



**Australian Government**  

---

**Medical Services Advisory Committee**

## **Public Summary Document**

### ***Application No. 1353 – Measurement of Calprotectin as marker of bowel inflammation***

**Applicant:** Taylor Bio-Medical Pty Ltd

**Date of MSAC consideration:** MSAC 72<sup>nd</sup> Meeting, 28-29 March 2018

Context for decision: MSAC makes its advice in accordance with its Terms of Reference, [visit the MSAC website](#)

#### **1. Purpose of application**

An application requesting Medicare Benefits Schedule (MBS) listing of measurement of faecal calprotectin (FC) as a marker of bowel inflammation was received from Taylor Bio-Medical Pty Ltd by the Department of Health.

The proposed medical service is FC testing for:

- differential diagnosis of irritable bowel syndrome (IBS) from inflammatory bowel disease (IBD) in the adult population, and differential diagnosis of IBD from non-IBD in the paediatric population; and
- monitoring disease activity in patients with known IBD.

#### **2. MSAC's advice to the Minister**

After considering the strength of the available evidence in relation to comparative safety, clinical effectiveness and cost effectiveness, MSAC did not support public funding of FC testing for the differential diagnosis of IBS from IBD in the adult population, for differential diagnosis of IBD from non-IBD in the paediatric population, or for monitoring disease activity in patients with known IBD.

MSAC considered that measurement of FC as a marker of bowel inflammation for IBD had reasonable diagnostic accuracy in a high prevalence population, but uncertain clinical effectiveness to improve health outcomes, unfavourable and highly uncertain cost-effectiveness, and likely large financial implications with little confidence that utilisation can be kept within the intended purposes and consequences.

MSAC acknowledged that there was a clinical need for a diagnostic triage test to differentiate IBD from IBS and thus to avoid some of the more invasive subsequent investigations such as colonoscopies, but advised that clinical benefit for monitoring IBD with FC had not been established. MSAC therefore advised that any resubmission should focus on the use of the test for diagnostic rather than monitoring purposes.

A resubmission for diagnostic testing should provide:

- a clearer definition of the appropriate target population;
- more relevant and robust test performance data in a lower prevalence setting similar to that of the proposed target population, or more appropriate modelling of the likely operating point with its prediction area in summary receiver operating characteristic (SROC) curves; and
- more realistic assumptions for the downstream consequences of FC testing versus no FC testing (such as for the estimated reduction in the extent of use of subsequent colonoscopy);
- appropriately amended economic model structure and inputs, which may simplify (if changes in health outcomes are estimated to be small) to a cost-minimisation analysis where the extra costs of add-on FC testing are equally offset by the estimated reduction in the proportion of patients undergoing colonoscopy.

If pursued, a resubmission for monitoring IBD should provide:

- a definition of current Australian practice to monitor IBD (e.g. frequency of colonoscopies), together with supportive data on the clinical utility of this practice both to establish any clinical need for FC monitoring in the first place, and then also to better identify the downstream management practices and health outcomes which might be improved by introducing FC monitoring;
- further information on the biological variability of inflammation in the stable IBD patient cohort in order to provide a better basis for when to initiate a change in clinical management and also a better basis for the requested frequency of FC testing for this purpose;
- appropriately amended economic model structure and inputs.

MSAC noted that any resubmission would need to be considered by ESC.

### **3. Summary of consideration and rationale for MSAC's advice**

MSAC noted that the FC test measures the levels of calprotectin (released by neutrophils involved in inflammation) in faeces. The test requires a patient to provide a stool specimen. Two test options were proposed, a laboratory-based test (enzyme-linked immunosorbent assay [ELISA]) and a point-of-care test (POCT).

MSAC noted that, based on the applicant's claim, the intention of the test is to reduce the need for more invasive investigations such as colonoscopies, endoscopies and associated pathology biopsies, and other diagnostic imaging procedures such as magnetic resonance imaging (MRI) and colonographies. MSAC noted that the estimated general population prevalence of IBD in Australia is 0.32% (Studd C et al 2016), and the prevalence of IBS in Australia is estimated at between 4.4% and 13% of the population (Canavan C et al 2014).

MSAC accepted that the comparators are usual care for the diagnostic population, and usual monitoring for the monitoring population. For both populations, it was proposed that any accurate test options would be replaced.

MSAC considered that the test is superior in terms of safety as it is non-invasive, although no comparative adverse event data were available.

MSAC noted that all of the included studies were undertaken in secondary settings among highly selected populations with a high prevalence of IBD (prevalence ranging from 20% to 69% in studies of adults and 36% to 80% in studies of children). MSAC noted that none of the included studies evaluated test performance in a general practitioner (GP) setting.

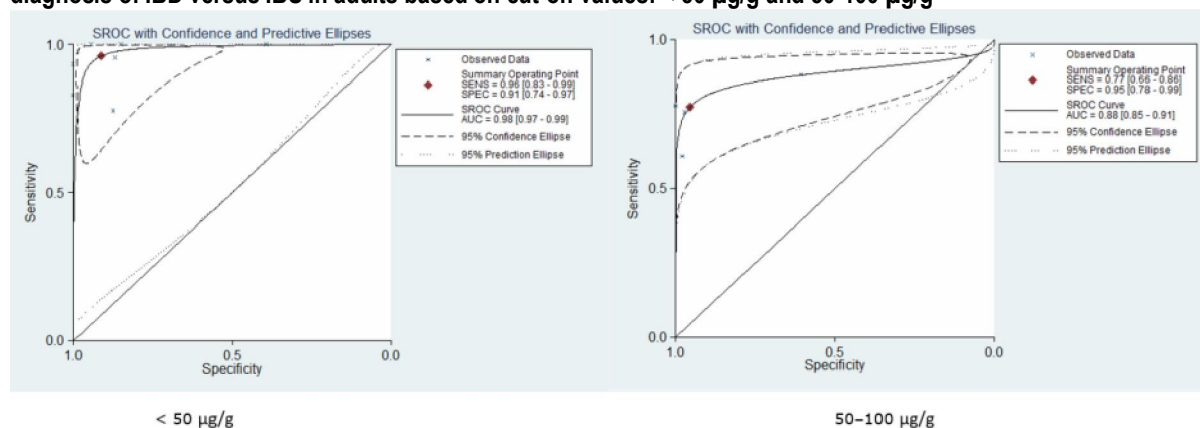
Consequently, MSAC considered that, although the test performance is good in high prevalence populations, this performance could not be extrapolated to the requested community setting of the MBS where the expected proportion of false positive test results would be greater (due to lower prevalence). In this context, MSAC noted that applying Fagan's nomogram using the pooled likelihood ratios and decreasing the pre-test probability (prevalence) from 50% (based on pooled estimates from included studies) to that estimated for the Australian community (0.32%) would substantially decrease the estimated post-test probability. MSAC considered that the nomogram's prerequisite assumption of constant likelihood ratios across the spectrum of disease between the study populations and the Australian community had not been demonstrated, noting particularly the large differences in prevalence involved.

MSAC noted that a linked evidence approach had been taken due to a lack of high level evidence. For diagnostic accuracy as a triage test compared with the reference standard of colonoscopy, MSAC noted the considerable variation in the evidence at the different test thresholds:

- in the adult population:
  - sensitivity of the test ranged from 78%–100% (<50 µg/g) and 61%–88% (50–100 µg/g);
  - specificity of the test ranged from 39%–100% (<50 µg/g) and 61%–100% (50–100 µg/g); and
  - pooled analysis (including all test thresholds) resulted in a sensitivity of 88% (95% CI 78%–94%) and a specificity of 91% (95% CI 81%–96%).
- in the paediatric population:
  - sensitivity ranged from 95%–100% (<50 µg/g), 82%–100% (50–100 µg/g), and 73%–100% (>100 µg/g);
  - specificity of the test ranged from 44%–93% (<50 µg/g), 59%–100% (50–100 µg/g), and 74%–100% (>100 µg/g); and
  - pooled analysis (including all test thresholds) resulted in a sensitivity of 95% (95% CI 90%–97%) and a specificity of 84% (95% CI 75%–90%).

MSAC considered that the variations in (a) the test thresholds of FC (from less than 15 µg/g to 487 µg/g), (b) the quantification methods (ELISA, semi-quantitative POCT, and quantitative POCT), and (c) the prevalence of IBD in the populations tested (specialist and GP setting) all resulted in heterogeneity in the generation and interpretation of test results. As a consequence, MSAC had low confidence in the likely operating point for the test performance of FC as a diagnostic triage test for IBD, noting that the confidence intervals for sensitivity and specificity underestimate the statistical uncertainty of the likely operating point, which is better represented in summary receiver operating characteristic (SROC) curves.

**Figure 1 SROC curves for studies investigating the sensitivity and specificity of faecal calprotectin for differential diagnosis of IBD versus IBS in adults based on cut-off values: < 50 µg/g and 50-100 µg/g**



MSAC considered that this heterogeneity would manifest in greater imprecision in terms of the numbers of false test results across both the adult and paediatric groups than suggested by the sensitivity and specificity data alone. In particular, MSAC noted that a false positive FC test may result in patients being placed onto an inflammatory pathway, which may lead to further testing including colonoscopy.

In addition, MSAC queried why, in studies of adult populations, raising the test threshold decreased the specificity of the test, which was unexpected. MSAC also considered that the assessment of the accuracy of the test for diagnostic purposes was subject to bias in the studies (from patient selection, flow and timing). MSAC also noted that interpretation of the FC test results is also subject to awareness of other confounding and environmental factors, such as varying age, body mass index, exercise, dietary fibre, smoking, and the presence of infection (Mendall et al. *Gastroenterology* 2016;16:216).

MSAC queried the proposed frequency of testing for diagnostic purposes, which appeared to be unsupported.

MSAC noted that the assessment of clinical utility of the test for diagnostic purposes was based on calculations of true and false negatives and true and false positives derived from the test performance results, applied to a theoretical population of 100 patients, and comparing FC testing with C-reactive protein (CRP) testing. MSAC noted that this approach assumed that these calculations would be the driver of clinical decisions and management, in a direct 1:1 manner. MSAC considered that this 1:1 assumption resulted in implausibly large effects on treatment and management, and that this method did not appropriately capture the downstream consequences of testing. In this context, MSAC noted that no evidence was provided to support whether a change in clinical management would occur as a result of using the FC test in the diagnostic population. Nevertheless, MSAC accepted that the clinical need for a triage test to differentiate IBD from IBS and thus avoid a proportion of the more invasive subsequent investigations, such as colonoscopies, had been identified, albeit that there was no direct evidence of overuse of colonoscopy for IBS.

MSAC noted that no direct evidence had been presented to support use of the FC test for monitoring of patients already diagnosed with IBD. MSAC noted that the results of diagnostic accuracy included studies providing head-to-head comparisons of FC with CRP tests versus the reference standard of colonoscopy. MSAC noted that meta-analysis of these studies showed that FC had higher sensitivity and similar specificity compared to CRP for detecting inflammation in patients with established IBD:

- FC: sensitivity 87% (95% CI 81%–91%); specificity 78% (95% CI 68%–86%);
- CRP: sensitivity 62% (95% CI 55%–68%); specificity 75% (95% CI 65%–83%).

MSAC considered that this evidence provided some confidence regarding the relative test performance compared to CRP, but absolute test performance over usual care was less certain. MSAC also noted that the pooled heterogeneity for FC was high at  $I^2 = 68\%$  (95% CI 51%–84%) and 77% (95% CI 66%–88%) for sensitivity and specificity, respectively, with unclear/high risk of bias in the selection of the study populations.

MSAC noted that the evidence provided to support the impact of FC testing on change in management in the monitoring population was based on four uncontrolled studies with a high risk of bias. MSAC considered that, given the study limitations, it was not clear whether the test results changed management, and that this remained a key area of uncertainty.

MSAC also considered that there was uncertainty around how the test might actually be used for monitoring of IBD, including noting that there was no clear rationale for the proposed frequency of testing in this context. MSAC noted that, although there may be clinical merit for use of the test for triage in the diagnostic setting, no evidence had been provided in the submission that the current treatment and monitoring of IBD was problematic, such as the overuse of colonoscopy, or under-treatment of IBD.

MSAC therefore considered that the clinical need for the test for monitoring IBD had not been established, noting that the test had not been approved for monitoring of IBD in the guidelines from the UK National Institutes of Health and Clinical Excellence (NICE; Quality Standard QS81; February 2015).

MSAC noted that, for monitoring, the clinical management algorithm indicated that funding the FC test would substantially increase testing frequency (up to five times per year) compared to current standard testing with erythrocyte sedimentation rate (ESR) and CRP. MSAC suggested that expert opinion would be helpful to define the current approach to monitoring of IBD in Australia (for example, frequency of colonoscopies) and that data were needed to determine the frequency of use and the clinical utility of the FC test for monitoring.

MSAC considered that the proposed fee for FC testing (\$80) required justification.

MSAC noted that the structure of all the economic models was based on a comparison of FC testing with usual care. MSAC noted that, for usual care, the test performance data for ESR/CRP was used to determine test outcome (true/false positive, true/false negative) and subsequent management. MSAC noted that this basis for comparison assumed replacement of FC testing for CRP testing (rather than adding FC testing to CRP testing), and again adopted the assumption that there was a 1:1 relationship between test performance and subsequent management. MSAC considered that this was an oversimplification which resulted in unreliable estimates of the comparative consequences of the changes in management in terms of subsequent healthcare resource use and health outcomes which favoured FC testing. For example, in terms of healthcare resource use, the models for differential diagnosis models overestimated the extent of colonoscopies avoided, but the model for monitoring underestimated the subsequent use of medicines to treat IBD.

MSAC noted that the economic models for differential diagnosis relied on costs per correct diagnosis which were offset by costs per colonoscopy avoided. MSAC noted the incremental cost effectiveness ratios (ICERs) for adults (\$761 per correct diagnosis and \$319 per colonoscopy avoided due to FC testing), and considered that the cost per quality adjusted life year (QALY) was high (\$56,675), and was dominated (FC testing was more costly and no more effective than usual care) in many sensitivity analyses. MSAC noted the paediatric model ICERs (\$1617 per correct diagnosis, \$730 per colonoscopy avoided) were associated

with a cost per QALY in which FC testing was dominated by usual care in the base case and most of the sensitivity analyses.

MSAC discussed other problems with the models for differential diagnosis:

- the ICER paradoxically increased in the adult model in response to an increase in prevalence, when a decrease would be expected due to better test performance;
- the assumption that an immediate negative test result would in fact improve quality of life beyond reassurance, as symptoms in these patients would continue; and
- uncertainty around the model inputs, such as the health benefit arising as a consequence of decreased delay in diagnosis (modelled as a QALY gain of between 0 and 11 hours).

MSAC noted that, while the incremental costs were relatively small (<\$100), the incremental benefits were also small, resulting in large ICERs for a number of outcomes. MSAC noted the extremely wide range of the ICERs, suggesting the cost-effectiveness of FC testing is highly varied.

MSAC considered that the submission failed to consider the consequences of a negative FC test, noting that the model had assumed 2.5% of this group would have a subsequent colonoscopy, but MSAC considered that this percentage was likely to be higher. MSAC discussed the absence of real-world data on the downstream care of patients who test negative for FC, which would be helpful to determine the impact of a negative test on patient management.

MSAC considered that there was insufficient information to confirm that diagnostic testing with FC improved outcomes for patients, and so suggested that a cost-minimisation approach comparing FC testing with no FC testing based on more realistic reductions in the extent of use of subsequent colonoscopy might be appropriate.

MSAC noted that, although the cost per QALY (\$10,093) in the economic model for monitoring appeared more favourable than for the differential diagnosis models over a 40-year time horizon, FC was dominated by usual care when the model assumed a 20-year time horizon. MSAC noted the uncertainty around the impact of FC monitoring on health outcomes, with a modelled QALY gain of 3 days over a 40 year time horizon. MSAC also noted that, as with the diagnostic model, the same 1:1 simplistic assumption based on test performance formed the basis of the management algorithm, but that this had not adequately captured the potential to lead to an increase in second and third line treatments that might arise from bringing forward a diagnosis of a worsening condition in a patient with IBD.

MSAC considered that as such, the economic model for monitoring required amendments to both its structure (for example to more accurately capture the downstream changes in healthcare resource use and health outcomes of the claimed earlier detection of worsening IBD) and its inputs (for example, the frequency of monitoring).

MSAC noted the high prevalence of IBS in Australia, estimated at approximately 10% of the population, and considered that, although the actual population presenting in primary care was estimated to be around 3%, the financial implications of funding were still substantial (~\$72 million and ~\$30 million for differential diagnosis and monitoring, respectively). MSAC considered that these may be underestimates. First, there is the potential for a range of tests to be ordered per sample, which may give rise to increased likelihood of co-claiming of items (such as, faecal occult blood tests). Second, in the primary care population, it may be difficult to manage leakage of the test item to other patients with minor abdominal symptoms.

MSAC noted the advice provided by the Royal College of Pathologists of Australasia (RCPA) that IBS is not a diagnosis of exclusion, and that a diagnosis of IBS should be based on the Rome IV criteria with a symptom onset of at least 6 months prior to diagnosis, which is discordant with the proposed item descriptor (abdominal symptoms of at least 6 weeks duration).

MSAC noted that there were no POCT items listed on the MBS, and as such a proposed listing for a POCT was out of scope for its consideration.

MSAC also discussed the use of different threshold values for monitoring as presented in the evidence and considered that measuring changes in FC levels in individual patients, where rising levels may prompt intervention, might be more useful. However, MSAC noted that no evidence had been provided as to the degree of change that would be biologically relevant. MSAC suggested that more information as to biological variability in the stable IBD patient cohort (FC test result <200 µg/g) would be helpful to more clearly define the target population for monitoring and its frequency, reducing the likelihood of unnecessary testing for consumers.

MSAC considered that consumers should be aware that the evidence did not currently support either use of the test in the general practice setting and that the potential for unnecessary testing was high.

MSAC advised that the item descriptor for differential diagnosis should be narrowed to target the patient population in which the test has real value (for example, patients without alarm symptoms, in whom colonoscopy could be avoided) if the submission is to be reconsidered.

MSAC noted that FC is widely used in private practice despite the poor evidence base. MSAC recommended that the Gastroenterological Society of Australia (GESA) and the Royal Australian College of General Practitioners (RACGP) be approached regarding the development of guidelines for FC testing. These organisations may also be able to assist in better defining the current clinical management algorithms which would be influenced by more clearly targeted FC testing for differential diagnosis and monitoring and thus support any resubmission.

#### **4. Background**

MSAC has not previously considered this application.

#### **5. Prerequisites to implementation of any funding advice**

Laboratory-based testing will be carried out in National Association of Testing Authorities (NATA)-accredited pathology laboratories. Point of care tests will only be performed and analysed by gastroenterologists. Testing requires operator training and compliance with the National Pathology Accreditation Advisory Council (NPAAC) guidelines for Point of Care testing.

The Applicant's clinical chemistry-specific protein IVDs (ARTG entry 223904) is registered on the Australian Register of Therapeutic Goods.

## 6. Proposal for public funding

The Applicant proposed two MBS items for differential diagnosis (**Table 1**), and two MBS items for monitoring (**Table 2**), allowing for laboratory-based testing (Category 6 – Pathology) and POCT (Category 2 – Diagnostic).

**Table 1 Proposed MBS item descriptors for differential diagnosis**

Category 2 – Diagnostic
MBS [item number] Faecal Calprotectin point of care testing of patients aged 3 years or older presenting with chronic (more than 6 weeks' duration) gastrointestinal symptoms which are suggestive of either inflammatory or functional bowel disease, where infectious causes have been excluded. A maximum of 3 tests may be performed in any 2-year period. Fee: [Applicant proposes a fee of \$80 per test] [Relevant explanatory notes]
Category 6 – Pathology
MBS [item number] Faecal Calprotectin testing of patients aged 3 years or older presenting with chronic (more than 6 weeks' duration) gastrointestinal symptoms which are suggestive of either inflammatory or functional bowel disease, where infectious causes have been excluded. A maximum of 3 tests may be performed in any 2-year period. Fee: [Applicant proposes a fee of \$80 per test] [Relevant explanatory notes]

**Table 2 Proposed MBS item descriptors for monitoring**

Category 2 – Diagnostic
MBS [item number] Faecal Calprotectin point of care testing of patients aged 3 years or older with an established Inflammatory Bowel Disease diagnosis such as ulcerative colitis or Crohn's disease, to monitor patients and evaluate disease activity. A maximum of 5 tests per year may be performed. Fee: [Applicant proposes a fee of \$80 per test] [Relevant explanatory notes]
Category 6 – Pathology
MBS [item number] Faecal Calprotectin testing of patients aged 3 years or older with an established Inflammatory Bowel Disease diagnosis such as ulcerative colitis or Crohn's disease, to monitor patients and evaluate disease activity. A maximum of 5 tests per year may be performed. Fee: [Applicant proposes a fee of \$80 per test] [Relevant explanatory notes]

## 7. Summary of public consultation feedback/consumer issues

Consultation feedback was received from specialists, consumers/care givers, and peak bodies. The feedback was overall supportive of using FC testing to differentiate between IBS and IBD, and monitor disease activity in people with IBD.

## 8. Proposed intervention's place in clinical management

### *Differential diagnosis*

In both the usual care and intervention pathways, the initial step involves the patient presenting to a GP, who assesses the patient for alarm features, and/or refers for a blood test, and/or a stool test for pathogens. In the intervention, the GP may also choose to refer for a FC laboratory-based test.



The results of all of these tests are used by the GP to determine the next stage in the pathway, which is either GP management or specialist referral. In the intervention pathway, a subsequent step may include a specialist undertaking a FC test, either as a POCT or in a laboratory setting, if the GP did not already refer the patient for a FC test.

The final stage in the pathway in both the usual care and intervention is either further clinical assessment (including colonoscopy/endoscopy for all patients, with some undergoing diagnostic imaging), or symptoms being controlled/resolved.

### ***Monitoring***

In the usual care pathway, monitoring is undertaken through blood tests, and colonoscopy/endoscopy and CT/MRI as required. In the intervention pathway, monitoring is undertaken through FC testing, either referred by a GP or specialist to a laboratory, or undertaken by a specialist as a POCT.

In both the usual care and intervention pathways, the end of the pathway for all patients is determined by their monitoring test results: either a continuation of standard care, or specialist review and further assessment.

The intention of the test is to reduce the need for more invasive investigations such as colonoscopies and endoscopies and associated pathology biopsies, and other diagnostic imaging procedures such as MRIs and colonographies.

## **9. Comparator**

### ***Differential diagnosis***

PASC advised that the comparator should be colonoscopy/endoscopy and biopsy. These investigations would be undertaken once in differential diagnosis. In addition to colonoscopy/endoscopy and biopsy, the Protocol identified some other diagnostic tests as comparators: full blood count (FBC), c-reactive protein (CRP), erythrocyte sedimentation rate (ESR), coeliac serology and iron tests if FBC results are suggestive, and/or stool test for pathogens.

The critique suggested that a more appropriate comparator may be “FC testing plus clinical workup compared to clinical workup without FC testing”. Clinical workup includes the requirement to exclude coeliac disease, pathogens, presence of anaemia as well as clinical history.

### ***Monitoring***

In the Protocol, PASC advised that the comparator for monitoring should include a combination of colonoscopy/endoscopy and biopsy, and blood tests.

For GP monitoring, the comparators are blood tests: for ulcerative colitis this includes FBC, CRP, and ESR, and for Crohn disease this includes FBC, CRP, iron, and B12. For specialist monitoring, the Protocol defined the comparators as colonoscopy/endoscopy and biopsy; and computed tomography/magnetic resonance imaging.

The critique noted that there was no discussion on which of the tests nominated is most likely to be replaced for patients seen by specialists, in particular, the use of colonoscopy/endoscopy or MRI. The critique noted that these are tests that provide additional information to that provided by a FC test. As such, FC testing is likely to be an adjunct to the

current testing regime for specialist, with the exception of ESR and CRP tests, tests for the presence of inflammation, which may be replaced.

## **10. Comparative safety**

### *Test adverse events*

No evidence of adverse events from the use of the FC test were identified in the literature.

### *Adverse events from change in management*

Adverse events were considered in terms of the impact of false positive and false negative test results on patient management. The rates of false negative and positive results change with respect to patient population (adults versus paediatric) and test thresholds of FC. False positives are more likely to result in unnecessary interventional treatment, and this was investigated with respect to perforation and bleeding following colonoscopies and sigmoidoscopies. False negative test results may result in delayed diagnosis and/or treatment. The length of diagnostic delay was positively correlated with complications such as the occurrence of bowel stenosis and intestinal surgery.

## **11. Comparative effectiveness**

Due to the absence of direct evidence on patient management and health outcomes from FC testing identified, a linked evidence approach was taken to both assessments on FC testing – differential diagnosis of suspected IBD and monitoring of disease activity in confirmed IBD.

### *Accuracy*

The reference standard for confirmation of inflammatory bowel disease (IBD) was endoscopy/colonoscopy ideally with histology of biopsies. Most studies used this as the reference standard, although some studies used other reference tests where appropriate, such as clinical follow up for paediatric patients with a negative FC result.

### *Differential diagnosis*

Diagnostic accuracy of FC was evaluated in adults with suspected IBD or IBS in six of the included studies. The pooled sensitivity of FC was 0.88 [95% CI: 0.78-0.94] and the pooled specificity was 0.91 [95% CI: 0.81-0.96], indicating a high level of sensitivity and specificity in this population. Diagnostic accuracy of FC was evaluated in paediatric patients with suspected IBD compared to non-IBD in nine of the included studies. Pooled sensitivity of FC was 0.95 [95% CI: 0.90-0.97] and the pooled specificity was 0.84 [95% CI: 0.75-0.90].

In subgroup analysis of FC at different test thresholds, compared to a FC threshold of <50 µg/g, at thresholds between 50-100 µg/g sensitivity remained high (96% compared to 95%) while specificity was reduced (91% compared to 77%). In paediatric patients, sensitivity reduced as the threshold increased but remained high at all levels; at <50 µg/g, 50-100 µg/g and >100 µg/g, pooled sensitivity was 99%, 94% and 91%, respectively. Pooled specificity increased with increasing test thresholds of FC; <50 µg/g, 50-100 µg/g and >100 µg/g, pooled specificity was 74%, 85% and 88%, respectively.

In adults, both testing methods, ELISA and POCT, showed similar and high sensitivity (86% and 93%, respectively) and specificity (90% and 94%, respectively), although it is noted that the pooled results for POCT relied on three separate observations in only one study. All paediatric studies investigated FC using the ELISA method.

In the few studies (four studies for CRP, and three studies for ESR) that compared FC to other standard laboratory parameters of inflammation, FC was shown to have higher sensitivity and comparable specificity to both CRP and ESR. Pooled sensitivity was 94%, 57% and 56%, for FC, CRP and ESR, respectively. Pooled specificity was 92%, 88% and 86%, for FC, CRP and ESR, respectively (Table 3).

**Table 3 Summary statistics for FC compared to CRP and ESR, against endoscopy/ colonoscopy for differential diagnosis of suspected IBD**

Accuracy (k=143 for CRP; 85 for ESR)	FC (n=234)	CRP (n=234)	ESR (n=140)
Sensitivity, % [95% CI]	0.94 [0.82-0.98]	0.57 [0.38-0.74]	0.56 [0.39-0.73]
Specificity, % [95% CI]	0.92 [0.70-0.98]	0.88 [0.76-0.95]	0.86 [0.69-0.95]
Positive likelihood ratio [95% CI]	11.39 [2.78-46.63]	4.93 [2.38-10.18]	4.13 [1.84-9.30]
Negative likelihood ratio [95% CI]	0.06 [0.02-0.20]	0.49 [0.32-0.75]	0.51 [0.35-0.73]
Diagnostic odds ratio [95% CI]	179.56 [43.27-745.16]	10.13 [3.87-26.48]	8.18 [3.08-21.70]
Area under the curve [95% CI]	0.98 [0.96-0.99]	0.86 [0.38-0.89]	0.81 [0.77-0.84]

k=number of participants with disease; n=number of participants in the studies

### **Monitoring**

Diagnostic accuracy of FC was evaluated in patients with confirmed IBD in twenty studies. The pooled sensitivity of FC in monitoring disease activity in patients with confirmed IBD was 0.88 [95% CI 0.84-0.91] and the pooled specificity was 0.72 [95% CI 0.65-0.78], indicating a high level of sensitivity but lower specificity in this population.

Subgroup analyses at different test thresholds of FC indicated that FC is sensitive but less specific at FC thresholds <100 µg/g; pooled sensitivity and specificity were 0.93 [95% CI 0.88-0.95] and 0.59 [95% CI 0.43-0.74], respectively. At thresholds between 100-200 µg/g, sensitivity remained high while specificity was improved; pooled sensitivity and specificity were 0.86 [95% CI 0.80-0.90] and 0.76 [95% CI 0.64-0.85], respectively. At thresholds >200 µg/g, sensitivity and specificity were similar to pooled results at thresholds between 100–200 µg/g; pooled sensitivity and specificity were 0.83 [95% CI 0.77-0.88] and 0.77 [95% CI 0.72-0.81], respectively.

Subgroup analyses were performed on the two main types of IBD: CD and UC. In CD, pooled sensitivity and specificity were 0.85 [95% CI 0.79-0.90] and 0.69 [95% CI 0.57-0.78], respectively. In UC, pooled sensitivity and specificity were 0.91 [95% CI 0.85-0.94] and 0.74 [95% CI 0.62-0.83], respectively. These findings indicate that FC has comparable sensitivity and specificity in patients with confirmed CD and UC, and to the broader IBD population.

Eleven studies investigating the diagnostic accuracy of FC in monitoring patients with confirmed IBD compared FC to a standard laboratory parameter of inflammation, C reactive protein (CRP). The pooled sensitivity and specificity of FC in these studies was 0.87 [95% CI 0.81-0.91] and 0.78 [95% CI 0.69-0.86], respectively. Pooled sensitivity and specificity of CRP was 0.62 [95% CI 0.55-0.68] and 0.75 [95% CI 0.65-0.83], respectively (Table 4). These findings suggest that FC has higher sensitivity and comparable specificity to CRP in monitoring disease activity in patients with confirmed IBD.

**Table 4 Summary statistics for FC compared to CRP, against endoscopy/ colonoscopy for monitoring disease activity in confirmed IBD**

Accuracy (k=741)	FC (n= 1,158)	CRP (n=1,158)
Sensitivity, % [95% CI]	0.87 [0.81-0.91]	0.62 [0.55-0.68]
Specificity, % [95% CI]	0.78 [0.68-0.86]	0.75 [0.65-0.83]
Positive likelihood ratio [95% CI]	3.98 [2.73-5.80]	2.48 [1.83-3.37]
Negative likelihood ratio [95% CI]	0.17 [0.13-0.24]	0.51 [0.45-0.57]
Diagnostic odds ratio [95% CI]	23.22 [13.96-38.63]	4.87 [3.38-7.02]
Area under the curve [95% CI]	0.90 [0.87-0.92]	0.71 [0.67-0.75]

Sub-group analysis of the diagnostic accuracy of FC compared to CRP in monitoring disease activity in confirmed CD showed that FC is more sensitive but less specific than CRP in patients with confirmed CD. Pooled sensitivity and specificity for FC were 0.86 [95% CI 0.81-0.90] and 0.69 [95% CI 0.53-0.81], respectively. Pooled sensitivity and specificity for CRP were 0.58 [95% CI 0.49-0.66] and 0.78 [95% CI 0.65-0.88], respectively.

Sub-group analysis of the diagnostic accuracy of FC compared to CRP in monitoring disease activity in confirmed UC indicated that FC is more sensitive and comparably specific than CRP in patients with confirmed UC. Pooled sensitivity and specificity for FC were 0.90 [95% CI 0.79-0.95] and 0.79 [95% CI 0.71-0.85], respectively while pooled sensitivity and specificity for CRP were 0.66 [95% CI 0.56-0.74] and 0.76 [95% CI 0.54-0.89], respectively.

#### ***Therapeutic efficacy (change in management)***

Clinical validity was measured in terms of the post-test probability of IBD in adult and paediatric patients tested with FC using Fagan's nomograms, which apply the pooled estimate of the positive and negative likelihood ratios to the pre-test probability (PTP) (prevalence) of IBD in the study populations to estimate the post-test probability of IBD.

In adults with suspected IBD, the PTP of IBD was determined to be 50%. This is based on the number of true positives in the overall pooled population in the included studies. Applying the pooled positive and negative likelihood ratios of 10 and 0.13, respectively, in adults with suspected IBD or IBS and a PTP of 50%, an abnormal test result for FC increases the probability of IBD to 91%, whereas a normal test result for FC reduces the probability to 12%.

In paediatric patients with suspected IBD, the PTP was estimated to be 54% in the study population. An abnormal test result for FC increases the probability to 87%, whereas a normal test result for FC reduces the probability to 7%.

In patients with confirmed IBD, the PTP of active IBD was determined to be 64%. Applying the pooled positive and negative likelihood ratios of 4 and 0.17, respectively, in patients with confirmed IBD and a PTP of 64%, an abnormal test result for FC increases the probability of active IBD to 88%, whereas a normal test result for FC reduces the probability to 23%. In patients tested with CRP, an abnormal test result increases the probability to 81%, whereas a normal test result for FC reduces the probability to 47%.

This analysis suggests that FC has good clinical validity in the prevalent population for detecting suspected IBD, and for detecting active IBD, particularly compared to CRP.

### ***Therapeutic effectiveness (health benefit from change in management)***

The clinical utility of FC was considered in terms of the impact of positive versus negative test results on patient management. This analysis showed that screening for suspected IBD by measuring FC levels will result in 52% of patients having a negative test result. Assuming that clinicians used the FC result to guide clinical decision-making, this would result in a 52% reduction in the number of adults requiring endoscopy. Of the remaining 48 patients with a positive test result who presumably undergo endoscopy, four will not have IBD.

The downside of this diagnostic screening strategy is delayed diagnosis in 12% (6 of 52) adults because of a false negative test result. Due to the differences in sensitivity and specificity at different test thresholds of FC, at a thresholds- of <50 µg/g, diagnosis would be delayed in 4% of adults, and 4% would be treated unnecessarily, compared to 5% delayed and 19% unnecessarily treated at thresholds between 50-100 µg/g.

In the population of children, 58 instead of 100 would undergo endoscopy. Nine of them will not have IBD, and diagnosis will be delayed in 8% of the affected children. At a threshold of <50 µg/g, diagnosis would be delayed in 3% of children, and 17% would be treated unnecessarily, compared to 7% delayed and 9% unnecessarily treated at thresholds between 50-100 µg/g. At thresholds >100 µg/g, diagnosis would be delayed in 11% of children, and 11% would be treated unnecessarily.

In monitoring disease activity in confirmed IBD, screening for IBD disease activity by measuring FC levels would result in a 36% reduction in the number of patients requiring interventional treatment (further investigation with endoscopy or an initiation or change in therapy). Eight out of 64 patients who undergo interventional treatment will not have increased disease activity. The downside of this monitoring strategy is delayed interventional treatment in 22% (8 of 36) patients because of a false negative test result.

Taken together, these findings indicate that FC has good clinical utility in helping to guide clinical decision making regarding patient care and would add to the suite of available tests for IBD. A summary of the findings for the linked evidence comparison of FC to CRP for differential diagnosis is shown in Table 5, and for monitoring, is shown in Table 6. Note that these findings relate to those studies in which a comparison of FC to CRP was made.

**Table 5 Summary of findings for the linked evidence comparison of faecal calprotectin relative to CRP, in differential diagnosis of patients with suspected IBD with assumed pre-test probability (prevalence) of 61%**

<b>Outcomes</b>	<b>Participants</b>	<b>Quality of evidence</b>	<b>No. per 100 patients with comparator</b>	<b>No. per 100 patients with intervention</b>	<b>Importance</b>	<b>Comments</b>
True positives	k=133 ; n=142	⊕⊕⊕⊙	35	57	Critical (8)	Earlier treatment
True negatives	k=85 ; n=92	⊕⊕⊕⊙	34	36	Critical (8)	Avoidance of unnecessary invasive procedure
False positives	k=9 ; n=142	⊕⊕⊕⊙	5	3	Critical (7)	Exposure to unnecessary invasive procedure
False negatives	k=7 ; n=92	⊕⊕⊕⊙	26	4	Critical (9)	Delayed diagnosis and treatment

**Table 6 Summary of findings for the linked evidence comparison of faecal calprotectin relative to CRP, in monitoring patients with confirmed IBD with assumed pre-test probability (prevalence) of 64%**

Outcomes	Participants	Quality of evidence	No. per 100 patients with comparator	No. per 100 patients with intervention	Importance	Comments
True positives	k=644 ; n=740	⊕⊕⊕⊕	40	56	Critical (8)	Earlier treatment
True negatives	k=326 ; n=418	⊕⊕⊕⊕	27	28	Critical (8)	Avoidance of unnecessary invasive procedure
False positives	k=96 ; n=740	⊕⊕⊕⊕	9	8	Critical (7)	Exposure to unnecessary invasive procedure
False negatives	k=92 ; n=418	⊕⊕⊕⊕	24	8	Critical (9)	Delayed diagnosis and treatment

<sup>a</sup> GRADE Working Group grades of evidence (Guyatt et al., 2013)

⊕⊕⊕⊕ **High quality:** We are very confident that the true effect lies close to that of the estimate of effect.

⊕⊕⊕⊕ **Moderate quality:** We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

⊕⊕⊕⊕ **Low quality:** Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

⊕⊕⊕⊕ **Very low quality:** We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

### ***Clinical claim***

The Protocol outlined the following clinical claims:

- differential diagnosis: that FC testing is cheaper, safer and more effective than the present alternative of performing colonoscopy; and
- monitoring: that FC testing is safer and more effective than the present alternative.

## **12. Economic evaluation**

### ***Differential diagnosis***

The modelled economic evaluation included a cost-utility and cost effectiveness analysis, which were conducted from the perspective of the healthcare payer, including costs related to the provision of health care resources and health outcomes. A summary of the key characteristics of the economic evaluation is given in **Table 7**.

**Table 7 Summary of the economic evaluation – diagnosis**

<b>Perspective</b>	Healthcare payer
<b>Comparator</b>	ESR/CRP testing
<b>Type of economic evaluation</b>	Cost utility/cost effectiveness
<b>Sources of evidence</b>	Selective review of literature and evidence in section B (systematic review)
<b>Time horizon</b>	Until confirmed diagnosis
<b>Outcomes</b>	Quality of life, procedures avoided, correct diagnoses, adverse events including perforations or bleeding from procedures.
<b>Methods used to generate results</b>	Expected value analysis
<b>Discount rate</b>	N/A
<b>Software packages used</b>	Microsoft Excel 2013

The overall costs and outcomes, and incremental costs and outcomes as calculated for the FC test and comparator in the model, with the base case assumptions, are shown in Table 8 and

Table 9 for the paediatric and adult population respectively. The ICERs for number of correct diagnoses and number of avoided procedures were calculated to be \$1,617 per correct diagnosis and \$730 per procedure avoided in the paediatric population, respectively. The ICERs for number of correct diagnoses and number of avoided procedures were calculated to be \$761 per correct diagnosis and \$319 per procedure avoided in the adult population, respectively.

Considering the average cost of colonoscopies, sigmoidoscopies and endoscopies, the ICER would be \$1,127 per procedure avoided, which based on an average cost of these procedures (approximately \$1,371) for the healthcare payer, would likely be cost effective or cost saving. In terms of quality of life, the ICER was dominated (driven by the rates of colonoscopy adverse events) However, there is a moderate degree of uncertainty surrounding the results.

**Table 8 Total and incremental costs and effectiveness of FC testing compared to ESR/CRP testing per person per year – diagnosis (paediatric population).**

	FC testing	ESR/CRP testing	Incremental cost/ effectiveness	ICER (\$ per unit)
Cost of testing (including GP/ specialist visits) (\$)	294	219	75	
Number of procedures	0.39	0.49	0.10	730
Number of adverse events	0.0008	0.0011	0.0002	324,272
Time active (weeks)	6.09	6.08	0.02	Dominated
QALYs	0.01	0.01	0.00	Dominated

ICER = Incremental Cost Effectiveness Ratio, QALY = quality adjusted life year, ESR = erythrocyte sedimentation rate, CRP = C-reactive protein, FC = faecal calprotectin. Dominated = greater cost and worse outcome. Note: components may not sum to totals due to rounding.

The modelled results for the paediatric population were most sensitive to changes in the sensitivity and specificity of the FC test and the underlying prevalence of IBD in the population. However, there is a high degree of uncertainty surrounding the results due to the lack of evidence in this population. There may be additional costs incurred for treating a paediatric population, which are not considered in the model structure or costings.

**Table 9 Total and incremental costs and effectiveness of FC testing compared to ESR/CRP testing per person per year – diagnosis (adult population)**

	FC testing	ESR/CRP testing	Incremental cost/ effectiveness	ICER (\$ per unit)
Cost of testing (including GP/ specialist visits) (\$)	290	219	71	
Number of procedures	0.30	0.53	0.22	319
Number of adverse events	0.0007	0.0012	0.0005	148,018
Time active (weeks)	5.6	6.5	0.839	84
QALYs	0.00839	0.00964	0.00125	56,675

ICER = Incremental Cost Effectiveness Ratio, QALY = quality adjusted life year, ESR = erythrocyte sedimentation rate, CRP = C-reactive protein, FC = faecal calprotectin. Dominated = greater cost and worse outcome. Note: components may not sum to totals due to rounding.

The modelled results for the adult population were most sensitive to changes in the sensitivity and specificity of the FC test, the underlying prevalence of IBD in the population.

## Monitoring

The modelled economic evaluation included a cost-utility and cost effectiveness analysis, and was conducted from the perspective of the healthcare payer (which due to available evidence was largely from the perspective of the MBS), including costs related to the provision of health care resources and health outcomes. A summary of the key characteristics of the economic evaluation is given in **Table 10**.

**Table 10 Summary of the economic evaluation – monitoring**

<b>Perspective</b>	Healthcare payer
<b>Comparator</b>	CRP testing
<b>Type of economic evaluation</b>	Cost utility/cost effectiveness
<b>Sources of evidence</b>	Selective review of literature and evidence in section B (systematic review)
<b>Time horizon</b>	1 period, 5 years and lifetime
<b>Outcomes</b>	Quality of life, procedures avoided, correct diagnoses, adverse events including perforations or bleeding from procedures.
<b>Methods used to generate results</b>	Decision analytic model using a decision tree with Monte Carlo simulation
<b>Health states</b>	Active Inactive
<b>Cycle length</b>	0.25 years to 1 year depending on monitoring frequency
<b>Discount rate</b>	5% annually default, sensitivity at 0% and 10% annually
<b>Software packages used</b>	Microsoft Excel 2013

Based on the clinical evidence and the modelled economic evaluation, FC testing appears to be cost effective – providing additional correct positive and negative diagnoses at relatively little cost. As shown in Table 11, the ICERs for the number of avoided procedures were calculated to be \$731 (610 – 913) per procedure avoided. Considering colonoscopies alone (noting there are other procedures) the ICER would be \$1,102 per colonoscopy avoided, which based on an average cost of a colonoscopy (almost \$2,400) for the healthcare payer is cost effective or cost saving.

The quality of life improvements would be additional to these results (noting they do not include the quality of life associated with avoided procedures, which is consistent with the approach taken by Waugh et al (2013)). However, there is a moderate degree of uncertainty surrounding the results.

**Table 11 Total and incremental costs and effectiveness of FC testing compared to ESR/CRP testing per person per year – monitoring**

	FC testing	ESR/CRP testing	Incremental cost/ effectiveness	ICER (\$ per unit)
Cost of testing (including GP/ specialist visits) (\$)	227	146	81	-
Number of procedures	0.34 (0.33, 0.34)	0.45 (0.42, 0.47)	0.11 (0.09, 0.13)	731 (610, 913)
Number of adverse events	0.0009 (0.0008, 0.0009)	0.0012 (0.0011, 0.0012)	0.0003 (0.0002, 0.0004)	255,056 (202,667, 343,971)
QALYs	0.381 (0.380, 0.381)	0.373 (0.372, 0.373)	0.008 (0.007, 0.009)	10,093 (8,980, 11,522)

ICER = Incremental Cost Effectiveness Ratio, QALY = quality adjusted life year. Dominated = greater cost and worse outcome. Note: components may not sum to totals due to rounding. ICER = Incremental Cost Effectiveness Ratio, QALY = quality adjusted life year, ESR = erythrocyte sedimentation rate, CRP = C-reactive protein, FC = faecal calprotectin. Dominated = greater cost and worse outcome. Note: components may not sum to totals due to rounding.



The modelled results for monitoring were most sensitive to the time horizon, probability of procedures or treatments with or without FC testing, thresholds (i.e. diagnostic accuracy) and the chosen discount rate.

### 13. Financial/budgetary impacts

An epidemiological approach was used to estimate the financial implications of the introduction of FC testing. The financial implications to the MBS resulting from the proposed listing of FC testing are summarised in **Table 12**. Introducing FC testing is expected to cost \$17.8 million in 2017-18. This is partly offset by savings of \$9.2 million due to a decrease in demand for other procedures. Therefore, the overall impact on the MBS is an increase in spending of \$8.6 million in 2017-18.

**Table 12 Total costs to the MBS associated with FC testing**

	2015-16	2016-17	2017-18	2018-19	2019-20
<b>FC test</b>					
Number of services	261,863	266,303	270,759	275,282	279,824
Sub-total cost (\$million)	17.8	18.1	18.4	18.7	19.0
<b>Associated interventions</b>					
Blood and biochemistry: services	-203,406	-206,856	-210,321	-213,841	-217,377
Blood and biochemistry: sub-total cost (\$ million)	-6.0	-6.1	-6.2	-6.3	-6.4
Sigmoidoscopy: services	-2,249	-2,287	-2,325	-2,363	-2,402
Sigmoidoscopy: sub-total cost (\$ million)	-0.2	-0.2	-0.2	-0.2	-0.2
Colonoscopy: services	-6,115	-6,218	-6,322	-6,426	-6,531
Colonoscopy: sub-total cost (\$ million)	-1.7	-1.8	-1.8	-1.8	-1.9
Endoscopy: services	-2,825	-2,873	-2,921	-2,969	-3,017
Endoscopy: sub-total cost (\$ million)	-0.3	-0.3	-0.3	-0.3	-0.3
MRI: services	-2,041	-2,076	-2,110	-2,145	-2,180
MRI: sub-total cost (\$ million)	-0.8	-0.8	-0.8	-0.8	-0.8
CT: services	-3,823	-3,888	-3,952	-4,018	-4,083
CT: sub-total costs (\$ million)	-1.7	-1.7	-1.7	-1.8	-1.8
Biopsy 1: services	-7,040	-7,159	-7,279	-7,399	-7,520
Biopsy 1: sub-total cost (\$ million)	-0.6	-0.6	-0.6	-0.6	-0.6
Biopsy 2: services	-3,221	-3,276	-3,330	-3,385	-3,441
Biopsy 2: sub-total cost (\$ million)	-0.4	-0.4	-0.4	-0.4	-0.4
GP: services	10,289	10,464	10,639	10,817	10,996
GP: sub-total cost (\$ million)	0.4	0.4	0.4	0.4	0.4
Specialist: services	28,157	28,635	29,115	29,604	30,095
Specialist: sub-total cost (\$ million)	2.0	2.1	2.1	2.2	2.2
<b>Total services</b>	69,590	70,769	71,954	73,157	74,365
<b>Total cost (\$ million)</b>	8.6	8.7	8.9	9.0	9.2

Note: biopsy 1 represents MBS item number 72823 while biopsy 2 represents MBS item number 72824.

### 14. Key issues from ESC for MSAC

This submission is a new application to support the listing on the MBS for FC testing. There are two target populations:

- people with symptoms that are suggestive of IBD (population 1); and
- people with a confirmed diagnosis of IBD (population 2).

FC is a test that is used to detect intestinal inflammation. For diagnosis, the proposal is that FC testing be used in the differential diagnosis of IBS from IBD (in the adult population), and IBD from non-IBD (in the paediatric population). The intention of the test is to reduce the need for more invasive investigations such as colonoscopies, endoscopies and associated pathology biopsies, and other diagnostic imaging procedures such as magnetic resonance imaging.

For monitoring purposes, the proposal is that FC testing be used to monitor the disease activity in people with known IBD, and monitor their response to treatment changes. The clinical claim is that FC testing allows for more frequent monitoring of disease activity, and an incipient relapse can be recognised earlier which allows for therapy adjustment, thus avoiding or ameliorating a recurrent acute episode.

ESC considered that the clinical algorithm is difficult to understand and was not applicable to general practice, where the eligible patient population is considered to be patients under the age of 45 years who do not have 'red flag' symptoms but have persistent abdominal symptoms and whom the general practitioner (GP) is considering sending to a specialist for further investigation.

ESC noted that a linked evidence approach had been taken due to a lack of direct evidence and that the evidence base was of low quality and had high heterogeneity.

ESC noted that with regards to the use of FC testing for differential diagnosis of suspected IBD from IBS (adults) or non-IBD (paediatric patients):

- evidence from pooled results showed reasonable sensitivity (88%) and high specificity (91%) compared to the reference standard (colonoscopy) in adults, and high sensitivity (95%) and reasonable specificity (84%) in the paediatric population;
- pooled sensitivity decreased with decreasing threshold of FC, although it remained high at all levels; at <50 µg/g, 50–100 µg/g and >100 µg/g, pooled sensitivity was 99%, 94% and 91%, respectively;
- pooled specificity increased with increasing threshold of FC; at <50 µg/g, 50–100 µg/g and >100 µg/g, pooled specificity was 74%, 85% and 88%, respectively; and
- FC was more sensitive and had comparable specificity relative to both CRP and ESR, based on limited evidence.

ESC noted that 11 studies compared FC to CRP for the monitoring of disease activity in confirmed IBD. However, the number of studies evaluating FC in comparison with other standard markers of inflammation or diagnostic imaging was insufficient to enable pooling of results.

ESC noted there were several limitations to the evidence for test accuracy, including:

- none of the included studies which assessed the performance of the test were conducted in a GP practice or in a setting with a prevalence of IBD that approximated the community prevalence;
- most of the included studies did not recruit patients consecutively and the time between faecal sample and endoscopy could be up to a month;
- variable test performance at different test thresholds which, given the high prevalence of IBD in the included studies (up to 80% in one paediatric group), makes it difficult to determine the applicability of this test to the wider community;
- the lack of data on the performance of the test in the general community samples where prevalence of IBD is 0.3% and incidence is 0.02%.

ESC noted that no conclusions could be drawn as to clinical utility and the effect of FC testing on change in management as no studies addressed this. ESC noted that the

intervention is completed on a stool sample, which is not associated with any risks, and the comparator for direct evidence was colonoscopy, with known side effects although no comparative adverse event data were available.

ESC noted that point-of-care testing (POCT) has similar diagnostic accuracy to the Enzyme-linked Immunosorbent Assay (ELISA) method, based on limited evidence from three studies. ESC queried the rationale for POCT, noting that immediate diagnosis is not clinically necessary and that in general; POCT is higher cost and potentially lower quality than laboratory testing.

In summary, ESC considered the diagnostic accuracy of the test as having high sensitivity and moderate specificity at the 50 µg/g test threshold. This may increase negative predictive value (NPV) and make the test more useful in ruling in IBD but less useful in ruling it out. ESC considered the applicability of the test in the real world where prevalence rates are much lower than those in the presented studies. ESC considered that the evidence in support of the FC test to change clinical management is uncertain given the lack of evidence in primary care.

ESC noted that the model results (for number of correct diagnoses, cost for number of avoided procedures, cost for number of adverse events avoided, cost for time in inactive IBD state and cost per additional QALY) varied widely, leading to a wide range of incremental cost-effectiveness ratios (ICERs) for the economic model.

For differential diagnosis, the ICERs were:

- correct diagnosis: \$1,617 for paediatric population, \$761 for adults;
- for procedures avoided (unnecessary colonoscopy): \$730 for paediatric population, <\$500 for adults;
- for adverse events avoided: over \$324,272 for paediatric population, \$148,018 for adults; and
- cost per QALY dominated for paediatric population and \$56,575 for adults.

For the monitoring population, the ICERs were:

- for correct diagnoses: \$267;
- for number of procedures avoided: \$731;
- for number of adverse events avoided: just over \$255,056; and
- cost per QALY: \$10,093.

ESC noted that both the incremental costs and the incremental benefits were relatively small leading to rather large ICERs for a number of outcomes (e.g. adverse events). ESC noted that based on the results of the evaluation, FC testing may be cost-effective for number of correct diagnoses and number of avoided procedures, but the same cannot be said for quality of life or number of adverse events avoided.

ESC noted that the sensitivity and specificity of each of the tests (FC and CRP/ESR) were key drivers of the model, and that the lower specificity of the FC test in a paediatric population contributed to the higher ICER. In the monitoring pathway, key drivers were sensitivity and specificity to determine active flares of IBD and hence increased GP and specialist visits for disease monitoring over the 40 year time frame.

ESC noted that assumptions around GP uptake were a second key driver, whereby the additional step of GP testing in the stepped evaluation increased the costs. ESC noted that the same rate of testing had been assumed for both the adult and paediatric populations, and that the economic model demonstrated considerable sensitivity to this variable, affecting the dominance of the intervention. ESC queried whether utility may be affected if FC testing resulted in a delay to diagnosis.

ESC queried the assumptions in the monitoring model where all parameters remain the same over the time period of 40 years, as this is unlikely to reflect clinical practice.

ESC noted that the key driver for the financial estimates is the incidence and prevalence of IBD and IBS in the population being tested (which was based on the adult population only).

This resulted in an estimated 27,392 to 29,249 tests per year over five years, and an estimated MBS cost for FC testing of between \$8 million and \$9 million per year, assuming 100% uptake and providing incidence and prevalence are stable. ESC considered that as the eligible population is a broader group of people with persistent gastrointestinal symptoms (as described in the proposed MBS descriptor); the actual population is likely to be significantly greater than the proposed estimates.

ESC advised that sensitivity analysis around the population prevalence for persistent gastrointestinal symptoms seen in general practice would be helpful to inform this issue. ESC was comfortable that the utilisation estimates for the monitoring test were reasonable given that the number of people with IBD is estimated by Crohns and Colitis Australia to be about 75000.

ESC considered that the population in the proposed item descriptor is too broad and advised that the item descriptor should specify patients who have similar symptoms between IBS and IBD where malignancy, infection, and rectal bleeding have been excluded, where other tests are normal and alarm symptoms (which should be listed in the descriptor) are absent. ESC considered that the negative predictive value (NPV) of the test is highest in this population.

ESC noted the difficulty in confining the tested population to the target group (those with reasonable pre-test probability) based on clinical criteria set out in MBS descriptors, particularly for GP requested tests. ESC discussed whether limiting the test to specialist providers (or in consultation with specialists) was an option, and considered that this may have some impact on the economic evaluation due to the offsets in the model (colonoscopies avoided), but noted that the model assumes a 90% rate of colonoscopy procedures following specialist referral.

ESC advised that age could be an exclusion and that the test be limited to the under 50 years age group, noting that guidelines from the National Institutes of Health and Clinical Excellence ([NICE; Quality Standard QS81; February 2015](#)) use 40 years of age as a recommended threshold (in some localised pathways).

ESC noted that no justification had been provided for the repeat testing intervals proposed in the item descriptor (up to three tests per two year period for differential diagnosis, and up to five tests per year for monitoring). ESC considered that as IBD is fluctuating and episodic, there was variation around the frequency of testing required and advised that sensitivity analysis around the frequency of testing would be helpful for further decision making.

ESC noted that no justification had been provided for the proposed fee of \$80 per test.

ESC considered that the specified minimum time for symptom duration (six weeks) is too broad to rule out other diseases. ESC considered that the model outcomes were based on the assumption that all other tests to rule out IBD had been performed, but this may not reflect clinical practice.

ESC acknowledged that the Gastroenterological Society of Australia (GESA) had provided extensive feedback on this submission.

ESC queried whether the patient experience of symptoms would be benefited by this test, as there is no consumer preference information available.

ESC Key ISSUES	ESC ADVICE
The population in the item descriptor is too broad	Narrow population down to where malignancy, infection and PR bleeding have been excluded and age less than 50. Narrow population down to where other tests are normal (coeliac disease etc.), and no alarm symptoms (and say what these are)
The included studies are heterogenous and not GP based	Accept effectiveness evidence and assume that negative predictive values will be higher in GP based population
POCT vs Lab testing	Should be addressed separately but rationale for POCT not clear.
FC test for monitoring	In monitoring, FC test is more sensitive but information about specificity is difficult to interpret Repeat testing intervals unclear in monitoring
Is there sufficient clinical utility in FC given the uncertainty around the evidence base?	Evidence-based guidelines to support this recommendation for use in primary and secondary care (as per NICE guidelines) <a href="https://www.nice.org.uk/sharedlearning/evaluation-of-guidelines-for-the-use-of-faecal-calprotectin-testing-in-primary-care">https://www.nice.org.uk/sharedlearning/evaluation-of-guidelines-for-the-use-of-faecal-calprotectin-testing-in-primary-care</a>

## 15. Other significant factors

Nil

## 16. Applicant's comments on MSAC's Public Summary Document

In response to the recommendations made, we will be resubmitting this application. This resubmission will focus on the Calprotectin (CP) test for diagnosis of IBD rather than for monitoring. As per MSACs recommendations, at this point, this will include:

1. a clearer definition of the appropriate target population;
  - a. Response: This will include Faecal Calprotectin testing of patients aged 3 years to 50 years of age (inclusive) presenting with chronic (more than 6 weeks' duration) gastrointestinal symptoms which are suggestive of either inflammatory or functional bowel disease, where infectious causes have been excluded.
  - b. Response: This will not include POCT.
2. more relevant and robust test performance data in a lower prevalence setting similar to that of the proposed target population, or more appropriate modelling of the likely operating point with its prediction area in summary receiver operating characteristic (SROC) curves;
  - a. Response: We will provide data from GP settings. This will be gained directly from GESA and via published literature since the original submission.
3. more realistic assumptions for the downstream consequences of FC testing versus no FC testing (such as for the estimated reduction in the extent of use of subsequent colonoscopy);
  - a. Response: As above this will be provided via updated research conducted by GESA.

4. appropriately amended economic model structure and inputs, which may simplify (if changes in health outcomes are estimated to be small) to a cost-minimisation analysis where the extra costs of add-on FC testing are equally offset by the estimated reduction in the proportion of patients undergoing colonoscopy.
  - a. Response: We would envision it will be a cost-minimisation in avoided colonoscopy

## **17. Further information on MSAC**

MSAC Terms of Reference and other information are available on the MSAC Website:  
[visit the MSAC website](#)