke

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CLINICAL UTILITY OF Fecal Calprotectin Monitoring in Asymptomatic Patients with Inflammatory Bowel Disease: A Systematic Review and Practical Guide

A. Heida, K.T. Park, P.F. van Rheenen

Inflamm Bowel Dis 2017;23:894-902
ABSTRACT

BACKGROUND: In asymptomatic patients with inflammatory bowel disease (IBD), “monitoring” involves repeated testing aimed at early recognition of disease exacerbation. We aimed to determine the usefulness of repeated fecal calprotectin (FC) measurements to predict IBD relapses by a systematic literature review.

METHODS: An electronic search was performed in Medline, Embase, and Cochrane from inception to April 2016. Inclusion criteria were prospective studies that followed patients with IBD in remission at baseline and had at least 2 consecutive FC measurements with a test interval of 2 weeks to 6 months. Methodological assessment was based on the second Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist.

RESULTS: A total of 1719 articles were identified; 193 were retrieved for full text review. Six studies met eligibility for inclusion. The time interval between FC tests varied between 1 and 3 months. Asymptomatic patients with IBD who had repeated FC measurements above the study’s cutoff level had a 53% to 83% probability of developing disease relapse within the next 2 to 3 months. Patients with repeated normal FC values had a 67% to 94% probability to remain in remission in the next 2 to 3 months. The ideal FC cutoff for monitoring could not be identified because of the limited number studies meeting inclusion criteria and heterogeneity between selected studies.

CONCLUSIONS: Two consecutively elevated FC values are highly associated with disease relapse, indicating a consideration to proactively optimize IBD therapy plans. More prospective data are necessary to assess whether FC monitoring improves health outcomes.
INTRODUCTION

Inflammatory bowel disease (IBD), consisting of Crohn’s disease (CD) and ulcerative colitis (UC), is a chronic, relapsing and remitting disorder of the gastrointestinal tract. The ultimate goal in IBD is to restore disease remission as early as possible and to prevent disease progression and resistance to pharmacotherapies. The concept of “monitoring” involves repeated testing aimed at early recognition of disease recurrence and timely adjustment of therapy plans.

The ideal monitoring test should be non-invasive, simple to conduct, and easily interpretable. It should detect an imminent disease flare – often undetectable by symptom-based reporting alone – and makes provision for proactive treatment optimization. In Table 1, several frequently used targets for disease monitoring are compared and evaluated for their suitability as a monitoring test in IBD. Although the gold standard for determining mucosal inflammation is endoscopy with histological confirmation, there is a need for clinically useful biomarkers for monitoring purposes since it is unrealistic, costly, and potentially harmful to perform regular, invasive endoscopies. This rationale is particularly true in children affected by IBD and patients with concomitant irritable bowel syndrome.

Calprotectin is a protein released by activated or damaged granulocytes, monocytes, macrophages and epithelial cells. It represents 60% of cytosolic protein in granulocytes and is resistant to metabolic degradation. Fecal calprotectin (FC) levels are related to neutrophil migration to the gastrointestinal tract. FC is a more sensitive marker of active disease compared to the other frequently used surrogate markers (C-reactive protein (CRP) and symptom-based clinical scoring systems, including Crohn’s Disease Activity Index (CDAI), Harvey Bradshaw Index, Pediatric CDAI, Simple Clinical Colitis Activity Index, and the Pediatric Ulcerative Colitis Activity Index (PUCAI). FC represents a practical monitoring test in IBD because testing can be done at home, and the protein is stable at room temperature for at least 3 days.

A general construct for FC-based disease monitoring in patients with IBD is shown in Figure 1, which illustrates the four phases of disease monitoring. Repeated FC measures are used to longitudinally track changes in a patient’s condition over time. In phase I, IBD is suspected, but neither endoscopically confirmed nor treated. In phase II, induction therapy is introduced to achieve disease control, resulting in patient response. Phase III begins with disease remission with continuation of maintenance therapy. The goal of monitoring in this phase is to detect deviations from the target range, indicating...
Table 1: Markers of disease activity used in IBD patients.

<table>
<thead>
<tr>
<th>Validity (correlation with gold standard)</th>
<th>Responsiveness to changes in condition</th>
<th>Signal-to-noise ratio (ability to differentiate changes in condition from background variability)</th>
<th>Practicality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoscopy</td>
<td>Gold standard</td>
<td>Gold standard</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low Requires bowel preparation and in children general anaesthesia</td>
<td></td>
</tr>
<tr>
<td>Symptom-based clinical indices</td>
<td>Poor (^3,13,35-37)</td>
<td>Moderate Affected by subjectivity(^6,7)</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moderate Risk of false positive results (irritable bowel syndrome) and false negative results (dissimulation)(^9,38)</td>
<td></td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>Moderate (^3,12,13,35)</td>
<td>Moderate Risk of false positive results (acute infections and other inflammatory conditions) and false negative results (normal CRP despite active disease)(^39)</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High Quick result; but requires venepuncture</td>
<td></td>
</tr>
<tr>
<td>Fecal calprotectin</td>
<td>Good (^12,29,38,41-43)</td>
<td>Good Risks quickly in case of relapse; falls rapidly with successful treatment(^28)</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moderate Risk of false positive results(^44,45)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>High Possible reluctance by patients for repeated stool collection(^46)</td>
<td></td>
</tr>
</tbody>
</table>

the start of phase IV. In phase IV, therapy is adjusted to re-establish disease control and bring FC levels back to the target range.

Given this background and clinical need for a standardized approach to non-invasive IBD monitoring, we performed a systematic review to evaluate whether FC monitoring could be used to detect imminent disease flares and sustained remission.
METHODS

Eligible studies were those that followed at least 10 patients with IBD in remission at baseline (monitoring phase III) and presented at least two consecutive FC measurements. We accepted FC test intervals between 2 weeks and 6 months. Studies that did not report the use of a FC cut-off (either predefined or based on receiver operating characteristic curves) were excluded from analysis.

Identification and selection of studies

We searched for studies published in Medline, Embase and the Cochrane Library. The search strategy for Medline was (“Leukocyte L1 Antigen Complex”[Mesh] OR “calprotectin”[tw] OR “calgranulin”[tw]) AND (“Inflammatory Bowel Diseases”[Mesh] OR “inflammatory bowel disease”[tw] OR “inflammatory bowel diseases”[tw] OR “IBD”[tw] OR “Crohn”[tw] OR “Colitis”[tw]). For Embase we used (“calgranulin”/exp OR “calprotectin”/exp) AND (“enteritis”/exp OR “inflammatory bowel disease”/exp OR “inflammatory bowel diseases”/exp OR “ibd” OR “crohn” OR “colitis”/exp). We restricted...
our search to studies published in English only. Duplicate articles were manually deleted using RefWorks. For further relevant studies, we checked the reference lists of identified papers. The first selection of studies was carried out by one reviewer (AH) on the basis of title and abstract. The full paper of each potentially eligible study was then obtained. Two authors (AH and PvR) independently assessed full manuscripts against the predefined inclusion criteria. Any disagreements were resolved by discussion, and consensus was reached with the third author (KTP).

Data extraction and management

The following characteristics were extracted from each selected study: name of first author, year of publication, country of origin, journal, study design criteria (prospective vs. retrospective design), sample size (the number of patients in follow-up), baseline characteristics (type of IBD, age group), FC test characteristics (including cutoffs tested), reference standard (endoscopy), other markers of disease activity used (including symptom-based clinical indices and CRP), prevalence of disease flares and the number of true positives, true negatives, false positives, false negatives. Pooling of data was greatly jeopardized due to heterogeneity between studies and was therefore not undertaken.

Assessment of risk of bias and applicability concerns

Study quality was assessed using the QUADAS-2 (QUality Assessment of studies of Diagnostic Accuracy included in Systematic reviews) checklist. In QUADAS four key domains are rated for risk of bias and concerns regarding applicability to the review questions. The signaling questions in each domain were specifically tailored to our review questions (Supplementary Table 1). We did not calculate summary scores because their interpretation is problematic and potentially misleading.

RESULTS

This review includes results of electronic searches up to 21 April 2016. A total of 1719 papers were identified, of which 193 were retrieved for full text review. Of these, 187 were excluded for not meeting the eligibility criteria. Six papers were included in the final analysis (Figure 2).

Study characteristics

Study characteristics of included studies are presented in Table 2. All studies were published in the most recent 3 years, and all except one were from European countries.
Systematic review: monitoring with fecal calprotectin

Records identified through database searching (n=2099)

Duplicates removed (n=380)

Records screened (n=1719)

Records excluded (n=1526)

Full-text articles assessed for eligibility (n=193)

Full-text articles excluded* (n=187)

Studies included in qualitative synthesis (n=6)

Figure 2: Flow diagram systematic literature search.

Sample size varied between 49 and 181 patients. All except one study included adult patients only. The mean proportion of patients experiencing a disease flare during the observation period was 33.3% (184 of 552; range 27 to 50%), and the total observation period was 10 to 18 months. All studies included patients with UC of which one followed patients with disease exclusively confined to the rectum. Two studies also included patients with CD. The time interval between consecutive FC tests varied between one and three months. One study compared control patients assigned to usual care with patients exposed to a FC-guided dose-escalation scheme with oral 5-aminosalicylates. For the sake of clarity we excluded the intervention group from our analysis, since the number of relapses in the intervention group was directly influenced by the therapeutic intervention.

Methodological quality of included studies

The methodological quality of the included studies is summarized in Table 3. All studies used a prospective design, enrolled patients with IBD in remission, used a commercially available FC assay, and tested FC during the initial remission period and periodically thereafter. One study used only clinical activity scores as reference standard instead of endoscopic evaluation. In half of the studies endoscopy was scheduled according to the protocol when relapse was suspected. Differential verification was evident in three studies. Substantial differences between studies were observed in clinical and endoscopic definitions of relapse and predefined FC cut-off levels.
<table>
<thead>
<tr>
<th>Study</th>
<th>N of patients in follow up</th>
<th>Age group</th>
<th>Study aim (prospective if not otherwise specified)</th>
<th>Type of IBD (number)</th>
<th>Proportion of patients with relapse</th>
<th>Median duration of follow-up (months)</th>
<th>Frequency of diagnostic testing (scoring method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dabritz 2013&lt;sup&gt;23&lt;/sup&gt; Germany</td>
<td>181</td>
<td>AC</td>
<td>Monitoring disease activity</td>
<td>UC (120); CD (61)</td>
<td>34%</td>
<td>10</td>
<td>Every 3 months or when suspicion of relapse</td>
</tr>
<tr>
<td>De Vos 2013&lt;sup&gt;28&lt;/sup&gt; Belgium, Norway</td>
<td>87</td>
<td>A</td>
<td>Monitoring disease activity</td>
<td>UC (87)</td>
<td>33%</td>
<td>12 or relapse</td>
<td>Every month Baseline, week 52 (sigmoidoscopy, Mayo endoscopic score)</td>
</tr>
<tr>
<td>Jauregui-Amazega 2014&lt;sup&gt;27&lt;/sup&gt; Spain</td>
<td>64</td>
<td>A</td>
<td>Evaluating accuracy of HR-rectosigmoidoscopy</td>
<td>UC (64)</td>
<td>27%</td>
<td>12 or relapse</td>
<td>Every 3 months Baseline, 12 months or relapse (HR-rectosigmoidoscopy)</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Number</td>
<td>Type of Study</td>
<td>Group Details</td>
<td>Follow-up</td>
<td>Outcome Measures</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>--------</td>
<td>---------------</td>
<td>---------------</td>
<td>-----------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>Lasson</td>
<td>Sweden</td>
<td>91</td>
<td>RCT comparing FC-based pharmacological intervention and usual care</td>
<td>UC (91); control group 40, intervention 51</td>
<td>Every 3 months</td>
<td>Baseline (sigmoidoscopy) Baseline (Mayo score)</td>
<td></td>
</tr>
<tr>
<td>Molander</td>
<td>Finland</td>
<td>49</td>
<td>Monitoring and predicting disease activity after stopping anti-TNF therapy</td>
<td>UC (28); CD (16); IBD-U (5)</td>
<td>12</td>
<td>0,1,2,3,4,5,6, 8,10,12 months or suspicion of relapse (ileocolonoscopy SES-CD or Mayo endoscopic subscore) 0,1,2,3,4,5,6, 8,10,12 months or suspicion of relapse (HBI (CD) or partial Mayo (UC)) 0,1,2,3,4,5,6, 8,10,12 months or suspicion of relapse</td>
<td></td>
</tr>
<tr>
<td>Yamamoto</td>
<td>Japan</td>
<td>80</td>
<td>Monitoring disease activity</td>
<td>UC proctitis (80)</td>
<td>10</td>
<td>Every 2 months Baseline and suspicion of relapse (endoscopy, UC-DAI score) Every 2 months (UC-DAI score, PGA) Every 2 months</td>
<td></td>
</tr>
</tbody>
</table>

Total 552 33.3%

Abbreviations: A = adults; C = children; CD = Crohn’s Disease; HBI = Harvey Bradshaw Index; IB = inflammatory bowel disease; IBD-U = IBD-unclassified; N = number of participants; (P)CDAI = (Pediatric) Crohn’s Disease Activity Index; PGA = Physicians Global Assessment; (P)UCAI = (Pediatric) Ulcerative Colitis Activity Index; SES-CD = Simple Endoscopic Score for Crohn’s disease; UC = ulcerative colitis; UC-DAI = Ulcerative Colitis Disease Activity Index.
Chapter 6

Table 3: QUADAS-2 checklist

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk of Bias</th>
<th>Applicability concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient selection</td>
<td>Index test</td>
</tr>
<tr>
<td>Dabritz 2013</td>
<td>✁</td>
<td>✁</td>
</tr>
<tr>
<td>De Vos 2013</td>
<td>✁?</td>
<td>✁</td>
</tr>
<tr>
<td>Jauregui-Amazega 2014</td>
<td>✁?</td>
<td>✁</td>
</tr>
<tr>
<td>Lasson 2015</td>
<td>✁</td>
<td>✁</td>
</tr>
<tr>
<td>Molander 2015</td>
<td>✁?</td>
<td>✁</td>
</tr>
<tr>
<td>Yamamoto 2015</td>
<td>✁</td>
<td>✁</td>
</tr>
</tbody>
</table>

= low risk of bias; ✁ = high risk of bias; ? = unclear risk of bias

Prognostic value of repeated FC measurements for relapse and sustained remission

All patients included in the final analysis collected the first feces sample while in remission. Most individual studies showed that asymptomatic patients with FC levels moving out of the normal range on the next measurement had higher risk of relapse within the next 2 to 3 months. When FC was elevated the probability of relapse increased to 53 - 83%, as is shown in Table 4.24–28 Consecutive normal FC values were associated with reduced risk of relapse, with 67 - 94% probability of remission in the next 2 to 3 months. One study investigated the prognostic value of ≥2 consecutive measurements above the upper limit of normal,28 while the others focused on an upward trend of FC between two measurements.23–27 As can be seen in Table 4, the former strategy resulted in the highest probability of relapse.

Optimal FC cut-off for monitoring disease activity

Probabilities of relapse and remission varied between studies, partly because different FC cut-offs were used. Variation in FC cut-offs could not explain all the difference. Patient variation, study design and type of FC assay may also have contributed to the heterogeneity of the test accuracy. Because of the limited number of studies included in this systematic review, we were not able to derive the ideal cut-off point.
<table>
<thead>
<tr>
<th>Study</th>
<th>FC assay</th>
<th>Upper limit of normal range (µg/g)</th>
<th>Basis of relapse diagnosis</th>
<th>Pretest probability of relapse</th>
<th>Post-test probability of relapse when upward trend in FC out of normal range (95%CI)</th>
<th>Post-test probability of relapse when consecutive values in normal range (95%CI)</th>
<th>Time between drift out of normal range to relapse</th>
<th>N per 100 patients</th>
<th>True Positives</th>
<th>True Negatives</th>
<th>False Positives</th>
<th>False Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dabritz</td>
<td>Immuno-diagnostic</td>
<td>15</td>
<td>C</td>
<td>34%</td>
<td>63% (55-71)</td>
<td>12% (8-19)</td>
<td>2-3 months</td>
<td>27</td>
<td>51</td>
<td>15</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>De Vos</td>
<td>PhiCal</td>
<td>300*</td>
<td>C&amp;E</td>
<td>33%</td>
<td>83% (61-94)</td>
<td>20% (15-27)</td>
<td>3 months</td>
<td>17</td>
<td>63</td>
<td>4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Jauregui-Amazega</td>
<td>Cerba Intern</td>
<td>250</td>
<td>E</td>
<td>27%</td>
<td>53% (33-73)</td>
<td>18% (12-26)</td>
<td>3 months</td>
<td>13</td>
<td>62</td>
<td>11</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Lasson</td>
<td>Bühlmann</td>
<td>300</td>
<td>C</td>
<td>50%</td>
<td>57% (47-67)</td>
<td>33% (15-58)</td>
<td>Unknown</td>
<td>40</td>
<td>20</td>
<td>30</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Molander</td>
<td>Calpro</td>
<td>200</td>
<td>E</td>
<td>31%</td>
<td>57% (36-76)</td>
<td>20% (12-30)</td>
<td>2-4 months</td>
<td>17</td>
<td>57</td>
<td>12</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Yamamoto</td>
<td>Calpro</td>
<td>55</td>
<td>E</td>
<td>30%</td>
<td>66% (52-77)</td>
<td>6% (2-16)</td>
<td>2 months</td>
<td>26</td>
<td>56</td>
<td>14</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI: confidence interval; C: relapse defined as clinical relapse; E: relapse defined as endoscopic relapse; C&E: relapse defined as both clinical relapse or endoscopic relapse.
* FC value above cut-off in two consecutive months.
** Only control group included in this table.


DISCUSSION

In this systematic review, we evaluated the utility of FC monitoring to detect imminent flares in asymptomatic patients with IBD. We identified only six studies meeting our inclusion criteria. Data collection was done prospectively in consecutive series of mostly UC patients with quiescent disease at baseline. We found that there was poor consistency of reference standard use and definition of relapse between the studies. Two consecutively elevated FC levels appeared to be the best predictor for relapse, but this was systematically investigated in only one study. An upward trend of FC out of the normal range was also prognostic for relapse, albeit with a lower probability of relapse.

Comparison with other reviews

We report the first systematic review that investigates the prognostic value of repeated FC measurements in asymptomatic patients with IBD. To date, there have been two meta-analyses of the diagnostic accuracy of a single FC measurement in almost exclusively symptomatic patients with previously diagnosed UC or CD. In these circumstances, symptom-based clinical indices and derangements in serological markers of inflammation would likely lead clinicians to intensify medical therapy. Inclusion of these studies may cause overestimation of the prognostic value of calprotectin relative to the practical situation, where a monitoring test is necessary to discriminate between those who have preclinical relapse and those with quiescent IBD. We moved away from single FC measurements that are read in isolation when relapse is suspected, and focused on repeated FC measurements in asymptomatic patients to predict relapse. Based on our review, we found that FC levels start rising 2 to 3 month before a relapse becomes apparent, and therefore support the biological implausibility that a single FC measurement at baseline can predict the clinical course over a 12 months period, as was suggested in a meta-analysis by Mao et al.

Cut-off levels

Furthermore, we were not able to identify the best FC cutoff for monitoring purposes. Currently, there is no consensus among IBD experts about the range of FC associated with mucosal healing, indicating a need for prospective and randomized studies comparing monitoring strategies that vary in thresholds.

Clinical Implications

Table 5 elaborates on the specific outcomes when FC monitoring strategy leads to effective adjustments in IBD therapy from a patients’ perspective. The underlying
assumption here is that FC monitoring serves to improve patient-centered outcomes, representing a proactive approach to detecting indolent disease activity. Of note, when adopting FC monitoring, key questions most relevant to decision making are whether the numbers of false negatives (missed cases with relapse) and false positives (cases without disease activity who may receive treatment intensification) are acceptable within the new monitoring paradigm.

Table 5: Implications of fecal calprotectin test results

<table>
<thead>
<tr>
<th>Consequences</th>
<th>Importance*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True positives</strong></td>
<td></td>
</tr>
<tr>
<td>Interpretation: Patient has active disease despite being symptom-free.</td>
<td></td>
</tr>
<tr>
<td>Presumed patient outcome: May benefit from shorter delay and potential early adjustment of therapy (intensity/switch/add)</td>
<td>CRITICAL</td>
</tr>
<tr>
<td><strong>True negatives</strong></td>
<td></td>
</tr>
<tr>
<td>Interpretation: Patient is in remission</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Presumed patient outcome: Benefit from reassurance</td>
<td></td>
</tr>
<tr>
<td><strong>False positives</strong></td>
<td></td>
</tr>
<tr>
<td>Interpretation: Patient is in remission, FC elevated</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Presumed patient outcome: Detriment from exposure to overtreatment</td>
<td></td>
</tr>
<tr>
<td><strong>False negatives</strong></td>
<td></td>
</tr>
<tr>
<td>Interpretation: Patient has active disease, but is not (yet) recognized</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Presumed patient outcome: Detriment from delayed diagnosis and delayed adjustment of therapy. False reassurance leading to ignoring symptoms</td>
<td></td>
</tr>
<tr>
<td><strong>Inconclusive results</strong></td>
<td></td>
</tr>
<tr>
<td>Interpretation: Not sure whether this increase in FC is clinically relevant</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Presumed patient outcome: Detriment from increased anxiety by uncertainty until next FC test result. May benefit from avoidance of overtreatment</td>
<td></td>
</tr>
<tr>
<td><strong>Complications of test</strong></td>
<td></td>
</tr>
<tr>
<td>May be perceived as unsanitary</td>
<td>NOT IMPORTANT</td>
</tr>
<tr>
<td><strong>Resource utilization (cost)</strong></td>
<td></td>
</tr>
<tr>
<td>Increases cost for ambulant diagnostic testing; however, endoscopy has much greater resource implications. FC-based home monitoring may reduce cost for out-patient health checks</td>
<td>IMPORTANT</td>
</tr>
</tbody>
</table>

* GRADE recommends classifying each outcome as either “critical for decision making”, “important but not critical for decision making”, or “not-important”.

Emerging evidence suggest that FC monitoring has the potential to result in less missed cases of asymptomatic IBD patients with on-going mucosal-level inflammation. In particular, IBD patients who underreport symptoms and pediatric patients requiring
anesthesia for each endoscopic evaluation are two subset of patients who may benefit from FC monitoring. From a patient perspective, bowel preparation for colonoscopy, repeated anaesthesia, and incurring indirect costs are practical and important considerations in favor of FC monitoring. Additionally, FC monitoring may serve as a feedback tool for better patient engagement, facilitating self-management strategies of their chronic condition.

Although there is no consensus on the optimal frequency of calprotectin retesting and cut-offs for treatment intensification, the authors of this paper routinely monitor children with IBD using an enzyme-linked immunosorbent assay (ELISA) allowing quantification. A practical cut-off range could be as follows: levels below 250 μg/g as indicative for disease remission (green), levels above 500 μg/g as indicative for disease flare (red), while levels between 250 and 500 μg/g indicating need for more frequent calprotectin monitoring (yellow), as shown in Figure 1. This “traffic light” is currently being evaluated in a prospective multicenter telemonitoring program. Future studies are needed to determine whether pre-emptive treatment intensification based on elevated FC levels will lead to long-term better patient outcomes, including reduction of hospitalizations, disability-associated costs and loss of productivity. The first prospective trials with mesalamine dose intensification and infliximab dose interval adjustment have already been performed with promising results.

Methodological limitations of the review

Although the methodology to conduct a systematic review and meta-analysis of diagnostic research is developed to a certain extent, at least for dichotomized tests, the systematic evaluation of a monitoring test is not bound to consensus guidelines. Although the papers we selected had to meet high methodological standards, we acknowledge several limitations. Significant heterogeneity in disease spectrum, study endpoints, FC cut-off levels, and quality of reporting are potentially confounding factors that may affect interpretation of the data and conclusions. Also, we restricted our search to studies published in English only, leading to potential bias.

CONCLUSION

This systematic review shows that the relapsing and remitting nature of IBD becomes less unpredictable with proactive FC monitoring in clinical practice, allowing early recognition of relapse prior to overt symptoms (or symptom reporting). While FC monitoring may represent a more proactive strategy for treatment modifications in a treat-to-target approach, more robust data are necessary to determine whether it will
improve decision-making and patient-centered outcomes.
REFERENCES

5. Loonen HJ, Derkx BHF, Koopman HM, Heymans HSA. Are parents able to rate the symptoms and quality of life of their offspring with IBD? Inflamm Bowel Dis 2002;8:270–6.


Systematic review: monitoring with fecal calprotectin

### Supplementary Table 1: QUADAS-2 criteria

**DOMAIN 1: Patient selection**

**RISK OF BIAS:** Could the selection of patients have introduced bias?

- Was a consecutive or random sample of patients enrolled?
  Score “yes” if the following words are stated in the article: consecutive, random or all patients were included in a defined time period.

- Did the study avoid inappropriate exclusions?
  Score “yes” if the study avoided inappropriate exclusions.

**APPLICABILITY CONCERNS:** Are there concerns that the included patients and setting do not match the review question?

- Was the diagnosis IBD previously confirmed by means of endoscopy?
  Score “yes” if patients had confirmed IBD

- Were all included patients in remission at the start of the observation period and were they followed over time or until disease relapse?
  Score “yes” if patients were in remission and consecutively followed up.

**DOMAIN 2: Index test**

**RISK OF BIAS:** Could the conduct or interpretation of the index test have introduced bias?

- Were the index test results interpreted without knowledge of the results of the reference standard?
  Score “yes” if the fecal calprotectin results were interpreted without knowledge of the results of the reference standard.

- If a cutpoint was used, was it pre-specified?
  Score “yes” if a fecal calprotectin cutpoint was pre-specified.

**APPLICABILITY CONCERNS:** Are there concerns that the index test, its conduct, or interpretation differ from the review question?

- Was fecal calprotectin measured with a common assay?
  Score “yes” if fecal calprotectin was measured with a commercially available assay.

- Was a fecal calprotectin test performed during initial remission?
  Score “yes” if FC test results were presented during initial disease remission.

**DOMAIN 3: Reference standard**

**RISK OF BIAS:** Could the reference standard, its conduct, or its interpretation have introduced bias?

- Is the reference standard likely to correctly classify the target condition?
  Score “yes” if at least one endoscopy was performed as reference standard.

- Were the reference standard results interpreted without knowledge of the results of the index test?
  Score “yes” if the reference standard was interpreted without knowledge of the fecal calprotectin test result.
**APPLICABILITY CONCERNS:** Are there concerns that the target condition as defined by the reference standard does not match the question?

- Was a disease flare assessed with endoscopy?
  Score “yes” if a disease flare was assessed with endoscopy in all cases.

**DOMAIN 4: Flow and timing**

**RISK OF BIAS:** Could the patient flow have introduced bias?

- Was the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between both tests?
  Score “yes” if the feces was collected within one month of the reference standard.

- Did all patients receive the same reference standard?
  Score “yes” if all patients received the same reference standard.

If all signaling questions within a domain were scored “yes”, the risk of bias was scored as “low”. If any signaling question within a domain was scored as “no”, the risk of bias was scored as “high”. When insufficient data was reported, the domain was scored as “unclear.”