Protocol for assessment of faecal calprotectin.

8th October 2013

HTA reference number 2012/48

Title: faecal calprotectin in the differential diagnosis of chronic bowel disease.

Aim: to review the clinical accuracy and cost-effectiveness of faecal calprotectin testing for distinguishing between inflammatory and non-inflammatory bowel disease in people with chronic lower GI symptoms.

NB. This protocol may evolve in the course of the assessment.

Assessment group – Warwick Evidence

Project lead: Norman Waugh

Professor of public health medicine and health technology assessment

Warwick Evidence

Division of Health Sciences

Warwick Medical School

Coventry CV4 7AL

norman.waugh@warwick.ac.uk 02476 151585

Project team;

Dr Ewen Cummins, health economist

Dr Pamela Royle, senior research fellow

Dr Deepson Shyangdan, research fellow

Dr Ramesh Arasaradnam, consultant gastroenterologist, UHCW and associate professor of gastroenterology, Warwick Medical School

Dr Ngianga-Bakwin Kandala, principal research fellow

Introduction

Chronic abdominal pain or discomfort, accompanied by diarrhoea or constipation, is common and the symptoms can be due to a number of different conditions, some more serious than others.

The conditions include irritable bowel syndrome (IBS) and inflammatory bowel disease (IBD). The commonest forms of the latter are ulcerative colitis and Crohn's disease, which make up about 90% (to be checked) of IBD.

Irritable bowel syndrome

The most common symptoms of IBS include recurrent colicky abdominal pain or cramping felt in the lower abdomen and relieved by defecation. There may be abdominal distension (bloating) and altered bowel habit – episodes of diarrhoea and constipation. Features supporting a diagnosis of IBS: symptoms >6 months, associated with other, non-GI problems, stress worsens symptoms. IBS is very common – perhaps 15% of the UK population, though many people who have it never consult their GPs about it. It is commonest in young women. The underlying mechanism is an alteration in the functioning of the muscle in the wall of the bowel. People who have it are constitutionally well and do not lose weight. It is a troublesome but not serious condition, in the sense that it does not lead to serious adverse events.

The cause of IBS is not known in most people, but it sometimes follows an episode of infectious gastroenteritis ("food poisoning").

Inflammatory bowel disease

Ulcerative colitis is characterised by inflammation of the colon, sometimes intense, with bloody diarrhoea, but often much milder.

Crohn's disease can present in different ways. It is also called "regional enteritis" but this is somewhat misleading because Crohn's disease can affect any part of the GI tract.

Both UC and Crohn's can cause autoimmune disorders in other parts of the body, including the eye (uveitis), the joints (arthritis), the skin (erythema nodosum) and the bile ducts (sclerosing cholangitis). The onset of Crohn's can be less obvious than that of UC. In children the first sign may be failure to grow.

Colorectal cancer may also cause inflammation.

The key point to note is that distinguishing amongst inflammatory and non-inflammatory diseases by purely clinical means – signs and symptoms – can be difficult. So many patients are referred to gastroenterology.

NICE Clinical Guideline 61 'Irritable Bowel Syndrome' recommends that people presenting with abdominal pain or discomfort, bloating or change in bowel habit for at least six months should be asked if they have any red flag indicators such as unexplained weight loss. They should also be clinically tested for red flag indicators including anaemia, rectal masses, inflammatory biomarkers for IBD (FC is not specifically mentioned) and late onset (>60 years) change in bowel habits. Presence of any of these indicators should result in a referral to secondary care for further investigation.

Therefore, patients presenting with symptoms/test results indicative of IBD are referred to secondary care for specialist investigation (most likely to a gastroenterology clinic).

If there are no red flag indicators to cause concern, the guideline states that patients who meet the IBS diagnostic criteria should receive the following laboratory tests to exclude other diagnoses:
- Full blood count (FBC)

- Erythrocyte sedimentation rate (ESR) or plasma viscosity

- C-reactive protein (CRP)

- Antibody testing for coeliac disease (endomysial antibodies [EMA] or tissue transglutaminase [TTG]).

Calprotectin

Calprotectin is a compound released by white blood cells. In people with bowel conditions that cause inflammation, the increased number of white blood cells in the bowel leads to an increase in faecal calprotectin (FC). There are now tests to detect or measure the level of calprotectin in faeces.

The proposed role of FC testing is, in people with lower gastrointestinal symptoms (pain, bloating, diarrhoea, change in bowel habit), to distinguish between those with inflammatory conditions and those with no inflammation. Many of those with inflammation will have IBD, but others may have cancer or other conditions. Most of those with no inflammation will have IBS.

Knowledge of the presence or absence of inflammation may affect the decision on referral for further investigation. The absence of inflammation may lead to a presumption of IBS, to be managed in primary care. The presence of inflammation would be likely to trigger referral to gastroenterology for further investigation, likely to include colonoscopy or sigmoidoscopy.

If calprotectin testing is cost-effective, the likely effect from testing being made more widely available would that it would become part of the primary care work-up pre-referral. So the main focus of this appraisal is expected to be on use of FC testing by primary care staff. However use of FC testing in secondary care will also be considered.

Hence there could be two benefits. Those with IBS would not be referred and might therefore escape further investigations especially colonoscopy. Those with inflammation might be referred and diagnosed sooner and receive appropriate treatment earlier.

Decision problem

As stated in the scope for this appraisal, the objective of the evaluation is to assess the clinical and cost-effectiveness of faecal calprotectin testing in distinguishing inflammatory from non-inflammatory diseases of the bowel. Scoping workshop feedback suggests that the following questions should be taken into account in guiding this evaluation:

- Is an FC test result a reliable way of identifying inflammation of the bowel?
- How do the different cut-off values used to interpret the results of quantitative FC tests affect their cost-effectiveness? What are the optimal cut-offs?
- What is the cost-effectiveness of the point-of-care tests? How does this compare to the fully quantitative FC tests?
- How will the performance of FC tests be affected when used in primary care, given the paucity of data on the use of these tests in primary care?
- How does performance of FC testing vary amongst primary and secondary care groups?

Methods

Population

Patients with lower gastrointestinal symptoms that are chronic, defined as persisting for at least 6- 8 weeks. All ages will be included. At the scoping workshop it was felt that a lower age of 12 years might be used, but preliminary investigation by the Warwick Evidence team suggests that studies in children and adolescents do not report results separately for the under 12s and over 12s. So we will have no lower age limit. The scope suggests an upper age limit of 60.

In adults, symptoms include abdominal pain or discomfort, bloating, diarrhoea or constipation.

The main focus will be those presenting in primary care but studies of hospital groups will also be included.

Paediatric and adult populations will be analysed separately.

Patients with red flag symptoms (as listed above) will be excluded since they should be referred without delay because such symptoms may be due to cancer.

Intervention

Faecal calprotectin tests. These are of two types;

- Laboratory testing, mostly using ELISA methods.
- Point of care testing (POCT), which can be used in primary care or in laboratories.

Lab methods are quantitative. POCT tests may be quantitative or semi-quantitative.

Some POCT testing may be used in smaller laboratories where throughput does not justify ELISA equipment

The scope envisages that the lab-based calprotectin tests can be treated as a group. We wil seek expert advice from Biochemistry on this. We may provide a narrative description of these tests in an appendix. We note that differences in extraction buffers might be important.

As the scope reports, cut-offs for FC may be a single point, such as 50 μg/g, so that values below indicate no inflammation and values equal to and above indicate inflammation is present, or multiple cut-offs may be used, with results classed as;

- no inflammation
- indeterminate result (likely resulting in the individual being re-tested at a later date)
- inflammation confirmed.

The scope cites anecdotal evidence suggesting that as many as 85 – 90% of individuals investigated using an FC test in a gastroenterology clinic will have an FC level of less than 50 μg/g (no inflammation). Of the remaining 10 – 15%, half will have an indeterminate result (50 – 200 μg/g). Some clinics use 50-100 μg/g and one study found that most of this group had no abnormal findings, so there may be a case for 100 being the cut-off.

The review will seek to determine the best cut-offs. However, it should be noted that decisions will not be made only on calprotectin levels, but on the whole clinical picture. This raises the question of whether there should be different cut-offs for different patient groups according to symptoms.

One question will be the role of POC testing. Our starting assumption is that a definitely negative POCT need not be checked by a quantitative lab method, but that borderline and positive ones will be re-tested by a lab method. The scope envisages repeat testing after borderline results. After positive

testing and referral to gastroenterology, we will assume that repeat testing by quantitative method (ELISA) will be done, partly as a baseline for future monitoring.

Comparators

In primary care, GPs suspecting inflammation can use ESR and CRP, which can indicate inflammation but not localise it. If GPs have access to faecal calprotectin testing, they would use that in people with suspected IBS. So FC would replace ESR and CRP testing.

However it might be more useful to compare pathways of care. A set of possible pathways is shown in appendix 2 in which the options include;

- No FC testing available. Clinical assessment and simple tests in primary care followed by decision on referral or symptomatic treatment/ therapeutic trial
- Lab testing available to GPs. Lab provides result.
- "Lab plus" where GP provides clinical details along with test request and gastroenterologist or clinical biochemist provides commentary and advice
- POCT available in primary care.

Outcomes

Depending on data availability, these may include;

- Referral rates
- Numbers of colonoscopies with/without FC testing
- Proportion of colonoscopies with no abnormal findings
- Duration from onset of symptoms
- Costs
- Adverse events such as complications of colonoscopy, late presentation of Crohn's disease
- Quality of life
- QALYs

Acceptance of the test will not be universal, and may vary amongst primary and secondary care – i.e some patients might decline to produce a sample of faeces for their GP, but might possibly for a gastroenterologist if the alternative is colonoscopy.

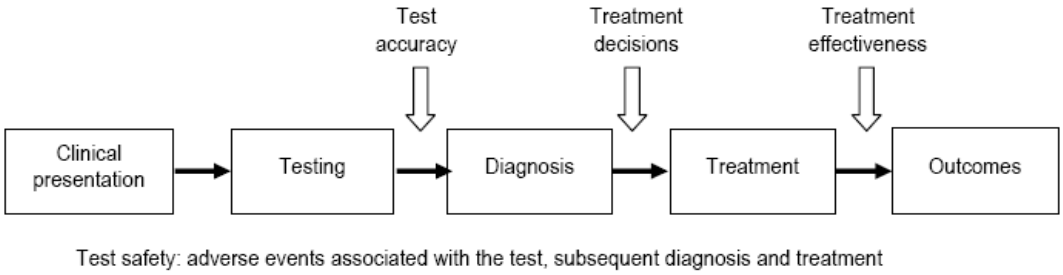
Methods

General approach

A framework of six stages has been used to describe the process of evaluating a diagnostic technology (Fryback and Thornbury, 1991):

1. Technical quality of test information – feasibility and optimisation.
2. Diagnostic accuracy
3. Diagnostic thinking impact – change in referring physician's diagnosis.
4. Therapeutic choice impact – change in patient management plan
5. Patient outcome impact
6. Societal impact – change in costs and benefits

Figure 1: Determinants of the clinical effectiveness of a diagnostic technologies (Medical Services Advisory Committee (MSAC), 2005)



We will use a similar approach. The key finding will not be whether the tests reliably measure faecal calprotectin, but whether FC testing improves patient outcomes.

Searches.

Our starting point will be the previous review by the Centre for Evidence-based Purchasing. This review will update that.

We will search MEDLINE, Embase, SCI and all sections of the Cochrane Library, for systematic reviews (including any previous health technology assessments) and primary studies.

The search strategy below will be used for Medline and adapted as appropriate for other databases. Searches will be not restricted to English language, in order to provide an impression of the total volume of literature. Some studies not in English may be translated if they look particularly useful, and if translation is available, but most will not.

1. exp Inflammatory Bowel Diseases/di [Diagnosis]

2. exp Irritable Bowel Syndrome/di [Diagnosis]
3. crohn's disease.tw.
4. ulcerative colitis.tw.
5. inflammatory bowel disease*.tw.
6. irritable bowel syndrome*.tw.
7. (IBS or IBD).tw.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. calprotectin.tw.
10. 8 and 9
11. limit 10 to english language

Searches will retrieve studies reporting for adverse events, especially associated with colonoscopy. We will seek data specific to diagnostic colonoscopy since rates are higher after therapeutic colonoscopy.

Selection of studies

Inclusion criteria: studies comparing faecal calprotectin as a guide to inflammation of the lower intestine, ideally with histology as the reference test. Initial searches reveal two problems with this. Firstly, some studies give numbers of patients with CD and UC, but do not give details on whether this is based on biopsy and histology. Secondly, some studies report colonoscopy as normal without further data on whether biopsy is done for e.g. microscopic colitis.

We will also seek follow-up studies of patients diagnosed as IBS.

Since the aim of the appraisal is to assess the usefulness of calprotectin for distinguishing between inflammatory and non-inflammatory bowel disease (in practice mostly between IBS and IBD), to be included studies should have;

- A mixed group of patients with symptoms but not yet diagnosed, and ideally a mix reflecting case mix in primary care. It is assumed that coeliac disease has been excluded by TTG testing.
- Calprotectin testing before, or blinded to the results of, endoscopy
- Endoscopy (usually colonoscopy but sigmoidoscopy only studies would not be ruled out in UC) for all patients, with the endoscopist blinded to the results of calprotectin testing
- the reference test of histology of biopsies taken during endoscopy

Hierarchies of evidence

1. The best evidence will come from studies in which FC is carried out in patients with symptoms of recent onset, lasting for at least 6 weeks, where the diagnosis is uncertain, and where colonoscopy is performed to provide a definitive histological diagnosis. We note from preliminary searches that some studies report colonoscopy but do not mention whether biopsies were obtained for histology. Some of these studies give details of UC and CD, and it is likely that histology was available but is simply not mentioned. Depending on numbers of studies, we may carry out a sensitivity analyses with and without studies with no mention of histology.
2. If at colonoscopy, the bowel appears normal, biopsies may still be taken, to check for microscopic colitis. However there may be cases where endoscopy is negative because in about 10% of cases of CD, it is limited to those parts of the small bowel that cannot be reached by colonoscopy or gastroscopy. There will also be cases where endoscopy is deemed to be too invasive. We will accept the following as proof of IBD;
    - Wireless capsule endoscopy with score indicating mild or worse activity (score >134)
    - Small bowel capsule biopsy
    - Radiological evidence of thickening of the bowel wall
    - Ultrasound evidence of thickening of the bowel wall
3. Some studies are simply series of patients with known IBS or IBD with FC test results but no recent endoscopy, or no endoscopy at all if IBS has been diagnosed on purely clinical grounds. We may use these as guides to thresholds in symptomatic cases of recent onset (within 6 months). We will exclude long-standing (over 12 months) cases of IBD and patients with IBD in remission.
4. Some studies report FC results in patients that have had multiple investigations without a definitive diagnosis. This could cause a problem of spectrum bias which is likely to mean that the patients are not representative of those with symptoms of recent onset presenting in primary care. Any such studies will be used only for assessing the value of FC testing in specialist care, or in sensitivity analysis
5. If data permit, we may carry out a sensitivity analysis using only studies that have more than 50% of patients with non-inflammatory conditions, as a guide to NPV and negative LR of calprotectin in primary care. (For adults and children separately.)

Exclusions

- Studies of faecal calprotectin for monitoring activity, or response to treatment, in people with known IBD.
- Patients with IBD in remission will be excluded by absence of symptoms.

- Studies of serum calprotectin.
- Short duration of symptoms (< 6 weeks).
- Patients with symptoms following an acute infectious episode, lasting for less than 3 months.
- Patients over 60
- Studies with more than 3 months interval between FC and colonoscopy.
- Studies where it is not clear whether symptoms are of recent onset.
- Patients taking NSAIDS or any other drug likely to results in raised FC levels. Low dose (75mg) aspirin will be allowed.
- Studies of patients with mix of long and short duration of symptoms may be useful if the majority (70-80%) are of short duration, or if the short duration group is reported separately. We may consider a sensitivity analysis including/excluding studies.

Where possible, data will be extracted from diagnostic studies for 2x2 tables, with FC as screening test and bowel histology as the reference test.

If data for 2x2 tables are not available, we will report what screening parameters are provided in studies.

We will rely mainly on studies published in full but may use evidence available in other forms for some purposes, such as identifying emerging research.

Assessment of methodological quality

We will use the QUADAS tool (see appendix 2), possibly modified. (www.bris.ac.uk/quadas/quadas-2)

Data collection, analysis and synthesis

We will use Review Manager, which now has a section for diagnostic reviews, and can generate coupled forest plots and ROC curves. We will also use MedCalc for producing figures. RevMan cannot do all the statistical analysis that is likely to be required and the statistical software package Stata will be used for more complex analysis.

If the main value of calprotectin testing is to rule out conditions causing inflammation, the key parameters will be negative predictive value (NPV) and negative likelihood ratio . Note that more than one test may be used, so if an initial test was negative but symptoms suggestive of IBD continued despite treatment for IBS, it could be repeated.

Results will need to take account of country of origin since the prevalences of CD and UC vary.

Heterogeneity will initially be examined by visual inspection of coupled forest plots of sensitivity and specificity using the reference standard of endoscopy, ideally with histology.

More variability among diagnostic accuracy study results is to be expected than with randomized trials. Some of this variability is due to chance, as many diagnostic studies have small sample sizes. The remaining heterogeneity may be due to differences in study populations, but differences in study methods are also likely to result in differences in accuracy estimates. Test accuracy studies with design deficiencies can produce biased results.

As recommended in Leeflang et al 2009, we will investigate and identify potential sources of bias and to limit the effects of these biases on the estimates and the conclusions of the test accuracy.

To address these sources of bias, we will use are sensitivity analysis, subgroup analysis or meta-regression analysis. The STATA software will be used since meta-regression cannot be performed using Review Manager.

We will also report statistics used in diagnostic test accuracy studies: the sensitivity and the specificity, the positive and negative predictive value, the likelihood ratios for the respective test results, or the Receiver Operating Characteristic (ROC) curve and quantities which are can be performed in STATA.

We will also explore two newly developed approaches to fitting random effects in hierarchical models overcome existing limitations: the hierarchical summary ROC model and the bivariate random effects model.

Both models give a valid estimation of the underlying ROC curve and the average operating point. Addition of covariates to the models, or application of separate models to different subgroups enables exploration of heterogeneity. Both models can be fitted with statistical STATA software that fits mixed models.

Cost-effectiveness analysis

This will include the following stages;

- Cost analysis. We note that the NHS Technology Adoption Centre (NTAC) calprotectin pilots are collecting data on referral rates, and that a cost-consequence analysis will be performed by NTAC. It is important that this analysis is available for this appraisal. We will also seek costs from other sources including University Hospital for Coventry and Warwickshire.
- Cost-effectiveness. We will start with the approach used by Hutton and colleagues in the CEP economic assessment, and summarised in their figure 1. However we expect to add another branch for indeterminate or borderline results. In addition, their analysis was largely a cost-consequence analysis, rather than a cost-effectiveness one. It is possible that introducing a calprotectin service for GPs would lead to better outcomes and cost savings, in which case a cost-minimisation analysis would be adequate. However if there are false negatives and false positives, we may need to analyse the trade-offs from adjusting sensitivity and specificity through cost-effectiveness modelling. The CEP report concluded that POCT dominated lab-based testing, but noted that fewer IBD cases were correctly identified.
- The relative cost-effectivenesses of different cut-off points will also be consideration.
- Final decisions on approach will be made in the light of the clinical effectiveness findings

Subgroups

- Children (under 14) vs adults
- IBD affecting only large bowel
- IBD affecting only small bowel
- . primary care vs secondary care groups as reflected in high proportions with IBS
- UC vs CD

Information from manufacturers

NICE will provide contact details for manufacturers and direct contact will be made as required. We note that there are several versions of some tests. When required, we will ask manufacturers to confirm which versions will continue to be marketed.

Data from manufacturers will not be used if received after 31st December.

Timelines

Progress report to NICE and NETSCC 7/1/13
Draft assessment report to NICE 21/2/13
Final assessment report to NICE 4/4/13
First AC meeting 8/5/13

Appendix 1

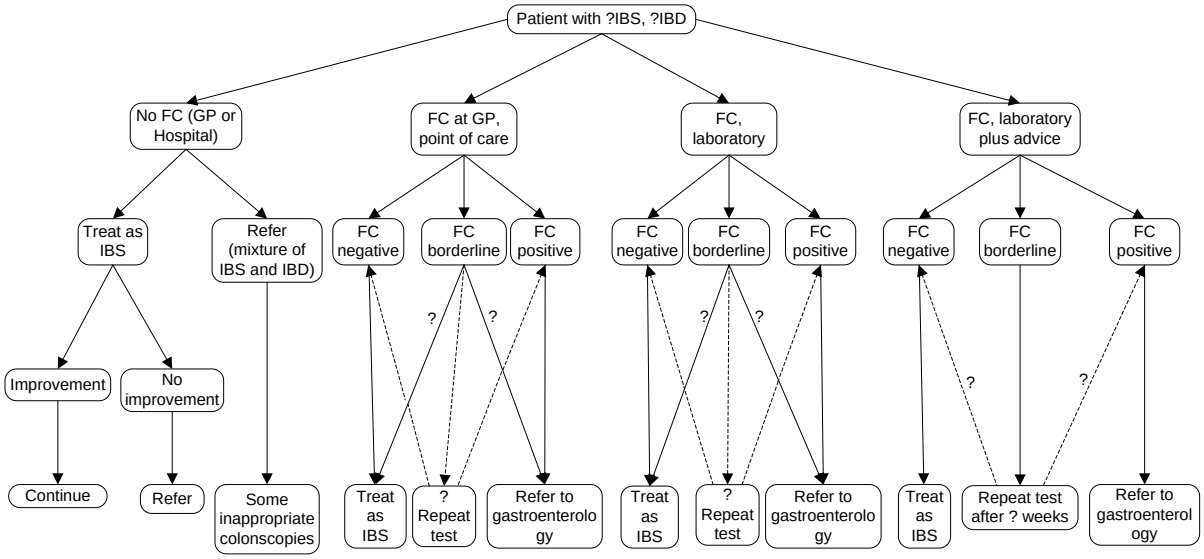
Quality assessment items derived from QUADAS tool (Whiting 2003)

1. Was the spectrum of patients representative of the patients who will receive the test in practice? (representative spectrum)
2. Is the reference standard likely to classify the target condition correctly? (acceptable reference standard)
3. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? (acceptable delay between tests)
4. Did the whole sample or a random selection of the sample, receive verification using the intended reference standard? (partial verification avoided)
5. Did patients receive the same reference standard irrespective of the index test result? (differential verification avoided)
6. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)? (incorporation avoided)
7. Were the reference standard results interpreted without knowledge of the results of the index test? (index test results blinded)
8. Were the index test results interpreted without knowledge of the results of the reference standard? (reference standard results blinded)
9. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? (relevant clinical information)
10. Were uninterpretable/ intermediate test results reported? (uninterpretable results reported)
11. Were withdrawals from the study explained? (withdrawals explained)

The term "quality assessment" is preferred to the more traditional "risk of bias" term because the latter, as used in systematic reviews such as Cochrane ones, is more associated with assessing internal validity of RCTs. We need to assess external validity through items such as spectrum bias.

Appendix 2. Possible service options

This is just for illustration and other options may be added.



References

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11(2):88-94.

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149(12):889-97.