Comparison of three tests for faecal calprotectin in children and young adults: a retrospective monocentric study

Christine Prell,1 Dorothea Nagel,2 Folke Freudenberg,1 Andrea Schwarzer,1 Sibylle Koletzko1

ABSTRACT

Objective: Faecal calprotectin is used as a sensitive marker for gastrointestinal mucosal inflammation. We compared the performance of three different assays in a large cohort of symptomatic paediatric patients.

Design: Retrospective monocentric study.

Setting: Inpatients and outpatients of a tertiary referral centre for paediatric gastroenterology.

Participants: 304 symptomatic patients (163 males, aged 2–20 years) with active inflammatory bowel disease (IBD/A, n=130), IBD in clinical remission (IBD/R, n=62), other intestinal diseases (n=45) and controls without identified intestinal disease (n=67).

Interventions: Calprotectin was measured in homogenised faecal samples with three tests (A: EliA Calprotectin, Phadia AB, Sweden; B: PhiCal, Calpro AS, Norway; C: EK-Cal, Bühllmann Laboratories, Switzerland).

Outcomes: Concordance between tests was calculated using Kendall’s τ coefficient.

Results: IBD/A and controls were correctly classified as 97.7%/82.1% (A), 97.7%/85.1% (B) and 98.4%/62.7% (C; not significant). Test C tended to have higher calprotectin values with a lower specificity compared to tests A and B. The concordance between two tests was 0.835 for tests A and B, 0.782 for tests A and C and 0.765 for tests B and C.

Conclusions: All three tests are very sensitive for detecting mucosal inflammation, but major differences exist between specificity and absolute values. It is highly advisable to use the test of the same manufacturer for follow-up and to monitor for disease activity.

INTRODUCTION

Faecal calprotectin (FC) is a sensitive, non-invasive marker for mucosal inflammation. It is a small calcium-binding protein belonging to the S100 family1 and accounts for 60% of the cytosolic proteins in neutrophilic granulocytes, but is also found in smaller parts in monocytes and macrophages.2 It is considered to have bactericidal and fungicidal properties.3 Elevated calprotectin concentrations are reported in plasma, synovial fluid, urine and faeces if there is an ongoing inflammation with recruitment of neutrophils.4 The advantage of FC measurement is the stability in stool samples of up to 7 days at room temperature which makes stool collection and transport simple and feasible.3 Increased FC concentrations have been found in several inflammatory gastrointestinal (GI) diseases including gastric cancer, colorectal cancer and colonic polyps, as well as in chronic inflammatory bowel disease (IBD),5 allergic colitis,6 inflammatory bowel disease (IBD),7–9
necrotising enterocolitis\(^6\) and cystic fibrosis.\(^7\) Increased FC concentrations can be considered as inflammation specific, but not disease specific. In paediatric patients FC has an excellent sensitivity for detection of new, untreated paediatric patients with active IBD.\(^8\) 9  A recent meta-analysis of the diagnostic accuracy of FC during the investigation of suspected paediatric IBD including eight studies with 394 children with IBD and 321 controls showed a pooled sensitivity of 0.978 (95% CI 0.947 to 0.996) with a negative likelihood ratio of 0.03. These excellent values have significantly reduced the number of invasive methods necessary for the diagnostic workup in children presenting with GI symptoms. In addition, FC is commonly used to monitor the success of treatment and to predict relapse in patients with IBD.\(^10\)–\(^12\) While several studies compared the performance of different faecal markers,\(^13\)–\(^15\) published data are scarce to compare the performance of different FC tests in the same faecal samples. The majority of commercially available tests use a cut-off of 50 µg/g stool between normal and elevated values, and values between 50 and 100 µg/g are considered as borderline. A low intertest variability regarding normal, borderline, moderate and high elevated levels is crucial when clinicians make decisions based on the results of FC concentrations. The aim of the study was to compare head to head the results of one fluorescence enzyme immunoassay (FEIA) and two ELISA in a large cohort of unselected paediatric patients who were investigated for FC in our institution because of different symptoms.

### PATIENTS AND METHODS

We retrospectively investigated stool samples of symptomatic children and adolescents that had been seen as inpatients or outpatients in the Dr. v. Hauner Children’s Hospital from December 2009 to September 2011 and in which the caring physician had ordered a test for FC. Patients with multiple measurements were only considered for their first sample. After the initial determination for clinical purposes faecal samples were frozen at –20°C up to 18 months before analysis. The patients’ final diagnoses were extracted from the hospital charts. Patients were excluded when there was insufficient clinical information on the final diagnosis or if they were younger than 2 years of age, because particularly infants are known to have high FC levels in the absence of disease.\(^16\) 17 In patients with proven IBD the disease activity at the time of stool sampling was judged by a paediatric gastroenterologist as clinical remission or active disease based on clinical history and physical examination. Patients were classified into four groups: patients with newly diagnosed or relapsed active IBD (group 1), IBD in remission (group 2), children with another confirmed intestinal pathology (group 3, mixed disease) and control children with no finding at the time of measurement or during follow-up indicating an organic GI disorder (group 4, controls).

Coded faecal samples were measured for calprotectin concentrations using one FEIA (test A: EliA Calprotectin, Phadia AB, Sweden) and two different ELISA tests (test B: PhiCal, Calpro AS, Norway; test C: EK-Cal, Bühllmann Laboratories, Switzerland). Faecal samples were homogenised prior to extraction. Homogenised samples were extracted separately for kits A, B and C according to the respective manufacturers’ directions for use, using the ‘Faecal sample preparation kit’ (Roche Diagnostics, Germany). Sample extracts were assayed according to respective manufacturers’ directions for use. The cut-off value between negative and positive given by the manufacturer was 50 µg/g for all three tests. The lowest and highest values measurable with tests A, B and C were 15, 19.5 and 30 µg/g and 3000, 2500 and 1800 µg/g, respectively. Values above the maximal measurable concentration were not further diluted in order to obtain the exact concentration.

Statistical analysis was performed using SAS V9.2 (SAS Institute Inc, Cary, North Carolina, USA); FC concentrations for the three tests are given as boxplots. Additionally, test results were divided according to their FC concentration into five clinical meaningful categories: normal (<50 µg/g), borderline (50–100 µg/g), moderately high (100–250 µg/g), high (250–1000 µg/g) and very high (>1000 µg/g). Concordance between two tests in these five categories was assessed using Kendall’s \(\tau\) coefficient. For comparison with clinical diagnoses, categories >100 µg/g were combined. Sensitivities referring to active IBD and specificities referring to healthy controls were calculated for the recommended cut-off of 50 µg/g. The diagnostic performance of the different tests was further analysed over the whole range of possible cut-off values by calculating the area under the curve (AUC) of receiver operating characteristic (ROC) curves.

### RESULTS

#### Patients

Of 344 samples measured, 21 were excluded because there was insufficient material available for all three tests; another 19 children were younger than 2 years of age. The final cohort included 304 patients (163 males, 141 females) with a median age of 14 years (range 2.0–20.8). The numbers of children classified to the four groups 1–4 (IBD active, IBD in remission, other GI diseases and controls) were 130, 62, 45 and 67 cases, respectively. Patient characteristics are given in table 1. The diagnoses in patients from group 3 (other GI disease) included challenge-proven food allergy (n=15), infectious gastroenteritis (n=8), bowel resection (short bowel syndrome, Hirschsprung’s disease, n=12), immuno-deficiency disorders with GI involvement (n=3), autoimmune enteropathy and untreated coeliac disease (n=5), vascular malformations (n=2) and ulcerations due to non-steroidal anti-inflammatory drugs intake (n=1). In the control group the majority of patients were finally
diagnosed to have functional disorders (irritable bowel syndrome, dyspepsia and constipation, n=55), carbohydrate maldigestion/absorption (lactose, fructose, n=5), Gilbert’s syndrome (n=2), children screened because of chronic recurrent multifocal osteomyelitis (n=3) and coeliac disease on a gluten-free diet (n=2).

Concentrations of FC
Distribution of FC values is shown in figure 1 for the three different assays (A, B and C) in the four different diagnostic groups. There was a clear tendency towards higher values in test C compared to tests A and B.

Table 2 gives the percentage of children in the different diagnostic groups (1–4) having FC values <50 µg/g (normal), 50–100 µg/g (borderline) and >100 µg/g (increased) for the three different tests. Four children with treated Crohn’s disease were classified by their physicians as having active IBD, but they had normal or borderline values for FC in all three tests (figure 1). In contrast, none of the 28 newly diagnosed patients with IBD had a value below 100 µg/g. Ten of 19 excluded infants and toddlers younger than 2 years had FC values <100 µg/g, and the remaining 9 children showed elevated values >100 µg/g, again with a clear tendency towards higher values in test C (data not shown).

Calculated sensitivity values to diagnose active IBD using the recommended cut-off of 50 µg/g were 97.7%, 97.7% and 98.4% for tests A, B and C, respectively (table 2). Specificity (calculated for symptomatic controls with no identified GI disease in group 4) using the recommended cut-off was 82.1%, 85.1% and 62.7% for tests A, B and C, respectively.

ROC analyses were performed to identify patients with active IBD compared to controls of group 4 for tests A, B and C and resulted in an AUC (95% CI) of 0.979 (0.956 to 100), 0.973 (0.949 to 0.998) and 0.970 (0.942 to 0.998), respectively (no significant difference between the 3 tests).

When we compared the test results in the five categories of FC concentrations we found a better concordance between tests A and B, compared to A and C and B and C (table 3). A difference of one category was found in 102, 103 and 138 samples, while a difference of more than one categories was observed in 11, 18 and 19 cases (table 3, panel D). The calculated concordance between two tests using the Kendall’s τ coefficient was 0.835 for tests A and B, 0.782 for tests A and C and 0.765 for tests B and C.

DISCUSSION
Detection of FC is an important diagnostic tool to screen for GI inflammation or to monitor success of therapy in patients with IBD. This non-invasive method is particularly suitable in paediatric patients in order to reduce the need for endoscopies. For clinical decision-making it is important to have a high sensitivity to detect children with undiagnosed IBD who present with unspecific symptoms such as abdominal pain or chronic diarrhoea. On the other hand, a low specificity will increase the number of endoscopies, follow-up investigations in children with self-limiting disease such as GI infections, or treatment escalation in patients with IBD. Several tests for the measurement of FC are commercially available. Studies so far on head-to-head comparisons of different calprotectin assays in symptomatic children are lacking. We compared the results of two conventional ELISAs and a newly developed FEIA in a large representative cohort of 304 paediatric patients. We excluded infants and young children below 2 years of age from our analysis, because increased FC concentrations have been reported in this age group in the absence of any disease.17 18
Our results showed that all three tests for the measurement of FC showed a high sensitivity (>97%, Table 2) to detect participants with active IBD. A recent meta-analysis investigated the diagnostic accuracy of FC in paediatric suspected IBD.8 Pooled data of 715 patients (394 with IBD, 321 controls) showed a high sensitivity of 98% with a modest specificity of 68% in paediatric patients with IBD at diagnosis. There is adequate evidence that in paediatric patients with IBD FC levels reflect disease activity as measured by different activity scores such as the modified Lloyd-Still score,19 the Paediatric Crohn’s Disease Activity Index (PCDAI)20 or the Paediatric Ulcerative Colitis Index (PUCAI).10 Tibble et al21 first described that an FC value of 50 µg/g was a sensitive and specific marker of relapse in Crohn’s disease (CD) and ulcerative colitis (UC).

### Table 2

<table>
<thead>
<tr>
<th>Group 1: IBD active</th>
<th>Group 2: IBD in remission</th>
<th>Group 3: other GI disease</th>
<th>Group 4: controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test value (µg/g)</td>
<td>&lt;50</td>
<td>50–100</td>
<td>&gt;100</td>
</tr>
<tr>
<td>Test A</td>
<td>2.3</td>
<td>2.3</td>
<td>1.5</td>
</tr>
<tr>
<td>Test B</td>
<td>2.3</td>
<td>2.3</td>
<td>1.5</td>
</tr>
<tr>
<td>Test C</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Gi, gastrointestinal; IBD, inflammatory bowel disease.

There is adequate evidence that in paediatric patients with IBD FC levels reflect disease activity as measured by different activity scores such as the modified Lloyd-Still score,19 the Paediatric Crohn’s Disease Activity Index (PCDAI)20 or the Paediatric Ulcerative Colitis Index (PUCAI).10 Tibble et al21 first described that an FC value of 50 µg/g was a sensitive and specific marker of relapse in Crohn’s disease (CD) and ulcerative colitis (UC).

### Table 3 Panels A-D: Comparison of the three different tests categorizing children according to their faecal calprotectin concentrations in five categories: normal: 0–50 µg/g, borderline: 50–100 µg/g, moderately high: 100–250 µg/g, high: 250–1000 µg/g, very high: >1000 µg/g.

#### Panel A: Comparison of test A vs. B

<table>
<thead>
<tr>
<th></th>
<th>0–50</th>
<th>50–100</th>
<th>100–250</th>
<th>250–1000</th>
<th>&gt;1000</th>
<th>Test A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–50</td>
<td>96*</td>
<td>6**</td>
<td>2**</td>
<td>104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–100</td>
<td>10*</td>
<td>6</td>
<td>8*</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100–250</td>
<td>3**</td>
<td>10*</td>
<td>20</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250–1000</td>
<td>1**</td>
<td>2**</td>
<td>17*</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;1000</td>
<td>1**</td>
<td>2**</td>
<td>47*</td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test B</td>
<td>110</td>
<td>25</td>
<td>49</td>
<td>100</td>
<td>20</td>
<td>304</td>
</tr>
</tbody>
</table>

#### Panel B: Comparison of test A vs. C

<table>
<thead>
<tr>
<th></th>
<th>0–50</th>
<th>50–100</th>
<th>100–250</th>
<th>250–1000</th>
<th>&gt;1000</th>
<th>Test A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–50</td>
<td>75</td>
<td>22*</td>
<td>5**</td>
<td>37</td>
<td>1**</td>
<td>104</td>
</tr>
<tr>
<td>50–100</td>
<td>5*</td>
<td>5</td>
<td>12*</td>
<td>2**</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>100–250</td>
<td>1**</td>
<td>2*</td>
<td>12</td>
<td>21*</td>
<td>1**</td>
<td>37</td>
</tr>
<tr>
<td>250–1000</td>
<td>1**</td>
<td>2**</td>
<td>8*</td>
<td>37</td>
<td>21*</td>
<td>69</td>
</tr>
<tr>
<td>&gt;1000</td>
<td>1**</td>
<td>3**</td>
<td>12*</td>
<td>54</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Test C</td>
<td>83</td>
<td>31</td>
<td>40</td>
<td>74</td>
<td>76</td>
<td>304</td>
</tr>
</tbody>
</table>

#### Panel C: Comparison of test B vs. C

<table>
<thead>
<tr>
<th></th>
<th>0–50</th>
<th>50–100</th>
<th>100–250</th>
<th>250–1000</th>
<th>&gt;1000</th>
<th>Test B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–50</td>
<td>77</td>
<td>26*</td>
<td>5**</td>
<td>1**</td>
<td>1**</td>
<td>110</td>
</tr>
<tr>
<td>50–100</td>
<td>5*</td>
<td>2</td>
<td>12*</td>
<td>5**</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>100–250</td>
<td>2*</td>
<td>18</td>
<td>26*</td>
<td>3**</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>250–1000</td>
<td>1**</td>
<td>4*</td>
<td>37</td>
<td>58*</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>&gt;1000</td>
<td>1**</td>
<td>1**</td>
<td>5*</td>
<td>13</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Test C</td>
<td>83</td>
<td>31</td>
<td>40</td>
<td>74</td>
<td>76</td>
<td>304</td>
</tr>
</tbody>
</table>

#### Panel D: Concordance and discrepancies in numbers of children between the three different tests

<table>
<thead>
<tr>
<th></th>
<th>Concordance</th>
<th>1 category discrepancy(*)</th>
<th>&gt;1 category discrepancy(++)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A vs. Test B</td>
<td>191</td>
<td>102</td>
<td>11</td>
</tr>
<tr>
<td>Test A vs. Test C</td>
<td>183</td>
<td>103</td>
<td>18</td>
</tr>
<tr>
<td>Test B vs. Test C</td>
<td>147</td>
<td>138</td>
<td>19</td>
</tr>
</tbody>
</table>

* 1 category discrepancy
** more than 1 category discrepancy
Differences of>1 category (**) between two tests are considered as clinically meaningful.
(sensitivity of 90%, specificity of 83%). Recently Aomatsu et al.\(^2\) also showed a good correlation of FC values and disease activity in paediatric patients determined by endoscopic activity scores. Four of 130 patients classified as having active disease based on non-bloody diarrhoea or abdominal pain had normal or borderline FC concentrations in all three tests. Pain and diarrhoea are unspecific and may be related to irritable bowel syndrome, lactose intolerance or loss of bile acids, conditions that commonly coexist in patients with IBD and are unrelated to disease activity.\(^2\)

In contrast patients with IBD judged by the clinician as ‘in remission’ showed a high percentage FC values above 100 µg/g, indicating subclinical intestinal inflammation. Although patients in clinical remission showed lower FC values than patients with active disease, only 45–56% of patients in remission showed FC values below the cut-off of 50 µg/g. Normal FC values indicate mucosal healing during endoscopy\(^2\) or microscopic healing.\(^2\) Sipponen and Kolho\(^1\) reported that only 35% of children with quiescent IBD had normal or borderline FC values. The low sensitivity of global assessments for CD compared to calprotectin has been shown in previous studies, in children\(^2\) and adults.\(^2\) A better correlation between the disease score and FC concentrations has been reported for patients with UC.\(^1,2\) It can be concluded that in treated patients with IBD FC seems more sensitive for the detection of intestinal inflammation compared to clinical judgement and activity scores.

Our study confirmed that elevated FC levels are not specific for IBD, as there were increased FC levels also in group 3 (other GI disorders). Elevated FC concentrations have been shown in adults with colorectal cancer or adenomatous polyps\(^2\) and in children with juvenile polyne\(^2\) as well as in some but not all children with untreated coeliac disease and cow’s milk or multiple food allergy\(^3\) or eosinophilic and lymphocytic colitis.\(^2\)

With respect to specificity two of the three tests had <10% of the control children with no obvious GI disease in the pathological range (>100 µg/g) and another 10% of the children with borderline values (50–100 µg/g). Elevated FC levels have been reported in 9 of 126 children (7%) with functional GI disorders\(^4\) based on the Rome III criteria.\(^5\) Since endoscopies had not been performed in most children in the control group, we cannot exclude that some of these children with elevated values had minor intestinal inflammation. Test C had more borderline and increased values in the control children which could lead to a higher proportion of endoscopies in the clinical setting. However, it cannot be excluded that the higher values in test C may be the result of a better recovery rate of calprotectin.

Stool samples were extracted separately for tests A, B and C from homogenised stool. Commercially available extraction devices used in FC tests, including the Roche kit used in the current study, may recover different rates from different tests not feasible. It is important to inform the caring physician and the parents/patient that the absolute values of different tests have to be interpreted with caution depending on the assay used. For follow-up measurements and monitoring of the efficacy of a therapeutic intervention it is highly advisable to use the test of the same manufacturer.

To investigate the concordance between the three tests we categorised the results in five categories with values of 2, 5 and 20 times the upper limit of normal. The cut-off for normal values for most tests is considered to be 50 µg/g, although some studies used 100 µg/g as upper limit of normal to calculate sensitivity and specificity.\(^4,9,38\) In agreement with other authors we characterised values >1000 µg/g as very high levels.\(^10,12\) and made another cut at 250 µg/g. We considered a difference of more than one category between two tests as clinically meaningful. There were fewer discrepancies between tests A and B, compared to A and C or B and C (table 3). Highly discrepant results cannot be explained by obtaining the faecal samples from different locations of the faeces, because a homogenised stool sample was the basis for all three tests. We did not perform duplicate measurements which eventually could have improved the concordance.

Since the cut-off values for the three tests are identical and the values are given as concentrations (µg/g), clinicians and patients may assume that the absolute values are comparable and reliable. In clinical practice a physician does not even know which test is performed in the laboratory and what the maximum detection value of the given test is. Our study showed that the absolute value for an individual sample markedly differed between the three tests, particularly for values in the pathological range. This makes a comparison of values from different tests not feasible. It is important to inform the caring physician and the parents/patient that the absolute values of different tests have to be interpreted with caution and decreasing value may not indicate improvement and vice versa.

In conclusion, measurement of FC in paediatric patients with unspecific symptoms is very helpful in order to avoid invasive procedure. Normal FC concentrations make the diagnosis of IBD very unlikely, regardless of which of the three tests have been used. However, with respect to specificity and the absolute values in the borderline and pathological range, we found major differences between the tests under investigation. Therefore, clinical interpretation of assay results should be made with caution depending on the assay used. For follow-up measurements and monitoring of the efficacy of a therapeutic intervention it is highly advisable to use the test of the same manufacturer.

**Contributors** SK designed the study, interpreted the data and wrote the manuscript with the help of CP. DN performed the statistical analysis and contributed to the interpretation of the data and writing of the manuscript. FF and AS collected clinical data of the patients, interpreted the data and critically reviewed the manuscript.

**Funding** SK received a speaker fee and financial support to perform this study from Phadia GmbH, part of Thermo Fisher Scientific, Freiburg, Germany.

**Competing interests** None.

**Ethics approval** The ethical committee of the clinical medical centre of the Ludwig Maximilians University approved the study protocol with the measurement of the anonymised faecal samples (project No. 402-11).

**Provenance and peer review** Not commissioned; externally peer reviewed.
REFERENCES


