CHAPTER 8

SUMMARY AND GENERAL DISCUSSION
The main objective of this thesis was to study the diagnostic strategies for inflammatory bowel disease (IBD) in children with chronic gastrointestinal symptoms, focusing on the value of testing with faecal calprotectin in primary care. This chapter summarises the main results, discusses clinical implications for primary care and methodological implications for diagnostic studies in primary care, and presents suggestions for further research.

**SUMMARY OF MAIN FINDINGS**

In chapter 2, a systematic review and meta-analysis is presented with an overview of the diagnostic accuracy of the signs, symptoms and diagnostic tests for IBD available in all healthcare settings. All 19 included studies were performed in secondary or tertiary care, which resulted in a high disease prevalence, ranging from 19% to 82%. It was shown, that for those children in whom a paediatrician considered endoscopy, symptoms were not sufficiently accurate to distinguish complaints caused by IBD from those related to other causes. Therefore, simple tests that produce minimal adverse effects were recognized as being essential when triaging for endoscopy. With a pooled negative likelihood ratio of 0.01 (0.0–0.1), faecal calprotectin had demonstrable ability in decreasing the probability of IBD. In addition, c-reactive protein and albumin had pooled positive likelihood ratios sufficiently high to indicate an increase of the probability of IBD that may be of clinical importance, with positive likelihood ratios of 5.1 (2.8–9.4) and 8.3 (3.7–18.7), respectively. Thus, we showed that a normal faecal calprotectin could safely rule out IBD, while a positive c-reactive protein or albumin could rule in IBD.

In chapter 3 we investigated the added diagnostic value of individual blood markers and faecal calprotectin beyond the value of symptoms alone. The results in this section were based on individual patient data in eight studies of referred children. Faecal calprotectin added the most diagnostic value to symptoms when compared against commonly used blood markers. When faecal calprotectin was added to symptoms, the proportion of the total number of patients assigned to the intermediate risk group, with an IBD probability ranging from 35% to 60%, decreased from 55% to 6%. Thus, adding faecal calprotectin to the diagnostic work-up of children with symptoms suggestive of IBD considerably decreased challenging diagnosis.

An important finding of the studies in previous chapters was that none was specifically performed among children in primary care. This is important because the diagnostic accuracies of tests performed in specialist care are not generalizable to primary care. In fact, differences in the patient spectrum and disease severity can affect pre-test probabilities and diagnostic test characteristics, including their sensitivity and specificity values. This highlighted the need for further studies among children in primary care.

In chapter 4 we described the methodological challenges of performing a diagnostic study in primary care. The DOK study was designed to evaluate the diagnostic accuracy of faecal calprotectin for IBD in children presenting with chronic gastrointestinal symptoms in primary care. Two prospective cohorts of children with chronic gastrointestinal symptoms were included: children presenting in primary care (primary care cohort) and children referred to specialist care (referred cohort). Faecal calprotectin was measured at inclusion and compared by endoscopic assessment or diagnosis at one-year clinical follow-up, which were used as reference standards for IBD diagnosis. Physicians were blinded to the faecal calprotectin results, and calprotectin values above 50 μg/g faeces were considered positive. Two important methodological
challenges were noted: not only does IBD have a low prevalence in primary care, but it is also diagnosed by invasive endoscopy under specialist care. A pragmatic design was therefore used.

Given the low prevalence of paediatric IBD in children with chronic gastrointestinal symptoms presenting in primary care, an ideal cohort study would require long term investigation in a very large population to identify a sufficient number of children with IBD. Given a prevalence of IBD of approximately 5% among children with chronic gastrointestinal symptoms presenting in primary care, we calculated that 7500 symptomatic children would be needed to estimate the sensitivity of faecal calprotectin with adequate precision. In such a case, the financial and logistic difficulties are prohibitive, making such a study infeasible. Therefore, we included a cohort of children in primary care and a cohort of referred children. The prevalence of IBD is known to be around 20% among children referred to specialist care, and this higher prevalence allowed the number of children needed to estimate the sensitivity to be reduced to 90. However, because a higher prevalence of IBD was anticipated (approximately 60%), the precision of the 95% confidence interval for the sensitivity decreased from 5% to 10%.

Children with a low likelihood of IBD were followed for 12 months because it was considered unethical to expose this group to invasive endoscopy. This choice could introduce a differential verification bias, where the two reference standards may have produced differences in classifications of IBD diagnosis that could influence the results for sensitivity and specificity. The verification pattern and the length of the follow-up are important aspects for deciding whether the verification introduced clinical relevant bias. Although the verification pattern was based on the clinical judgements of paediatric gastroenterologist, and we may assume that these assessments were at least somewhat subjective, our results showed that children who underwent endoscopy at baseline were at a higher risk for IBD than children who received follow-up. A 12-month follow-up period was considered appropriate to observe the appearance of new alarm symptoms suggestive for IBD because it is very rare for untreated cases to remain symptom-free beyond that time. Although follow-up was the best achievable option given the reality of clinical care, there remains a small chance that we misclassified cases. Therefore, in chapter 5, we provide insight into pattern and extent of differential verification by comparing the contingency tables by reference standard.

In chapter 5 we reported the test characteristics of faecal calprotectin for IBD in children presenting with chronic gastrointestinal symptoms in primary care. In the primary care cohort, 24 children had more than one red flag and were referred to a paediatric gastroenterologist. None of the children in the primary care cohort was diagnosed with IBD. Faecal calprotectin was elevated (>50 μg/g faeces) in 15 of 114 children, yielding an overall specificity of 0.87 (0.80–0.92). The referred cohort consisted of 24 children with red flags presenting in primary care, plus 66 children referred to specialist care by a general practitioner or general paediatrician. In these 90 referred children, IBD was confirmed by endoscopy in 17 cases (19%). All children with IBD had elevated faecal calprotectin levels (range 53–2823 μg/g faeces), giving a sensitivity of 0.99 (0.81–1.00) for faecal calprotectin. We concluded that a positive faecal calprotectin result in children with chronic gastrointestinal symptoms presenting in primary care was unlikely to indicate IBD by itself. However, in children in whom a general practitioner considers referral for further diagnostic assessment of IBD, a normal faecal calprotectin result could safely rule out IBD and reduce the number of referrals.

In chapter 6 we evaluated the added diagnostic value of c-reactive protein and faecal calprotectin beyond evaluation of alarm symptoms, to determine the optimal diagnostic test strategy before referral to specialist care among children with symptoms suggestive of IBD. We compared three test strategies: 1) alarm symptoms alone; 2) alarm symptoms plus c-reactive protein; and 3) alarm symptoms plus faecal calprotectin. Of the 90 included children, 17 (19%) had IBD. Adding the faecal calprotectin result to the presence of alarm symptoms increased the area under the curve significantly from 0.80 (0.69–0.90) to 0.97 (0.93–1.00) (P = 0.002). However, adding c-reactive protein to alarm symptoms did not increase the area under the curve significantly. Decision curves confirmed these patterns and showed that the addition of faecal calprotectin to the clinical evaluation of alarm symptoms provided the diagnostic test strategy with the highest net benefit at all reasonable threshold probabilities. We concluded that the optimal strategy, for further stratifying children who have already been identified as at risk for IBD by their general practitioner, was to evaluate the presence of alarm symptoms and faecal calprotectin together.

In chapter 7 we explored the test characteristics of a point-of-care test measuring both faecal calprotectin and lactoferrin for IBD in children presenting in primary care with chronic gastrointestinal symptoms. In the primary care cohort, calprotectin and lactoferrin had specificities of 0.93 (0.89–0.98) and 0.98 (0.93–0.99), respectively. In the referred cohort, the specificities of calprotectin and lactoferrin were both 0.94 (0.72–0.99). Surprisingly, use of the tests together provided no substantial added value over the use of either in isolation. The use of point-of-care calprotectin testing could reduce the referral rate by 76%, while the use of the lactoferrin test did so by 81%, but at the expense of missing one child with IBD (6%). Thus, point-of-care testing has the potential to reduce the need for referral for further diagnostic work-up in specialist care, with a low risk of missing children with IBD.

**CLINICAL IMPLICATIONS FOR PRIMARY CARE**

It is a diagnostic challenge in primary care to differentiate between functional gastrointestinal disorders, in which diagnostic testing should be minimized, and organic pathology that should not be missed. IBD and coeliac disease are the most important diagnoses a general practitioner needs to eliminate before considering a functional disorder. At the start of this thesis, several guidelines that were designed to assist the general practitioner in this diagnostic dilemma were reviewed. These ranged from general guidance for the management of abdominal pain in children to more specialized guidance for the diagnosis and treatment of children with IBD. Although small discrepancies exist between these guidelines, they consistently suggest, that when evaluating the risk of IBD in children with chronic gastrointestinal symptoms, the general practitioner should be guided by the presence or absence of alarm symptoms and the results of blood tests.

The main results of this thesis, that faecal calprotectin had good test characteristics in primary care and added value over alarm symptoms alone, suggest that faecal calprotectin can be used as a triage test to reduce unnecessary referrals among symptomatic children. However, this is contingent on the faecal calprotectin test not being used as a replacement for a thorough history and physical examination. We also found that c-reactive protein had no additional value over alarm symptoms, and that faecal calprotectin showed the
highest discriminative power of all laboratory markers, indicating that faecal calprotectin is presently the best triage test for IBD in primary care. In theory, it is possible that the faecal calprotectin test could replace the need for invasive blood tests in children with suspected IBD. When coeliac disease is suspected, antibody tissue transglutaminase (anti-TTG) and immunoglobulin A tests can also be performed. In this study all anti-TTG tests were normal, including that in the case of a child with coeliac disease and immunoglobulin A deficiency.

Given the very low probability of IBD and the considerable risk of false positives, faecal calprotectin cannot be recommended for use in the screening of all children with chronic gastrointestinal symptoms. The number of false positives (5%) in our study was comparable to the previously reported number of false positives (50%) in children with functional gastrointestinal disorders presenting to specialist care. Referral of children with functional gastrointestinal disorders should be avoided because this approach does not lead to the intended reassurance for either child or parent, and it may in fact lead to a chronic course of abdominal pain. In addition, the decision curve analysis showed that testing for faecal calprotectin in children with a threshold probability below 5% barely improved the net benefit compared to the net benefit of referring all children. This highlights the importance of testing only in the context of specific clinical features, and of adopting a sequential strategy to ensure appropriate referral for further diagnostic assessment. Testing with faecal calprotectin can only be recommended in children in whom the general practitioner has identified alarm symptoms; that is, those children in whom they consider referral for further assessment of possible IBD.

To maintain the high negative predictive value, the diagnostic threshold might need to be higher in primary care. Although, in chapter 5, we found that an increase in the threshold from 50 μg/g to 250 μg/g faeces reduced the referral rate by 14% (with a drop from 32% to 18% referred), this threshold also led to false-negative results and missed cases with IBD (with an increase from 0% to 18% cases missed). Therefore, in children for whom the general practitioner considers referral for diagnostic evaluation for IBD, they can reassure those with a calprotectin value below 50 μg/g and refer to specialist care those with a value above 250 μg/g faeces. In children with a value between 50 μg/g and 250 μg/g faeces, the calprotectin value alone gives no direct justification for referral, but this does not preclude the chance that the child may have or may develop IBD or another organic disease. A safety net policy could be applied in this situation, with clear communication given to children and their parents about the situation, including advice on the need to re-contact the general practitioner if specific signs or symptoms develop and reassurance that children with persistent symptoms will be re-tested. Children whose symptoms persist and whose calprotectin values remain high can then be referred during the course of follow-up in primary care. A disadvantage of this strategy is, that while unnecessary referral may be prevented in children with functional gastrointestinal disorders, diagnosis may be delayed in children with IBD. Figure 1 shows a hypothetical diagnostic strategy for IBD in children presenting with chronic gastrointestinal symptoms in primary care.

The faecal calprotectin test is limited by the fact that levels increase after the use of non-steroidal anti-inflammatory drugs and antibiotics. In our study, children using this medication for more than 3 months were excluded, and children using the medication for <3 months had stool collection postponed until the end of treatment. Three children without IBD used antibiotics (n = 2) or NSAIDs (n = 1) for 7 days at baseline and had their stool samples collected at the end of the treatment. Two of these children displayed elevated calprotectin values, which resulted in false-positive results. More studies are required to evaluate the optimal interval needed between the end of treatment and testing with faecal calprotectin.

In a correspondence letter, Poullis et al. suggested that proton-pump inhibitors also affect the faecal calprotectin level. Since we did not include any children using proton-pump inhibitors, we could not evaluate this hypothesis. In total, 43 children reported chronic medication use, typically the oral contraceptive pill or asthma medication (salmeterol/fluticasone), but we did not evaluate whether these affected faecal calprotectin levels because none of the recorded medications was associated with false-positive results in previous studies. Moreover, although the calprotectin level is known to be higher in healthy children younger than four years and in children with bacterial gastroenteritis, only one of the eight children with gastroenteritis in the DOK study had an elevated calprotectin value. Nevertheless, it is important that the general practitioner be aware of these factors that
influence calprotectin levels.

There is presently insufficient evidence on either the diagnostic value or the cost-effectiveness of point-of-care testing with faecal calprotectin to recommend it over the enzyme-linked immunosorbent assay (ELISA) test. However, the point-of-care test showed comparable test characteristics as the ELISA test in the same patient population, though more studies are needed to confirm this result. The characteristics of the ELISA test are reported in chapter 5, while those of the point-of-care test are reported in chapter 6. An important benefit of the point-of-care test is that stool samples can be brought from home and tested by the general practitioner in clinic, with results available within 15 minutes. This can expedite the diagnostic trajectory for children with chronic gastrointestinal symptoms. However, the labour and equipment costs need to be compared between the point-of-care test and the ELISA test. A pitfall of point-of-care testing is that general practitioner may come to rely on this simple test rather than evaluating alarm symptoms, which might lead to the unintended consequence of an increased number of false-positive results and, therefore, an increased number of unnecessary referrals.

METHODOLOGICAL IMPLICATIONS FOR DIAGNOSTIC STUDIES IN PRIMARY CARE

Patient selection

The characteristics of a test are evaluated by comparing the results of the test to be evaluated with the results of a reference standard in the same patients. Ideally, this should involve a cohort of consecutive patients with clinical suspicion of having the condition of interest, and in which both the test under evaluation and the reference standard are performed in all patients. Studies among patients with low prior disease probabilities have important methodological challenges, specifically that the inclusion of a proper amount of patients with the disease requires the inclusion of a large number of symptomatic patients. It is important to identify methodological solutions that reduce the sample size and costs of the investigation, deal with ethical challenges, yet yield unbiased estimates of test accuracy.

So, what are the advantages and disadvantages of the pragmatic design used in the DOK study of a primary care and referred cohort? An obvious advantage is that we could compare the specificity in both cohorts. Moreover, by performing subgroup analysis in children who were referred by their general practitioner and paediatrician, we could gain insight into the influence of prevalence on the sensitivity and specificity of faecal calprotectin. We showed that spectrum bias did affect the specificity of faecal calprotectin, but that it did not substantially affect the sensitivity. The tendency of the specificity to be higher at a lower disease prevalence, and for a sensitivity that did not change by disease prevalence, was comparable with the result of a meta-analysis investigating the association between test sensitivity and specificity with disease prevalence. This implies that differences in prevalence particularly reflect differences in the spectrum of patients without disease. Two of the twenty-three included meta-analyses showed significant associations between prevalence and sensitivity, with sensitivity being higher among patient populations where the disease prevalence was higher. Therefore, the authors of the meta-analysis concluded that clinicians can use the prevalence as a guide to select studies that closely match their setting.

A disadvantage of the patient selection method was that the inclusion of children with IBD was lower than expected. If we had included patients with IBD at a prevalence of 20%, the precision of the 95% confidence intervals would have decreased from 5% to ±10%. Because the prevalence of IBD in referred children was 16%, the sensitivity had wide confidence intervals ranging from 0.81 to 1.00. The implications of a sensitivity of 0.81 might be of concern to patients and clinicians. Further research in primary care might have an important impact on our confidence in the effect estimate, causing it to change.

Although diagnostic studies concerning rare diseases in primary care are difficult, other solutions are available to deal with the methodological challenge of patient selection. This includes case-control studies with reversed-flow or two-gate design with representative sampling (i.e., a nested case-control design). Using registration data for routine care might be a solution to help determine the diagnostic accuracy of faecal calprotectin in symptomatic children presenting in primary care. But, such a study would only be feasible if calprotectin was used in routine care. Also, registration data often does not reliably include the details about symptoms and tests necessary for high-quality diagnostic research. However, solutions for the missing data and other methodological challenges have been proposed in diagnostic test research. A nested case-control design, using registration data, might be an understudied solution for determining the diagnostic accuracy of studies in patients with a low prior disease probability in primary care. In such a nested case-control design, the cases and controls are extracted from a source population with a known sample size. Compared with the traditional case-control design, it has the advantage of being suitable for calculating the absolute disease risk next to predictive values and odds ratios; thus, it might be useful in diagnostic accuracy studies among patients with low probabilities of having the disease. Biases associated with these solutions have been extensively reported, but an overview describing methodological challenges linked to practical recommendations, and how to deal with them, is lacking for diagnostic studies among patients with low prior disease probabilities. Recommendations for researchers on design, analysis, reporting and interpreting of results are needed.

Multiple reference standards

Verification is a critical step in any study of diagnostic accuracy. Ideally, the preferred reference standard should be performed in all symptomatic patients. However, diagnosis is not always possible with the preferred reference standard for practical or ethical reasons. In these situations, an alternative reference standard can be used. Although this method of differential verification can be a useful solution, bias can be introduced if the inferior reference standard incorrectly classifies patients as diseased or non-diseased; the resulting diagnostic accuracy would then be systematically different from that in an ideal study using only the preferred reference standard. This systematic deviation introduced by the use of two reference standards is called differential verification bias. Naaktgeboren et al. 2013 provided practical recommendations on how to report results when differential verification occurs. Our presentation of results in chapter 5 was based on these recommendations. A flow diagram, such as that presented in chapter 5, gives an example of how to report the verification process and contingency tables for the index test by each reference standard. Moreover, they provide additional questions related to the QUADAS-2...
Summary and General Discussion

We did not evaluate the population perspective, so we did not assess how the test would affect health care budgets. Both these factors should be addressed in future research. Furthermore, it would be interesting to see research evaluating the diagnostic value of the Rome III criteria in combination with the exclusion of alarm symptoms and normal faecal calprotectin when evaluating functional gastrointestinal disorders, such as irritable bowel syndrome.

An impact or implementation study is needed to determine whether using faecal calprotectin in the presence of alarm symptoms actually improves decision making and cost-effectiveness in daily practice in primary care. This strategy could be evaluated by a stepped wedge cluster randomized trial or a before–after trial in which the diagnostic strategy is evaluated before and after implementing the strategy in a large number of general practices. Because the low prevalence of IBD could make a randomized trial difficult to achieve, the use of before and after registration data might be a solution. Other relevant solutions have been proposed for methodological problems associated with the use of routine care databases in observational studies on treatment interventions (such as missing data and confounding by indication), and these may also apply to diagnostic test research.

Overall, impact studies have confirmed whether faecal calprotectin is a useful test in primary care, it will be possible to include it in primary care guidelines, which should use the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach to assessing the quality of evidence and strength of recommendations.

The GRADE approach allows accurate research and accurate debate concerning judgements about diagnostic tests, and the evidence profiles provide simple transparent summaries that can be used by clinicians.

Overall Conclusion

In this thesis, we have shown that faecal calprotectin is a useful test for ruling out IBD in children with chronic gastrointestinal symptoms in whom the general practitioner considers referral for further diagnostic work-up, e.g., in the presence of alarm symptoms. Our data indicated that optimal outcomes in the management of children presenting with chronic gastrointestinal symptoms in primary care were achieved by adopting a sequential diagnostic strategy. Specifically, we demonstrated that a strategy of performing faecal calprotectin testing when alarm symptoms had been identified was associated with a reduction in unnecessary referrals. However, an impact study is now needed to determine whether this approach might actually improve both the diagnostic decision making of general practitioners in daily practice and the cost-effectiveness of diagnostic assessment in primary care.
REFERENCES


