Author’s response to reviews

Title: Is there an added value of faecal calprotectin and haemoglobin in the diagnostic work-up for primary care patients suspected of significant colorectal disease? A cross-sectional diagnostic study

Authors:
Sjoerd Elias (s.elias@umcutrecht.nl)
Liselotte Kok (l_kok_@hotmail.com)
Niek de Wit (n.j.dewit@umcutrecht.nl)
Ben Witteman (wittemanb@zgv.nl)
Jelle Goedhard (goedh536@planet.nl)
Mariëlle Romberg-Camps (m.camps@orbisconcern.nl)
Jean Muris (jean.muris@maastrichtuniversity.nl)
Karel Moons (k.g.m.moons@umcutrecht.nl)

Version: 1 Date: 02 Aug 2016

Author’s response to reviews:

Reviewer 1

The authors have performed a very interesting analysis to try and distinguish patients with SCD from those without in patients referred for colonoscopy by GP.

We have the following comments:

1. We have questions about the choice of FIT and its characteristics. Why was this FIT chosen and not one of the more established FIT used in the different screening programs such as OC-Sensor and FOB-Gold? Is the chosen FIT quantitative or qualitative? From the text it is not exactly clear whether the FIT results were used in a dichotomous fashion (positive/negative) or as a quantitative outcome in ug hb/g feces. The latter could significantly improve the discriminatory performance of the model, and if was not used should definitely be considered a serious
limitation of this work and be discussed. If the FIT was used dichotmously: what was the cut-off for a positive test?

We agree with the reviewer that a more detailed description of the FIT test is needed in the manuscript. At the time of the planning of the project there were (and still are) many FITs available, and the Alere Clearview® iFOBT One Step Faecal Occult Blood Test Device was one of the occult blood tests on the market that was well introduced in clinical practice, that is intended for use in physicians offices, and that allowed a rapid turnaround, which we considered important for implementation in general practice. The OC-Sensor and/or the FOB-Gold are indeed good alternatives.

The Clearview® iFOBT One Step Occult Blood test is a rapid chromatographic immunoassay for qualitative detection of human faecal haemoglobin and its immediate degradation products. The lower detection limit as stated by the manufacturer is 6 g haemoglobin/g faeces. The test is inherently dichotomous (positive/negative), and we consequently analysed it as such. At this threshold, the FIT showed a sensitivity of 88.6% (95%CI: 74.2-95.9) and a specificity of 77.9% (95%CI: 74.3-80.9%) for colorectal cancer in our data. This remarkably agrees with a reported sensitivity for colorectal cancer of 89.3% and a specificity of 79.1% for the OC-Sensor at a threshold of 10 g haemoglobin/g faeces that was reported in a UK-based study of 755 primary care patients referred for endoscopy.1

One of the main reasons we chose this qualitative POC FIT for our study is its ease of use in a primary care situation. Nevertheless, using a qualitative FIT certainly introduces limitations that indeed need to be discussed in more detail. According to the above we have made several changes in the text.

Changes in the text in red (background section):
We specifically focused on POC tests as these can be easily executed at the time and place of patient care.

Changes in the text in red (method section):
We analysed the faecal samples for calprotectin concentration by a quantitative POC test (Quantum Blue®; dynamic range 30-300 g/g) and by an enzyme-linked immunosorbent assay
(ELISA; EK-CAL Calprotectin ELISA, both from Bühlmann Laboratories), both yielding estimates of g calprotectin/g faeces, and for faecal Hb by a qualitative POC FIT (Clearview® iFOBT One Step Faecal Occult Blood Test Device, Alere Health), yielding either a positive or negative test result (lower detection limit of 6 g/g).

We then added the faecal biomarker tests to this basic diagnostic model (the calprotectin tests continuously and the POC FIT dichotomously) (...)

Changes in the text in red (discussion section):

Finally, the use of a qualitative POC FIT in the way that we did in this study, although easily implemented in primary care, also has limitations. First, as the qualitative POC FIT yields a positive or a negative test result (with a detection limit of 6 g Hb/g faeces), the diagnostic information that would be available by quantitatively assessing the amount of Hb present in faeces is lost. (...) ...the POC FIT performed well in our study despite these limitations, and the sensitivity and discriminatory performance of faecal Hb testing in primary care will thus likely be even better when using Hb stabilizing buffers in faecal sample collection devices and using a quantitative FIT.

2. The authors state that they consider blood tests to be burdensome and have therefore used more strict analysis for including these type of tests in the model. However, no such criteria have been used for stool-based tests. Given barriers that people face against stool-based testing I don't feel that stool-based testing is any less burdensome to patients than blood-based testing.

This is indeed an excellent remark that was not addressed sufficiently in the main text of our manuscript. Acceptance of diagnostic tests is probably culturally determined, and in daily practice most Dutch patients do not have problem with faecal sampling, and prefer this to a painful prick. However, we agree with this reviewer that faecal testing may be burdening to patients too, just as burdensome as blood testing. We therefore also have used a more stringent selection criterion for the faecal tests. This information was included in Supplementary Table S1 that contains more details about the model development strategy. All five faecal biomarker extended models fulfilled the more stringent selection criterion. To make this issue of faecal marker selection more clear we have now addressed it in the main text also.
Changes in the text in red (method section):

We then added the faecal biomarker tests to this basic diagnostic model (...) As faecal testing may also be burdensome, we used the same stringent selection criterion for each faecal biomarker test as for the blood analyses (i.e. p<0.05 for model improvement).

3. In the discussion the authors suggest that the miss rate of SCD may not be so bad, because it will be merely a delay in diagnosis, rather than missing the diagnosis. However, we are not talking about a screening situation, but a clinical situation where sensitivity is more important than specificity. I agree with the authors that a reduction in colonoscopy of 30% means a substantial reduction in burden for patients and health care resources. However, I do not feel that the authors can so easily dismiss the miss rate of SCD. A recent study showed that delay in diagnosis may result in an increase of advanced disease (Meester, Clin Gastro Hepatol 2016), and the delay in diagnosis of stage I cancer may very well cause this cancer to progress to stage II or III. In addition there is the patient distress, that he was first denied colonoscopy, to be later detected with CRC anyways.

We agree that the issue of missing a cancer diagnosis is something that should not be taken too easily. However, exclusion of any risk of missing cancer is impossible, not in the least because cancers are missed during colonoscopy as well. The challenge is to find an acceptable balance in cost-effectiveness of care. We think we found this balance in the present approach, in which we restrict the number of colonoscopies required at the price of delaying the diagnosis of missing a small number of patients with a serious disease. The legitimation of delaying instead of missing is that we are talking about symptomatic patients. One of the security measures that primary care physicians use in daily practice is that they after explaining that according to their assessment there is no reason for further work-up at this moment advise patients to re-consult in case symptoms persist longer than 2 weeks. In practice this means that patients with persisting complaints after a false negative decision not to refer for colonoscopy will in majority return because of persistence of symptoms. This will then lead to a positive decision for colonoscopy, which will be explained by the argument that the persistence of symptoms is the reason to refer in second phase. In daily practice this may lead to a delay of 2-3 weeks, which is unlikely to affect the outcome of care. Although this may be a difficult message, this delayed diagnostics proves acceptable for patients if adequately explained.
Given the natural history of advanced adenomas and colorectal cancer, it is indeed possible that even a small delay in diagnosis could result in stage advancement. But a delay of several weeks will probably increase this risk infinitesimally. As estimated in the paper by Meester et al. by microsimulations, delays of 2-3 weeks in diagnostic colonoscopy following a positive FIT will have likely very little impact on colorectal cancer mortality. Even if referral is postponed by two months, the projected adverse effects on a population level seem marginal from that study (very small absolute increased risk of adenomas progressing to invasive cancer and very small absolute increased risk of stage-advancement of invasive cancers). Of note, this simulation study focussed on delaying a diagnosis in a high-risk population (i.e. FIT positives), whereas in our study the concern pertains to diagnostic delay in a low-risk population, likely further curbing the potential adverse effects of diagnostic delay on a population level. We have added text to the discussion in line with this reasoning.

Changes in the text in red (discussion section):

With keen attention in case of non-referral at first consultation to persisting symptoms over a time frame of 2-3 weeks, we think this will result in delaying, but not missing such diagnoses. Such a limited delay will also not likely advance the disease stage substantially for SCD patients initially non-referred.

4. What is the explanation for the increase in detection rate of advanced adenomas after 2011?

This is indeed an intriguing observation that we cannot fully explain. The increase in advanced adenoma detection was independently present in each of the two largest endoscopy centres that together evaluated 95% of the CEDAR patients (one of these centres started participating in 2009 and the other in 2010). Furthermore, the enrolled patient mix did not change substantially during the study period (i.e. the age, gender, and sign- and symptom distribution was not different before and since 2011), and when adjusted for these patient characteristics by logistic regression, the odds ratio of advanced adenoma comparing the study period since 2011 with that before 2011 was 2.7 (95%CI: 1.4-5.4). So geographic variation and patient mix do not seem to explain this trend.

One of the theoretical explanations we discussed is the introduction of the CRC screening program. Around 2011, Dutch gastroenterologist were obliged to take a practical and theoretical exam to qualify for participation in the colorectal cancer screening program that started in 2014. As part of the practical exam, gastroenterologists needed to show adequate adenoma detection
rates, which probably encouraged polypectomy of adenomas previously left untouched, and may have lead to an implicit lowering of the threshold of judging small adenomas as larger than 1 cm. Our data however show that overall, 128 adenomas of any size were detected in CEDAR for an overall prevalence of 16%, which was 14% before and 18% since 2011; 81 adenomas were judged to be 1cm or less for a small adenoma prevalence of 10%, which was also 10% before 2011 as well as since 2011. So although the increase in larger adenoma rate is in line with the above-suggested explanation, the unaltered rate of small adenomas is not.

Changes in the text in red (discussion section):

Related to this, the prevalence of AAs in our study almost doubled from February 2011 onwards (from 4.2 to 7.7%, comprising 25.8% versus 41.8% of SCD cases an increase that could not be explained by changes in patient mix throughout the study period, nor by differences in detection rates between endoscopy centres, but may have been introduced by increased awareness of gastroenterologist who around that time started preparing for the introduction of the colorectal cancer screening program in 2014).

5. The introduction, methods and results are very well written. However, it seems that the discussion is written in a somewhat ad-hoc fashion. It lacks explanation of results, and immediately jumps into strengths and limitations of the study. Also, the discussion starts of very forcefully, while the conclusion in the end is quite weak. This needs more balancing.

We have reorganized the discussion according to these suggestions and have changed the conclusion.

Changes in the text in red (conclusion section):

A simple model including information from history taking, physical examination, and a POC FIT may safely rule out SCD and prevent unnecessary endoscopy referral in approximately one third of SCD-suspected primary care patients. Adding a calprotectin test to such a strategy has limited value.
The authors are correct that the majority of patients referred from primary care for endoscopic bowel examination do not have significant colorectal disease. The development of diagnostic strategies to better exclude SCD in suspected primary care patients is timely. The authors further evaluated the added value of a faecal calprotectin point-of-care (POC) and a POC faecal immunochemical test for haemoglobin (FIT). They extended the use of data from a now rather old prospective diagnostic study in patients with persistent lower abdominal complaints referred for endoscopy. They concluded that FIT - and to a much lesser extent calprotectin - POC testing showed incremental value for SCD diagnosis beyond standard clinical information. A diagnostic strategy with routine clinical data and a POC FIT test may safely rule out SCD and prevent unnecessary endoscopy referral in approximately one third of SCD-suspected primary care patients.

This is an interesting study that adds to current knowledge on triage of patients with lower abdominal symptoms. It is reasonably well written and easy to follow. However, there are a number of major and minor flaws in the manuscript submitted. These follow in the order in which they appear in the manuscript. This list is not exhaustive.

The abstract should be more explicit. There are many faecal biomarkers. The title should read calprotectin and haemoglobin instead of faecal biomarkers.

According to this reviewers useful comments (also below) we have made several changes to the abstract. Furthermore, we changed the title according to the above suggestion.

Changes in the text in red (abstract):

Background: The majority of primary care patients referred for bowel endoscopy do not have significant colorectal disease (SCD), and are in hindsight unnecessarily exposed to a small but realistic risk of severe endoscopy-associated complications. We developed a diagnostic strategy to better exclude SCD in these patients and evaluated the value of adding a faecal calprotectin point-of-care (POC) and/or a POC faecal immunochemical test for haemoglobin (FIT) to routine clinical information.
Methods: We used data from a prospective diagnostic study in patients from 266 Dutch primary care practices with persistent lower abdominal complaints referred for endoscopy to develop a multivariable diagnostic model for SCD with routine clinical information, which we extended with faecal calprotectin POC (quantitatively in g/g faeces) and/or POC FIT results (qualitatively with a 6 g/g faeces detection limit). We defined SCD as colorectal cancer (CRC), inflammatory bowel disease, diverticulitis, or advanced adenoma (>1cm; AA), detected at endoscopy.

Results: Of 810 patients, 141 (17.4%) had SCD. A diagnostic model with routine clinical data discriminated between patients with and without SCD with an area under the receiver operating characteristic curve (AUC) of 0.741 (95%CI: 0.694-0.789). This AUC increased to 0.763 (95%CI: 0.718-0.809; p=0.078) when adding the calprotectin POC test, to 0.831 (95%CI: 0.791-0.872; p<0.001) when adding the POC FIT, and to 0.837 (95%CI: 0.798-0.876; p<0.001) upon combined extension. At a 5.0% SCD probability threshold for endoscopy referral, 30.4% of the patients tested negative based on this combined POC-tests extended model (95%CI: 25.7-35.3%), with 96.4% negative predictive value (95%CI: 93.1-98.2%) and 93.7% sensitivity (95%CI: 88.2-96.8%; missing one [stage 1] CRC, four diverticulitis, and four AA patients). Excluding the calprotectin POC test from this model still yielded 30.1% test negatives (95%CI: 24.7-35.6%) and 96.0% negative predictive value (95%CI: 92.6-97.9%), with 93.0% sensitivity (95%CI: 87.4-96.4%; missing one additional AA patient).

Conclusions: FIT and to a much lesser extent calprotectin POC testing showed incremental value for SCD diagnosis beyond standard clinical information. A diagnostic strategy with routine clinical data and a POC FIT test may safely rule out SCD and prevent unnecessary endoscopy referral in approximately one third of SCD-suspected primary care patients.

Changes in the text in red (title page):

Is there an added value of faecal calprotectin and haemoglobin in the diagnostic work-up for primary care patients suspected of significant colorectal disease? A cross-sectional diagnostic study

The internationally recommended and accepted term is faecal immunochemical test for haemoglobin (FIT) and not faecal immunochemical haemoglobin test (FIT).
We have changed this throughout the manuscript (changes are highlighted in red in the main documents).

In the Abstract and elsewhere, the term advanced adenoma (<1 cm) is used. This is an unusual and simplistic definition. Most would also tell that an advanced adenoma is present if there are three or more adenoma, if any adenoma had a villous characteristic and if there was a high degree of dysplasia. This requires explanation and elucidation.

We concur with this reviewer and reviewer 3 that classifying adenomas as advanced based on their size (>1cm) has limitations, although this definition has been used before. Adenomas one centimetre or smaller were all classified as non-SCD in our analyses, regardless of their histology. Accordingly, 81 adenomas were classified as non-SCD. Of these, we currently have unfortunately no information on histological features, and we can therefore not include small adenomas with a villous component, high-grade dysplasia or serrated polyps in our definition of SCD. Certainly, these lesions are clinically relevant and would best be diagnosed. Nevertheless, the number of such lesions now misclassified as non-SCD in our data is likely very small. Based on their prevalence in <1 cm adenomas (e.g. according to Lieberman et al, Gastroenterology 2008;135:1100-1105), this number will be about 2 to 3 (i.e. about 2% of all SCD cases in CEDAR). This possible misclassification is therefore unlikely to have affected the conclusions of our study.

Changes in the text in red (discussion section):

When defining SCD, we only included adenomas >1cm as AA, without taking histologic high-risk features such as the presence of high-grade dysplasia or villous components in smaller adenomas into account. However, such high-risk features are seldom present in small adenomas,35 and we estimate that about 2 to 3 of the small adenomas we have considered non-SCD are actually high-risk lesions. This amount of misclassification (i.e. only ~2% of all SCD cases in CEDAR) will likely not have importantly influenced the results.

In the Abstract and elsewhere in the paper, the authors use the term c-index: this is an unusual term for what is the Area Under the Curve (AUC) in RIOC analysis. The use of c-index and AUC is inconsistent in text and tables. I strongly recommend using the AUC only. Moreover, AUC should always be accompanied with the essential 95% CI.
We agree and have changed this throughout the manuscript (changes are highlighted in red in the main documents).

Keywords should be alphabetical and it is suggested that "point-of-care" be included.

This has also been amended accordingly.

Changes in the text in red (title page):

Keywords: calprotectin, diagnosis, faecal immunochemical test, point-of-care, primary care, significant colorectal disease

Very importantly, it is convention that all percentages are reported to one significant number after the decimal point. This must be done throughout this work. This also applies to 95%CI, of course.

As requested by the reviewer, we now report all percentages to one significant number after the decimal throughout the entire manuscript text, tables, figures, and supplements. We have done the same for the reported AUC estimates, except in the discussion for more easy comparison with the literature (where AUCs are often not reported with one significant number after the decimal point). These changes are highlighted in red in the main documents.

Introduction - Background: the data presented is termed Fraser's Rule of Sixths - https://scpnblog.wordpress.com/2015/06/12/investigating-bowel-symptoms-remember-the-rule-of-sixths/ 

The Frasers Rule of Sixths is indeed an interesting and useful interpretation of the literature regarding the yield of endoscopy referral from primary care in terms of the amount of serious colorectal disease that is detected (i.e. 1/6th). We have directly cited what is undoubtedly some of the literature that underlies this Rule of Sixth to inform the readers of the manuscript about the expected number of in hindsight - unnecessary endoscopy referrals from primary care.
Methods. This reviewer is very surprised that iron deficiency anaemia for which there is no explanation is not considered. This symptom is certainly very important in the assessment of suspected colorectal cancer - see NICE NG12 - https://www.nice.org.uk/guidance/ng12?unlid=9874814692015129115552

We very much agree with the reviewer that a low blood level of Hb is an important candidate predictor of significant colorectal disease in suspected primary care patients. We therefore did include Hb levels when developing our multivariable diagnostic model. We however modelled Hb levels continuously instead of using a pre-set threshold for anaemia, in order to preserve as much diagnostic information from this variable as possible. As explained in the methods section, we evaluated for all continuous candidate predictors whether a transformation was necessary to maintain linearity. When we evaluated the relation between Hb levels and SCD risk, we found that this relation was best described as a U-shape (by modelling it both with a linear and with a quadratic term: adding this quadratic term to a linear-term only model improved model fit with p=0.038). Without adjustment for overlapping information, this U-shaped relation looks like this (figure not provided in manuscript):

Thus in our data, low as well as high Hb levels are associated with increased SCD risk, information that would be lost when simply analysing the presence of anaemia. A similar U-shape relation between haematocrit and CRC risk has been described before by Brazer et al. 2 In our manuscript, this u-shaped relation between Hb and SCD risk was mentioned in the footnotes of Table 1 and Table S1 in the supplement.

For Hb to be included in our diagnostic model, it had to improve a model based on patient history and physical examination predictors with a p-value of <0.05 (we used a more stringent selection criterion for blood tests to limit patient burden as much as possible, see statistical analysis section). When adding Hb modelled as a U-shape to this patient history/physical examination model, it did not improve the model fit (p=0.23, see supplementary Table S1). In response to this reviewers question, we have reanalysed our data using the presence of anaemia instead of Hb as a continuous variable, but again, this did not improve model fit (p=0.27). These analyses indicate that in our data, Hb levels, either modelled continuously or dichotomously using a threshold for anaemia, do not add diagnostic information beyond the variables that were selected for the diagnostic model from patient history and physical examination (i.e. age, abdominal pain, rectal blood loss, rectal mucus, weight loss, change in bowel habit, abdominal bloating, constipation, and digital rectal examination).
We have made several changes in the manuscript addressing the above issues.

Changes in the text in red (methods section):

Subsequently, Hb and/or CRP were only selected if they significantly improved the patient history/physical examination model. We deliberately used a more stringent selection criterion for the blood analyses (p<0.05 instead of AIC-based) in view of the patient burden associated with obtaining this information. Blood Hb and CRP were modelled continuously instead of using a threshold for abnormal values (e.g. defining anaemia), to preserve as much diagnostic information as possible.

In all modelling, continuous predictors were included as such, using transformations if necessary to maintain linearity, while truncating outliers. Transformations were necessary for blood Hb (U-shape relation with SCD risk), and for duration of abdominal pain and CRP (logarithmic relations). See Supplementary File for further model development details.

Changes in the text in red (results section):

Nine of the 15 candidate predictors from patient history and physical examination were selected for the basic diagnostic model, to which blood Hb did not significantly contribute (p=0.23) but CRP did (p=0.03; see Table 2 for specification of the basic diagnostic model).

History taking and examination: do not use the term stools, please: use faeces and fecal throughout. Moreover, there is a measure of faecal consistency - the Bristol Scale of Hardness: this might have been applied, with advantage.

We have changed the term stools into faeces throughout the manuscript. These changes are highlighted in red in the main documents. The Bristol Scale of Hardness is indeed an interesting and potentially valuable way to assess faecal consistency, that unfortunately not available in CEDAR. We hope this reviewer can agree with the way we defined constipation and diarrhoea.

Blood and faecal SCD markers: this section is inadequate. This reviewer recognises that there are details in the paper of Kok et al cited, but more is required here to make this paper stand alone. While it may not be necessary to do a full STARD guideline explanation - see Bossuyt PM, et al.
STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015 Oct 28;351:h5527, the essential items should be included. The C50 (threshold) for the POC-FIT MUST be given in this paper in µg Hb/g faeces.

This certainly is a valid remark, and we have described the assessment of the faecal markers in more detail in order to make this paper stand-alone from our previous publication.

Changes in the text in red (methods section):

A pre-endoscopy venous blood sample was drawn to estimate Hb and C-reactive protein (CRP) concentrations according to routine clinical practice. Directly following study inclusion, patients provided faeces samples collected before bowel preparation for endoscopy in a plain blue-capped faecal container, and kept refrigerated (4°C) for a maximum of two days before handing in. Study protocol allowed freezing (−20°C) of faecal samples before processing (this occurred in 67.9% of samples; median days between collection and processing: 10; 10th-90th percentile: 4-21). If not frozen, the refrigerated faecal samples needed to be processed for calprotectin testing within six days (adherence 96.3%; median days: 2; 10th-90th percentile: 0-3), and needed to be tested for Hb within three days of collection (adherence 94.5%; median days: 2; 10th-90th percentile: 0-3).

We analysed the faecal samples for calprotectin concentration by a quantitative POC test (Quantum Blue®; dynamic range 30-300 g/g) and by an enzyme-linked immunosorbent assay (ELISA; EK-CAL Calprotectin ELISA, both from Bühlmann Laboratories), both yielding estimates of g calprotectin/g faeces, and for faecal Hb by a qualitative POC FIT (Clearview® iFOBT One Step Faecal Occult Blood Test Device, Alere Health), yielding either a positive or negative test result (lower detection limit of 6 g/g). Laboratory technicians performed the ELISA, and trained research nurses the POC tests, blinded for clinical information and according to the manufacturers instructions. Briefly, for the calprotectin assays 80 mg homogenized faeces was centrifuged and the supernatant was tested for calprotectin (1:16 diluted for the POC test and undiluted for the ELISA; supernatant for the ELISA was stored at -20°C for maximally four months before analysis); for the POC FIT three separate random areas of the faecal sample were stabbed by the specimen collection stick and transferred to the collection tube, and two drops of extracted specimen were then applied to the test device. For more details see Kok et al.11
Faecal haemoglobin is very unstable in native faeces. The authors tell that the patients brought refrigerated faecal samples. The time difference between faecal collection and analysis MUST be documented and evidence for stability presented.

We thank the reviewer for this important remark that is well taken. We have expanded the method section about the assessment of the faecal biomarkers, which now includes the key (time) aspects regarding the handling of the faecal samples. 67.9% of samples were frozen before analysis, and these samples were processed for testing after a median of 10 days after collection (90% within 21 days). Samples that were not frozen were processed after a median of 2 days after collection (90% within 3 days; see our response to this reviewers previous question for the corresponding changes in the manuscript).

We have now also investigated the relation between freezing, the number of days between faecal sample collection and analysis, and the test results of the POC FIT. Below figure (not included in the manuscript) shows the proportion of POC FIT positive test results according to the number of days between collection and analysis, and whether the faecal samples had been frozen before analysis or not (error-bars are 95% confidence intervals). We truncated outliers in time between collection and testing within groups of freezing at the top 10% (see x-axis labels for respective values for frozen (blue) and non-frozen (red)).

The red and blue lines show the predicted probability of a positive POC FIT according to freezing status and time (derived from a logistic regression model with the POC FIT test result as dependent variable and the number of days between collection and testing (continuously) and whether a sample was frozen or not (dichotomously) as main effects an interaction term between freezing and time was not significant). With increasing time between collection and testing, the chance of a positive POC FIT result decreased, albeit non-significantly (p=0.19). Frozen samples were more likely to be POC-FIT test-negative than non-frozen samples (p=0.017).

Both for calprotectin POC and the ELISA test results (threshold for positivity >50 g/g), we did not find a significant relation between freezing of samples or storage time, as shown in the next two plots (not included in the manuscript); p-values for calprotectin test results in relation to freezing status, days of storage, or their interaction were all >0.18).
So, as expected from the literature, we indeed found evidence of faecal Hb instability in our data resulting in slightly less positive POC FIT test results with increasing number of days between collection and testing, and, more profoundly, between frozen and non-frozen samples. The calprotectin test results seemed rather stable to these sample-processing issues.

This decrease in POC FIT test positivity due to sample-processing issues likely has led to a lower sensitivity for SCD, and perhaps a higher specificity. To investigate the impact of freezing on the diagnostic performance of the POC FIT, we added a [POC FIT*freezing status] interaction term to the combined POC-test and the POC FIT only extended models. The results of these analyses showed that freezing status did not significantly modify the diagnostic performance of the POC FIT in these models (p-values for interaction were 0.48 and 0.38 respectively).

In conclusion, we did find evidence of faecal Hb instability in our data, but not for faecal calprotectin. This Hb instability was largely due to freezing and may have led to a lower SCD sensitivity of the POC FIT. We could however not confirm that the diagnostic odds ratio for SCD of the POC FIT was different in patients whose faecal samples were frozen compared to those patients whose samples were not frozen. We have addressed this issue in the discussion section.

Changes in the text in red (discussion section):

Second, patients collected faecal samples in regular blue-capped containers without Hb stabilizing buffer (so each patient needed to fill only one faecal container for both calprotectin and Hb analysis). Samples were kept refrigerated, and if not frozen before further processing 90% was tested within 3 days of collection. Additional data-analysis showed that the chance of a positive POC FIT slightly decreased with increasing time between collection and testing (0.3% absolute decrease per day; p=0.19), and that frozen samples were more likely to be POC FIT negative than non-frozen samples (absolute 8.6% decrease in POC FIT positivity; p=0.017; calprotectin results seemed not to be affected). Some patients have thus likely tested falsely negative for the POC FIT because of Hb degradation in our study. However, in none of the models with POC FIT its odds ratio for SCD significantly differed in patients whose faecal samples were and were not frozen. Furthermore, the POC FIT performed well in our study despite these limitations, and the sensitivity and discriminatory performance of faecal Hb testing in primary care will thus likely be even better when using Hb stabilizing buffers in faecal sample collection devices and using a quantitative FIT.
I am sure blood was not analysed for CRP: surely it was serum separated from the blood collected. Thus, blood tests is slang and must not be used as a term.

Indeed, CRP was analysed in serum as per routine clinical practice. We have changed the term blood test in blood analysis throughout our revised manuscript. These changes are highlighted in red in the main documents.

The authors have CRP and haemoglobin: they also have haemoglobin and CRP. The order of variables must be consistent throughout text and tables.

According to this reviewers request we now consistently addressed Hb and CRP throughout the manuscript in that order. These changes are highlighted in red in the main documents.

Diagnostic outcomes (and later sections); the authors have defined CRC as the abbreviation for colorectal cancer. Once an abbreviation is defined, it should be used ubiquitously. Moreover, if the authors had defined an abbreviation (AA) for advanced adenoma, and used this throughout, the paper would be easier to follow perhaps. Similarly, IBD could be defined at first mention of inflammatory bowel disease and used throughout and not randomly as in this manuscript.

We have changed the use of abbreviations (also including Hb and CRP) according to these suggestions. These changes are highlighted in red in the main documents.

Statistical analyses: levels are not measured for CRP and Hb - concentrations are estimated.

We have changed the wording accordingly.

Changes in the text in red (methods section):

(...) we first developed a basic diagnostic model for SCD considering 15 patient history and physical examination predictors (listed in Table 1), and simple blood analyses (Hb and CRP concentrations).
The paper is inconsistent with respect to the order of calprotectin and faecal haemoglobin: the order in which the material on these measurands is presented should be consistent.

We have changed the order in presenting calprotectin first and then faecal haemoglobin throughout the manuscript. These changes are highlighted in red in the main documents.

Statistical analysis: as above, please eliminate c-index and use the better understood AUC throughout.

We agree that the term c-index is less established than AUC, and have therefore, as requested by this reviewer, now used AUC throughout the manuscript. These changes are highlighted in red in the main documents.

The authors have, page 6, ..different SCD probability thresholds: 2.5%, 5% and 7.5%). And this should read 2.5, 5.0 and 7.5 since the number of significant figures must be consistent - 5 is not the same as 5.0!

This was changed throughout the manuscript. These changes are highlighted in red in the main documents.

Importantly, the Discussion section is lacking with regard to the literature on the use of FIT in the assessment of patients presenting in primary care with lower abdominal symptoms. The study by Mowat et al examined FUT and calprotectin in this clinical setting and is cited (although the reference given is simply inadequate). However, the following papers must be discussed and cited. Moreover, the authors think that a diagnostic strategy with routine clinical data and a POC FIT test may safely rule out SCD - can they discuss why clinical data are required since quantitative estimates of faecal haemoglobin concentration seem to work just as well, if not better.

Godber IG, Todd LM, Fraser CG, MacDonald LR, Ben Younes H. Use of a faecal immunochemical test for haemoglobin can aid in the investigation of patients with lower abdominal symptoms. Clinical Chemistry and Laboratory Medicine 2016;54:595-602.


We have added above references to the manuscript. This reviewers question about why to include clinical data when evaluating the POC FIT is a question we often encounter. In the discussion we have therefore addressed why we think that evaluating the POC FIT (and the calprotectin tests) requires the use of clinical data in adjunct to the POC FIT test.

Changes in the text in red (discussion section):

Other studies have also advocated quantitative faecal Hb testing for ruling out SCD,30,31 or advanced neoplasia,32-34 in symptomatic patients.

Previous results suggest that using a single test could in fact be sufficient in deciding whom to refer for endoscopy. Indeed, our results also underscore that a positive POC FIT already implies the need for referral by itself (at the 5.0% SCD probability threshold; see nomogram in Supplementary File). Here the clinical data do not add much, but they do when the POC FIT returns negative. Also, in daily clinical practice, and certainly in primary care, it is rare that except in a screening situation physicians would immediately apply such test in suspected patients presenting with symptoms and signs of SCD without even considering any other pre-test diagnostic information from history taking and physical examination. The diagnostic process ion
primary care is sequential, starting with history taking and physical examination and follow-up testing only in case the first provide indications that legitimates additional testing. To adhere as much as possible to primary care practice, we therefore explicitly evaluated first the diagnostic value of history taking, physical examination, and simple blood testing, and subsequently the added value of the POC FIT test, rather than the other way around. Obviously, in unsuspected people, in the realm of screening, a single-test approach using first and foremost the POC FIT test, seems a very reasonable approach, but in our view not for diagnostic work-up of clinically suspected patients which was the focus of this paper.

Discussion: limitations of the study must include the fact that the fecal samples for FIT were not collected into specimen collection devices provided by the manufacturer that ensure Hb stability. Moreover, the authors should address the limitations of POC FIT as compared to high quality automated quantitative estimates using immunoturbidimetry.

Indeed, as this reviewer justly points out, patients collected the faecal samples in standard blue-capped faecal containers without Hb stabilizing buffer (we added this information to the methods section, see above). The protocol however required the samples to be refrigerated and handed in for processing within maximally two days from collection. For patients, this procedure was more convenient as it required the use of only one faecal specimen container. For primary care physicians, the use of a POC FIT has advantages as they can be used in their offices. But these procedures certainly have limitations, which we now have addressed in the discussion.

Changes in the text in red (discussion section):

Finally, the use of a qualitative POC FIT in the way that we did in this study, although easily implemented in primary care, also has limitations. First, as the qualitative POC FIT yields a positive or a negative test result (with a detection limit of 6 g Hb/g faeces), the diagnostic information that would be available by quantitatively assessing the amount of Hb present in faeces is lost. Second, patients collected faecal samples in regular blue-capped containers without Hb stabilizing buffer (so each patient needed to fill only one faecal container for both calprotectin and Hb analysis). Samples were kept refrigerated, and if not frozen before further processing 90% was tested within 3 days of collection. Additional data-analysis showed that the chance of a positive POC FIT slightly decreased with increasing time between collection and testing (0.3% absolute decrease per day; p=0.19), and that frozen samples were more likely to be POC FIT negative than non-frozen samples (absolute 8.6% decrease in POC FIT positivity;
p=0.017; calprotectin results seemed not to be affected). Some patients have thus likely tested falsely negative for the POC FIT because of Hb degradation in our study. However, in none of the models with POC FIT its odds ratio for SCD significantly differed in patients whose faecal samples were and were not frozen. Furthermore, the POC FIT performed well in our study despite these limitations, and the sensitivity and discriminatory performance of faecal Hb testing in primary care will thus likely be even better when using Hb stabilizing buffers in faecal sample collection devices and using a quantitative FIT.

Table 1: Accuracy as a heading is unsatisfactory since this word has many meanings: it should be Diagnostic Accuracy. Please get rid of c-index. Percentages here must be to one figure after the decimal point. As documented above, the authors have measured blood Hb and defined anaemia, so why is this not included as an indicator of SCD? The threshold for the POC FIT must be given here.

We have changed Table 1 according to these comments. These changes are highlighted in red in the main documents. With regard to Hb and anaemia, see our response to that particular remark above.

Table 2: NRO and IDI should be defined in the legend to the Table since tables should stand alone. Use AUC and include 95%CI for all AUC.

We also have changed Table 2 according to these comments. These changes are highlighted in red in the main documents.

Table 3: Use 5.0, as above and not 5. The abbreviations used - CRC, IBD, Div. and Aden. should either be spelled out or the abbreviation defined in the legend to the Table.

Again, changes have been made accordingly. These changes are highlighted in red in the main documents.
Table 4 include OR in legend.

This is included in the legend of Table 4.

Figure 2 - the material on FIT should come before the material on calprotectin - this applies to the legend embedded in the Figure as well. Why not get rid of c-index throughout, as recommended above and simply use AUC to enhance readability and comprehension.

For the sake of consistency, we did not change the order of FIT and calprotectin in this figure, as we have now consistently addressed calprotectin before FIT throughout the manuscript (see above).

Reviewer 3

The Authors evaluated new diagnostic strategies in patients with persistent suspected symptoms within a primary care setting in order to avoid unnecessary endoscopic examinations.

The paper is a high quality level one and worth certainly to be published.

I have only minor remarks:

* The Authors defined advanced adenoma as adenoma larger than 10 mm. Usually the definition of these lesions includes also histological parameters. The Authors should explain why they used this definition.

We thank this reviewer for this valid point that indeed needs more explanation. We concur with this reviewer and reviewer 2 that classifying adenomas as advanced based on their size (>1cm) has limitations, although this definition has been used before. Adenomas one centimetre or smaller were all classified as non-SCD in our analyses, regardless of their histology. Accordingly, 81 adenomas were classified as non-SCD. Of these, we currently have unfortunately no information on histological features, and we can therefore not include small
adenomas with a villous component, high-grade dysplasia or serrated polyps in our definition of SCD. Certainly, these lesions are clinically relevant and would best be diagnosed. Nevertheless, the number of such lesions now misclassified as non-SCD in our data is likely very small. Based on their prevalence in <1 cm adenomas (e.g. according to Lieberman et al, Gastroenterology 2008;135:1100-1105), this number will be about 2 to 3 (i.e. about 2% of all SCD cases in CEDAR). This possible misclassification is therefore unlikely to have affected the conclusions of our study.

Changes in the text in red (discussion section):

When defining SCD, we only included adenomas >1cm as AA, without taking histologic high-risk features such as the presence of high-grade dysplasia or villous components in smaller adenomas into account. However, such high-risk features are seldom present in small adenomas,35 and we estimate that about 2 to 3 of the small adenomas we have considered non-SCD are actually high-risk lesions. This amount of misclassification (i.e. only ~2% of all SCD cases in CEDAR) will likely not have importantly influenced the results.

* The Authors have used the diagnostic accuracy of point-of-care (POC) calprotectin and faecal immunochemical haemoglobin tests, giving for more information the reference n. 14. They should explain better this new concept.

Indeed, the method section about the faecal calprotectin and Hb tests was too brief in the initially submitted version for this manuscript for it to stand alone. We therefore have added more detail to this section.

Changes in the text in red (methods section):

A pre-endoscopy venous blood sample was drawn to estimate Hb and C-reactive protein (CRP) concentrations according to routine clinical practice. Directly following study inclusion, patients provided faeces samples collected before bowel preparation for endoscopy in a plain blue-capped faecal container, and kept refrigerated (4°C) for a maximum of two days before handing in. Study protocol allowed freezing (-20°C) of faecal samples before processing (this occurred in 67.9% of samples; median days between collection and processing: 10; 10th-90th percentile: 4-21). If not frozen, the refrigerated faecal samples needed to be processed for calprotectin testing within six days (adherence 96.3%; median days: 2; 10th-90th percentile: 0-3), and needed to be
tested for Hb within three days of collection (adherence 94.5%; median days: 2; 10th-90th percentile: 0-3).

We analysed the faecal samples for calprotectin concentration by a quantitative POC test (Quantum Blue®; dynamic range 30-300 g/g) and by an enzyme-linked immunosorbent assay (ELISA; EK-CAL Calprotectin ELISA, both from Bühlmann Laboratories), both yielding estimates of g calprotectin/g faeces, and for faecal Hb by a qualitative POC FIT (Clearview® iFOBT One Step Faecal Occult Blood Test Device, Alere Health), yielding either a positive or negative test result (lower detection limit of 6 g/g). Laboratory technicians performed the ELISA, and trained research nurses the POC tests, blinded for clinical information and according to the manufacturers instructions. Briefly, for the calprotectin assays 80 mg homogenized faeces was centrifuged and the supernatant was tested for calprotectin (1:16 diluted for the POC test and undiluted for the ELISA; supernatant for the ELISA was stored at -20°C for maximally four months before analysis); for the POC FIT three separate random areas of the faecal sample were stabbed by the specimen collection stick and transferred to the collection tube, and two drops of extracted specimen were then applied to the test device. For more details see Kok et al.11

* The type of FIT used in the study is a qualitative one, but no information were available about cut off of sensitivity.

The Clearview® iFOBT One Step Occult Blood test is a rapid chromatographic immunoassay for qualitative detection of human faecal haemoglobin and its immediate degradation products. The lower detection limit as stated by the manufacturer is 6 g haemoglobin/g faeces. We have included this information in the tables, the method section (see previous answer), and in the discussion. These changes are highlighted in red in the main documents.

Editorial Requests

Please note that all submissions to BMC Medicine must comply with our editorial policies. Please read the following information and revise your manuscript as necessary. If your manuscript does not adhere to our editorial requirements this will cause a delay whilst the issue is addressed. Failure to adhere to our policies may result in rejection of your manuscript.
Ethics:

If your study involves humans, human data or animals, then your article should contain an ethics statement which includes the name of the committee that approved your study.

If ethics was not required for your study, then this should be clearly stated and a rationale provided.

Consent:

If your article is a prospective study involving human participants then your article should include a statement detailing consent for participation.

If individual clinical data is presented in your article, then you must clarify whether consent for publication of these data was obtained.

Availability of supporting data:

BioMed Central strongly encourages all data sets on which the conclusions of the paper rely be either deposited in publicly available repositories (where available and appropriate) or presented in the main papers or additional supporting files, in machine-readable format whenever possible. Authors must include an Availability of Data and Materials section in their article detailing where the data supporting their findings can be found. The Accession Numbers of any nucleic acid sequences, protein sequences or atomic coordinates cited in the manuscript must be provided and include the corresponding database name.

Authors Contributions:

Your 'Authors Contributions' section must detail the individual contribution for each individual author listed on your manuscript.

We have addressed all above editorial requests.