Diagnostic strategies in children with chronic gastrointestinal symptoms in primary care
Holtman, Geeske

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:
2016

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.

Download date: 22-12-2018
CHAPTER 4

CHALLENGES IN DIAGNOSTIC ACCURACY STUDIES IN PRIMARY CARE: THE FAECAL CALPROTECTIN EXAMPLE

Gea A Holtman
Yvonne Lisman-van Leeuwen
Boudewijn J Kollen
Johanna C Escher
Angelika Kindermann
Patrick F van Rheenen
Marjolein Y Berger

ABSTRACT

BACKGROUND
Low disease prevalence and lack of uniform reference standards in primary care induce methodological challenges for investigating the diagnostic accuracy of a test. We present a study design that copes with these methodological challenges and discuss the methodological implications of our choices, using a quality assessment tool for diagnostic accuracy studies (QUADAS-2).

DESIGN
The study investigates the diagnostic value of faecal calprotectin for detecting inflammatory bowel disease in children presenting with chronic gastrointestinal symptoms in primary care. It is a prospective cohort study including two cohorts of children: one cohort will be recruited in primary care and the other in secondary/tertiary care. Test results of faecal calprotectin will be compared to one of the two reference standards for inflammatory bowel disease: endoscopy with histopathological examination of mucosal biopsies or assessment of clinical symptoms at 1-year follow-up.

DISCUSSION
According to QUADAS-2 the use of two reference standards and the recruitment of patients in two populations may cause differential verification bias and spectrum bias, respectively. The clinical relevance of this potential bias and methods to adjust for this are presented. This study illustrates the importance of awareness of the different kinds of bias that result from choices in the design phase of a diagnostic study in a low prevalence setting. This approach is exemplary for other diagnostic research in primary care.

INTRODUCTION
In primary care, patients often present with non-specific symptoms and the incidence of severe illnesses is low. Differentiating between innocent symptoms and a rare, but serious organic disease is a diagnostic dilemma for the general practitioner (GP). Unnecessary referrals and diagnostic testing need to be balanced against the risk of missing a diagnosis and introduction of an unacceptable long diagnostic delay. In primary care, both the GP and the patient would greatly benefit from simple, non-invasive and specific triage tests. However, many of these tests are not validated in primary care.

An example of such a diagnostic dilemma are children presenting with chronic or recurrent gastrointestinal symptoms. This clinical picture is common, but few children will actually have inflammatory bowel disease (IBD), which includes Crohn’s disease and ulcerative colitis. The incidence of non-specific abdominal pain in Dutch children is 2500/100,000 per year, while the incidence of IBD is 5.2/100,000 per year. Clinical symptoms in children with IBD are often non-specific and show substantial overlap with functional gastrointestinal disorders. In European secondary and tertiary care facilities the measurement of calprotectin in stool is used as an effective triage method for endoscopy, which is the reference standard for the diagnosis of IBD. Calprotectin is a marker of inflammation that can be measured by using a simple non-invasive test, but has never been evaluated in children in a primary care setting. The different patient spectrum in primary care has consequences for the pre-test probability and test characteristics. Before calprotectin testing can be recommended to distinguish functional from organic gastrointestinal disorders at the primary care level, information is required on the positive predictive value of faecal calprotectin at the primary care level.

The preferred design to evaluate the diagnostic value of faecal calprotectin in children with chronic gastrointestinal symptoms would be a cross-sectional study. Such a design has two methodological challenges. Firstly, the design of a diagnostic study for rare diseases requires a large population in order to identify a sufficient number of children with IBD; the financial and logistic exercise involved makes such a study infeasible. Secondly, the preferred reference standard to detect IBD is endoscopy, but it is unethical to perform this invasive test in children with a low likelihood of organic gastrointestinal disease.

Here we present an example of a design that copes with these methodological challenges. The methodological implications of applied design choices are examined using an evidence-based quality assessment tool for diagnostic accuracy studies (QUADAS-2).

DESIGN
DESIGN AND SETTING
The DOK (Darm Onderzoek bij Kinderen; translated as bowel research in children) study is a prospective cohort study with a follow-up period of one year, also known as a delayed type cross-sectional study. The study consists of two prospective cohorts. We will recruit a primary care cohort of children presenting consecutively in primary care in the northern part of the Netherlands (primary care cohort). A second cohort consists of children that will
be referred to secondary and tertiary care facilities across the Netherlands (referred cohort). The index test is faecal calprotectin and the two reference standards for IBD are endoscopy with histopathological examination of mucosal biopsies, or assessment of clinical symptoms at 1-year follow-up (Figure 1). The DOK study was approved by the Medical Ethics Review Committee of the University Medical Center Groningen. Written informed consent will be obtained from the parents and from the child if aged ≥12 years. Inclusion started in June 2011.

**STUDY POPULATION**

Children aged 4-18 years presenting with chronic diarrhoea (≥2 weeks diarrhoea or ≥2 episodes of diarrhoea in the past 6 months) or recurrent abdominal pain (≥2 episodes of abdominal pain in the past 6 months) will be eligible for participation. Diarrhoea was defined as moderately to watery loose stools matching score 5, 6 or 7 of the Bristol Stool Form Scale. One episode is defined as 3 days or more.

Exclusion criteria are: a previously established diagnosis of chronic organic gastrointestinal disease; a complete evaluation in the past 6 months for abdominal symptoms including endoscopy; chronic use of antibiotics, non-steroidal anti-inflammatory drugs (NSAIDs) or oral corticosteroids (defined as daily use during ≥3 months/year); faecal calprotectin test in the past 6 months; and difficulty in understanding questionnaires. The number of patients not participating due to the exclusion criteria or refusal are anonymously recorded, including the patient characteristics and, if available, the reason for non-participation.

**MEASUREMENTS**

**Physical examination**

The GP or paediatric gastroenterologist performs a structured physical examination and assesses extra-intestinal symptoms and peri-anal lesions according to the Dutch diagnostic guideline. The participating GPs receive training on structured physical examination of children with symptoms suggestive of IBD.

**Questionnaire on Paediatric Gastrointestinal Symptoms**

The Dutch version of the Questionnaire on Paediatric Gastrointestinal Symptoms ROME III (QPGS-RIII) is completed, by the patient or a parent at baseline and at 12 months follow-up. The QPGS-RIII consists of two reports, a parent report for children aged 4-18 years and a self-report for children aged ≥10 years. The questionnaire has been translated into Dutch. The English version of QPGS has good content validity and test-retest reliability.

**Blood and faecal tests**

In the blood sample haemoglobin, erythrocyte sedimentation rate, C-reactive protein, platelet count and serology tests for celiac disease (IgA tissue transglutaminase antibodies) are measured. Faeces is tested for colon pathogens (Salmonella enterica, Campylobacter jejuni, Shigella spp/EIEC, STEC) and parasites (Giardia lamblia, Cryptosporidium spp, Dientamoeba fragilis, Entamoeba histolytica) with the real-time multiplex PCRs. Blood and faeces tests are performed at local certified laboratories. If a child is using NSAIDs, antibiotics or oral corticosteroids for short-term use (<3 months), the collection and testing of faeces is postponed until the end of that treatment.

---

Figure 1. Flow chart of the DOK study.

The GP or paediatric gastroenterologist selects eligible children. At baseline inclusion, exclusion criteria and red flags are determined. The parents and child ≥10 years complete two questionnaires, i.e. a Questionnaire on Paediatric Gastrointestinal Symptoms (QPGS) and a symptoms questionnaire, in addition faeces (parasites and colon pathogens) are obtained. Children meeting ≥1 red flags are evaluated for eligibility for endoscopy by a paediatric gastroenterologist. Children without red flags receive a 1-year follow-up. The arrows indicate that the GP can refer a child during follow-up for endoscopic evaluation and the children who are not eligible for endoscopy receive a follow-up. After 1 year, information about diagnosis and clinical symptoms is collected based on the two above-mentioned questionnaires.
Faecal calprotectin

After baseline assessments the patients send the faeces sample by pre-stamped return envelope to the laboratory where the samples are stored at -80°C. At the end of the data collection period the samples are defrosted before analysis. Faecal calprotectin is measured by means of a commercially available quantitative enzyme-linked immunosorbent assay (ELISA). In accordance with the manufacturer’s guidelines, values above 50 μg/g faeces are regarded as positive.

Red flags

In all children red flags of IBD will be searched for using a structured evaluation form (Table 1). Children who fulfil the inclusion criteria and have ≥1 red flags are referred to a paediatric gastroenterologist who will decide whether the child requires endoscopic examination. This decision will be based on the medical history, physical examination and blood testing. Children without red flags, or those who are not eligible for endoscopy will be followed for one year.

### Table 1. Definitions of red flag symptoms for inflammatory bowel disease.

<table>
<thead>
<tr>
<th>Red flag symptom</th>
<th>Measurement</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth failure</td>
<td>History and physical examination</td>
<td>Target height range &gt; -1 SDS</td>
</tr>
<tr>
<td>Involuntary weight loss</td>
<td>History</td>
<td>Involuntary decrease in weight of &gt; 1 kg</td>
</tr>
<tr>
<td>Rectal blood loss</td>
<td>History</td>
<td>Rectal blood loss with defecation without constipation according to ROME III criteria</td>
</tr>
<tr>
<td>Positive family history of IBD</td>
<td>History</td>
<td>Affected first-degree relative(s)</td>
</tr>
<tr>
<td>Extra-intestinal symptoms</td>
<td>Physical examination</td>
<td>Eyes (episcleritis, scleritis, uveitis), skin (erythema nodosum, pyoderma gangrenosum, psoriasis), mouth ulcers, finger clubbing, arthritis</td>
</tr>
<tr>
<td>Peri-anal lesions</td>
<td>Physical examination</td>
<td>Skin tags, haemorrhoids, fissures, fistulas, or abscess</td>
</tr>
<tr>
<td>Anaemia (Hb)</td>
<td>Local laboratory</td>
<td>4-12 years &lt; 7.1 mmol/l, boy 12-18 years &lt; 8.1 mmol/l, girl 12-18 years &lt; 7.4 mmol/l</td>
</tr>
<tr>
<td>CRP</td>
<td>Local laboratory</td>
<td>&gt; 10 mg/l</td>
</tr>
<tr>
<td>ESR</td>
<td>Local laboratory</td>
<td>&gt; 20 mm/h</td>
</tr>
<tr>
<td>Platelets</td>
<td>Local laboratory</td>
<td>&gt; 450 x 10^9/l</td>
</tr>
</tbody>
</table>

SDS = standard deviation score; Hb = haemoglobin; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate.

### Endoscopy

Endoscopy is performed under full anaesthesia or deep sedation by an experienced paediatric gastroenterologist and entails esophagogastroduodenoscopy and ileocolonoscopy. Two biopsies of each intestinal segment are taken. The histopathological examination will be performed by an experienced gastrointestinal histopathologist. IBD is classified according to the Paris classification.

### Follow-up

Follow-up is done using a symptom questionnaire that was developed for the study in cooperation with paediatric gastroenterologists and GPs. This questionnaire will be completed by the parent or child (if aged ≥ 210 years) at 3, 6, 9 and 12 months follow-up. The GP will perform a structured physical examination to assess red flags in children with clinical symptoms at 12 months. Those with ≥1 red flags at 12 months will be referred to a paediatric gastroenterologist to determine a diagnosis.

### Blinding

The paediatric gastroenterologists, pathologists, GPs and researchers will be blinded to the outcome of the faecal calprotectin test. The laboratory technician will be blinded for the clinical characteristics of the child and the result of endoscopy.

### OUTCOME

IBD is confirmed when the endoscopic picture and the histopathological picture match. Absence of IBD is defined as a negative endoscopic and histopathological examination, or when there was no indication to perform endoscopy at all during the 12 months follow-up. Besides, all children without red flags at 12 months follow-up are considered not to have IBD.

### SAMPLE SIZE

Based on available literature we expect to find a specificity of 93% in the primary care cohort. To estimate the specificity and a 95% confidence interval (CI) spanning 5%, we assume a maximum IBD incidence of 5 per 100 children with gastrointestinal complaints and a loss to follow-up of 10%, we will then need a sample size of 118 children in the primary care cohort. In a worst case scenario with a specificity of 75%, a sample size of 118 children will widen the 95% CI to 8%.

Sensitivity was calculated in children with red flags (primary care cohort and referred cohort). Based on an expected sensitivity of 95% we need to include 73 children with IBD in order to estimate the sensitivity and a 95% CI spanning 5%. With a IBD prevalence of 8% and a loss to follow-up of 10% we need to include 100 children with red flags. The prevalence of IBD is difficult to estimate; with a prevalence of 20% the spanning of the 95% CI of the sensitivity will widen to 10%.

### STATISTICAL ANALYSES

Specificity of faecal calprotectin for IBD in primary care will be calculated by dividing the number of negative faecal calprotectin tests by the total number of children without IBD.
included in the primary care cohort. Sensitivity will be calculated by dividing the number of positive faecal calprotectin tests by the total number of children with IBD in children with red flags of both the primary care and referred cohort. The estimates of specificity and sensitivity will be reported as percentages with 95% CIs.

**DISCUSSION**

**ASSESSING THE RISK OF BIAS**

To address the risk of bias in our study design and the applicability of the results we applied the QUADAS-2 checklist\(^{11}\) that includes four domains: patient selection, index test, reference standard and flow and timing (flow of patients through the study and timing of the index test and reference standard). Each domain was scored as low or high risk of bias, based on the answers to the signalling questions. If all answers concerning a domain are “yes”, the risk of bias can be judged as low. If any signalling question is answered “no” the risk of bias can be judged as high. The first two domains were scored as low or high concerns regarding applicability. Two items were excluded because one item assessed heterogeneity between studies, which is only applicable in systematic reviews. The second item asked whether all patients are included in the analysis, which can only be assessed after completion of the study. The results of the QUADAS-2 assessment are shown in Table 2.

**RISK OF BIAS**

*Problems with the reference standard*

A perfect reference standard in a diagnostic accuracy study is said to fulfil three criteria: 1) The reference standard provides error-free classification of all subjects. 2) The same reference standard is used to verify all index results. 3) The index test and reference standard can be performed within a short interval to avoid changes in target condition status.\(^{31}\)

Risk of bias in the DOK study is related to the choice of the reference test, which is not the same for all included patients (differential verification bias).\(^{29}\) In addition, follow-up is not considered a reference standard for IBD in daily practice. This choice may lead to missed diagnoses and will influence the estimates of sensitivity and specificity. We chose a differential verification design, because it is unethical to perform endoscopy in children with a low likelihood of organic gastrointestinal disease. Therefore, children who have a low IBD risk receive a follow-up of one year, which is considered to be a suitable period.\(^{13}\) On the opposite side, it might be possible that even more children will be identified because, using a 1-year follow-up, children with initially mild IBD can be detected when they have an aggravation of symptoms later in time. These children could have been missed when endoscopy was performed at initial presentation. Children in whom endoscopy was not indicated during the 1-year follow-up (either because they no longer have symptoms or because their red flags are not suggestive for IBD) are considered not to have IBD. The probability that we will miss a child with IBD is considered to be extremely low.\(^{13}\) Adjustment for differential verification bias will be made, if possible, using a Bayesian approach.\(^{29,30}\)

The patient flow of the DOK study could introduce bias. A delay of one month between

---

### Table 2. Quality assessment of the DOK study design (QUADAS-2).

<table>
<thead>
<tr>
<th>Signalling questions</th>
<th>Answer</th>
<th>Risk of Bias/ Applicability</th>
<th>Planned adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1: Patient selection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is a consecutive sample of patients enrolled?</td>
<td>Yes</td>
<td>Low Risk</td>
<td></td>
</tr>
<tr>
<td>Is a case-control design avoided?</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the study avoid inappropriate exclusions?</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applicability</td>
<td></td>
<td>High Concern</td>
<td>Magnitude will be evaluated</td>
</tr>
<tr>
<td>Are there concerns that the included patients and setting do not match the topic of our study (patients had symptoms suggestive of inflammatory bowel disease in primary care)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Domain 2: Index test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of bias</td>
<td></td>
<td>Low Risk</td>
<td></td>
</tr>
<tr>
<td>Are the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If a threshold was used, is it pre-specified?</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applicability</td>
<td></td>
<td>Low Concern</td>
<td></td>
</tr>
<tr>
<td>Are there concerns that the index test, its conduct, or interpretation differ from the topic of our study (faecal calprotectin was measured with ELISA)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Domain 3: Reference standard</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of bias</td>
<td></td>
<td>High Risk</td>
<td>Probably not clinically relevant</td>
</tr>
<tr>
<td>Is the reference standard likely to correctly classify the target condition?</td>
<td>No</td>
<td></td>
<td>Adjustment in analysis (^{29,30})</td>
</tr>
<tr>
<td>Are the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Domain 4: Flow and timing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of bias</td>
<td></td>
<td>High Risk</td>
<td>Represents care as usual (^{29,30})</td>
</tr>
<tr>
<td>Is there an appropriate interval between index test and reference standard?</td>
<td>No</td>
<td></td>
<td>Repeated measurement index test before endoscopy</td>
</tr>
<tr>
<td>Do all patients receive a reference standard?</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do all patients receive the same reference standard?</td>
<td>No</td>
<td></td>
<td>Adjustment in analysis (^{29,30})</td>
</tr>
</tbody>
</table>
stool sample collection and reference standard is considered to be an appropriate time period. In children of the referred cohort the interval between faecal sampling and endoscopy will generally be less than one month. For referred children in the primary care cohort this interval is likely to exceed the period of one month. To investigate whether the concentration of faecal calprotectin accurately measures the same outcome as endoscopy, faeces will be collected again shortly before endoscopy. In children not referred to a secondary or tertiary care facility the period between faecal sampling and reference test will be one year. During this period the calprotectin concentration may change and, therefore, the initial test result will no longer be related to the outcome of endoscopy. This will underestimate sensitivity and specificity. Here we adopt a pragmatic approach. We want to establish whether faecal calprotectin can serve as a triage test in children who are presenting for the first time to their GP. A negative faecal calprotectin value at the start of the study, and a positive endoscopic result at the end of the study, should be considered as a false-negative test result.

APPLICABILITY OF STUDY RESULTS

Problems with the patient selection

Test characteristics should be evaluated in a clinically relevant population.\(^3\) In the DOK study the patients with symptoms suggestive of IBD will be recruited in both primary and secondary/tertiary care. Spectrum bias is to be expected as our patient cohorts will have different characteristics.\(^3\) To reduce the risk of spectrum bias one should ideally only include children who initially presented at the primary care level. The low prior probability in this setting makes such a study design infeasible with considerable financial and logistic problems.\(^3\) We decided to use a pragmatic design, based on the following assumptions: in case of a very low prior probability of IBD, a GP wants to avoid unnecessary referrals. The false-positive rate thus needs to be low. Therefore, we will evaluate specificity of faecal calprotectin in children presenting in primary care. In children with red flags, a GP wants to rule out IBD and minimize false-negative results. Sensitivity will thus be evaluated in children referred to secondary or tertiary care (children with red flags in primary care cohort and referred cohort).

We assume that this sensitivity is a representative estimate for sensitivity measured in children with red flags in primary care. This implies two additional assumptions: 1) in both cohorts the ratio IBD/non-IBD in children with red flags will be comparable (which we will test); 2) children with red flags of both cohorts are comparable (which we will test by comparing the clinical characteristics). In case the children from the referred cohort are more severely ill, sensitivity will be overestimated. Heterogeneity can then be assessed by subgroup analyses of the test performance.

CONCLUSION

Low disease prevalence and lack of uniformity in reference standard in primary care creates methodological challenges in primary care level diagnostic accuracy studies. We presented a pragmatic design in which the magnitude of potential bias will be assessed and controlled. Awareness of the potential biases and its implications allows to discuss possible solutions and to overcome such bias. The validity of diagnostic research at the primary care level may be considerably improved with the proposed design.

REFERENCES


