PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effect of faecal calprotectin testing on referrals for children with chronic gastrointestinal symptoms in primary care: study protocol for a cluster randomised controlled trial
AUTHORS	Ansems, Sophie; Berger, Marjolein; Rheenen, Patrick; Vermeulen, Karin; Beugel, Gina; Couwenberg, Maria; Holtman, G.

VERSION 1 – REVIEW

REVIEWER	Batra, Akshay University Hospital Southampton NHS Foundation Trust
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	It is a very important study which hopefully would positively impact the management of Functional GI disorders in primary care. The study is well designed and planned. I only have a few minor points 1. Please highlight that the value of Faecal calprotectin is because of it's negative predictive value rather than positive. 2. I do not think the list of warning signs is complete. Especially if thsi is being used to guide the use of Fcal it might miss a few patients with IBD. e.g. it does not include common symptoms like reduced energy or appetite, nocturnal symptoms of diarrhoea and pain. 3. Fcal should be used with caution in children presenting with Bleeding PR and ca be falsely elevated. 4. I was unclear from from methodology if all patients with warning signs will have Fcal-POCT tested or if it woudl be at discretion on GP> 5. Please use the term organic disease instead of somatic disease.

REVIEWER	Adler, Jeremy University of Michigan, Child Health Evaluation and Research Center
REVIEW RETURNED	18-Dec-2020

GENERAL COMMENTS	The Investigators submitted a protocol for a pragmatic trial investigating the value of fecal calprotectin testing for children with abdominal symptoms. This is a very well thought out and well-designed study. I have several comments and a couple of concerns outlined below. GENERAL

	1. A number of the references in this protocol are quite old. INTRODUCTION 1. According to the Rome criteria, celiac disease and IBD do not need to be excluded in order to diagnose functional disorders. While I believe the current trial is still appropriate and has important value, this sentence should probably be modified to be consistent with the literature (Hyams JS, Gastroenterol 2016;150:1456–68). 2. The investigators noted important adverse disease outcomes that are associated with delay in diagnosis and treatment. However, the reference for IBD is a dated review article. Consider referencing more recent primary sources which directly address serious outcomes from delayed treatment. For growth failure, consider Walters TD, Gastroenterol 2014;146:383-91. For stricturing complications, consider Kugathasan S, Lancet 2017;389:1710-18. For perianal fistulas, consider Adler J, JAMA Netw Open 2020;3(6):e207378. For surgery consider Nahon S, Dig Dis Sci 2016;61:3278-84. 3. It would be helpful to include the specific details of the Dutch Society recommendations in this protocol. What specific testing is recommended for anyone suspected of having celiac disease or IBD? METHODS AND ANALYSIS 1. It appears that the FCal practices will receive teaching, and the control groups will not. This will introduce an important difference in the practices that will not be accounted for in the analyses. If referral patterns differ between the two groups as anticipated, it will be impossible to distinguish how much of the difference can be attributed to FCal testing vs. the education the physicians received. To avoid this issue, it would be preferable to have similar teaching in the control group, but without FCal. 2. Are there any circumstances in which a physician works in more than one practice in the Netherlands? If so, those practices should either be excluded or allocated to the same treatment group. 3. It will be important to track and report the cross-overs. How many patients in the control group have FCal tested? How many patients in the FCal group do not return the stool samples? 4. It appears that in the intervention group, FCal will be recommended “instead of blood tests when IBD is suspected”. FCal is not sensitive to celiac disease. Does that mean the GP needs to differentiate whether they are concerned about celiac disease vs IBD? This differentiation is not asked of the control group, and will introduce further discrepancies between groups. 5. If FCal is done in place of blood tests, recommend instructions for a stepwise approach. For example, if FCal is negative, then do blood tests. Or if celiac is considered, and blood tests are negative, then do FCal.? If there is no stepwise approach, and if both blood tests and FCal are not tested, then there may be greater number of missed diagnoses.

	6. In my experience far fewer than 90% of patients return stool samples. It is probably worthwhile anticipating a higher rate of drop-out in the power calculation. 7. I am also concerned that evaluating 6 month outcomes is too short a timeframe. Ricciuto et al. (Arch Dis Child 2018;103:319-26) reported IQR of diagnostic delay of 2.9-12.5 months, with 20% delay >1 yr. Nahon et al. (Dig Dis Sci 2016;61:3278-84) reported median delay of 5 months. If the proposed study limits observations to 6 months, they are likely to miss quite a few referrals and diagnoses. This is particularly important because diagnostic delay is associated with worse outcomes (Schoepfer AM, J Pediatr Gastroenterol Nutr 2017;64:245-7). 8. It is unclear how the retrospective search for eligible children will be implemented. Is this meant to identify children recently seen in practice? How much of a time lag will be allowed? Will those families be contacted after the visit for enrollment and lab testing? 9. The protocol refers to “alarm symptoms”, but this is never defined in the protocol. These should be explicitly defined, since the GP’s base their decisions on this (Figure 2). I recommend including both alarm symptoms and signs. Consider including more complete list than that displayed in Supplementary File 3. 10. Alarm signs should be included in addition to symptoms (decreased growth velocity, abdominal mass, perianal lesions, digital clubbing, etc.). See comments under Supplementary File info below. 11. In the economic evaluation, it will be important to consider the cost of a missed, or delayed diagnoses. Diagnostic delay is associated with increased cost and resource utilization (Burisch J, Inflam Bowel. Dis 2015;21:121-31). This should be includd in the proposed cost modeling. 12. Recent evidence suggests a lower range of FCal for younger children (Roca M, Scientific Reports 2020;10:20565). I recognize this was not used in the preliminary studies. The investigators may consider altering the cut-off by age as a sensitivity analysis. DISCUSSION 1. The investigators note the concern of selection bias with GPs in the intervention group gaining greater knowledge of alarm symptoms. This is a significant concern that could easily be avoided by providing parallel training to the control group as suggested above. 2. Monitoring whether GPs adhere to the protocol is appropriate. It will also be important to monitor if patients adhere, such as returning stool sample, doing blood tests, etc. FIGURES 1. Figure 2, clearly describes the decision process for the intervention group. This not only reminds of how to interpret FCal, it also will remind the GP about alarm symptoms. There should be a similar algorithm to remind the GPs about alarm symptoms for the control group (without including FCal).

	2. In Figure 2, the “Alarm symptoms” should include signs as well as symptoms. 3. In Figure 2, the alarm symptoms are vague and inconsistent with the text and supplementary materials. They should include delayed puberty, early satiety, borborygmi, etc. Please also define “perianal abnormalities”, “skin abnormalities” etc. This should be more clearly specified. SUPPLEMENTARY MATERIALS 1. In Supplemental Figure 1, it is important to specify which anti-TTG to test (IgA is more reliable). It is also important to check total serum IgA. If IgA deficiency, anti-TTG is not reliable, and referral to GI would be appropriate. 2. Also, in Supplemental Figure 1, standard fecal culture does not identify parasites and other organisms (like giardia, cryptosporidium, etc.). 3. Supplemental File 1 makes recommendations to obtain medical history and perform physical examination. But alarm signs and symptoms are not included. This would be a good place to include these. 4. Supplementary File 3, should include alarm “signs” in addition to “symptoms.” 5. Recommend including other important alarm symptoms such as arthralgia, early satiety, nocturnal bowel movements, borborygmi, dysphagia. 6. Please clarify what is meant by skin abnormalities. Recommend explicitly including erythema nodosum and pyoderma gangrenosum for IBD, and dermatitis herpetiformis for celiac disease. 7. Also it is important to clarify what is meant by perianal abnormalities. Perianal symptoms (discharge, pain) may not be present in patients with perianal Crohn’s disease (Korelitz BI, Inflam Intest Dis 2018;3:40-2). So it is important to include perianal exam findings including skin tags, deep fissures, fistulas, abscesses (Singer A, Clin Gastroenterol Hepatol, ePub April 30 2020). 8. Supplementary File 2 makes it clear that GPs in the FCal group receive teaching about differential diagnosis, prevalence, and definitions of chronic GI symptoms. It very important the GPs in the control group also receive similar teaching as noted above. 9. Giving explicit guidance would be helpful for GPs (in both groups) because perianal exam findings commonly are mistaken for conditions other than Crohn’s disease (Singer A, Pediatr 2016;137:e20152878). 10. Also consider including digital clubbing, delayed puberty, and nocturnal bowel movements as alarm symptoms (Jiménez Treviño S, Frontiers in Pediatrics 2020;8:1–9)

	SPIRIT ITEMS 1. Item 11. Interventions were clearly described. 2. Item 12. The primary outcome of referral to secondary care is not collected directly from the trial participant. However, I agree that this is an appropriate primary outcome. 3. Item 12. The timepoint of the primary outcome is reported. However, as noted above, I believe the timepoint should be extended from 6 to 12 months (or possibly longer). 4. Item 14. Sample size estimates and outcome used were appropriately reported. 5. Item 14. Please see comment above about losses to follow-up. 6. Item 15. Location of recruitment was described. 7. Item 15. The expected recruitment rate was not described. 8. Item 15. There is no comment in this protocol about identifying a person or people at each site who will identify and recruit patients. 9. Item 16c. Allocation implementation is only referred to as being done by computer-generated list by independent researcher. SPIRIT appears to require identifying a specific person. Not sure if this is actually necessary at this stage. 10. Item 17a. Blinding is well described including explaining why some aspects of the study cannot be blinded. 11. Item 18a. It is not clear who will be collecting the data for the primary outcome, or how it will be obtained. 12. Item 18b. There is no mention of strategies to encourage participant retention. This will be particularly important if the timeframe of outcome is increased as I recommend doing. 13. Item 20a. The statistical methods do not discuss the effect measure for the primary outcome. Confidence intervals were only mentioned in the methods for the economic analyses. 14. Item 20c. It appears that patients who did not return FCal sample are not going to be included. But this is not clear. Intention to treat analyses are generally preferred. However, if the investigators feel otherwise, it would be best to explicitly state inclusion plan. 15. Item 29. There is no statement about sharing data.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Akshay Batra, University Hospital Southampton NHS Foundation Trust Comments to the Author:

It is a very important study which hopefully would positively impact the management of Functional GI disorders in primary care. The study is well designed and planned. I only have a few minor points

We would like to thank the reviewer for the positive remarks about our study. We are happy to hear that there are only minor points.

1. Please highlight that the value of Faecal calprotectin is because of it's negative predictive value rather than positive.

We have added the following sentence to our introduction:

'However, we emphasize that FCal is especially appropriate for use in primary care due to its high negative predictive value (1.00; 95% CI, 0.94-1.00) rather than its positive predictive value (0.59; 95% CI, 0.41-0.75).' (page 5)

2. I do not think the list of warning signs is complete. Especially if this is being used to guide the use of Fcal it might miss a few patients with IBD. e.g. it does not include common symptoms like reduced energy or appetite, nocturnal symptoms of diarrhoea and pain.

We thank you for your suggestion. We acknowledge that our list of alarm symptoms might not be complete. While designing the trial, we based our alarm symptoms on the IBD Guideline of the Dutch Society of Pediatricians and the FCal diagnostic strategy of the North Netherlands Pediatric IBD Consortium (Escher et al. 2018; Van Vijver et al. 2012). We selected the most discriminating alarm

symptoms in consultation with an expert panel consisting of, among others, two pediatric gastroenterologists and two GPs. Whereas (for example) reduced energy and appetite are common in children with IBD, the symptoms in itself are non-specific and highly prevalent in other common health problems. This phenomenon may unintentionally increase the number of children that will be tested and therewith increase the number of false positives. Additionally, our FCal cut-off values are based on study cohorts from diagnostic accuracy studies in primary care settings (Holtman et al. 2016; Ramraj et al. 2018).

We now address our selection of alarm symptoms in the Discussion:

'In a consensus meeting with pediatricians and GPs we selected the alarm symptoms for IBD with the highest discriminatory power. We are of the opinion that adding less discriminating alarm symptoms will unintentionally increase the number of false positive findings. ' (page 17)

3. Fcal should be used with caution in children presenting with Bleeding PR and can be falsely elevated.

To the best of our knowledge, we are not aware of publications describing rectal bleeding causing falsely elevated FCal. We do warn GPs in the online training about the following causes of elevated FCal: age <4 years, chronic use of NSAIDs, chronic use of oral corticosteroids, chronic use of antibiotics, gastroenteritis, food intolerance (celiac disease), food allergy and juvenile intestinal polyps.

4. It was unclear from methodology if all patients with warning signs will have FCal-POCT tested or if it would be at discretion on GP.

We acknowledge this was unclear from the protocol. Therefore we have changed the sentence '…and

each GP will be free to decide whether to use the FCal-POCT during baseline or follow-up consultation(s)' to

Although GPs are instructed to only use FCal when the child presents with alarm symptoms (Table 1), FCal use will be at their own discretion. Consequently, children without FCal testing may also be included and GPs may also test and include children with alarm symptoms other than in the online training or children without alarm symptoms.' (page 7)

5. Please use the term organic disease instead of somatic disease.

Thank you for your suggestion. We now use the term organic disorders instead of somatic disorders.

Reviewer: 2

Dr. Jeremy Adler, University of Michigan, University of Michigan Comments to the Author:

The Investigators submitted a protocol for a pragmatic trial investigating the value of fecal calprotectin testing for children with abdominal symptoms. This is a very well thought out and well-designed study. I have several comments and a couple of concerns outlined below.

We would like to thank the reviewer for the positive comments about our study and for his time to take a careful look at our protocol.

GENERAL

1. A number of the references in this protocol are quite old.

If appropriate we have replaced references with more recent references.

INTRODUCTION

2. According to the Rome criteria, celiac disease and IBD do not need to be excluded in order to diagnose functional disorders. While I believe the current trial is still appropriate and has important value, this sentence should probably be modified to be consistent with the literature (Hyams JS, Gastroenterol 2016;150:1456–68).

We agree that the definition of the ROME IV criteria should be used. Therefore we have changed the sentence 'At least 90% of these children will have functional gastrointestinal disorders (FGID), but before this diagnosis.' can be made, inflammatory bowel disease (IBD) celiac disease, and other causes must be excluded

to 'However, before this diagnosis can be made, the GP should ascertain after appropriate medical evaluation that the symptoms cannot be attributed to inflammatory bowel disease (IBD), celiac disease, and other causes.' (page 4)

3. The investigators noted important adverse disease outcomes that are associated with delay in diagnosis and treatment. However, the reference for IBD is a dated review article. Consider referencing more recent primary sources which directly address serious outcomes from delayed

treatment. For growth failure, consider Walters TD, Gastroenterol 2014;146:383-91. For stricturing complications, consider Kugathasan S, Lancet 2017;389:1710-18. For perianal fistulas, consider Adler J, JAMA Netw Open 2020;3(6):e207378. For surgery consider Nahon S, Dig Dis Sci 2016;61:3278-84.

Thank you very much for your suggestions and taking the time to find appropriate references. Please see our response per suggested reference below.

- Walters TD, Gastroenterol 2014;146:383-91

o Thank you for your suggestion, we have added this reference to the protocol.

- Kugathasan S, Lancet 2017;389:1710-18

o Thank you for your suggestion, we have added this reference to the protocol.

- Adler J, JAMA Netw Open 2020;3(6):e207378

o Thank you for your suggestion, we have added this reference to the protocol.

- Nahon S, Dig Dis Sci 2016;61:3278-84

o This article describes an increased risk of early surgery in case of a long diagnostic delay (>13 months). However, this cohort consists of adult patients and therefore we do not find it appropriate to use this article as a reference

We acknowledge that our reference (Kim and Ferry 2004) is a dated publication. We have replaced this with a slightly more recent and primary resource (Sawczenko et al. 2006) . Additionally, we have added a more recent publication reporting growth failure and stricturing and internal fistulising complications as a consequence of diagnostic delay in children with IBD (Ricciuto et al. 2020).

We have changed the sentence ‘At the same time, it is critical that we avoid delaying the diagnosis and treatment of IBD and celiac disease to minimise complications such as anaemia and growth failure, and in the case of IBD, delayed sexual maturation (6,7) to

‘At the same time, it is critical that we avoid delaying the diagnosis and appropriate treatment of IBD and celiac disease to minimise complications such as anaemia and growth failure (7-10), and in the case. of IBD, delayed sexual maturation (7), stricturing (10) and internal fistulising complications (10-12) ’ (page 4)

4. It would be helpful to include the specific details of the Dutch Society recommendations in this protocol. What specific testing is recommended for anyone suspected of having celiac disease or IBD?

Supplementary File 1 ‘Information leaflet control group’ shows the recommendations of the Dutch Society of GPs. The last column shows the recommended testing for suspected celiac disease and IBD. We have changed the recommendation for suspected celiac disease from ‘anti-TTG’ to ‘tTGA and total serum IgA’ as the guideline prescribes (see revised Supplementary File 1).

In addition, we changed the sentence ‘The Dutch Society of General Practitioners (Nederlands Huisartsen Genootschap; NHG) recommends additional testing for suspected celiac disease and blood tests for suspected IBD (e.g., haemoglobin, leukocytes and ESR).’ to:

‘The Dutch Society of General Practitioners (Nederlands Huisartsen Genootschap; NHG) recommends testing tissue transglutaminase IgA (tTGA) and total serum IgA for suspected celiac disease and testing haemoglobin, leukocytes and ESR for suspected IBD.’ (page 4)

We have also changed the following sentence: ‘All GPs will receive an information leaflet about what is considered care as usual (Supplementary File 1)’

to ‘All GPs will receive an information leaflet about what is considered care as usual per the guideline of the Dutch society for GPs (Supplementary File 1).’ (page 7)

METHODS AND ANALYSIS

5. It appears that the FCal practices will receive teaching, and the control groups will not. This will introduce an important difference in the practices that will not be accounted for in the analyses. If referral patterns differ between the two groups as anticipated, it will be impossible to distinguish how much of the difference can be attributed to FCal testing vs. the education the physicians received. To avoid this issue, it would be preferable to have similar teaching in the control group, but without FCal.

We combined FCal testing with an educational intervention because it is essential that new test users are trained about its indications (risk stratification), interpretation, and possible drawbacks (false positive and false negative rates). Implementation of a new test without teaching about the test strategy could lead to missed diagnosis, over-diagnosis, and unnecessary costs for patients and wider society (André et al. 2004; Schols, Dinant, and Cals 2018). We therefore think that the educational element is an integral part of this RCT in order to increase the compliance to the new test strategy. This is of similar importance when a test is eventually implemented in a real-world-setting.

We agree that it will be impossible to distinguish whether the effect can be attributed to FCal testing or to the educational element, but we think that these two elements should go hand-in-hand. We will conduct however, a qualitative evaluation exploring GPs’ experiences with the different elements of the intervention in order to identify barriers and facilitators for adherence to the test strategy.

We now highlight the above considerations in our Discussion:

‘This study evaluates the impact of a test strategy in which FCal testing is a major component. It will be

impossible to distinguish whether a possible effect can be attributed to FCal testing or to the training. However, we think that these two elements should go hand-in-hand in order to increase the compliance to the new test strategy and prevent missed diagnoses, over-diagnosis and unnecessary costs for patients and wider society (22,66). This is of similar importance when a test is implemented in a real world setting. (page 16)

6. Are there any circumstances in which a physician works in more than one practice in the Netherlands? If so, those practices should either be excluded or allocated to the same treatment group.

In The Netherlands, it is rare for GPs to work in two or more practices. However, in case a GP works in multiple practices not allocated to the same study arm, this GP will only include children in the practice that first started participation in this study.

We have added the following sentence to the section 'Randomization and Blinding': 'On the rare occasion that a GP works in multiple practices not allocated to the same study arm, this GP will only include children in the practice that first participated in this study.' (page 7)

7. It will be important to track and report the cross-overs. How many patients in the control group have FCal tested? How many patients in the FCal group do not return the stool samples?

We thank you for your suggestion.

As stated in Supplementary File 3 'Data collection from medical records over 6 months' we will collect whether FCal is tested in both study arms. In addition, we will also collect whether children did not return their stool samples when the GP intended to test FCal. We have clarified this by adding the following sentence to the paragraph 'Outcomes': 'For FCal in specific, we will also collect whether children return their stool samples.' (page 11)

Additionally, children in the intervention group who do not return their stool sample will not be included in the per protocol analysis. We have amended the following sentence in the 'Analysis':

'Analyses will then be repeated for both the primary and secondary outcomes on a per protocol basis. In the intervention group, we will only include children who receive the intended diagnostic strategy (per the indications explained in the online training and with a returned stool sample) or rightfully did not receive FCal testing (children without alarm symptoms) and in the control group, we will only include children who did not undergo FCal testing.' (page 13)

8. It appears that in the intervention group, FCal will be recommended "instead of blood tests when IBD is suspected". FCal is not sensitive to celiac disease. Does that mean the GP needs to differentiate whether they are concerned about celiac disease vs IBD? This differentiation is not asked of the control group, and will introduce further discrepancies between groups.

We ask both groups to differentiate whether they suspect the child of IBD or other causes of chronic gastrointestinal symptoms, such as celiac disease.

The control group receives an information leaflet (Supplementary File 1) containing recommendations according to the guideline of the Dutch Society of GPs. This leaflet recommends separate testing for celiac disease, IBD and gastroenteritis. In the intervention group, we advise GPs

to 'Also consider other organic diseases, such as gastroenteritis and celiac disease, and perform appropriate diagnostic testing if indicated' (Revised Figure 2).

We acknowledge that the sentence "However, this leaflet will be amended to recommend FCal instead of blood tests when IBD is suspected' is confusing. Therefore, we have changed this to 'However, this

leaflet will be amended to recommend FCal instead of ESR, leukocytes and haemoglobin when IBD is suspected.' (page 7)

9. If FCal is done in place of blood tests, recommend instructions for a stepwise approach. For example, if FCal is negative, then do blood tests. Or if celiac is considered, and blood tests are negative, then do FCal.? If there is no stepwise approach, and if both blood tests and FCal are not tested, then there may be greater number of missed diagnoses.

Please see our response to comment #8.

We do not recommend testing FCal in place of blood tests. Since this is a pragmatic trial, the order of testing is on discretion of the GP. The same applies to which tests the GP uses.

10. In my experience far fewer than 90% of patients return stool samples. It is probably worthwhile anticipating a higher rate of drop-out in the power calculation.

Please see our response to comment #4 from reviewer #1.

Since children without FCal testing may also be included in the intervention group, the stool sample return rate will not influence the drop-out rate. However, as explained in our response to comment #7, we will document whether stool samples are returned and will account for this in the per-protocol analysis.

11. I am also concerned that evaluating 6 month outcomes is too short a timeframe. Ricciuto et al. (Arch Dis Child 2018;103:319-26) reported IQR of diagnostic delay of 2.9-12.5 months, with 20% delay >1 yr. Nahon et al. (Dig Dis Sci 2016;61:3278-84) reported median delay of 5 months. If the proposed study limits observations to 6 months, they are likely to miss quite a few referrals and diagnoses. This is particularly important because diagnostic delay is associated with worse outcomes (Schoepfer AM, J Pediatr Gastroenterol Nutr 2017;64:245-7).

We thank you for your comment and the suggested articles. We would also like to draw your attention to two other papers describing rather shorter time intervals between first physician visit and referral / diagnosis (Schoepfer et al. 2017; median 3 months (IQR 1-9 months) and Sergi et al. 2020; median 1.8 months (IQR 0.66-5.1 months)).

We have chosen for a follow -up of 6 months because we expect the majority of referrals to take place within this period. The guideline of the Dutch Society of GPs emphasizes that follow-up of children with chronic gastro-intestinal symptoms has to be planned carefully in consultation with the parents. With proper instructions and follow-up, most children with IBD will be recognized within 6 months. Furthermore, we believe that the FCal test strategy will have the largest effect on patient important outcomes (e.g. referrals, parental worries) within the first 6 months after GP consultation. Additionally, we would like to emphasize that the expected number of children with IBD in our cohort will be low, supported by our preliminary study with almost identical in- and exclusion criteria (Holtman et al. 2016). Due to these expected small numbers, a possible difference in IBD diagnoses between the intervention and control group is more likely to be attributable to chance than to the intervention.

However, we do acknowledge that it is important to explore the reasons for diagnostic delay in children with IBD originating from the primary care physician. Therefore, we are planning to conduct a qualitative analysis describing the clinical presentation and diagnostic trajectory of the children diagnosed with IBD in our cohort. This will take place separately from this RCT and will therefore not be included in this protocol.

12. It is unclear how the retrospective search for eligible children will be implemented. Is this meant to identify children recently seen in practice? How much of a time lag will be allowed? Will those families be contacted after the visit for enrollment and lab testing?

We have changed the following section:

‘Additionally, research staff will retrospectively search for eligible children in GP registration databases using a search strategy based on International Classification of Primary Care (ICPC) codes (Supplementary File 4). Together with a short introduction about the study provided by the GP, all included children and/or parents will receive a patient information letter and will be asked to provide informed consent for completing questionnaires.’

To:

‘Additionally, research staff will retrospectively search for eligible children seen in practice in the previous 3 months. They will search in GP registration databases using a search strategy based on International Classification of Primary Care (ICPC) codes (Supplementary File 4). Together with a short introduction about the study provided by the GP, All included children and/or parents (regardless of the recruitment strategy) will receive a patient information letter and will be asked to provide written informed consent for completing questionnaires.’ (page 11)

13. The protocol refers to “alarm symptoms”, but this is never defined in the protocol. These should be explicitly defined, since the GP’s base their decisions on this (Figure 2). I recommend including both alarm symptoms and signs. Consider including more complete list than that displayed in Supplementary File 3.

Please see our answer to comment #2 regarding the selection of alarm symptoms.

We acknowledge that the term ‘alarm symptoms’ could confuse readers. In the online training, we also instruct GPs to look for alarm signs during physical examination. We recognize this was not clear from the protocol. To accentuate this, we have added Table 1 ‘Definitions of alarm symptoms for Inflammatory Bowel Disease’. This table shows whether alarm symptoms are ascertained at medical history or physical examination.

Table 1. Definitions of alarm symptoms for Inflammatory Bowel Disease

Alarm symptom	Method of ascertainment	Definition of positive finding
		erythema nodosum
		perianal fissures, perianal abscesses

These definitions apply to the alarm symptoms mentioned in the protocol, figures and all supplementary files.

IBD; Inflammatory Bowel Disease.

14. Alarm signs should be included in addition to symptoms (decreased growth velocity, abdominal mass, perianal lesions, digital clubbing, etc.). See comments under Supplementary File info below.

Please see our answer above (comment #13).

15. In the economic evaluation, it will be important to consider the cost of a missed, or delayed diagnoses. Diagnostic delay is associated with increased cost and resource utilization (Burisch J, Inflam Bowel. Dis 2015;21:121-31). This should be included in the proposed cost modeling.

We thank you for the suggested paper. We acknowledge it is important to include the costs of missed and delayed diagnoses in the cost analysis. However, we considered this too detailed information to include in this protocol. We will address this topic in our planned publication about the cost effectiveness of the intervention.

16. Recent evidence suggests a lower range of FCal for younger children (Roca M, Scientific Reports 2020;10:20565). I recognize this was not used in the preliminary studies. The investigators may consider altering the cut-off by age as a sensitivity analysis.

We acknowledge that younger children tend to have higher FCal values. This was one of our grounds to recommend GPs to only refer children with an FCal value >250 mg/kg and to monitor children with a value between 50 and 250 mg/kg. Roca et al. describe that about 20% of healthy children had FCal concentrations above 50 mg/kg. However, the 95th percentile for FCal was 104.5 mg/kg and they only report one child with a value >250 mg/kg. Therefore, we think our recommendations are adequate.

We would like to thank you for informing us about this recent report and have added the paper as a reference to the following sentence on page 8: ‘However, an FCal value >50 µg/g also has a high false-positive rate (13%) when tested in a population of children both with and without alarm symptoms.’

DISCUSSION

17. The investigators note the concern of selection bias with GPs in the intervention group gaining greater knowledge of alarm symptoms. This is a significant concern that could easily be avoided by providing parallel training to the control group as suggested above.

We would like to refer to our answer to comment #5.

We would like to emphasize that GPs in the control group will receive an information leaflet about standard care according to the guideline of the Dutch Society of GPs (Supplementary File 1). The guideline as well as the leaflet contain information about alarm symptoms for IBD. We have elucidated this in the revised version of Supplementary File 1. We would also like to stress that the guidelines of

the Dutch Society of GPs are an integral part of primary care in The Netherlands and are used on a regular basis by Dutch GPs.

18. Monitoring whether GPs adhere to the protocol is appropriate. It will also be important to monitor if patients adhere, such as returning stool sample, doing blood tests, etc.

Unfortunately, we cannot monitor whether all blood and urine samples are returned, since this information is not reliably extractable from the children's medical files. In addition, we believe that non-returned samples are part of a real-world-setting, and should not be accounted for in the intention-to-treat analysis. However, we will account for non-returned stool samples in the per-protocol analysis, as noted in our answer to comment #7.

FIGURES

19. Figure 2, clearly describes the decision process for the intervention group. This not only reminds of how to interpret FCal, it also will remind the GP about alarm symptoms. There should be a similar algorithm to remind the GPs about alarm symptoms for the control group (without including FCal).

Please see our answers to Comment #5 and #17.

21. In Figure 2, the "Alarm symptoms" should include signs as well as symptoms.

Please see our response to comment #13 regarding the addition of alarm signs. Table 1 now shows whether the GP ascertains alarm symptoms at medical history or physical examination.

22. In Figure 2, the alarm symptoms are vague and inconsistent with the text and supplementary materials. They should include delayed puberty, early satiety, borborygmi, etc. Please also define "perianal abnormalities", "skin abnormalities" etc. This should be more clearly specified.

Please see our response to comment #2 from Reviewer 1 regarding the choice of alarm symptoms. Table 1 now shows the definitions of the alarm symptoms. We have ensured that the alarm symptoms mentioned in revised Figure 2 and revised supplementary files 1 and 3 are consistent with Table 1.

SUPPLEMENTARY MATERIALS

23. In Supplemental Figure 1, it is important to specify which anti-TTG to test (IgA is more reliable). It is also important to check total serum IgA. If IgA deficiency, anti-TTG is not reliable, and referral to GI would be appropriate.

Please see our response to comment #4 and revised Supplementary File 1. We now specify that the recommendation is to test tissue transglutaminase IgA and total serum IgA.

The guideline of the Dutch Society of GPs does indeed recommend referral in case of an IgA deficiency. We did not adopt this in the leaflet because we considered this too detailed information that GPs can look up in the guideline themselves.

24. Also, in Supplemental Figure 1, standard fecal culture does not identify parasites and other organisms (like giardia, cryptosporidium, etc.).

We apologize for the misinterpretation. We have added 'fecal ova and parasite test' to the last column of Supplementary File 1.

25. Supplemental File 1 makes recommendations to obtain medical history and perform physical examination. But alarm signs and symptoms are not included. This would be a good place to include these.

Please see our response to comment #17 and revised supplementary file 1.

26. Supplementary File 3, should include alarm "signs" in addition to "symptoms."

Please see our response to comment #13.

27. Recommend including other important alarm symptoms such as arthralgia, early satiety, nocturnal bowel movements, borborygmi, dysphagia.
Please see our response to comment #2 from Reviewer 1 regarding the selection of alarm symptoms.

28. Please clarify what is meant by skin abnormalities. Recommend explicitly including erythema nodosum and pyoderma gangrenosum for IBD, and dermatitis herpetiformis for celiac disease.

Table 1 now shows the definition of skin abnormalities (including erythema nodosum and pyoderma gangrenosum). We recognize that dermatitis herpetiformis is associated with celiac disease. However, this dermatological condition is not mentioned in the guideline of the Dutch Society of GPs and therefore not adopted in the information leaflet.

29. Also it is important to clarify what is meant by perianal abnormalities. Perianal symptoms

(discharge, pain) may not be present in patients with perianal Crohn's disease (Korelitz BI, Inflam Intest Dis 2018;3:40-2). So it is important to include perianal exam findings including skin tags, deep fissures, fistulas, abscesses (Singer A, Clin Gastroenterol Hepatol, ePub April 30 2020).

Table 1 now shows the definition of perianal abnormalities, including skin tags, fissures, fistulas and abscesses ascertained at a perianal exam.

30. Supplementary File 2 makes it clear that GPs in the FCal group receive teaching about differential diagnosis, prevalence, and definitions of chronic GI symptoms. It very important the GPs in the control group also receive similar teaching as noted above.

We would like to refer to our answer to comment #5. Supplementary File 1 'Information leaflet control group' also contains information about the differential diagnosis and prevalence of chronic gastrointestinal symptoms (column 'Epidemiology'). In addition, the guideline of the Dutch Society of GPs contains information about the definition of chronic abdominal pain and chronic diarrhea. This guideline is an integral part of primary care in The Netherlands an used on a regular basis by Dutch GPs.

31. Giving explicit guidance would be helpful for GPs (in both groups) because perianal exam findings commonly are mistaken for conditions other than Crohn's disease (Singer A, Pediatr 2016;137:e20152878).

Thank you for your suggestion. We instruct GPs in the intervention group to conduct a perianal exam and look for skin tags, perianal fistulas, haemorrhoids, perianal fissures, perianal abscesses. The online training shows a picture of each of the aforementioned.

As mentioned in our response to comment #5, we do not believe we should explicitly instruct GPs in the control group besides what is considered standard care according to the guideline of the Dutch Society of GPs. This guideline recommends inspecting the perianal region searching for perianal fissures, scars, abscesses, and fistulas.

32. Also consider including digital clubbing, delayed puberty, and nocturnal bowel movements as alarm symptoms (Jiménez Treviño S, Frontiers in Pediatrics 2020;8:1–9)

We would like to refer to our answer to comment #2 from reviewer 1 regarding our selection of alarm symptoms.

SPIRIT ITEMS

33. Item 11. Interventions were clearly described.

Thank you for your comment.

34. Item 12. The primary outcome of referral to secondary care is not collected directly from the trial participant. However, I agree that this is an appropriate primary outcome.

Thank you for your comment.

35. Item 12. The timepoint of the primary outcome is reported. However, as noted above, I believe the timepoint should be extended from 6 to 12 months (or possibly longer).

Please see our response to comment #11.

36. Item 14. Sample size estimates and outcome used were appropriately reported.

Thank you for your comment.

37. Item 14. Please see comment above about losses to follow-up.

Please see our response to comment #10.

38. Item 15. Location of recruitment was described.

Thank you.

39. Item 15. The expected recruitment rate was not described.

We have adjusted the following sentence in 'Sample size': 'Given a mean cluster size of 3 and an intraclass correlation coefficient of 0.06, we would need 366 children (183 per arm).'

To 'Given a mean cluster size (expected recruitment rate per practice) of 3 and an intraclass correlation coefficient of 0.06, we would need 366 children (183 per arm).' (page 12)

We have also adjusted this in the revised SPIRIT checklist.

40. Item 15. There is no comment in this protocol about identifying a person or people at each site who will identify and recruit patients.

We have changed the following sentence from: 'GPs will recruit consecutive eligible children during baseline consultations for one year (Figure 3).'

To: 'GPs will identify and recruit consecutive eligible children during baseline consultations for one year (Figure 3). ' (page 11)

41. Item 16c. Allocation implementation is only referred to as being done by computer-generated list by independent researcher. SPIRIT appears to require identifying a specific person. Not sure if this is actually necessary at this stage.

We have added the name of the independent researcher in 'Randomisation and blinding': 'GP practices will be randomised by a computer-generated list using varying block randomisation in 1:1 ratio by an independent researcher (H van der Worp, PhD) not involved in the project.' (page 7)

42. Item 17a. Blinding is well described including explaining why some aspects of the study cannot be blinded.

Thank you.

43. Item 18a. It is not clear who will be collecting the data for the primary outcome, or how it will be obtained.

We have changed the sentence ‘The trial inclusion form is then sent to the researchers, and for all included children, data will be retrieved from their medical files for each consultation (including baseline) over a 6-month follow-up period (Supplementary File 3)’ to

‘The trial inclusion form is then sent to the researchers, and for all included children, data will be retrieved from their medical files for each consultation (including baseline) over a 6-month follow-up period by research team in a standardized online data-entry form (Supplementary File 3). ’ (page 12)

44. Item 18b. There is no mention of strategies to encourage participant retention. This will be particularly important if the timeframe of outcome is increased as I recommend doing.

We do not expect large problems in patient retention for the primary outcome (referral), since this will be retrieved from the medical files by the research team. However, we acknowledge that patient retention for the secondary outcomes evaluated by questionnaires could be a problem. We have added the following sentence to ‘Data collection’:

The estimated time to complete each questionnaire is 15–20 minutes, and if they are not completed, the child and/or parents will automatically receive reminders via e-mail after 7 and 14 days. If not completed after two reminders, we will call the child and/or parents by phone. (page 12)

45. Item 20a. The statistical methods do not discuss the effect measure for the primary outcome. Confidence intervals were only mentioned in the methods for the economic analyses.

We have added the following sentence to the paragraph ‘Analysis’: All analyses will be presented as estimates of intervention effects (adjusted mean differences or odds ratios, as appropriate), with associated 95% CIs and p values. (page 12)

46. Item 20c. It appears that patients who did not return FCal sample are not going to be included. But this is not clear. Intention to treat analyses are generally preferred. However, if the investigators feel otherwise, it would be best to explicitly state inclusion plan.

Please see our answer to comment #7.

47. Item 29. There is no statement about sharing data.

We have added the following sentence to the paragraph ‘Dissemination’: The data of this study will be available on request. (page 14)

REFERENCES

André, Malin, A ˚ Ke Schwan, and Inga Odenholt. 2004. “The Use of CRP Tests in Patients with Respiratory Tract Infections in Primary Care in Sweden Can Be Questioned.” Scandinavian Journal of Infectious Diseases 36(3): 192–97.

Escher, Johanna C et al. 2018. “Inflammatoire Darmziekten (IBD) Bij Kinderen En Adolescenten.” Nederlandse Vereniging voor Kindergeneeskunde: 295.

Holtman, Gea A et al. 2016. “Diagnostic Accuracy of Fecal Calprotectin for Pediatric Inflammatory Bowel Disease in Primary Care: A Prospective Cohort Study.” Annals of family medicine 14(5): 437–45.

Kim, Sandra C, and George D Ferry. 2004. “Inflammatory Bowel Diseases in Pediatric and Adolescent Patients: Clinical, Therapeutic, and Psychosocial Considerations.” Gastroenterology 126(6): 1550–60.

Ramraj, Ramya et al. 2018. “Utility of Fecal Calprotectin in Evaluation of Chronic Gastrointestinal Symptoms in Primary Care.” Clinical Pediatrics 57(9): 1058–63.

Ricciuto, Amanda et al. 2020. “Diagnostic Delay Is Associated With Complicated Disease and Growth Impairment in Paediatric Crohn’s Disease.” Journal of Crohn’s and Colitis 2020: 1–13.

Sawczenko, Andrew, Anne B. Ballinger, Martin O. Savage, and Ian R. Sanderson. 2006. “Clinical Features Affecting Final Adult Height in Patients with Pediatric-Onset Crohn’s Disease.” Pediatrics 118(1): 124–29.

Schoepfer, Alain M. et al. 2017. “Systematic Evaluation of Diagnostic Delay in Pediatric Inflammatory Bowel Disease.” Journal of Pediatric Gastroenterology and Nutrition 64(2): 245–47.

Schols, Angel MR, Geert-Jan Dinant, and Jochen WL Cals. 2018. “Point-of-Care Testing in General Practice: Just What the Doctor Ordered?” British Journal of General Practice 68(673): 362–63.

Sergi, Consolato et al. 2020. “Spanish Pediatric Inflammatory Bowel Disease Diagnostic Delay Registry: SPIDER Study From Sociedad Española de Gastroenterologıa, Hepatologıá y Nutrición Pediátrica.” Frontiers in Pediatrics | www.frontiersin.org 8: 584278.

Van Vijver, Els De et al. 2012. “Safely Ruling out Inflammatory Bowel Disease in Children and Teenagers without Referral for Endoscopy.” Archives of Disease in Childhood 97(12): 1014–18.

VERSION 2 – REVIEW

REVIEWER	Adler, Jeremy University of Michigan, Child Health Evaluation and Research Center
REVIEW RETURNED	21-Mar-2021

GENERAL COMMENTS	I would like to thank the Investigators for addressing my questions and concerns. The revised protocol document has been improved, and I have no other additional comments.