National Institute for Health and Care Excellence

DIAGNOSTICS ASSESSMENT PROGRAMME

Evidence overview

PredictSURE IBD and IBDX to guide personalised treatment of Crohn's disease

This overview summarises the key issues for the diagnostics advisory committee's consideration. It is intended to be used with NICE's final scope for the assessment and the diagnostics assessment report. A glossary of terms can be found in appendix B.

1 Background

1.1 Introduction

The purpose of this assessment is to evaluate the clinical and cost effectiveness of using PredictSURE IBD or IBDX to predict disease course (severe or mild) and guide personalised treatment of Crohn's disease. The anticipated use of these tests in the NHS would be in people with Crohn's disease who:

- have newly or recently diagnosed disease
- have moderate to severe active disease
- are currently not receiving any concomitant steroids, immunomodulators or biologic therapies
- would not receive top-down therapy with current standard care in the NHS.

Crohn's disease

Crohn's disease is a relapsing-remitting form of inflammatory bowel disease (IBD). Some people with Crohn's disease have frequent flares and relapses. Their disease often does not respond to standard drug therapy, despite

multiple treatment escalations, for example dose escalations and add-on therapies including immunosuppressants and biologics. These people are at a high risk of severe complications (such as intestinal obstruction, fistulae or perianal disease), progressive disability and surgery.

Other people with Crohn's disease can have a prolonged remission without the need for treatment escalations and have good long-term outcomes.

Treatment

People with Crohn's disease usually follow a 'step-up' therapy. This involves titrating and escalating drugs on a trial and error basis. Therapy starts with corticosteroids, then immunosuppressants, then biological therapies if the disease does not respond, or loses response, to treatment.

However, this strategy is not suitable for everyone with Crohn's disease. It may take a long time for people with refractory or relapsing disease to have adequate control of their disease. This puts them at increased risk of irreversible bowel damage, serious complications and the need for surgery. Some suggest these people could benefit from top-down treatment: early treatment with biological therapies such as tumour necrosis factor (TNF)-alpha inhibitors. This early treatment could achieve a faster and higher rate of mucosal healing and has the potential to modifying the natural course of disease. Early adequate control of disease activity could lead to fewer disease flare-ups, prevent bowel damage, and limit the need for surgery in patients with severe disease.

Top-down treatment reverses the order of treatment in step-up treatment, starting with biologics, then immunosuppressants then steroids. Top-down treatment is not routinely used in the UK. But it is sometimes offered in specialist centres to people thought to have particularly severe disease.

There is a high unmet need to better predict disease course and guide personalised therapy in Crohn's disease, because known risk factors have limited predictive value. This would help people with severe Crohn's disease

get the right kind of treatment and monitoring to adequately control their disease and avoid over-treating and over-monitoring people who do not have severe disease.

PredictSURE IBD and IBDX

The diagnostics assessment report considers 2 tests that could help predict disease course and guide therapy: PredictSURE IBD and IBDX, which have different mechanisms of action.

PredictSURE IBD is a blood test that detects the genomic signature of CD8+ T-cell exhaustion. CD8+ T cells are part of the immune system that regulates immune response. T-cell exhaustion is a state characterised by T-cell dysfunction, which lowers T-cell-related immune response. People with a non-exhausted CD8+ T-cell signature were linked to a higher risk of frequently relapsing disease course than people with an exhausted signature. The CD8+ T-cell exhaustion signature was first identified in isolated peripheral blood mononuclear cells (PBMC). The blood-based PredictSURE IBD test was developed to avoid the need for cell separation, to make it suitable for routine clinical use.

IBDX is a panel of 6 indirect solid-phase enzyme-linked immunosorbent assay (ELISA) kits measuring levels of anti-glycan antibodies in human serum:

- IBDX anti-Saccharomyces cerevisiae (gASCA) IgG ELISA kit
- IBDX anti-laminaribioside (ALCA) IgG ELISA kit
- IBDX anti-chitobioside (ACCA) IgA ELISA kit
- IBDX anti-mannobioside (AMCA) IgG ELISA kit
- IBDX anti-chitin (anti-C) IgA ELISA kit
- IBDX anti-laminarin (anti-L) IgA ELISA kit

Anti-glycan antibodies are markers of seroreactivity to microbial antigens, which were reported to be linked to Crohn's disease prognosis.

Provisional recommendations on these technologies will be made by the diagnostics advisory committee at the committee meeting on 6 August 2020.

1.2 Scope of the assessment

Table 1 Scope of the assessment

Decision question	Is testing with PredictSURE IBD or IBDX in people with active Crohn's disease a clinically and cost-effective use of NHS resources?
Populations	People with active Crohn's disease who are: • currently not receiving any concomitant steroids, immunomodulators or biologic therapies • have newly or recently diagnosed disease • have moderate to severe active disease • would not receive top-down therapy with current standard care in the NHS.
Interventions	PredictSURE IBD and IBDX
Comparator	Current clinical practice in the NHS: most people with Crohn's disease are offered an accelerated step-up treatment (also referred to simply as step-up therapy). No test or algorithm is currently used to predict disease course.
Healthcare setting	Secondary and tertiary care
Intermediate outcomes	Intermediate measures for consideration may include: • time to result • number of test failures • number of inconclusive test results • percentage of patients classified as high and low risk of frequently relapsing disease course • percentage of patients who were offered top-down therapy • test accuracy (sensitivity, specificity, positive predictive value, negative predictive value and hazard ratios for predicting severe disease course).
Clinical outcomes	Clinical outcomes for consideration may include: • rates and duration of response and remission • rates and duration of flare-ups and relapses • rates and duration of remission free from steroids and surgery • cumulative steroid exposure • measures of mucosal healing • rates of and time to treatment escalation • rates of and time to hospitalisation • rates of and time to surgical intervention

	• rates of and time to serious complication • composite outcomes such as major adverse outcomes (hospitalisation, surgery or serious complication) • adverse effects of treatment.
Patient-reported outcomes	Patient-reported outcomes for consideration may include health-related quality of life.
Costs	Costs will be considered from an NHS and personal social services perspective. Costs for consideration may include: • cost of testing (including the cost of sample collection, processing, transport, and the testing service) • cost of treatment (including biologics) • costs of other resource use (for example associated with managing active disease states, flare-ups or complications) • outpatient appointments • hospitalisation • additional tests • surgery.
Time horizon	The time horizon for estimating clinical and cost effectiveness should be long enough to reflect any differences in costs or outcomes between the technologies being compared.

The cost effectiveness of interventions should be expressed in terms of incremental cost per quality-adjusted life year.

Further details including descriptions of the interventions, comparators, care pathway and outcomes can be found in the final scope.

2 The evidence

This section summarises data from the diagnostics assessment report compiled by the external assessment group (EAG).

2.1 Clinical effectiveness

The EAG did a systematic review based on the review protocol, to identify evidence on the diagnostic performance and clinical effectiveness of PredictSURE IBD and IBDX tests to predict severe disease course and guide personalised treatment in people with Crohn's disease. Full details of the inclusion and exclusion criteria start on page 18 of the diagnostics assessment report. The comparator was defined as current standard care, in

which no tool is used to predict disease course and most patients receive standard ‘step-up’ therapy. The methodological quality of the included studies was assessed using the QUIPS (Quality of Prognosis Studies in Systematic Reviews) tool.

The EAG found 8 primary studies (reported in 12 publications) that met the inclusion criteria, 1 reporting the diagnostic performance of PredictSURE IBD, and 7 reporting the performance of IBDX. In addition, the EAG found 3 literature reviews (reported in 4 publications) whose citation lists were searched. All primary studies were observational in design, and all but one were done in an adult population (1 study enrolled children). The PredictSURE IBD study enrolled mostly newly diagnosed patients, while IBDX studies enrolled people with long-lasting Crohn’s disease. Different studies used different definitions of severe disease course. The studies reported the accuracy for predicting severe disease course (according to whether patients classified as high and low risk subsequently developed a severe disease course), hazard ratios (HRs) for following a severe disease course for people classified as high compared with low risk, or both outcomes.

Evidence for IBDX

A total of 7 studies reported on IBDX. Two were prospective prognostic studies. Five reported on the correlation between positive status for IBDX biomarkers and a presence or history of complications or surgery. None of the studies was done in people with newly diagnosed Crohn’s disease. Median duration of disease at the time of testing ranged from 10.6 (interquartile range [IQR] 1.7 to 52.3) months to 9.4 (IQR 1 to 44) years (duration of disease was not reported in 2 studies). Severity of disease activity at the time of testing was not reported in any study. All studies were considered to be at a moderate or unclear risk of bias in the measurement of confounding factors domain. Three studies were considered to be at a moderate risk of bias in the participation domain of the QUIPS tool.

A prospective German study by Rieder et al. (2010c) assessed the IBDX panel for predicting complications (fistulas or stenoses) and Crohn's disease-related surgery (Table 2). The median duration of disease at baseline for 76 people with no prior complication or surgery was 10.6 (IQR 1.7 to 52.3) months. The median follow up for this group was 53.7 months. During this time, 20 people developed a complication and 14 people needed surgery (23 had either surgery or complication). People who tested positive for 2 or more out of 6 IBDX markers had a significantly higher risk of complications (HR 2.5; 95% confidence intervals [CI] 1.03 to 6.1; p=0.043), or surgery (HR 3.6; 95% CI 1.2 to 11.0; p=0.023). The EAG highlighted the small sample size informing the estimate of risk.

A study by Wolfel et al. 2017 (reported as a conference abstract only) enrolled 118 people who had undergone a surgical intestinal resection related to Crohn's disease. Most people (92%), had undergone surgery because of internal penetrating, stricturing disease or both. Median duration of disease was not reported. All 6 IBDX markers were measured after surgery. After a median follow up of 100 months, neither the number of positive biomarkers combined nor the quartile sum score (definition not provided) predicted a shorter time to repeat intestinal surgery.

Table 2 Key study characteristics and results: IBDX prospective studies (all 6 IBDX markers)

Study	Population	Outcomes
Rieder et al. 2010c (full publication) Germany	76 people with Crohn's disease and no prior complication or surgery. Median duration of disease at baseline: 10.6 (IQR 1.7 to 52.3) months. Median follow up: 53.7 months.	20 people had a complication, 14 people had surgery (23 people had a complication or surgery). Analyses were adjusted for age, sex, BMI, disease activity and duration, age at diagnosis and disease location. HR (95% CI) for complication, by the number of positive markers: • 1 or more: 1.8 (0.61 to 5.4), p=0.29 • 2 or more: 2.5 (1.03 to 6.1); p=0.043 • 3 or more: 2.6 (0.92 to 7.2); p=0.072 HR (95% CI) for surgery, by the number of positive markers: • 1 or more: 2.6 (0.58 to 12.0); p=0.21 • 2 or more: 3.6 (1.2 to 11.0); p=0.023 • 3 or more: 2.8 (0.80 to 9.6); p=0.11 HR (95% CI) for complication or surgery, by the number of positive markers: • 1 or more: 2.2 (0.74 to 6.5); p=0.16 • 2 or more: 2.8 (1.2 to 6.4); p=0.016 • 3 or more: 3.1 (1.2 to 8.1); p=0.019
Wolfel et al. 2017 (conference abstract) Location unclear	118 people with Crohn's disease who had one surgical resection. Duration of disease not reported. Median follow up 100 months.	Neither the quartile sum score nor the number of positive markers combined predicted a shorter time to surgery.

Abbreviations: BMI, body mass index; HR, hazard ratio; IQR, interquartile range.

An additional 5 studies reported on correlations between the IBDX markers and a history of surgery, complications, or both. Four studies were cross-sectional and 1 had a prospective design but reported a correlation with complications or the need for surgery before or during follow up combined.

A German study by Rieder et al. (2010b) enrolled 363 people with Crohn's disease, with a median duration of disease of 66.8 (IQR 11 to 141) months. At baseline, 186 (51.2%) people had fistulas, and 90 (24.8%) had strictures. An

additional 28 people developed a complication during the median follow up of 59 months. 257 (70.8%) people needed inflammatory bowel disease-related surgery (224 at baseline and 33 during follow up). Based on an analysis of 75 people who had at least 4 serum samples during follow up, the study reported that people who had surgery (before or during the follow up) had a higher number of positive IBDX markers (median 2.0 [1.0 to 3.0]) than those who did not (median 1.0 [0.0 to 2.0]; odds ratio [OR] 1.5 [95% CI 1.3 to 1.8]; p<0.001). Similarly, people with a complication (before or during follow up) had a higher number of positive IBDX markers (median 2.0 [1.0 to 3.0]) than those who did not (median, 0.0 [0.0 to 2.0]; OR, 1.5 [95% CI: 1.3 to 1.9]; p<0.001).

A cross-sectional Canadian study by Seow et al. (2009) enrolled 517 people with Crohn's disease who had a median duration of disease of 8.9 (IQR 0.02 to 46.30) years. The study reported a correlation between the increasing rate of surgery and the increasing number of positive markers (from 51.64% of people with 1 positive marker, to 76.67% of people with 5 to 6 positive markers; p<0.0001 across the categories).

A cross-sectional US study by Harrell et al. (2010; reported as a conference abstract only) enrolled 172 people with Crohn's disease (duration of disease was not reported). 113 (66.1%) patients had complicated disease behaviour, defined as intestinal fistula, stricture, or both, with or without the need for surgery. A higher number of positive IBDX markers (7-marker test: IBDX markers plus IgG ASCA) was significantly associated with complicated disease behaviour, need for surgery or both (OR 3.3; 95% CI not reported; p=0.0005).

A cross-sectional French study by Paul et al. (2015) enrolled 107 people with Crohn's disease who had a diagnosis for more than 1 year (median 9.4 years; IQR 1 to 44 years). The study did not assess the diagnostic accuracy of the entire IBDX panel and instead assessed the diagnostic accuracy of the 6 individual markers. Positivity for the ASCA and AMCA antibodies had the

best validity for differentiating severe from non-severe course of Crohn's disease, with an area under the curve (AUC) of 0.63 and 0.65, respectively. Combining ASCA and AMCA made differentiating severe from non-severe course of Crohn's disease more precise, with an AUC of 0.71. Of the panel of tested markers, only AMCA antibodies tended to be associated with a higher risk of Crohn's disease-related surgery, with an OR of 2.1 (95% CI 0.8 to 5.1; p=0.10) but this was not statistically significant.

A cross-sectional German study by Rieder et al. (2012) enrolled 59 children (under 18) with Crohn's disease who had a median duration of disease of 18.0 (IQR 12.0 to 43.0) months. The authors concluded that results for the younger cohort were aligned with those from an adult cohort. A higher number of positive serum biomarkers was associated with an increased risk of complicated Crohn's disease and with needing Crohn's disease-related surgery (estimates of effect not reported).

Rieder et al. (2011) was a longitudinal analysis that did not meet the inclusion criteria for the review but was thought to provide useful information. This study assessed whether levels of the individual biomarkers fluctuate over time. The authors reported that, despite marked changes in overall immune response and in levels of individual biomarkers over a median follow up of 17.4 months (IQR 8.0 to 31.6 months), the status of positivity or negativity for an individual biomarker remained mostly stable over time.

Evidence for PredictSURE IBD

One study reported on the prognostic ability of PredictSURE IBD (Biasci et al. 2019). The study was considered to be at low or unclear risk of bias according to the QUIPS tool.

The study enrolled people aged 18 years and over with active Crohn's disease or ulcerative colitis who were not receiving concomitant glucocorticosteroids, immunomodulators or biological therapies at 4 UK centres. Active disease was confirmed by at least 1 objective marker (raised C-reactive protein [CRP], raised calprotectin or endoscopic signs of active

disease) in addition to active symptoms. All people were tested with PredictSURE IBD and classified into either a high or low-risk group by the algorithm. All people received conventional step-up therapy in accordance with national and international guidelines, regardless of the test results and risk categorisation (clinicians were blinded to gene expression analyses). The study reported results for the following cohorts (all had the outcome data prospectively collected):

- CD8 T-cell cohort (66 people with Crohn's disease): patients who were classified into IBD1 (corresponding to low T-cell exhaustion) and IBD2 (corresponding to high T-cell exhaustion) based on testing isolated PBMC. These patients were not tested using PredictSURE IBD so results reported for this cohort are not relevant to this assessment.
- Training cohort (39 people with Crohn's disease): patients whose data were used to develop a whole blood gene signature underlying the PredictSURE IBD test (designed to give the same results as the test on PBMC, but to be more suitable for routine laboratory testing). All patients were then retested using PredictSURE IBD to classify patients as high and low risk.
- Validation cohort (66 people with Crohn's disease): patients who were classified as high and low risk based on the results of the PredictSURE IBD test.

Most people in the validation cohort (61 of 66 [92.4%]) had newly diagnosed Crohn's disease. 27 (40.9%) were categorised as at high risk of severe Crohn's disease (IBDHi), and 39 (59.1%) were categorised as low risk (IBDLo). Median duration of follow up was 1.6 (IQR 1.0 to 3.7) years in the high-risk group and 2.4 (IQR 1.8 to 3.8) years in the low-risk group. People categorised as IBDHi had a statistically significantly higher risk of first treatment escalation compared with those designated as IBDLo, with a HR of 2.65 (95% CI 1.32 to 5.34; p=0.006). Sensitivity and specificity for predicting the need for multiple (2 or more) escalations within the first 12 months were 77.8% and 70.6%, respectively, and within 18 months, 72.7% and 73.2%, respectively. Negative predictive value for predicting multiple escalations

within the first 18 months was 90.9%. Positive predictive value was 42.1% (calculated by the EAG based on a 2X2 table provided by the company).

The EAG highlighted a number of limitations and key considerations in the study, discussed further in section 4.

Comparative evidence for IBDX and PredictSURE IBD

A study comparing PredictSURE IBD and IBDX's abilities to predict the need for multiple treatment escalations (Lyons 2020) was presented at the European Crohn's and Colitis Organisation's (ECCO) conference in February 2020.

A subset of people from the training and validation cohorts from the PredictSURE IBD study, all recruited from a specialist IBD clinic at Addenbrooke's Hospital, Cambridge, were tested for the panel of IBDX markers. 74 had active Crohn's disease, of whom 59 (80%) were newly diagnosed. As per the study design described in section 2.1 (evidence for PredictSURE IBD), all enrolled patients had active disease at enrolment, and all received accelerated step-up treatment. IBDX and PredictSURE IBD tests were done according to manufacturers' instructions, using blood serum compared with whole blood RNA samples.

43 (58%) patients tested positive for at least 1 IBDX marker and 14 (19%) patients tested positive for 2 or more markers (when treating equivocal results as negative). A smaller proportion of people with newly diagnosed disease were positive for any IBDX marker than people with an established diagnosis (30 out of 59 [51%] compared with 13 out of 15 [87%]; p=0.018, Fisher's exact test).

The cohort was then stratified into 2 groups based on the number of positive markers: people positive for 2 or more markers compared with people positive for only 1 or 0 markers. No significant differences between the 2 groups were identified in terms of time to, or frequency of, treatment escalations. In comparison, when this cohort was stratified by PredictSURE IBD, people

classed as high risk had a significantly shorter time to treatment escalation than people classed as low risk (p=0.001).

2.2 Costs and cost effectiveness

The EAG did a search to identify existing studies investigating the cost effectiveness of PredictSURE IBD and IBDX in Crohn's disease and economic evaluations of treatments for people with newly diagnosed, moderate to severe Crohn's disease. The EAG also constructed a de novo economic model to assess the cost effectiveness of PredictSURE IBD and IBDX to guide treatment choices in Crohn's disease.

Systematic review of cost-effectiveness evidence

The EAG did not identify any published economic studies for PredictSURE IBD or IBDX but found 11 relevant evaluations for treatment options in Crohn's disease.

One study (Marchetti et al. 2013) specifically compared the cost effectiveness of top-down (1st step infliximab plus azathioprine, 2nd step additional infliximab plus azathioprine, 3rd step methylprednisolone plus azathioprine) and step-up (1st step methylprednisolone, 2nd step methylprednisolone plus azathioprine, 3rd step infliximab plus azathioprine) approaches in Italy. The base-case analysis of this study showed that a top-down strategy was associated with a quality-adjusted life year (QALY) gain of 0.14 and savings of €773, making it dominant against the step-up strategy in newly diagnosed Crohn's disease patients. The treatment strategy modelled in Marchetti is not representative of UK NHS practice.

One UK study compared 9 induction treatment sequences (composed of 4 treatment lines) for Crohn's disease (see the health economics report for NICE's guideline on Crohn's disease), and the remaining 9 studies compared individual treatment steps.

PredictImmune also submitted the abstract of an economic study for PredictSURE IBD in Crohn's disease and ulcerative colitis. The study results

were presented at ECCO conference in February 2020. Study methods are presented in section 4.1.2 of the diagnostics assessment report, starting on page 57. The results of the study show that, over a 15-year time horizon, top-down treatment guided by PredictSURE IBD produced an incremental cost-effectiveness ratio (ICER) of £7,179 per QALY gained when compared with standard care (Buchanan et al. 2020).

Economic analysis

The external assessment group (EAG) developed a de novo economic model, as described from page 60 of the diagnostics assessment report, to assess the cost effectiveness of PredictSURE IBD and IBDX tests to guide therapy in adults (16 and over) with Crohn's disease who have:

- been newly diagnosed with Crohn's disease,
- moderate to severe active disease, and
- not been offered biologics under current standard care.

The comparator was standard care, in which no test or algorithm is available to determine the long-term disease course. Instead, prognosis is based on clinical judgement of presenting signs and symptoms and known clinical risk factors for complicated disease course.

The population in the economic model was based on the cohort of patients with Crohn's disease from Biasci et al. (2019), for which anonymised individual patient data were shared by the manufacturer. The cohort consisted of 105 people with Crohn's disease of whom 88 were newly diagnosed. 39 people (high and low risk split not reported) were from the training cohort and 66 people (41% high risk and 59% low risk) were from the validation cohort. The EAG's analysis was based on 40 patients ([redacted] [58%] high-risk patients and [redacted] [42%] low-risk patients) whose treatment matched the standard definition of step-up therapy, that is, people who received first-line therapy with corticosteroids, and second-line treatment with immunomodulators (after failure of corticosteroids). So the EAG excluded

people who received other first-line therapies [redacted], did not receive any treatment escalation from corticosteroids to immunomodulators, or received other treatment as their second-line treatment (such as [redacted].

No detailed data were available for IBDX and therefore its cost effectiveness was assessed as an exploratory scenario analysis only. This analysis assumed the same efficacy as PredictSURE IBD, which may be unlikely because the 2 tests have different mechanisms of action and used different definitions of severe disease in their supportive studies.

Model structure

The model took the perspective of the NHS and personal social services. Both costs and benefits were discounted at 3.5% per year. The model had a lifetime time horizon (65 years) with a cycle length of 2 weeks.

The EAG assumed that PredictSURE IBD (and IBDX in the scenario analysis) categorises patients into high and low risk for following a severe disease course. People categorised as high risk receive a top-down treatment strategy, while those classified as low risk receive step-up therapy aligned with current standard care (that is, without a prognostic test all people receive step-up therapy). Therefore, only people classified as high risk receive differential treatment compared with the 'no testing' strategy, and the model effectively is comparing the benefits of a top-down strategy with a step-up strategy in the high-risk population. The low-risk populations in the step-up and top-down arms of the model cancel out because they receive the same treatment.

The EAG assumed that both top-down and step-up treatment strategies start with prednisolone, a corticosteroid (Figure 1) but this has not been modelled for either arm. Clinical advice received by the EAG suggests that the moderate to severe form of Crohn's disease is highly unlikely to respond to treatment with corticosteroids. So the EAG made a simplifying assumption that corticosteroids fail for all high-risk patients. When induction therapy with

steroids fails, people in the step-up strategy have their treatment escalated to immunomodulators, and when this treatment fails, to anti-TNF therapy. In comparison, people in the top-down strategy escalate directly to anti-TNF therapy once the treatment with corticosteroids fails. The EAG gathered from clinical experts that immunomodulators would not be given after biologics in the top-down arm. However, it did explore a scenario in which immunomodulators were included as the last treatment option after biologics in the top-down arm.

The EAG adopted a hybrid modelling approach, with a decision tree for the induction treatment (Figure 2), and Markov transition model for the maintenance treatment (Figure 3). In the induction model, patients whose disease does not respond (deterioration; no change; or an improvement of less than 70 in Crohn's Disease Activity Index [CDAI] score) receive second-line treatment, according to their treatment allocation (top down or step up). Patients whose disease responds to the induction treatment (an improvement in CDAI score above 70) move to the maintenance model. They can enter the maintenance model either in remission, mild, or moderate to severe health states. Patients can then move between these states during maintenance therapy, reflecting the different levels of response to maintenance therapy. Patients in the mild and moderate to severe states are at risk of relapse and escalating to the next treatment step. The probability of patients transitioning between health states depends on the treatment class received (see details starting on page 97 of the diagnostics assessment report).

Surgical events are modelled as a stand-alone outcome. This means that patients do not explicitly leave their health state in a specific cycle to move to the surgery state. Instead, in every model cycle, a proportion of surgery is estimated, and the associated costs and temporary impact on patients' quality of life is calculated (see page 126 of the diagnostics assessment report for details). The model does not account for any long-term effects of surgery on patients' quality of life.

Adverse events were not included in the model but the EAG considered that this should have a limited effect on the model results.

Death is the absorbing state in the model (see page 109 of the diagnostics assessment report for details of mortality assumptions).

Figure 1 Top-down treatment strategy compared with step-up treatment strategy

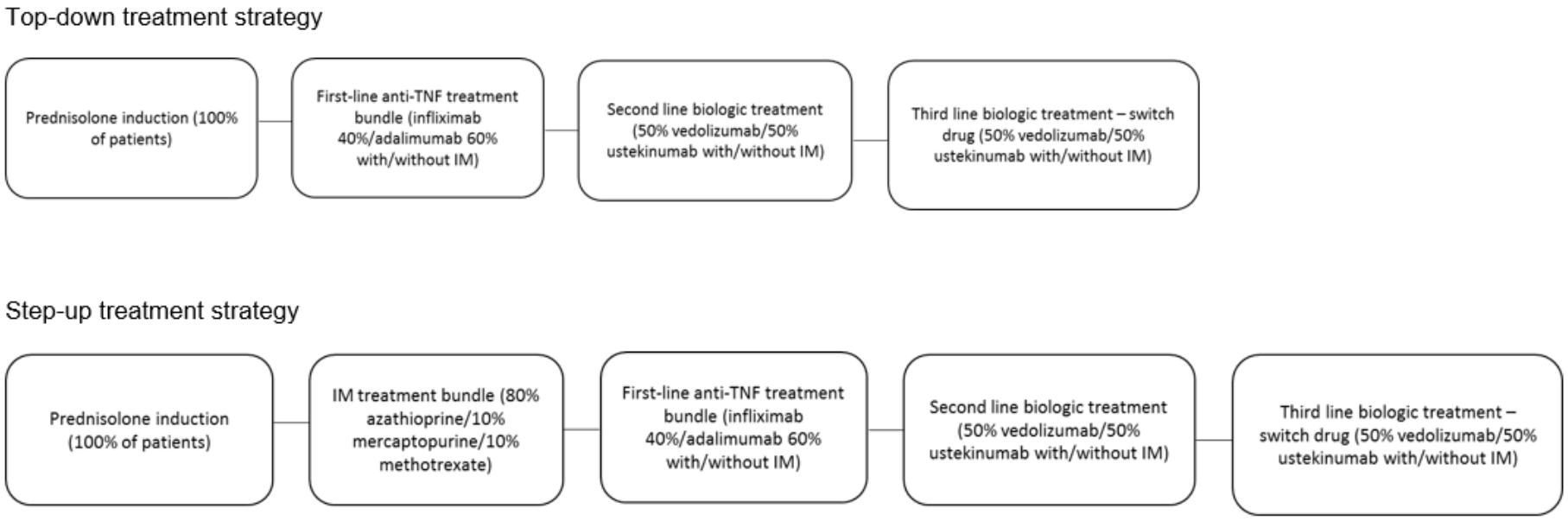
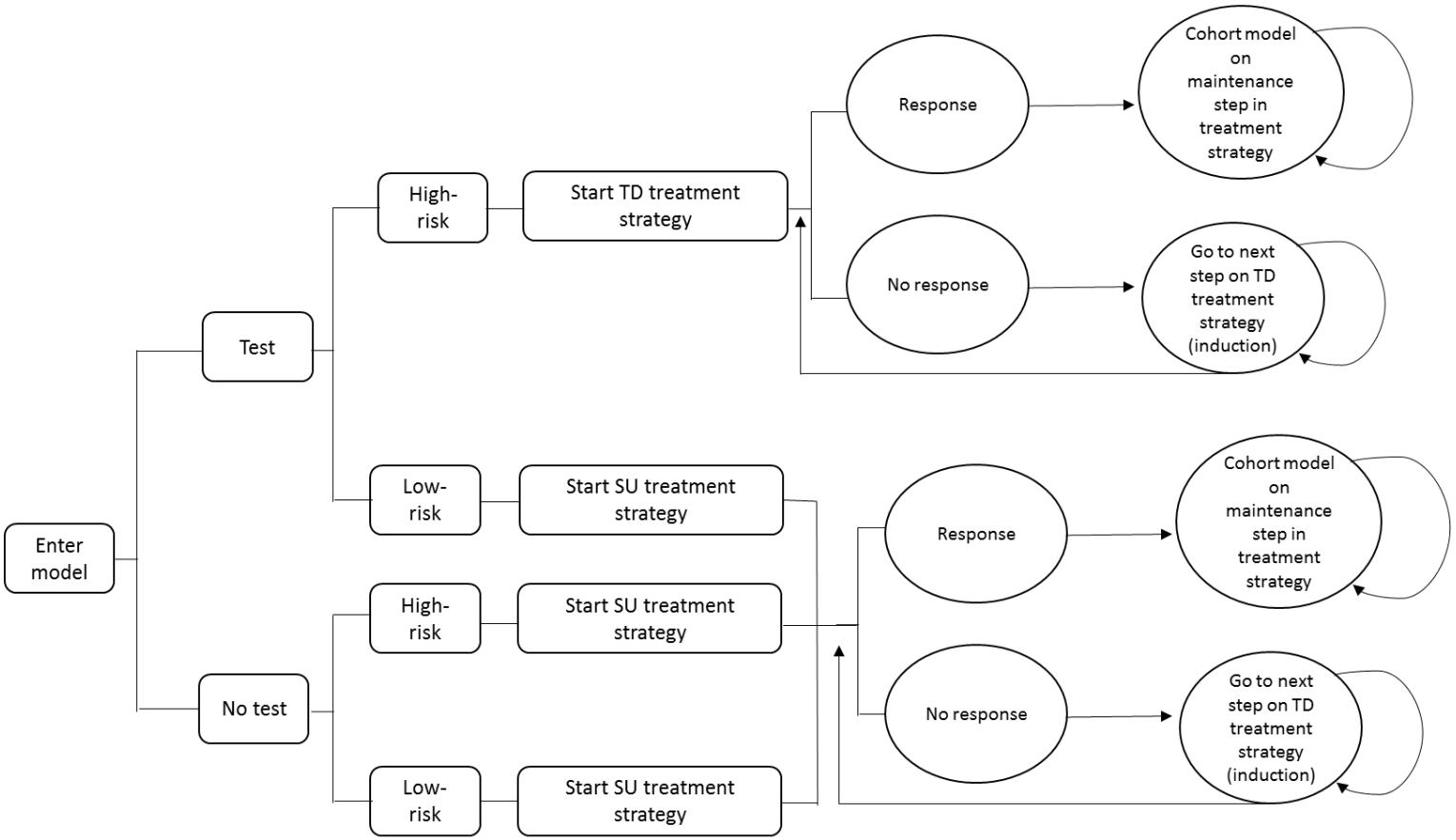
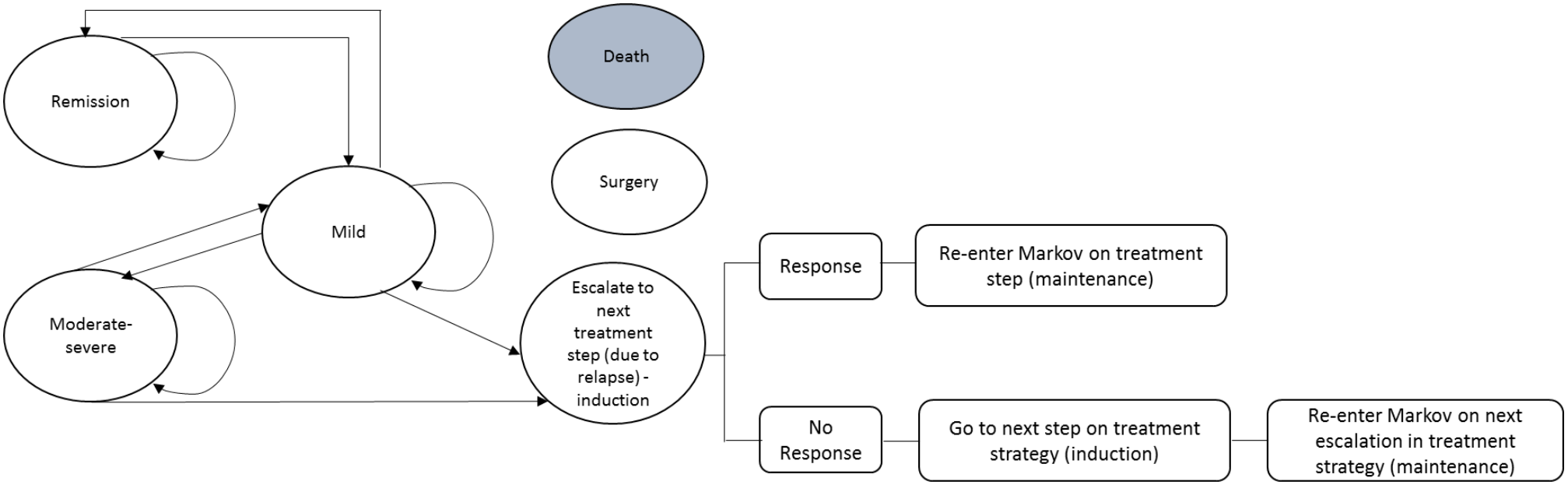


Figure 2 Model for induction treatment



Abbreviations: SU, step up; TD, top down.

Figure 3 Cohort model for top down and step up maintenance steps



Remission = CDAI of 150 or below; mild disease activity = CDAI between 150 and 220; moderate to severe disease activity = CDAI between 220 and 600.

Clinical inputs

No evidence was identified on the clinical effectiveness of top-down compared with step-up therapy in high-risk patients. Therefore, a linked evidence approach was taken, with the following studies informing key clinical inputs:

- Biasci et al. (2019; described in section 2.1) for time to treatment escalation (as described in the model; corresponding to time to second treatment escalation in the individual patient data) and time to surgery in high-risk compared with low-risk patients. Time to treatment escalation in the initial base case did not reset with every escalation, however, the revised base case allows the time to treatment escalation to reset with every escalation (see page 7 of the diagnostics assessment report addendum for details).
- D'Haens et al. (2008) and its 10-year follow-up study by Hoekman et al. (2018) for effectiveness of top-down compared with step-up therapy.

The study by D'Haens et al. (2008) was a 2-year, multicentre, open-label randomised trial that evaluated the clinical efficacy of early combined immunosuppression (top-down therapy; n=67) compared with conventional treatment (step-up therapy; n=66) in people with newly diagnosed Crohn's disease. People randomised to top-down therapy received induction treatment with infliximab and azathioprine. Patients received no infliximab maintenance but were allowed infliximab as needed and, if necessary, corticosteroids, to control disease activity. People randomised to step-up therapy received corticosteroids, followed, in sequence, by azathioprine and infliximab. The primary outcome measures were remission without corticosteroids and without bowel resection at weeks 26 and 52. The Hoekman et al. (2018) study was a retrospective review of the medical records of patients included in the D'Haens et al. (2008) trial, which collected data on hospitalisation, flares, surgery, clinical activity, and other outcomes, for a median follow up of 10 years. The top-down therapy in this study differed from that assumed by the EAG in the model (people in the top-down group did not receive initial corticosteroids in the study). Also, people were not stratified by their risk of

following a severe disease course, and the treatment sequence did not include maintenance treatment with infliximab, which is now part of standard infliximab treatment. Therefore, the study has only limited applicability to this assessment.

The EAG made the following assumptions in their base-case analysis to use the D'Haens et al. (2008) data:

- Randomisation resulted in balanced populations of high and low-risk patients within each treatment group.
- The relative treatment effect of top-down and step-up strategies in a mixed-risk population is the same as the relative treatment effect in a high-risk population.
- Time to relapse is a proxy measure for time to next treatment escalation.
- The effectiveness of the treatment strategies in D'Haens et al. (2008) is a proxy for the treatment effectiveness of the first step in the top-down and step-up strategies modelled.

Time to treatment escalation

The EAG assumed that treatment escalations correspond to a relapse (or a flare) to patients' current treatment. Time to first treatment escalation in the step-up treatment strategy was based on the time to second treatment escalation (from immunomodulator to anti-TNF) from the individual patient data for [redacted] high-risk patients (who had [redacted] escalation events) and [redacted] low-risk patients (who had [redacted] escalations). Patients who did not have the treatment escalation were censored (Figure 4, Table 3). Results were extrapolated, and lognormal and Gompertz curves provided the best fit for the high-risk and low-risk cohorts, respectively. The EAG recognises that applying the lognormal curve to the Kaplan–Meier data from Biasci decreases the probability of treatment escalation as time passes. This implies that the probability of treatment escalation decreases with time but only within a specific treatment step. When a patient moves on to the next treatment in the sequence the probability of treatment escalation resets to be the same as it was when a

patient started their first treatment. Therefore, in the base case, the EAG assumed that patients have the same baseline probability of escalating to the next step in the step-up treatment sequence regardless of the number of previous escalations. This assumption means that the probability of treatment escalation from anti-TNF to other biologics is the same regardless of when in the treatment path it was received. The EAG acknowledged that these assumptions are a simplification of clinical reality, in which time to escalation is likely to depend on the number of previous treatments. The EAG also noted that both the number of patients and events in the analysis were very small, so results of the analysis need to be interpreted with caution.

Figure 4 Time to second treatment escalation estimated by the EAG: Kaplan–Meier and extrapolated data

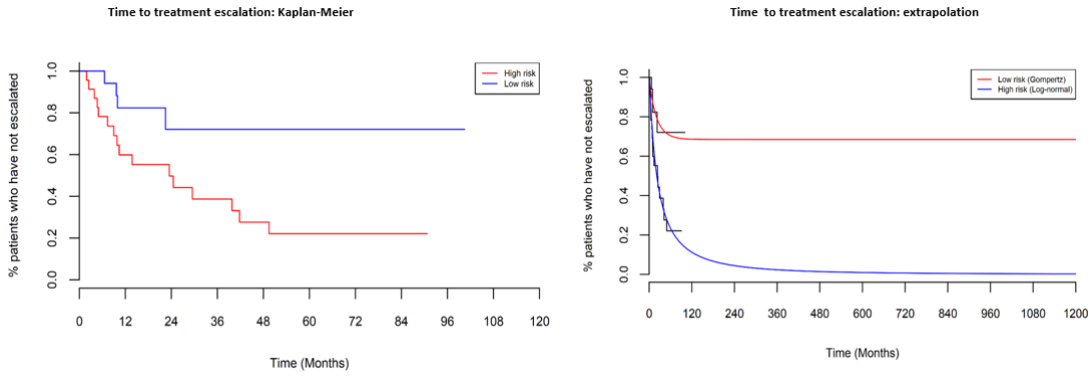


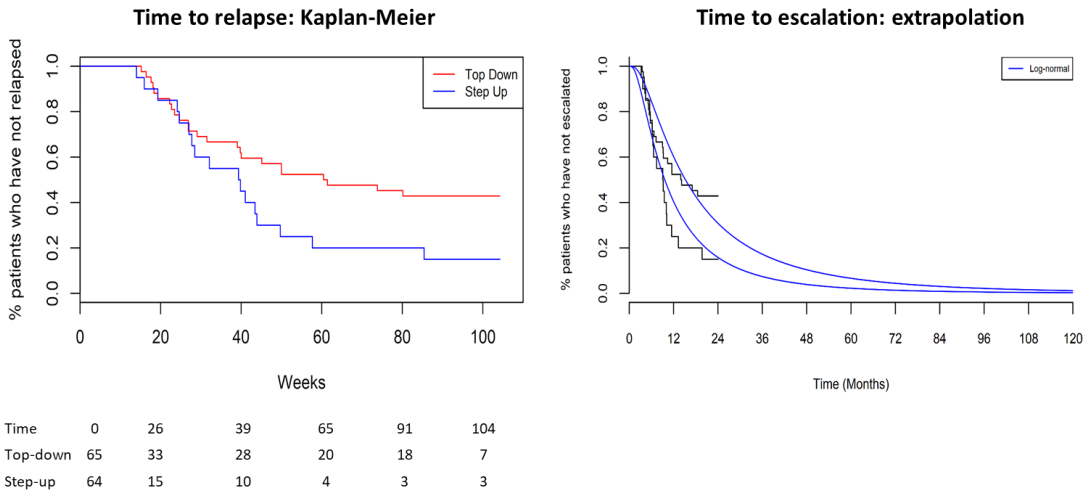
Table 3. Time to treatment escalation for Kaplan-Meier (figure 4 left)

Time (in months)	0	12	24	36	48	60
Low-risk	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
High-risk	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

To estimate the relative treatment effect of top-down and step-up therapy for time to treatment escalation, the EAG digitised the time to relapse Kaplan–Meier data from D'Haens et al. (2008;Figure 5) and used the number of patients at risk provided in the study to simulate the pseudo-individual patient-level data. The EAG chose the lognormal curve to extrapolate the data and estimate the hazard function to apply to the high-risk top down arm of the model (Figure 5; right panel). The relative hazard function was then applied to

time to treatment escalation curves for the first step of the top-down arm (anti-TNF) and the first step of the step-up arm (immunomodulator) of the model. Subsequent treatment steps in both the top-down and step-up strategies were assumed to have the same time to treatment escalation as anti-TNF in the top-down arm. In the revised base case, baseline time to treatment escalation data for step-up resets with every treatment step. There was no change made to measures of relative treatment effectiveness for top down compared with step up.

Figure 5 Kaplan–Meier and lognormal extrapolation of time to relapse (time to treatment escalation)



Time	0	26	39	65	91	104
Top-down	65	33	28	20	18	7
Step-up	64	15	10	4	3	3

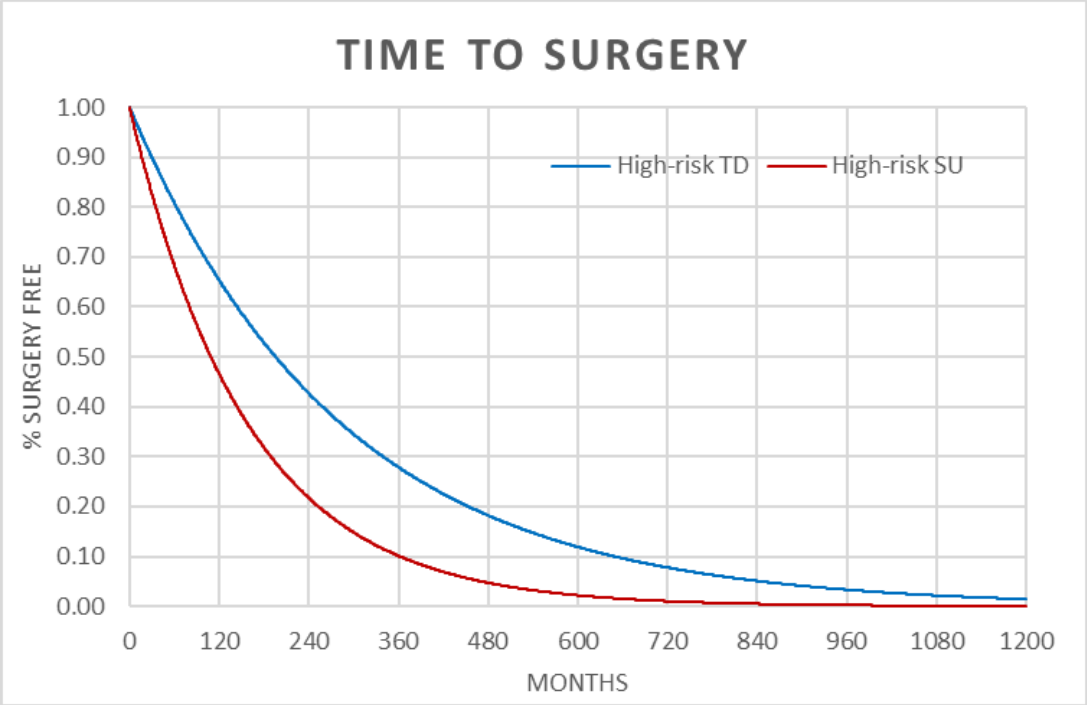
Time to surgery

The model aimed to capture the impact of top-down therapy on the need for surgery in high-risk patients. Clinical expert opinion provided to the EAG indicated that patients with Crohn's disease may undergo surgery for multiple reasons, including as a final treatment option or to treat complications (for example, fistulas or strictures). The EAG acknowledged that surgery might have a beneficial impact on patients' quality of life, which would not be captured in the model. On the other hand, the long-term negative consequences of surgery, such as the need for repeat surgery (potentially resulting in short bowel syndrome) and for total parenteral nutrition, were also not captured in the model.

Time to surgery was based on data from 87 newly diagnosed patients ([redacted] operations) with Crohn's disease from the Biasci et al. (2019) cohort, because 1 patient who had surgery as a first-line treatment option was excluded from the analysis. Because of the small number of operations captured in the low-risk and high-risk cohorts, and no significant difference in surgery between the 2 groups, the EAG combined results from both cohorts to estimate time to surgery. Time to surgery was estimated as a stand-alone outcome in the model (that is, patients do not explicitly leave their health state in a specific cycle to move to the surgery state) because of the lack of data to inform transition probabilities. To avoid double-counting, the EAG adjusted treatment costs, based on the assumption that patients receiving surgery stop their current treatment in the model, and applied a surgery-related disutility to patients' total utility in that model cycle.

To estimate time to surgery in the high-risk top-down cohort, the EAG digitised the Kaplan–Meier data for time to surgery from Hoekman et al. (2018) and extrapolated the results using an exponential model (most plausible clinical fit, based on clinical expert opinion that 50% of patients with Crohn's disease are expected to undergo surgery during the first 10 years from diagnosis, and 25% would receive surgery in the subsequent 5-year period). The study showed no significant differences in the time to surgery across treatment arms. According to study authors, this could be linked to several aspects of study design (lack of statistical power and treatment regimens used). Therefore, these results should be interpreted with caution. The hazard function taken from Hoekman et al. (2018) was applied to Biasci to estimate time to surgery in high-risk top-down cohort (Figure 6).

Figure 6 Time to surgery in high-risk patients: top-down compared with step-up therapy



Abbreviations: SU, step up; TD, top down.

Effectiveness of induction and maintenance therapies

The EAG did a pragmatic literature review to estimate the effectiveness of the therapies included in the model. Probabilities of response and remission with induction and maintenance therapies (see Table 4) were based on data from the pragmatic search and from NICE's technology appraisal guidance on vedolizumab for treating moderately to severely active Crohn's disease after prior therapy. The EAG assumed the same rate of response and remission to biologics and anti-TNF regardless of prior lines of therapy (that is, regardless of if people who had never had immunomodulators or for whom immunomodulators had failed).

Based on NICE's guidance on vedolizumab, the EAG estimated that 21.2% of responders remained in the moderate to severe disease state. Further details on how transition probabilities for induction and maintenance therapy were

estimated can be found starting on page 91 of the diagnostics assessment report.

Table 4 Probabilities of response and remission with induction and maintenance therapies

Treatment strategy	Induction: response	Induction: remission	Maintenance: response	Maintenance: remission
Top down: biologics	32%	13%	2%	28%
Top down: anti-TNF	26%	37%	10%	33%
Step up: biologics	32%	13%	2%	28%
Step up: anti-TNF	26%	37%	10%	33%
Step-up: immunomodulator	23%	16%	15%	25%

Mortality

The EAG assumed that Crohn's disease does not directly impact patients' mortality and used UK general population data, matched for sex and age, to estimate patients' survival in the economic model (Office for National Statistics national life tables for the UK). The EAG assumed that surgery was associated with a 0.0015 increase in the probability of dying per month, based on the study by Silverstein et al. (1999). The EAG acknowledged that the study is old and so surgery procedures and surgery-related mortality might have improved since then. However, the EAG did not identify more recent sources to populate this parameter in the model.

Costs

The following costs are considered in the model (all valued in 2019 GBP):

- diagnostic test costs
- treatment costs
- acute and chronic care costs of Crohn's disease (including costs of surgery).

The total cost of testing charged by the laboratory was £1,250 for PredictSURE IBD and £347 (estimated) for IBDX (Table 5). The cost of sample collection (including the special tube needed for PredictSURE IBD) and transport to the central laboratory were not considered in the model. No training costs are anticipated for either technology.

Table 5 Cost of testing services for both technologies

Device name	Unit cost	Source
PredictSURE IBD	£1,250	Company's reply to request for information: price set by the Cambridge laboratory for the testing service.
IBDX	£347 (using HMRC exchange rate USD/GBP 1.2483)	Company's reply to request for information and EAG's assumptions (estimate cost).

Table 6 shows doses, prices and induction dosages for induction therapies in top-down and step-up treatment strategies taken from BNF and NHS reference costs, and maintenance therapy dosages based on clinical opinion.

Table 6 Treatment doses and costs for induction and maintenance therapies

Treatment	Dose per unit (mg)	List price per unit	Induction dosages	Maintenance dosages
Ustekinumab	130	£2,147.00	Induction dose at week 0 depends on body weight: 260 mg for 56 kg 390 mg for 56 kg to 85 kg 520 mg for 86 kg or over	90 mg every 8 weeks
Vedolizumab	300	£2,050.00	300 mg at week 0, 2 and 6	300 mg every 8 weeks
Infliximab	100	£377,66	5 mg/kg at week 0, 2 and 6	5 mg/kg every 8 weeks
Adalimumab	40	£308.13	160 mg at week 0; 80 mg at week 2	40 mg every 2 weeks
Azathioprine	50	£0.04	2.5 mg/kg/week for 8 weeks	2.5 mg/kg/week
6-MP	50	£1.97	1.25 mg/kg/week	1.25 mg/kg/week
Methotrexate	25/15	£16.64 /£14.92	25 mg/week for 8 weeks	15 mg/week
Prednisolone	2.5	£0.04	40 mg; tapered by 5 mg per week – 8 weeks total	No maintenance with prednisolone
Intravenous administration (outpatient)	1	First: £199 Follow up: £212	Not applicable	Not applicable

The total 2-week costs of managing health states (see Table 7) include the following:

- outpatient costs (inflammatory bowel disease consultant, dietician, inflammatory bowel disease nurse, helpline, pharmacist and nutritional support)
- radiology (plain X-ray, CT scan of abdomen or pelvis, MRI scan of abdomen or pelvis, DEXA scan, MRI small bowel)
- endoscopies (oesophagogastroduodenoscopy, sigmoidoscopy, colonoscopy, double balloon enteroscopy, wireless capsule endoscopy)
- hospitalisations.

Table 7 Summary of health state costs per 2-week cycle, excluding surgery

Health state	Total cost in the model
Remission	£17
Mild	£27
Moderate to severe	£122

The average cost of surgery was estimated as £8,813, based on the number of occurrences of each type of surgery reported in Biasci et al. (2019).

Utility values

The EAG used the utility values from NICE's guidance on vedolizumab (based on EQ-5D data from GEMINI studies;Table 8) in the base-case analysis and a mapping algorithm based on NICE's technology appraisal guidance on ustekinumab for moderately to severely active Crohn's disease after previous treatment in a scenario analysis. This was because utility values from the vedolizumab guidance were considered 'theoretically superior to the values estimated from the mapping algorithm because they are directly elicited' by the EAG and the committee during the later appraisal of ustekinumab.

All utilities were adjusted to account for the age and sex of the modelled population, according to Ara and Brazier 2010.

The disutility values for surgery were based on values reported in Marchetti et al. that patients having surgery retained 0.5 of their utility estimate for 1 month.

Table 8 Utility values used for remission, mild, moderate to severe health states

Health state	NICE guidance on vedolizumab	NICE guidance on ustekinumab
Remission	0.820	0.820
Mild disease	0.730	0.700
Moderate to severe	0.570	0.550

Summary of key original base-case assumptions

The following assumptions were applied in the original base-case analysis, as summarised starting on page 119 of the diagnostics assessment report:

- PredictSURE IBD categorises people into high and low-risk groups (assumed 100% accuracy). High-risk patients in the test arm receive top-down therapy. High-risk patients in the no test arm receive step-up therapy, which is current standard care. The low-risk populations in the step-up and top-down arms receive the same treatment so they cancel out.
- All patients start on corticosteroid therapy and follow up with either immunomodulators (step-up cohort) or anti-TNF therapy (top-down cohort). Escalations from corticosteroids to immunomodulators and from corticosteroids to anti-TNF were not modelled because in both strategies (the top-down and the step-up arms) 100% of patients would receive initial induction treatment with corticosteroids.
- 30% of people receiving anti-TNF and 20% of people receiving non-anti-TNF biologics receive combination treatment with immunomodulators.
- Response to anti-TNF does not depend on the prior lines of therapy. Therefore people starting anti-TNF in the step-up strategy catch up with those who received an early anti-TNF in the top-down strategy. People in the step-up cohort can also benefit from having an additional treatment step, because some people will respond to immunomodulators.

- People in the top-down strategy have a longer time to treatment escalation and a longer time to surgery than people in the step-up strategy, based on extrapolation of results from D'Haens et al. (2008) and Hoekman et al. (2018).
- Surgery was considered a stand-alone outcome in the model, that is, patients did not leave their respective health states to enter a surgery health state. Instead, in every model cycle, a proportion of operations was estimated, and the associated costs and impact on patients' quality of life were calculated. The EAG applied a surgery-related disutility to patients' total utility in that model cycle and assumed a small increase in probability of mortality per month.
- The model does not capture the impact of complications, long-term consequences of surgery or adverse events on patients' quality of life or costs.

Additional assumptions in the revised base-case analysis

The EAG produced a revised base case (described in the diagnostics assessment report addendum) which uses the same assumptions listed above. An additional assumption in the revised base case is that the time to treatment escalation restarts on each new treatment rather than decreasing over time and as treatment sequences progress.

Revised base-case analysis results

The results of the deterministic base-case analysis comparing PredictSURE IBD with standard care shows that the top-down arm (using PredictSURE IBD) costs more and has less QALYs (see Table 9).

Table 9 Base case deterministic cost effectiveness results (discounted)

Intervention	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard care	£201,925	15.86	–	–	–
PredictSURE IBD	£211,009	15.79	£9,084	-0.08	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

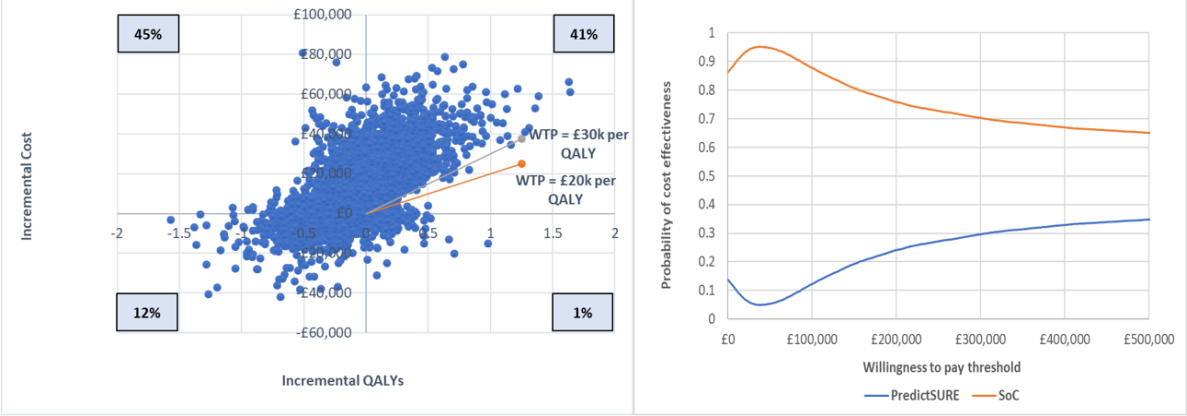
The probabilistic revised base-case analysis results (Table 10) were aligned with the deterministic ones, that is, testing and using top-down strategy in high-risk patients was less clinically effective and more costly than current standard care (no testing, using step-up strategy for all patients). The testing strategy had a less than 10% probability of being cost effective against standard care at the maximum acceptable ICERs of £20,000 and £30,000 as shown by the cost-effectiveness acceptability curve (Figure 7). The final ICER is mainly driven by the differences in cost of top-down compared with step-up strategies because the difference in incremental QALYs is very small (Table 10).

Table 10 Revised base-case analysis probabilistic cost effectiveness results (discounted)

Intervention	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard care	£224,904	15.70	–	–	–
PredictSURE IBD	£237,036	15.67	£12,132	-0.03	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

Figure 7 Cost-effectiveness plane (left) and cost-effectiveness acceptability curve (right)



Scenario analyses

Results of the deterministic fully incremental cost-effectiveness analysis show that PredictSURE IBD and IBDX have higher costs and lower QALYs (that is, they were dominated) compared with standard care of no testing (Table 11). In the absence of robust evidence on the prognostic accuracy of both tools, the cost-effectiveness analysis only differs in the cost of the tests.

Table 11 Revised base case fully incremental cost effectiveness results (discounted)

Intervention	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard care	£201,925	15.86	–	–	–
IBDX	£210,106	15.79	£8,181	-0.08	Dominated
PredictSURE IBD	£211,009	15.79	£903	0	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

In the scenario with immunomodulators as the last treatment option in the top-down arm, the EAG assumed that the probability of remission and relapse was the same as those for people who had immunomodulators as first treatment step in the step-up arm. Deterministic base case results for this scenario showed that PredictSURE IBD (top-down strategy) generated 0.07 additional QALYs compared with the step-up strategy, at an additional cost of

£7,502, producing an ICER of £105,148 per QALY gained (see page 32 of the addendum to the diagnostics assessment report for further details).

Results of all but one of the other scenarios that explored uncertainties in assumptions made in the model (Table 12), showed that PredictSURE IBD was dominated by standard care. In assuming that no high-risk patient in the step-up arm derived benefit from immunomodulators PredictSURE IBD had an ICER of £170,180 per QALY gained. Clinical equivalence between the 2 strategies was reached when it was assumed that 97% of high-risk patients in the step-up arm do not benefit from immunomodulators.

Table 12 Results of scenario analyses

Intervention	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Scenario 1: Applying IBDX cost					
Standard care	£201,925	15.86	–	–	–
IBDX	£210,106	15.79	£8,181	-0.08	Dominated
Scenario 2: Applying utilities from NICE's guidance on ustekinumab					
Standard care	£201,925	15.57	–	–	–
PredictSURE IBD	£211,009	15.50	£9,084	-0.08	Dominated
Scenario 3: Applying induction vectors and transition probabilities based on studies used in NICE's guidance on vedolizumab					
Standard care	£201,695	15.86	–	–	–
PredictSURE IBD	£210,841	15.78	£9,146	-0.08	Dominated
Scenario 4: Applying equivalent TTS curves for top down and step up					
Standard care	£201,925	15.86	–	–	–
PredictSURE IBD	£211,575	15.78	£9,650	-0.08	Dominated
Scenario 5: Removing Ara and Brazier utility adjustment					
Standard care	£201,925	15.92	–	–	–
PredictSURE IBD	£211,009	15.84	£9,084	-0.08	Dominated
Scenario 6: Use the minimum induction period from the treatment class to estimate induction costs					
Standard care	£196,077	15.84	–	–	–
PredictSURE IBD	£204,704	15.76	£8,627	-0.08	Dominated

Scenario 7: 100% of high-risk patients who receive step-up therapy do not respond to immunomodulator treatment					
Standard care	£209,797	15.78	–	–	–
PredictSURE IBD	£211,009	15.79	£1,212	0.01	£170,180

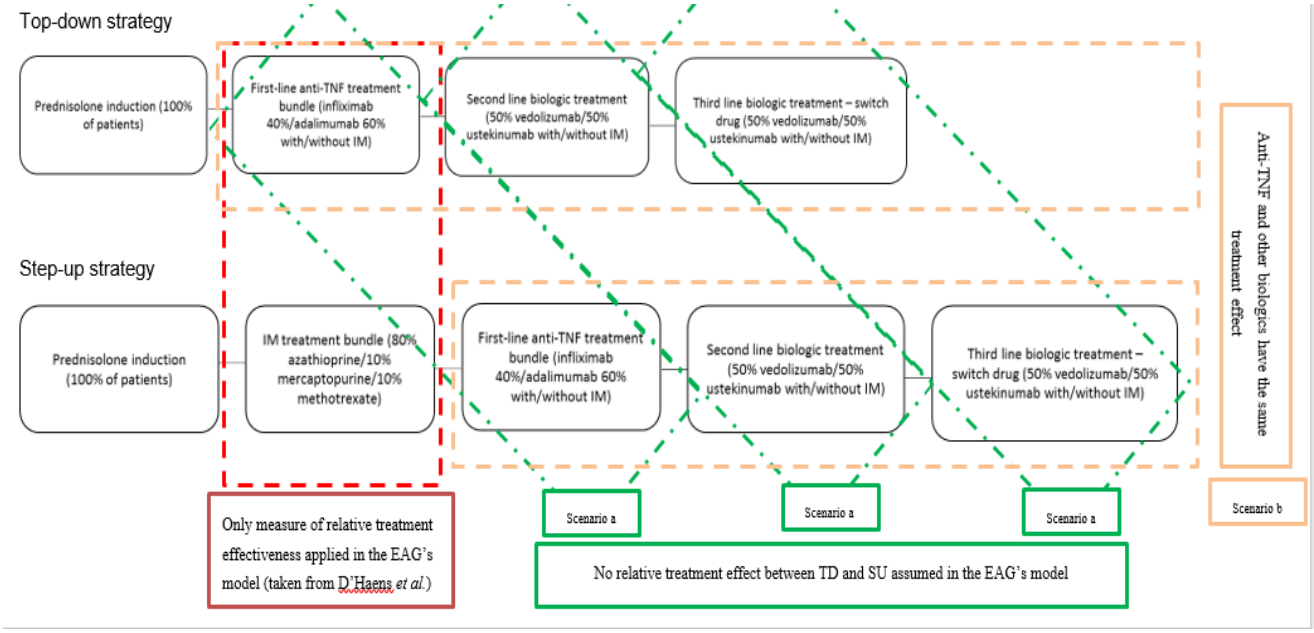
Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; TTS, time to surgery.

Individual scenario analysis

The EAG ran individual scenario analyses, results of which show that in most cases dominance of the step-up strategy did not change (Table 13). The base-case assumption was that the PredictSURE IBD test has 100% prognostic accuracy, which is quite unlikely in practice. The EAG instead assumed a 25% misdiagnosis of the course of the disease. An accuracy of 75% was arbitrarily chosen in the absence of robust evidence to inform this scenario. The QALY gain associated with PredictSURE IBD in this scenario can be attributed to assigning misdiagnosed low-risk patients to receive ant-TNF in the top-down arm of the model. It is assumed that these patients will not need any further treatment escalation.

In the base case some high-risk patients are thought to benefit from immunomodulators in the step-up arm and then receive biologics, which have same benefits as biologics in the top-down arm. The relative treatment effectiveness in the base case was applied only to the immunomodulator compared with anti-TNF step as seen in Figure 8. The EAG explored a scenario that assumed the benefit of the biologics in the top-down arm compared with biologics in the step-up arm extends to the next treatment steps. This means that people receiving second-line biologic treatments in the top-down arm have an added treatment benefit compared with those receiving a second-line biologic treatment in the step-up arm (Figure 8). Another scenario assumed that once patients move on to second or third-line biologics there is no further benefit of top down over step up (see further details starting on page 22 of the addendum to the diagnostics assessment report).

Figure 8 Assumptions around measure of relative treatment effect



Abbreviations: EAG, external assessment group; IM, immunomodulator; SU, step up; TD, top down; TNF, tumour necrosis factor.

Assumptions around treatment discontinuation were based on Marchetti 2013, which reported that 76% and 40% of people had mucosal healing after 2 years in remission with biologic treatments in the top-down and step-up strategies, respectively. The EAG explored having the same probability of mucosal healing in both the step-up and top-down arms, that is, either 76% or 40% in both arms. A separate scenario assessed the impact of discontinuing treatment after 12 months of continuous remission with maintenance treatments of biologics.

To explore the impact of the assumptions made about surgery, the EAG assumed that patients who try all biologic treatments without their disease being adequately controlled go on to have surgery and are temporarily cured for 2 years, after which they return to a moderate or severe state. A separate scenario excluded surgery from the model.

Table 13 Results of individual scenario analyses

Intervention	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Scenario 2.1.1: Misdiagnosis					
Standard care	£201,925	15.86	–	–	–
PredictSURE IBD	£211,782	16.01	£9,856	0.15	£64,876
Scenario 2.1.2a i: Assuming half of the base case risk of relapse (in the first treatment steps) for TD vs SU for second and subsequent treatment steps					
Standard care	£197,986	15.78	–	–	–
PredictSURE IBD	£208,878	15.75	£10,892	-0.04	Dominated
Scenario 2.1.2a ii: Assuming the same as base case risk of relapse (in the first treatment steps) for TD vs SU for second and subsequent treatment steps					
Standard care	£193,282	15.70	–	–	–
PredictSURE IBD	£205,961	15.70	£12,679	-0.002	Dominated
Scenario 2.1.2b i: Assuming half of the base case risk of relapse (in the first treatment steps) for TD vs SU for anti-TNF vs biologics in TD					
Standard care	£197,986	15.78	–	–	–
PredictSURE IBD	£207,699	15.73	£9,713	-0.06	Dominated
Scenario 2.1.2b ii: Assuming the same as base case risk of relapse (in the first treatment steps) for TD vs SU for anti-TNF vs biologics in TD					
Standard care	£193,282	15.70	–	–	–
PredictSURE IBD	£203,599	15.66	£10,317	-0.04	Dominated
Scenario 2.1.3a i: Assuming discontinuation of biologic treatment for 76% TD; 40% SU					
Standard care	£181,522	15.86	–	–	–
PredictSURE IBD	£178,016	15.79	-£3,506	-0.08	£46,263*
Scenario 2.1.3a ii: Assuming discontinuation of biologic treatment for 76% TD; 76% SU					
Standard care	£163,159	15.86	–	–	–
PredictSURE IBD	£169,238	15.79	£6,079	-0.08	Dominated

Scenario 2.1.3a iii: Assuming discontinuation of biologic treatment for 40% TD; 40% SU					
Standard care	£181,522	15.86	–	–	–
PredictSURE IBD	£189,024	15.79	£7,502	-0.08	Dominated
Scenario 2.1.3b: Assuming discontinuation of biologic treatment for 100% TD; 100% SU					
Standard care	£150,917	15.86	–	–	–
PredictSURE IBD	£156,047	15.79	£5,130	-0.08	Dominated
Scenario 2.1.4a: Assuming surgery as last treatment step					
Standard care	£203,916	16.13	–	–	–
PredictSURE IBD	£213,060	16.06	£9,144	-0.07	Dominated
Scenario 2.1.4b: Removing surgery from the model					
Standard care	£197,827	15.88	–	–	–
PredictSURE IBD	£207,497	15.80	£9,670	-0.08	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; TTS, time to surgery; TD, top-down; SU step-up

*This ICER is for standard care compared with PredictSURE IBD, meaning that the prognostic tool is cheaper than standard care but also less effective.

Combined scenario analysis

Some of the individual scenarios described in table 13 were combined, to explore the impact of increasing the effectiveness of the top-down strategy while decreasing the treatment cost of biologics. The results of these combined scenarios varied (Table 14). PredictSURE IBD (top-down strategy) was dominant (scenario 2.3.1a) when the probability of misdiagnosed cases was combined with the reduced costs associated with discontinuing biologic treatments (probability of discontinuing treatment because of mucosal healing was assumed to be higher in top-down treatment). PredictSURE IBD was also dominant (scenario 2.3.4a) when the top-down strategy was assumed to be more beneficial than the step-up strategy; when a higher proportion of patients achieved mucosal healing with top down treatment and no high-risk patient

responded to immunomodulators in the step up arm. Scenario 2.3.4b differs from scenario 2.3.4a in assuming that same proportion of patients achieve mucosal healing in both strategies. This scenario produced an ICER of £29,225 per QALY gained (in favour of top-down) which is less than the £30,000 threshold (see further details starting on page 35 of the addendum to the diagnostics assessment report).

Table 14 Results of combined scenario analyses

Intervention	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Scenario 2.3.1a (Scenario 2.1.1 misdiagnosis plus assuming discontinuation of biologic treatment for 76% TD; 40% SU)					
Standard care	£181,522	15.86	–	–	–
PredictSURE IBD	£176,541	16.01	-£4,981	0.15	Dominant
Scenario 2.3.1b (Scenario 2.1.1 Misdiagnosis plus assuming discontinuation of biologic treatment for 76% TD; 76% SU)					
Standard care	£163,159	15.86	–	–	–
PredictSURE IBD	£168,153	16.01	£4,995	0.15	£32,875
Scenario 2.3.1c (Scenario 2.1.1 Misdiagnosis plus assuming discontinuation of biologic treatment for 40% TD; 40% SU)					
Standard care	£181,522	15.86	–	–	–
PredictSURE IBD	£188,819	16.01	£7,298	0.15	£48,034
Scenario 2.3.2a (Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming discontinuation of biologic treatment for 76% TD; 40% SU)					
Standard care	£174,162	15.70	–	–	–
PredictSURE IBD	£173,517	15.70	-£645	-0.002	£330,616*
Scenario 2.3.2b Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming discontinuation of biologic treatment for 76% TD; 76% SU)					
Standard care	£156,954	15.70	–	–	–
PredictSURE IBD	£165,233	15.70	£8,279	-0.002	Dominated
Scenario 2.3.2c (Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming discontinuation of biologic treatment for 40% TD; 40% SU)					
Standard care	£174,162	15.70	–	–	–
PredictSURE IBD	£184,525	15.70	£10,363	-0.002	Dominated
Scenario 2.3.3 (Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming that 100% of SU patients do not respond to IM)					
Standard	£201,178	15.61	–	–	–

care					
PredictSURE IBD	£205,961	15.70	£4,782	0.08	£57,757
Scenario 2.3.4a (Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming discontinuation of biologic treatment for 76% TD; 40% SU plus assuming that 100% of SU patients do not respond to IM)					
Standard care	£180,986	15.61	–	–	–
PredictSURE IBD	£173,517	15.70	-£7,469	0.08	Dominant
Scenario 2.3.4b (Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming discontinuation of biologic treatment for 76% TD; 76% SU plus assuming that 100% of SU patients do not respond to IM)					
Standard care	£162,813	15.61	–	–	–
PredictSURE IBD	£165,233	15.70	£2,420	0.08	£29,225
Scenario 2.3.4c (Assuming the same as base case risk of relapse for second and subsequent treatment steps plus assuming discontinuation of biologic treatment for 40% TD; 40% SU plus assuming that 100% of SU patients do not respond to IM)					
Standard care	£180,986	15.61	–	–	–
PredictSURE IBD	£184,525	15.70	£3,539	0.08	£42,740

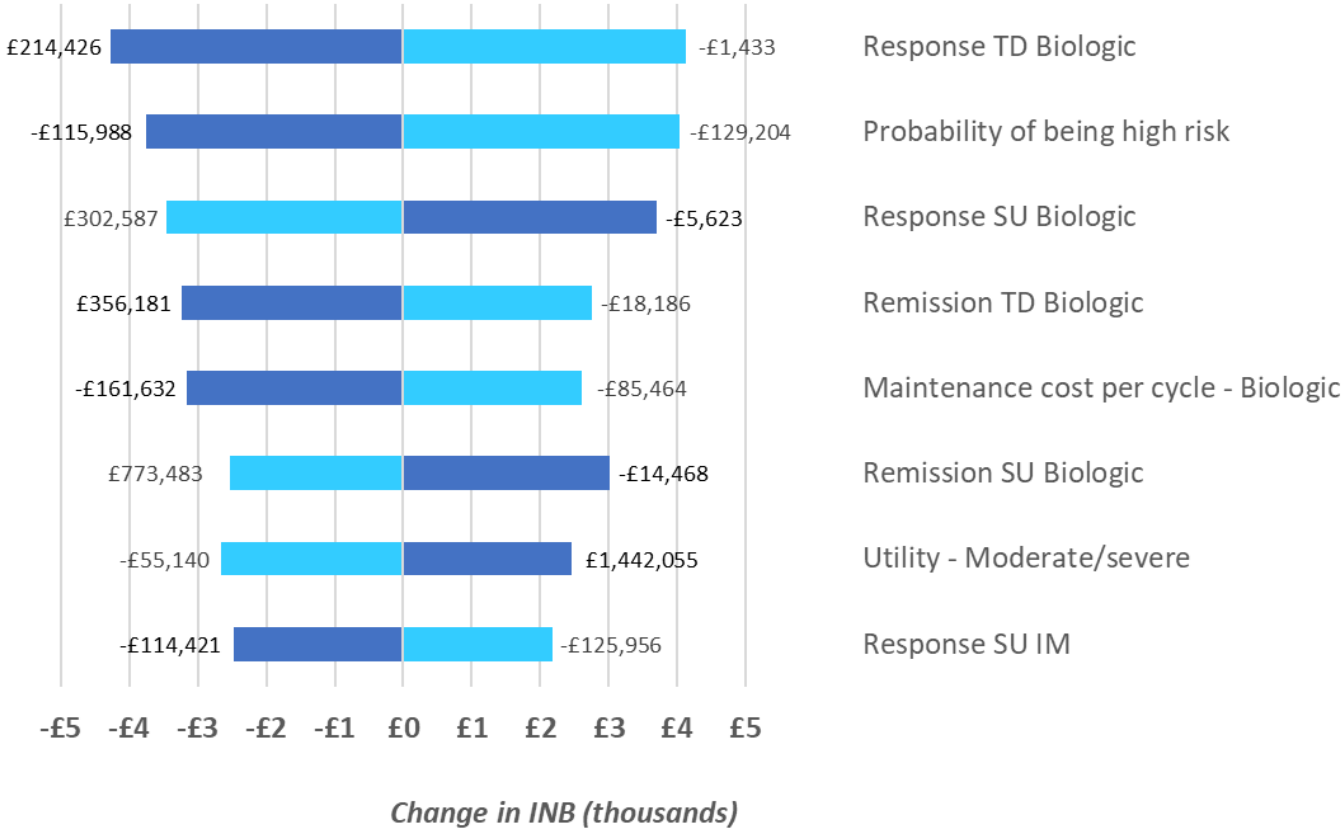
Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; TTS, time to surgery TD, top-down, SU step-up.

*This ICER is for standard care compared with PredictSURE IBD, meaning that the prognostic tool is cheaper than SC but also less effective.

One-way sensitivity analysis

A range of one-way sensitivity analyses explored model results based on the upper and lower bounds for model inputs (listed on page 14 of the addendum to the diagnostics assessment report). A tornado plot (Figure 9) shows that response to biologics in the top-down arm of the model is a key driver of the deterministic ICER.

Figure 9 One-way sensitivity analyses that have the greatest impact on incremental net monetary benefit (ICERs given at the top and lower end of bars)



Abbreviations: IM, immunomodulator; INB, incremental net benefit; SU, step up; TD, top down.

3 Summary

Clinical effectiveness

The EAG found 8 primary studies (reported in 12 publications) that met the inclusion criteria, 1 reporting the diagnostic performance of PredictSURE IBD, and 7 reporting the performance of IBDX.

None of the studies for IBDX was done in people with newly diagnosed Crohn's disease; median duration of disease at the time of testing ranged from 10.6 (IQR 1.7 to 52.3) months to 9.4 (IQR 1 to 44) years. Only 1 prognostic study (Rieder et al. 2010c) assessed the ability of the IBDX test to

predict future negative outcomes (developing a complication or needing surgery) in people with no prior complication or surgery at baseline. People who tested positive for 2 or more out of 6 IBDX markers had a significantly higher risk of complications (HR 2.5; 95% CI 1.03 to 6.1; p=0.043), or surgery (HR 3.6; 95% CI 1.2 to 11.0; p=0.023) during the median follow up of 53.7 months than people who tested positive for 0 or 1 markers. The study was limited by the small sample size (n=76). Another prognostic study (Wolfel et al. 2017) showed that IBDX markers did not predict time to repeat surgery (n=118; people had had Crohn's disease-related surgery; median follow up of 100 months). One study reported that IBDX markers correlated with either a history of complications or surgery at baseline or their occurrence during follow up (Rieder et al. 2010b). The remaining 4 IBDX studies were cross-sectional, reporting the correlation between IBDX markers and disease phenotype at the time of testing.

One UK study (Biasci et al. 2019) reported on the prognostic ability of PredictSURE IBD in adults with inflammatory bowel disease, including 66 people with Crohn's disease in the validation cohort and 38 in the training cohort. Most people had newly diagnosed disease. In the validation cohort, people categorised as high risk (n=27; 40.9%) had a statistically significantly higher risk of at least 1 treatment escalation compared with those designated as low risk (n=39; 59.1%), with a HR of 2.65 (95% CI 1.32 to 5.34; p=0.006). Median duration of follow up was 1.6 (IQR 1.0 to 3.7) years in the high-risk group and 2.4 (IQR 1.8 to 3.8) years in the low-risk group. Sensitivity and specificity for predicting the need for 2 or more escalations within the first 12 months were 77.8% and 70.6%, respectively, and within 18 months, 72.7% and 73.2%. Escalations within the first 18 months were 90.9% and 42.1%, respectively.

Cost effectiveness

The EAG did not identify robust economic evidence on the prognostic accuracy of PredictSURE IBD and IBDX. Given the lack of evidence, the

EAG's economic analysis focused on assessing the cost effectiveness of the top-down and step-up treatment strategies in people at high risk of a severe disease course. Eleven economic studies of treatments for Crohn's disease were identified. One Italian study (Marchetti et al. 2013) specifically assessed the cost effectiveness of top-down compared with step-up treatment strategies in newly diagnosed Crohn's disease patients over a 5-year time horizon. The authors concluded that the top-down strategy was dominant. Results of the economic model done by PredictImmune (Buchanan et al. 2020) showed that using PredictSURE IBD to guide top-down treatment in Crohn's disease produced an ICER of £7,179 per QALY gained (£1,852 incremental cost and 0.258 incremental QALYs) when compared with standard care over a 15-year time horizon. The key drivers of the model results were the time horizon, rates of mucosal healing in top-down compared with step-up therapy, the costs of hospitalisation and costs and quality of life in the severe disease health state.

The EAG's de novo economic model was adapted from the model submitted by PredictImmune. The population in the model is largely based on the population in the Biasci study complemented by individual patient data supplied by PredictImmune. Clinical outcome data for the group treated with the step-up strategy (no test strategy) informed the baseline outcomes. Relative treatment effectiveness for top-down compared with step-up strategies for time to treatment escalation was quantified by applying an estimate of treatment effect (in the form of relative hazard functions) to the first step of the top-down arm (anti-TNF) compared with the first step of the step-up arm (immunomodulators). There was no robust evidence to inform subsequent treatment escalations. The EAG assumed that high-risk patients who start treatment with immunomodulators (step-up strategy) escalate treatment faster than high-risk patients who start treatment with anti-TNF (top-down strategy). Surgical events were estimated as a stand-alone outcome with relative treatment effectiveness for time to surgery taken from the Hoekman study.

The results of the EAG's economic analysis suggests that the step-up strategy may be more clinically beneficial than the top-down strategy. The deterministic and probabilistic revised base case results show that top-down strategy is associated with additional costs (£9,084 and £12,132 respectively) and a loss of QALYs (0.08 and 0.03 respectively). Key drivers of the economic results were assumptions around stopping biologics (resulting in reduced costs), the benefit of top-down compared with step-up treatment (and its impact on time to treatment escalation) and assumptions around the number of people who respond to treatment with immunomodulators in the step-up arm.

The cost effectiveness of IBDX was explored in a scenario analysis, in which input parameters for PredictSURE IBD and IBDX only differed in the cost of the tests.

4 Issues for consideration

Clinical effectiveness

There was no evidence on the prognostic accuracy of IBDX in people with newly diagnosed Crohn's disease. All IBDX studies enrolled people with long-lasting disease:

- Only 1 study (Rieder et al. 2010c) assessed the prognostic ability of IBDX in people with no prior complication or surgery. The study was limited by the small sample size (n=76).
- Only 1 study (Wolfel et al. 2017) assessed whether IBDX test results can predict the time to next surgery in people who had a Crohn's disease-related surgery at baseline.
- One study reported a correlation of IBDX with either a history of complication or surgery at baseline, or their occurrence during follow up (Rieder et al. 2010b).
- The remaining 4 IBDX studies were cross-sectional, reporting the correlation between IBDX markers and disease phenotype at the time of testing. Although 1 study (Rieder et al. 2011) showed the expression of

markers remains stable over time, the study enrolled people with long-lasting disease and had a relatively short follow-up time. So correlation between IBDX markers at the time of diagnosis and future outcomes is unclear.
- There is no evidence on whether the classification of patients into high and low-risk categories by IBDX can help to guide treatment choices and improve patient outcomes.

There was limited evidence on the prognostic ability of PredictSURE IBD:

- The only study used the need for multiple treatment escalations as a proxy for severe disease course. The clinical relevance of this is not clear because it's not clear if people classed as high risk by PredictSURE IBD are at an increased risk of developing complications or needing surgery.
- There is no evidence on the clinical utility of the test, that is, whether the classification of patients into high and low-risk categories by PredictSURE IBD can help to guide treatment choices and improve patient outcomes.
- The PROFILE study is currently ongoing and will provide information on the clinical validity and utility of the test. It is a biomarker-stratified trial of top-down compared with step-up therapy in people classed as high or low risk of severe disease. Results are expected in 2022.
- Another 2 prognostic studies are ongoing for PredictSURE IBD: a multicentre observational study based in the US (PRECIOUS) and a prospective study in children. How the test performs in children is currently unknown.

There was very limited evidence on how IBDX and PredictSURE IBD and compare in terms of performance. A sub-study (Lyons et al. 2020, based on the same cohort as Biasci et al. 2019) compared PredictSURE IBD and IBDX.

Crohn's disease is a heterogenous disease. The performance of the 2 tests and potential clinical utility across all disease subtypes may need to be

explored. Furthermore, people with Crohn's disease often present with comorbid conditions; the impact of comorbidities on the analytical and clinical utility of PredictSURE IBD and IBDX may need to be explored.

Cost effectiveness

Model structure

- Clinical practice may vary across the UK. Therefore the treatment pathway in the model may not fully reflect the treatment pathway of all patients.
- The EAG assumed anti-TNF therapy would be combined with immunomodulators in 30% of patients but the clinical experts do not agree on this.
- In the base case, the model predicted fewer QALYs with top-down than with step-up therapy in high-risk patients. This is contrary to the clinical expectation that top-down therapy has more benefits for high-risk patients, as noted in the D'Haens study (the treatment sequence modelled in D'Haens is different from the sequence modelled by the EAG) and the value proposition put forward by the companies.
- Once patients have tried all biologic treatment options, they have surgery. Surgery was assumed to have a temporary curative effect of 2 years, after which patients are assumed to revert to moderate to severe active disease health state for their remaining life, which may not be clinically plausible.
- The long-term impacts of surgery was not considered (a transient improvement in health due to surgery was explored in sensitivity analyses).
- The impact of complications or adverse events on costs or QALYs is not considered in the model.

Clinical inputs

- Modelling was restricted to 40 patients whose treatment matched the standard definition of step-up therapy, that is, people who received first-line therapy with corticosteroids, and second-line treatment with immunomodulators. The EAG excluded people who received other first-line therapies and people who did not have a treatment escalation.

- The efficacy of a top-down approach in people who are expected to follow a severe disease course is currently unknown. A search to identify studies comparing top-down with step-up therapy was done and the study by D'Haens was identified. In this study:
  - An unselected population was used (not stratified by their risk of following severe disease course).
  - Top-down and step-up therapies differed from the sequence assumed by the EAG in the model (people in the top-down group did not receive initial corticosteroids; people in the step-up group received an immunomodulator only after failure of corticosteroids).
  - Top-down treatment consisted of induction therapy with infliximab only (no maintenance therapy), then as-needed infliximab – standard care now is to offer maintenance with infliximab.
  - A large proportion of people randomised to step-up therapy received biologics quite early on in the pathway (an accelerated standard step-up).
- The efficacy of anti-TNF therapy was assumed to be exactly the same regardless of the line of therapy, and so people in having step-up therapy were assumed to catch up with those having top-down (higher efficacy of early biologics was tested in sensitivity analyses).

Costs

- The modelling assumed all people receive an accelerated adalimumab course; in clinical practice, a proportion of patients is expected to be on a standard adalimumab course.
- The modelling did not account for dose adjustments (for example, dose increases in nonresponding patients or dose reductions in responding patients).
- All responding patients are assumed to continue the treatment throughout their life (discontinuation after 2 years was explored in sensitivity analyses).
- The costs of post-surgical care were not included, which could be substantial.

Other

- The final population in the model (n=40) had 58% high risk and 42% low-risk patients (with [redacted] escalations in the high-risk group and [redacted] escalations in the low-risk group). In contrast, the validation cohort in the Biasci study (n=66) had 41% high risk and 59% low-risk patients.
- IBDX was assumed to have the same efficacy as PredictSURE IBD in an exploratory analysis, which is unlikely to be valid because the 2 tests have different mechanisms of action and used different definitions of severe disease course in their supportive studies.
- An economic model developed for adults with Crohn's disease may not be generalisable to children, given the differences in natural history and treatment pathways between adults and children with Crohn's disease.

5 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

PredictSURE IBD has not yet been validated in children (17 or under); therefore, the prognostic accuracy of PredictSURE IBD in children is unknown. However, evidence on children may become available while the guidance is in development. IBDX has been studied in both adults and children.

Crohn's disease can have a substantial and long-term adverse effect on a person's ability to carry out normal day-to-day activities. Therefore, people with Crohn's disease may be covered under the disability provision of the Equality Act (2010).

6 Implementation

The NICE adoption team identified potential factors that could encourage implementation of PredictSURE IBD and IBDX to guide personalised treatment of Crohn's disease:

- A high unmet need for a prognostic test in Crohn's disease to guide personalised therapy.
- Sending samples from referring trusts to central laboratories is well established in the NHS.
- Secure, online reporting of results.
- The 7 to 10-day turnaround is acceptable.

Potential adoption barriers they identified include:

- Uncertainty about the clinical utility of the test.
- It is not clear who in the NHS would pay for the test.
- PAXgene Blood RNA tubes are not routinely used in the NHS.
- Potential delays if testing is in a single central laboratory.
- The need for training and quality assurance systems if testing is expanded to other laboratories.

7 Authors

Ewa Rupniewska
Topic lead (until December 2019)

Tosin Oladapo
Topic lead (from February 2020)

Frances Nixon
Technical adviser

August 2020

Appendix A: Sources of evidence considered in the preparation of the overview

A. The diagnostics assessment report for this assessment was prepared by BMJ Technology Assessment Group:

Edwards SJ, Barton S, Bacelar M, Karner C, Cain P, Wakefield V, Marceniuk G. PredictSURE IBD and IBDX to guide personalised treatment of Crohn's disease in adults: A Diagnostics Assessment Report. BMJ Technology Assessment Group, 2019.

B. The following organisations accepted the invitation to participate in this assessment as stakeholders. They were invited to attend the scoping workshop and to comment on the diagnostics assessment report.

Manufacturers of technologies included in the final scope:

- Glycominds LLC
- PredictImmune Ltd

Other commercial organisations:

- AbbVie
- Janssen

Professional groups and patient/carer groups:

- British Society of Gastroenterology
- Royal College of Physicians
- Crohn's & Colitis UK

Research groups:

- None

Associated guideline groups:

- None

Others:

- Department of Health and Social Care
- Healthcare Improvement Scotland
- NHS England
- Welsh Government
- The Association of the British Pharmaceutical Industry

Appendix B: Glossary of terms

Adalimumab: a recombinant human anti-TNF-alpha IgG1 monoclonal antibody.

Anti-glycan antibodies: antibodies directed against microbial cell wall surface components.

CD8+ cytotoxic T cells: cells of the immune system that are involved in the regulation of immune response.

CD8+ T-cell exhaustion: a state characterised by the stepwise and progressive loss of T-cell functions.

Immunosuppressants: a class of drugs used to supress or prevent an immune response.

Inflammatory bowel disease: a group of inflammatory conditions of the colon and small intestine, the 2 most common being Crohn's disease and ulcerative colitis.

Infliximab: a chimeric (human-murine) anti-TNF-alpha IgG1 monoclonal antibody.

Seroreactivity: the reactivity of the blood serum, that is, the presence of specific antibodies (for example, against an infectious or non-infectious microorganism), in the serum of a patient.

TNF-alpha inhibitors: biological therapies that target the TNF-α protein with the aim of modifying the inflammatory disease process.

Ustekinumab: a fully human monoclonal antibody that targets interleukin-12 (IL-12) and IL-23.

Vedolizumab: a humanised IgG1 monoclonal antibody that targets alpha 4 beta 7 integrin.