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

DIAGNOSTIC ASSESSMENT REPORT

This report was commissioned by the NIHR Systematic Reviews Programme as project number 128968/T

BMJ Technology Assessment Group

Title: PredictSURE-IBD and IBDX to guide personalised treatment of Crohn's disease in adults

Produced by: BMJ Technology Assessment Group (BMJ-TAG)

Authors: Steven J Edwards, Director of Health Technology Assessment, BMJ-TAG, London

Samantha Barton, Principal Health Technology Assessment Analyst, BMJ-TAG, London

Mariana Bacelar, Principal Health Economist, BMJ-TAG, London

Charlotta Karner, Health Technology Assessment Analysis Manager, BMJ-TAG, London

Peter Cain, Senior Health Economist, BMJ-TAG, London

Victoria Wakefield, Principal Health Technology Assessment Analyst, BMJ-TAG, London

Gemma Marceniuk, Associate Health Economist, BMJ-TAG, London

Correspondence to: Steve Edwards, [redacted]

Date completed: 17th December 2019

Source of funding: This report was commissioned by the NIHR Systematic Reviews Programme as project number 128968/T.

Declared competing interests of the authors:

No competing interests were declared which affect the impartiality of this report. BMJ Technology Assessment Group (BMJ-TAG) and the editorial team of The BMJ work independently to one another. The views and opinions expressed in this report are those of the BMJ-TAG.

Acknowledgements:

The External Assessment Group (EAG) would like to thank Dr Gordan Moran (Clinical Associate Professor of Gastroenterology, University of Nottingham, Nottingham), Dr John Saunders (Consultant Gastroenterologist, Royal United Hospitals Bath, Bath), and Dr Sami Hoque (Consultant Gastroenterologist, Barts NHS Health Trust, London) for providing clinical advice throughout the project and Dr Christopher Stinton (Senior Research Fellow, Warwick Medical School, Warwick) for providing advice on interpretation of prognostic accuracy. The EAG would also like to thank Mr Adil Butt for his contribution to the running of searches to retrieve records on health-related quality of life, to the data extraction of economic and health-related quality of life studies and to the quality assessment of studies related to cost-effectiveness.

Rider on responsibility for protocol

The views expressed in this report are those of the protocol and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows: 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 Diagnostic Assessment Report. BMJ Technology Assessment Group, 2019.

Contributions of Authors

Steve Edwards	Project lead: supervised the production of the final report; critical appraisal of the clinical evidence; and critical appraisal of the economic evidence
Samantha Barton	Devised and carried out the literature searches on prognostic accuracy and clinical impact of the tools; contributed to study selection, data extraction, critical appraisal of the clinical evidence; and wrote the clinical sections of the report.
Mariana Bacelar	Development of the conceptual economic model and report writing.
Charlotta Karner	Study selection; and critical appraisal of the clinical evidence.
Peter Cain	Development of the economic model; carried out the economic analysis and report writing.
Victoria Wakefield	Study selection; data extraction; and critical appraisal of the clinical evidence.
Gemma Marceniuk	Devised and carried out the economic literature searches; study selection; data extraction; critical appraisal of the economic evidence and wrote the systematic literature review for economic evidence.

ABSTRACT

Background

Crohn's disease (CD) is a lifelong condition that can affect any segment of the gastrointestinal tract. People with CD may be at high- or low-risk of developing complications and being able to identify what level of risk a patient has could lead to personalised management.

Objectives

To assess the prognostic test accuracy, the clinical impact and the cost effectiveness of two tools for the stratification of people with a diagnosis of CD by risk of following a severe course of disease.

Methods

A systematic review of MEDLINE, EMBASE, CENTRAL, and CDSR was carried out from inception to June 2019 for studies assessing prognostic accuracy and clinical impact and to July 2019 for studies reporting on cost effectiveness. Two reviewers agreed studies for inclusion and assessed quality of included studies. One reviewer carried out data extraction from studies, with data validated by a second reviewer. Clinical and methodological heterogeneity across studies precluded synthesis of data.

The EAG developed a *de novo* economic model consisting of a decision tree designed to allocate patients to a response category after initial induction therapy in either the top-down (TD) or step-up (SU) treatment arms. The decision tree was followed by a cohort model, where patients' level of response to maintenance therapy was assessed. The goal of the economic model was to assess the cost-effectiveness of TD therapy vs SU therapy in high-risk patients.

Results

Searches of electronic databases identified 16 publications, including systematic reviews, that were deemed to be relevant to the review of prognostic accuracy. Additionally, documents supplied by the companies marketing the prognostic tools were reviewed. Included studies were assessed for risk of bias and applicability using the QUIPS (QUality In Prognosis Studies) tool. No study meeting eligibility criteria reported on the prognostic accuracy of the IBDX® biomarker-stratification tool as assessed using the full panel (six biomarkers), whereas one observational study provided estimates of sensitivity, specificity, and negative predictive value for the PredictSURE-IBD™ tool. All identified studies were considered to be of low quality.

As no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, the development of the economic model sets a structural framework for analysing future available data on prognostic accuracy and assesses the costs and consequences of treating high- and low-risk patients with both TD and SU strategies.

In the base case economic analysis, due to a paucity of data, the accuracy of PredictSURE IBD was assumed to be 100%. A similar assumption was made for IBDX in a scenario analysis, with the only difference between the two tests in the scenario being the cost of the tests.

The incremental analysis of cost-effectiveness demonstrates that the TD strategy (via the use of PredictSURE IBD in the model) is dominated by SU (via the SC arm of the model), with an additional cost of £9,526 and a QALY loss of 0.06.

Conclusions

Despite extensive systematic searches of the literature, no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, IBDX and PredictSURE-IBD.

While the model indicates that SC dominates the tests, the lack of evidence for prognostic accuracy with the two tests and the uncertainty around the benefits of TD and SU treatment approaches means that these results should be interpreted as indicative rather than definitive.

Study registration

The protocol for the review is registered on PROSPERO as CRD42019138737.

Funding

This report was commissioned by the NIHR Systematic Reviews Programme as project number 128968/T.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TABLE OF ABBREVIATIONS

Abbreviation	In full
5-ASA	5-Aminosalicylate
ACCA	Anti-chitobioside antibodies
AIC	Akaike Information Criterion
ALCA	Anti-laminaribioside antibodies
AMCA	Anti-mannobioside antibodies
ASCA	Anti-*Saccharomyces cerevisiae* antibodies
Anti-C	Anti-chitin antibody
Anti-L	Anti-laminarin antibody
BIC	Bayesian Information Criterion
BNF	British National Formulary
CASP	Critical Appraisal Skills Programme
CD8	Cluster of differentiation 8
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
cDNA	Complementary deoxyribonucleic acid
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CGQL	Cleveland Global Quality of Life
CI	Confidence interval
CRD	Centre for Reviews and Dissemination
CT	Computed tomography
DAR	Diagnostics assessment review
DARE	Database of Abstracts and Reviews of Effects
DEXA	Dual Energy X-ray Absorptiometry
DNA	Deoxyribonucleic acid
DSU	Decision Support Unit
DTA	Diagnostic test accuracy
EAG	External assessment group
ELISA	Enzyme-linked immunosorbent assay
EQ-5D	EuroQol 5-dimensions

FCP	Faecal calprotectin
gASCA	anti-Saccharomyces cerevisiae antibodies
HBI	Harvey–Bradshaw Index
HR	Hazard ratio
HRQoL	Health-related quality of life
HSUV	Health state utility values
HTA	Health technology assessment
IBD	Inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
IBDX	Crohn's disease Prognosis Test
ICER	Incremental cost-effectiveness ratio
IM	Immunomodulator
IPD	Individual patient data
ISRCTN	International Standard Randomised Controlled Trials Number
ITT	Intention-to-treat
KM	Kaplan–Meier
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
NHS	National Health Service
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
OD	Optical density
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PGWBI	Psychological General Well-Being Index
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROFILE	Predicting outcomes for Crohn's disease using a molecular biomarker
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life year
QUIPS	Quality of Prognosis Studies in Systematic Reviews
RCT	Randomised controlled trial

RNA	Ribonucleic acid
ROBINS-I	Risk Of Bias In Non-randomised Studies – of Interventions
RT	Reverse transcriptase
SC	Standard care
SLR	Systematic literature review
SU	Step up
TA	Technology Appraisal
TD	Top down
TMPT	Thiopurine methyltransferase
TNF	Tumour necrosis factor
TSD	Technical Support Document
TTE	Time to treatment escalation
TTS	Time to surgery
UK	United Kingdom
WBC	White blood cell

GLOSSARY

Accuracy	The ability of a test to identify positive and negative cases correctly. Calculated as the proportion of true positives and true negatives in all evaluated cases.
Cost effectiveness analysis	An economic analysis that converts effects into health terms and describes the costs per additional health gain.
False negative	An incorrect negative test result for an affected individual.
False positive	An incorrect positive test result for an unaffected individual.
Incremental cost-effectiveness ratio	The difference in the mean costs of two interventions in the population of interest divided by the difference in the mean outcomes in the population of interest.
Markov model	An analytical method particularly suited to modelling repeated events or the progression of a chronic disease over time.
Meta-analysis	Statistical techniques used to combine the results of two or more studies and obtain a combined estimate of effect.
Negative predictive value	Probability that people with a negative test result truly do not have the target condition.
Opportunity costs	The cost of forgone outcomes that could have been achieved through alternative investments.
Positive predictive value	Probability that people with a positive test result truly have the target condition.
Probabilistic sensitivity analysis	A method of quantifying uncertainty in a mathematical model, such as a cost-effectiveness model.
Reference standard	The best currently available test against which the index test is compared.
Sensitivity	Proportion of people with the target condition who test positive.
Specificity	Proportion of people without the target condition who test negative.
True negative	A correct negative test result for an unaffected individual.
True positive	A correct positive test result for an affected individual.

PLAIN ENGLISH SUMMARY

Crohn's disease (CD) is a condition in which parts of the digestive system become inflamed. CD affects people of all ages and is a lifelong condition for which there is no cure. Any part of the digestive system can be affected, and severity of disease can vary from person to person. Symptoms come and go and there can be times when there are no symptoms at all. Common symptoms of CD are diarrhoea, stomach ache, and blood in faeces. Medicines are given to reduce or control symptoms, and to try to stop inflammation from coming back. Some people with CD are at a higher risk than others of having more relapses and of developing complications of CD that might require surgery. This study aimed to see how effective two tools are at identifying people who might develop a complication or need surgery, which could help in choosing a person's treatment with the goal of reducing number of relapses and risk of surgery in the longer term. In addition, the review assesses the cost-effectiveness of the tools in terms of their value for money. We found limited evidence on how accurate the tools are in identifying people at high risk of complications. The lack of evidence on the tools meant that the cost-effectiveness analysis was simply assessing the value for money of standard care or a more aggressive treatment pathway for people with CD at higher risk of complications. The results of this analysis was that standard care was found to offer more value for money than a more aggressive treatment pathway for people with CD at higher risk of complications.

SCIENTIFIC SUMMARY

Background

Crohn's disease (CD) is characterised by inflammation of the gastrointestinal tract. CD is a lifelong condition for which there is no cure. The course of CD is characterised by recurring cycles of exacerbation (also referred to as flare) and remission, with the frequency of flare and duration of remission being highly variable across those with the condition. Some people are at a higher risk of following a more aggressive course of disease, which is typified by more frequent relapses and manifestation of penetrating or stricturing complications. For those with active disease, the National Institute for Health and Care Excellence (NICE) recommends treatment of CD with a step-up (SU) approach, which involves initial treatment with a glucocorticosteroid and stepwise progression through a pathway of immunomodulator (IM) and, finally, biological therapy with or without IM, as determined by response at each treatment step. However, research suggests that earlier aggressive treatment with the potent combination of biological therapy and IM could improve clinical outcomes for those at high risk of developing complicated CD. No test is available in the National Health Service (NHS) to stratify people with CD by risk of following a severe course of CD. Identification of those at a higher risk of developing complications of CD could lead to personalised management of an individual's condition.

Objectives

The aim of the diagnostic assessment review reported here was to assess the prognostic test accuracy, the clinical impact, and the cost-effectiveness of two prognostic tools for inflammatory bowel disease (IBD) in identifying those at a high risk of following a severe course of CD. To achieve the goal of the project:

- systematic reviews of the literature were carried out to identify evidence on prognostic accuracy and clinical impact of the Crohn's disease Prognosis Test (IBDX®) and PredictSURE-IBD™ in stratifying those with CD by risk of following a severe course of disease;

- an economic model was developed to assess the cost effectiveness of using the IBDX and PredictSURE-IBD tools.

Methods

Assessment of prognostic accuracy and clinical impact

Electronic databases (MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews, the Cochrane Central Register of Controlled Trials) were searched from inception to June 2019. Eligible studies assessed the prognostic accuracy or clinical impact of the IBDX (panel of six biomarkers) and PredictSURE-IBD tools in stratifying people at higher risk of following a severe course of CD. Two reviewers independently screened potentially relevant studies for inclusion against prespecified criteria, and assessed the quality of studies reporting prognostic accuracy using the Quality of Prognosis Studies in Systematic Reviews (QUIPS) tool. One reviewer extracted data from the included studies, with a second reviewer validating the data.

Assessment of cost effectiveness

The EAG developed a *de novo* economic model consisting of a decision tree designed to allocate patients to a response category after initial induction therapy in either the top-down (TD) or SU treatment arms. The decision tree is followed by a cohort model, where patients' level of response to maintenance therapy was assessed.

Patients enter the decision tree model after being allocated to the test (with either PredictSURE IBD™ in the base case or IBDX® in as scenario analysis) or no test (SC) arm. In both test and no test arms, patients are categorised as high-risk or low-risk patients, according to test results; or clinical judgment alone, depending on the model arm. Given that patients in the SC arm of the model can only receive the SU treatment approach and that the TD treatment approach is assumed to be received only by high-risk patients, the economic model is ultimately assessing the cost-effectiveness of TD therapy vs SU therapy in high-risk patients.

After induction therapy patients are classified as responders (improvement in CDAI score higher than 70) or non-responders (deterioration; no change; or an improvement of less than 70 in CDAI score). Duration of induction therapy differs by class of treatments (i.e., IM, anti-TNF, and second-line biologic). If patients respond to induction therapy, they move to the maintenance cohort model, while non-responders escalate to the next step on their allocated treatment strategy.

Responders to their first induction therapy enter the maintenance cohort model in the remission (CDAI<=150); mild (CDAI 150-220); or moderate to severe (CDAI 220-600) health states. Patients can move between these states during maintenance therapy, reflecting the different levels of response to treatment. The probability of patients transitioning between these states is also dependent on the

treatment class received. Patients in the mild and in the moderate to severe states are at risk of escalating to the next treatment step.

The EAG estimated surgical events as a stand-alone outcome in the model. This modelling simplification 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 surgeries is estimated, and the associated costs and impact on patients' quality of life is calculated. Patients who receive surgery in the model have an increased probability of dying associated with the procedure.

The economic assessment was taken from the perspective of the NHS and Personal Social Services and both costs and benefits were discounted at 3.5% per annum. Cycle length in the model was 2 weeks, and the time horizon of the model was 65 years.

Results

Searches of electronic database searches retrieved 6,258 unique records. Initial screening of titles and abstracts led to the identification of 36 publications for review of full texts. Of the 36 articles evaluated, 16 publications, including systematic reviews, were deemed to be relevant to the review of prognostic accuracy. Additionally, documents supplied by the companies marketing the prognostic tools were reviewed. Included studies were assessed for risk of bias and applicability using the QUIPS (QUality In Prognosis Studies) tool. Most studies reporting results for the IBDX tool were determined to be at moderate risk of bias for the population domain as the studies included those with a recent diagnosis and those with an established diagnosis of CD, and, in some studies, those with presence of complicated disease at baseline. Data were not analysed separately for the individual subgroups. Most studies were considered to be at a low risk of bias for attrition and for measurement of prognostic factors because all samples taken were analysed with the relevant tool and results generated as per the company's individual protocols. Additionally, outcome assessment was deemed to be a low risk of bias across many studies as the clinicians were masked to the results of the biomarker assessment.

Prognostic test accuracy

Twelve publications describing eight studies were included in the assessment of the prognostic accuracy of the tests. Seven of the studies reported results on utility of the IBDX kit and one study provided data on PredictSURE-IBD in stratifying those at high-risk of following a severe course of CD. Limited evidence is available from the included full-text publications on the prognostic accuracy of PredictSURE-IBD, and none is available on prognostic accuracy of IBDX, as determined by measures such as sensitivity and specificity. Most evidence on the utility of the two tools is derived from

observational studies that report estimates of risk of experiencing a clinical outcome associated with an aggressive course of CD, for example, need for treatment escalation, development of a complication or surgery. No study retrieved reported on the clinical impact of use of IBDX or PredictSURE-IBD in terms of influencing the treatments given in the management of active CD.

IBDX

Two studies reported an effect estimate for the risk of experiencing a complication and need for surgery by number of biomarkers testing positive. Both studies prospectively followed a cohort of people with an established diagnosis of CD. The two studies reported an increased risk of experiencing a complication or requiring surgery in those with positive status for at least two or three biomarkers out of the six comprising the IBDX panel. A third study identified a trend towards a larger proportion of people requiring surgery with increasing number of biomarkers testing positive, with a statistically significant difference across the categories assessed ($p<0.0001$).

PredictSURE-IBD

One observational study (prospective cohort) reported a sensitivity and specificity for predicting the need for multiple escalations within the first 18 months of 72.7% and 73.2%, respectively, where a cut-off of two or more treatment escalations was applied to categorise people as having followed a more aggressive course of CD. A negative predictive value of 90.9% was reported for PredictSURE-IBD of predicting multiple escalations within the first 18 months. The study additionally reported that those categorised as high risk of following a severe course of CD had a statistically significantly higher risk of first treatment escalation compared with those designated as low risk, with a hazard ratio of 2.65 (95% CI: 1.32 to 5.34; $p=0.006$).

Cost effectiveness

As no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, the development of the economic model sets a structural framework for analysing future available data on prognostic accuracy and assesses the costs and consequences of treating high- and low-risk patients with both TD and SU strategies.

The clinical input parameters in the base case economic model for PredictSURE IBD™ and in the scenario analysis for IBDX® are the same. The only difference in the cost-effectiveness analyses of the two diagnostic tests is the cost of the test.

The EAG found two main sources of evidence that could be used to model time to treatment escalation (TTE) and time to surgery (TTS). Nevertheless, each source could only partially inform the TTE and TTS analyses in the model. Therefore, clinical data informing the analysis had to be derived from multiple sources. This approach is not ideal and creates a patchwork network of evidence, introducing uncertainty in the economic results. It is anticipated by the EAG that this problem will be (at least partially) overcome when results from the PROFILE trial are available to populate the economic model.

The incremental analysis of cost-effectiveness demonstrates that the TD strategy (via the use of PredictSURE IBD™ in the model) is dominated by SU (via the SC arm of the model), with an additional cost of £9,526 and a QALY loss of 0.06.

Conclusions

Despite extensive systematic searches of the literature, no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, IBDX and PredictSURE-IBD. In terms of sensitivity and specificity for estimate of prognostic accuracy, the External Assessment Group (EAG) is unaware of a validated definition for determination of whether a person has followed a severe course of CD, for example, a set number of treatment escalations or development of a complication or need for surgery. Thus, the EAG considers the criterion required for a true positive or false positive for IBDX and PredictSURE-IBD to be unclear. The EAG considers it would be challenging to ascertain an accurate estimate of prognostic accuracy of the tools in stratifying course of CD and to do so would require carrying out a prospective study that included a group or groups that received only "step-up" (SU) treatment after determination of risk of course of CD. The ongoing PROFILE RCT randomises people to accelerated SU or TD treatment after determination of high or low risk of following a severe course of CD and so will provide additional data to inform estimates of prognostic accuracy.

One of the key underlying assumptions in the EAG's base case economic analysis is that high-risk patients who initiate treatment with IMs escalate treatment quicker than high-risk patients who initiate treatment with anti-TNF (supported by the data presented in D'Haens *et al*.). However, once these patients initiate subsequent treatment with an anti-TNF (their second treatment step), they "catch-up" with patients on the TD treatment strategy. As some high-risk patients who receive SU treatment respond to IM treatment, having the additional IM step in the SU strategy is advantageous to patients in the EAG's base case analysis as patients still subsequently receive treatment with biologics, which are assumed to have the same effect as biologics is the TD arm. Given the paucity of data to substantiate any further benefits in subsequent treatment steps in the TD vs SU approaches, the EAG considered this to be the most conservative modelling approach.

The EAG's analysis has shown that too much uncertainty remains around the potential benefits of TD treatment for high-risk patients. The cost-effectiveness of a TD strategy compared with a SU strategy in high-risk patients is highly dependent on two unanswered questions: 1) do some high-risk patients derive a benefit from receiving IM treatment before moving to biologic treatment? 2) do SU high-risk patients have the same benefits as TD high-risk patients once they initiate the TD treatment pathway (i.e. treatment with anti-TNF). In the EAG's model, the potential disadvantage of waiting to initiate treatment with anti-TNF was only based on the increased risk of surgery in the SU arm, however, the negative impact of surgery in the analysis was not enough to offset the advantages of initial treatment with IM for SU patients.

The EAG conducted a range of analyses to test extreme scenarios around increasing the relative treatment effectiveness of the TD approach while decreasing the relative costs associated with TD. The ICERs for PredictSURE IBD™ (and TD) compared with SC (and SU) fell below £30,000 in the analysis. However, the EAG notes that these results need to be interpreted with extreme caution as the assumptions made in these scenarios were designed to test extreme clinical scenarios and were not evidence-based. The EAG concludes that its base case analysis showing that TD is dominated by SU remains the most conservative assessment of the relative cost-effectiveness of these treatment strategies.

Study registration

The protocol for the review is registered on PROSPERO as CRD42019138737.

Funding

This report was commissioned by the NIHR Systematic Reviews Programme as project number 128968/T.

1 BACKGROUND AND DEFINITION OF THE DECISION PROBLEM

1.1 Description of Crohn's disease

Crohn's disease (CD) is one of the two primary types of inflammatory bowel disease (IBD), with the other being ulcerative colitis.[1-3] The symptoms of CD and ulcerative colitis are similar, and both types of IBD are characterised by inflammation of the gastrointestinal tract. CD is a lifelong condition that is characterised by recurring cycles of exacerbation (also referred to as flare) and remission, and for which there is no cure. The frequency of flare and the duration of remission are highly variable across those affected by CD. Some people are at a higher risk of following a more aggressive course of disease, typified by more frequent relapses and manifestation of penetrating or stricturing complications.[1-3] Identification of those at a higher risk of developing complications of CD could lead to personalised management of an individual's condition and improvement in clinical outcomes.

1.1.1 Aetiology, pathology and prognosis

Neither the underlying aetiology of CD nor the factors that determine the course and prognosis of the disease are fully understood. Environmental factors (e.g., smoking), genetic predisposition, and dysregulation of the immune system are thought to have a role in the development of and course of CD.[2, 4]

CD can affect any segment of the gastrointestinal tract, from the mouth to the anus, but the most commonly affected areas are the distal ileum (the last part of the small intestine) and the colon.[5] CD that is primarily located in the colon often has a high symptom burden, whereas disease affecting the ileum can be extensive but be associated with relatively few symptoms.[6] Diseased segments are frequently separated by intervening areas of healthy bowel tissue.[2, 4] The size of the inflamed area may be limited to a few centimetres, or could affect an extensive part of the bowel. As well as affecting the lining of the gastrointestinal tract, CD may also penetrate through the wall of the bowel.[2, 4]

As CD can affect any part of the gastrointestinal tract, and to differing extents, symptoms experienced by people with the disease vary markedly, sometimes making recognition and diagnosis difficult.[2, 4] Moreover, symptoms and severity of disease can change over time. People with CD most commonly present with:[2, 4, 7]

- abdominal pain;
- diarrhoea (mucus, pus or blood may be mixed with the diarrhoea);

- tiredness and fatigue;
- loss of appetite and weight loss;
- anaemia.

CD can also lead to signs and symptoms outside the gastrointestinal tract, which are known as extraintestinal manifestations and have been reported to be more common with CD primarily located in the colon.[6, 7] Associated conditions typically occur during flare, but can also manifest during remission or before development of any signs of IBD. Conditions developing as a result of CD include:[7]

- arthritis (more commonly of the large joints of the arms and legs, including the elbows, wrists, knees and ankles);
- skin problems, most commonly erythema nodosum;
- eye problems (episcleritis, scleritis and uveitis);
- liver problems (e.g., primary biliary cholangitis).

Flares of IBD indicate a return to active disease and, potentially, symptoms for an individual. Several factors have been proposed as triggers for flare, including poor adherence to treatment, certain medications (e.g., antibiotics and non-steroidal anti-inflammatory drugs), infection, smoking, and emotional stress.[8, 9] As has been noted in other immune-mediated diseases, the course of CD varies widely among affected individuals, making it challenging to predict the severity or frequency of occurrence of flare.

As CD is not curable, the goal of management of the condition is to induce and maintain remission. Population-based studies investigating long-term prognosis of CD reported that, within the first year of diagnosis, 50–65% of people achieved remission, and 15–25% experienced low level of disease activity.[10-12] However, 10–30% of people with CD had a relapse or exacerbation of their condition in the first year. Long-term follow-up (10–15 years) indicated that 67–73% of people with CD experienced a chronic relapsing course and 13–20% had a chronic disease course with continuous activity. By contrast, 10–13% of those with CD achieved remission for several years. For those with CD in remission after treatment, relapse rates at 1, 2, 5, and 10 years are estimated at 20%, 40%, 67%, and 76%, respectively.[13]

Those who develop CD that follows a non-severe course might achieve prolonged remission with no treatment. In contrast to a non-complicated course of CD, cases characterised as following a severe course are likely to experience more frequent flares, and typically require early aggressive treatment strategies, including multiple treatment escalations and augmentation. People with severe forms of CD are at high risk of complications of disease, including intestinal obstruction, fistulae and perianal disease, and progressive disability and need for surgery.[2, 4, 7]

Prognostic factors associated with a more complicated, severe course of CD include bowel damage, extraintestinal manifestations of disease, higher number of flares, need for glucocorticoids, and resultant hospitalisations.[14] Other risk factors for a worse course of disease include smoking, and fistula formation. Factors present at diagnosis of CD found to be associated with a worse prognosis were young age (<40 years), presence of perianal disease, and initial need for glucocorticosteroid treatment.[15] Presence of known risk factors for flare and for complications in CD could influence the treating clinician's management of the condition, but consensus on use of risk factors to determine prognosis of disease has yet to be achieved and treatment can vary.

1.1.2 Epidemiology

CD can appear at any age, but is most often diagnosed in adolescents and adults between the ages of 20 and 30, with a second peak in diagnosis, albeit smaller, between the ages of 60 and 80 years.[16] In the UK, it is estimated that CD affects one in every 650 people[7] and that there are at least 115,000 people with the condition.[4] Incidence and prevalence of CD have been rising since the mid-1970s, with highest rates observed in Northern Europe and North America.[17] Incidence of CD in the UK is reported to be about 8 per 100,000 people per year,[18, 19] with an age–sex adjusted point prevalence of 144.8 per 100,000 people.[19]

1.1.3 Impact of Crohn's disease

Affecting men and women equally, CD is a debilitating disease that has a marked impact on physical and emotional health, as well as quality of life. Additionally, CD is associated with high economic burden due to disability, loss of work productivity, surgery and hospitalisation.[20] A UK study published in 2015 estimated the annual cost of care for a person with CD to be £6,156 (£1,800 for those in remission compared with £10,513 for those experiencing relapse), which translated to a total annual cost of ~£700 million.[21] Five years after onset, 15% to 20% of people are affected to some degree by their disease, and between 50% and 80% of people with CD will eventually need surgery as a result of, for example, development of strictures, perforation of the bowel, or failure of drug therapy.[22]

1.1.4 Current diagnostic and treatment pathways

1.1.4.1 Identification of those at risk of following severe course of Crohn's disease

As highlighted earlier, the symptoms of CD are common to various conditions, which makes diagnosis challenging. The diagnosis and determination of extent of CD is reached through a combination of clinical examination, laboratory tests, radiological imaging, and endoscopy.[23] Furthermore, once a diagnosis of CD has been made, there is no validated test or algorithm available to stratify people with CD by risk of developing complications of disease.

Standard laboratory investigations for a person suspected of having CD include assessment of full blood count, inflammatory markers (e.g., C-reactive protein [CRP] and faecal calprotectin [FCP]), electrolytes, and liver enzymes, as well as microbiological analysis of a stool sample.[23] Although raised inflammatory markers are not specific to IBD, and identification does not differentiate IBD from infectious colitis, high CRP levels are broadly correlated with severity of disease activity in CD, and can be used to monitor disease progression.

Once a diagnosis of CD has been established, guidelines suggest that subsequent investigations focus on assessing level of activity of disease, as well as risk of complications in the longer term.[24] Three key areas are assessed when determining the severity of CD: impact of the disease on the individual (e.g., clinical symptoms, quality of life, fatigue, and disability); burden of disease (e.g., mucosal lesions, upper gastrointestinal involvement, and disease extent); and course of disease (e.g., structural damage, perianal disease, number of flares, and extraintestinal manifestations).[25]

Two clinical tools available to assess level of disease activity are the Crohn's Disease Activity Index (CDAI)[26] and the Harvey–Bradshaw Index (HBI).[27] The HBI is a simple derivative of the CDAI, and the two tools are correlated, with a change in the CDAI of 100 points corresponding to a 3-point change in the HBI.[28] Clinical experts commented that, in clinical practice, their preference is for the HBI tool, as the CDAI is impractical for routine clinical assessment and its use is typically limited to clinical trials. Severity of disease activity is categorised as:[15]

- clinical remission: CDAI score of ≤150, which corresponds to a HBI score of ≤4;
- mild: CDAI score of 150–220, which corresponds to a HBI score of 4–8;
- moderate–severe: CDAI score of 220–450, which corresponds to a HBI score of ≥8;
- severe fulminant disease: CDAI score of ≥450, which corresponds to a HBI score of ≥15.

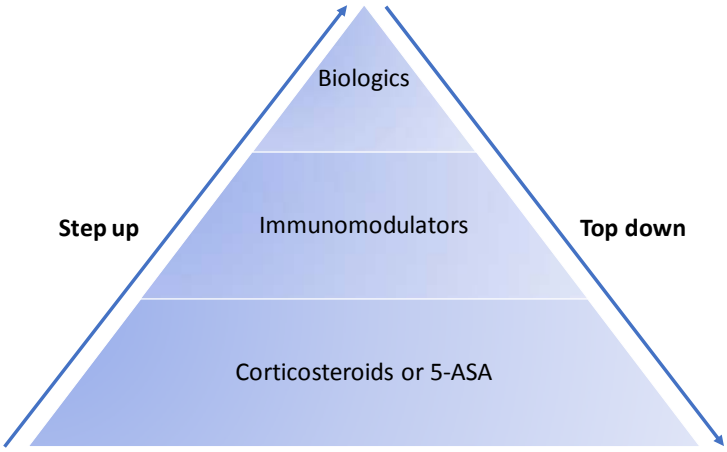
CD activity and severity could be considered a continuum, and some people might not easily be categorised based on their symptoms. Moreover, the CDAI and HBI are based on subjective measures, and there is a move to using more objective parameters and the presence or absence of bowel destruction to assess severity.[24] Using patient-reported outcomes to assess disease activity in CD is also becoming more common. Often used to guide treatment recommendations, the CDAI and HBI scores represent status of activity at a point in time and do not account for long-term prognosis or course of disease.[14]

Endoscopic assessments and biopsies provide data on the level of disease activity in CD but do not give an insight into factors associated with relapse and course of disease. Evaluating blood- and stool-based biomarkers of inflammation, such as CRP and faecal calprotectin (FCP), respectively, is less invasive than endoscopy, and such laboratory tests provide reproducible, quantitative results that, together with clinical assessment, can aid clinicians in the diagnosis and management of CD. However, as noted earlier, serum and faecal biomarkers are not necessarily specific to CD, and they have limited applications in the prediction of the severity of the course of IBD, including CD, in the longer term.[29] There is no consensus or algorithm available outlining how to combine known risk factors to determine long-term prognosis of CD, and estimation of risk of following a severe course of disease is based on subjective clinical judgement, together with input from the patient.

1.1.4.2 Management of Crohn's disease

The goal of treatment in CD is to initially control or reduce symptoms to induce remission.[30] Once symptoms are under control, maintenance treatment might be given to prolong remission and minimise risk of relapse. Globally, there are two pharmacological treatment algorithms followed in the management of active CD – the "step up" (SU) and "top down" (TD) approaches (Figure 1) – both of which involve several tiers of medication, and, as the names suggest, are the inverse of each other.[31] Additionally, surgery might be necessary at any stage of the disease but can be considered as an alternative to medical treatment in some people, particularly in the setting where the disease is limited to the distal ileum.[30]

Figure 1. “Step up” versus “top down” treatment algorithms for Crohn’s disease



Abbreviation: 5-ASA, 5-aminosalicylate.

Note: In the treatment hierarchy, the more potent drug therapies are placed at the top of the pyramid.

Currently, NICE guidance (NG129) recommends a SU approach for the medical management of CD.[30] The SU algorithm (Section 1.1.4.2.1) involves starting treatment with the least aggressive of the medical options available and escalating therapy in reactive stepwise stages in response to recurrent flares or persistently active disease. An alternative treatment path involves an “accelerated SU” plan in which those patients considered to have more severe disease or with clinical markers of poor outcome advance rapidly up the treatment ladder, receiving earlier aggressive therapy. The Evidence Assessment Group’s (EAG’s) clinical experts advised that, for those people judged to be at risk of a more severe clinical course (e.g., extensive small bowel disease, perianal disease or upper gastrointestinal disease), most clinicians would prefer an “accelerated SU” approach rather than follow the slower conventional “SU” algorithm.

The TD approach (Section 1.1.4.2.2) is not recommended by NICE at the time of writing.[30] The strategy involves treatment earlier in the pathway with biological therapies, which are more clinically effective but are also potentially associated with a greater risk of adverse effects (e.g., increased rate of infection and malignancy).[32] Early use of biological therapies in a TD approach is thought to modify the course of CD, to increase the possibility of mucosal healing (preventing structural damage of the bowel), and to be more effective than the SU approach at inducing and prolonging remission:[31] the goal of achieving mucosal healing during treatment is gaining acceptance but is not yet part of standard care in the UK.

Another challenge in the management of CD is timing of de-escalation of treatment, which can be defined as either decreasing the dose of a drug or complete cessation of therapy. De-escalation of therapy in both SU and TD strategies is typically considered when a person achieves deep remission,

which comprises clinical and biological remission. De-escalation is proposed for those at highest risk of potential complications of treatment, such as infection or malignancy, or for those at lowest risk of relapse after cessation of treatment. De-escalation might not be appropriate for all those achieving deep remission. Factors that need to be accounted for when considering de-escalation of therapy include age, gender, treatments given and severity of CD.[33] A systematic review evaluating de-escalating anti-TNF or IM therapy in people with CD who were in deep remission for at least 6 months found that de-escalating medical therapy in this cohort of people was appropriate for only a small proportion of carefully selected people, predominantly those with uncomplicated disease and the elderly.[33]

Neither the SU nor TD approach is suitable for all people with CD. Considering the risk–benefit profile of the TD approach, some clinicians could be reticent to expose those with mild activity of CD at time of assessment, or those thought to be at low risk of experiencing a relapse, to the unnecessary risk of an adverse effect. Conversely, those assessed as potentially being at risk of having a severe course of disease are at risk of undertreatment if the conventional SU approach is followed, with consequent prolonging of symptoms and of inadequate control of disease activity, and the associated long-term risks. Another consideration is cost of treatment, with the TD approach typically more expensive compared with the SU approach.[32]

Ability to easily stratify those with CD by risk of course of disease could help identify the most appropriate treatment strategy for the patient.

1.1.4.2.1 "Step-up" approach

NICE NG129[30] advises starting treatment with a glucocorticosteroid (prednisolone, methylprednisolone or intravenous hydrocortisone [for in patients]) to induce remission in those with a first presentation or a single inflammatory exacerbation of CD in a 12-month period. For those with mild disease who cannot tolerate or who are contraindicated to the recommended glucocorticosteroids, alternative treatments for first presentation or a single inflammatory exacerbation in 12 months are budesonide (another glucocorticosteroid) and 5-aminosalicylate (5-ASA). Additionally, budesonide can be considered for those who have one or more of distal ileal, ileocaecal or right-sided colonic disease. For children or young people for whom there is a concern about growth or adverse effects, NICE advises considering enteral nutrition as an alternative to a conventional glucocorticosteroid.[30]

Both budesonide and 5-ASA are less effective than the preferred initial glucocorticosteroids, but they might be associated with fewer adverse effects: clinical experts advised that increasingly 5-ASAs are considered to have a limited role in the management of CD. Budesonide should not be considered for those presenting with severe disease activity or exacerbations.

Should remission not be achieved after induction therapy, the next step in the treatment pathway is addition of an immunomodulator (IM; azathioprine, mercaptopurine or methotrexate) to conventional glucocorticosteroid or budesonide, specifically in cases where:[30]

- a person experiences two or more inflammatory exacerbations in a 12-month period;

or

- the glucocorticosteroid dose cannot be tapered.

NICE cautions that, before offering azathioprine or mercaptopurine, thiopurine methyltransferase activity (TPMT) should be assessed. Azathioprine or mercaptopurine should not be offered in cases when TPMT activity is deficient (very low or absent), and a lower dose of both IMs should be considered if TPMT activity is below normal but not deficient (according to local laboratory reference values). Alternatively, if it is thought that the person would be unable to tolerate mercaptopurine or azathioprine, addition of methotrexate could be considered.

For adults with severe active CD whose disease has not responded to conventional therapy (including IM and/or glucocorticosteroid treatments), or who are intolerant of or have contraindications to conventional treatment, recommended therapy is escalation to infliximab or adalimumab within their licensed indications, both of which are tumour necrosis factor (TNF)-alpha inhibitors.[30] Biosimilars of infliximab and adalimumab are available and can be used interchangeably with originator anti-TNFs in clinical practice. Infliximab and adalimumab can be given alone or in combination with an IM, and the therapies should be given as a planned course until treatment failure (including the need for surgery), or until 12 months after the start of treatment, whichever is shorter. Treatment with infliximab or adalimumab could be continued if there is clear evidence of ongoing active disease as determined by clinical symptoms, biological markers and further investigation, including endoscopy, if necessary. However, NICE advises that disease activity should be reassessed at least every 12 months to determine whether continued treatment with infliximab or adalimumab is still clinically appropriate. People whose CD relapses on cessation of treatment with biological therapy should have the option to recommence infliximab or adalimumab.

For those with moderately to severely active CD and who have failed treatment with a TNF-alpha inhibitor (i.e., disease has responded inadequately or lost response to treatment) or who are intolerant to conventional therapies and are contraindicated to anti-TNFs, other biologics, such as vedolizumab and ustekinumab, are additional treatment options.[30]

Once a person achieves remission, NICE advises discussing with people affected by CD, together with their family members or carers, options for managing their condition, one of which may be no further treatment.[30] For those who choose to proceed with therapy to maintain remission, available options are:

- azathioprine or mercaptopurine as monotherapy to maintain remission when previously used with glucocorticosteroids (including budesonide) to induce remission and for those who have not previously received these drugs;
- methotrexate:
    - for people who required methotrexate to induce remission;
    - for people who tried but could not tolerate azathioprine or mercaptopurine for maintenance;
    - for people contraindicated to azathioprine or mercaptopurine.
- continued treatment with biologic, if appropriate.

1.1.4.2.2 "Top-down" approach

Although the "top-down" approach is not recommended by NICE, clinicians in specialist centres might choose to offer the strategy as an option to those they consider to have a poor prognosis in terms of outcomes, for example, those with complex perianal disease, significant fistulising disease or those with multiple risk factors. No accepted treatment strategy is available for the TD approach, with disparity across studies in the definition of "aggressive" therapy. TD can involve early use of biological therapies, or of IMs, or a combination of biological therapy and IM. In two landmark studies evaluating the clinical efficacy of early aggressive therapy in those with CD, "top-down" treatment comprised infliximab in combination with azathioprine.[34, 35] However, evidence in support of the effectiveness of the TD approach when compared directly with the "step-up" approach is inconsistent,[32] with two studies finding a benefit of early treatment with biologics[35, 36] and one reporting no benefit over the less aggressive strategy.[37] Variation in results across studies could be related to differences in, for example, the definition of 'early' intervention and in trial design, outcomes measured, population, and trial duration.

Being able to better predict the course of CD would help clinicians identify those who could benefit most from early use of aggressive treatments (IMs and biological therapies) and decide on the most appropriate treatment to manage symptoms. Tools such as the PredictSURE-IBD (PredictImmune) and

IBDX® (Glycominds International) could potentially help achieve the goal of personalising treatment in CD.

1.2 Description of the technologies under assessment

1.2.1 Crohn's disease Prognosis Test

Glycominds International envisage that the IBDX tool can be implemented at three key stages in the management of CD:

- differential diagnosis of CD from ulcerative colitis;
- to assess risk of developing more aggressive disease course in those diagnosed with CD and not having yet experienced complications and/or undergone surgery;
- to predict risk for future events in those who have experienced a first CD complication or surgery.

The IBDX tool detects serum levels of specific anti-glycan-antibodies, which are a set of serological biomarkers reported to be highly specific for Crohn's disease, with potential predictive value for prediction of complicated course of disease.[38] Glycans are saccharides that can be attached to various biological molecules through an enzymatic process called glycosylation. Glycans are usually found on the exterior of cell walls, and they form the main components of the cell wall surface in many microbes, including fungi, yeast, and bacteria.[38]

An atypical interaction of environmental, genetic and microbial factors with the immune system is thought to lead to the production of antibodies against intestinal microorganisms in those with CD that results in the gastrointestinal inflammation typical of the condition.[39, 40] Examples of microbial antibodies include anti-Saccharomyces cerevisiae antibodies (ASCA; also referred to as gASCA), antibodies against Pseudomonas-associated sequence I2 (anti-I2), and against the bacterial flagellin cBir1 (anti-cBir1).[41] Anti-glycan antibodies comprise antibodies against ASCA, anti-mannobioside antibodies (AMCA), anti-laminaribioside antibodies (ALCA), anti-chitobioside antibodies (ACCA), anti-laminarin antibody (anti-L) and anti-chitin antibody (anti-C).

Antibodies detected by the IBDX tool include:[42]

- ACCA;
- ALCA;

- AMCA;
- gASCA;
- anti-L;
- anti-C.

The IBDX tool is supplied as a set of six biomarker kits (listed above), each of which detects a circulating antibody against the kit-specific antigen in patient serum or plasma by an indirect solid-phase enzyme-linked immunosorbent assay (ELISA). Individual kits contain the relevant anti-glycan 96-well microwell plate (12 X 8 well strips), ELISA reagents, negative control, positive control, and calibrators:[43] each kit can assess up to 90 samples excluding controls, but the company recommends running samples in duplicate (i.e., maximum of 45 assays per kit accounting for controls). The microwell plates, conjugates and controls are specific for each kit, but all other reagents are the same. All kits follow the same procedure (including incubation times), so they can easily be processed at the same time, if desired. On completion of incubation, absorbance of the calibrator, controls and samples can be evaluated spectrophotometrically. Optical density (OD) is directly proportional to the amount of bound antibody. Arbitrary units are calculated based on sample OD and calibrator serum sample OD.[43] For each biomarker, positivity is assessed based on the cut-off values presented in Table 1.

Those with CD are considered to be at greater risk for disease complication (stricturing or penetrating) or surgery intervention if they are positive for two or more serological markers.[42] Figure 2 presents a flowchart (adapted from that available in the instructions for the IBDX kit[43]) summarising how to interpret the complete panel of results from the individual biomarkers.

The company highlights that anti-glycan antibodies are also detected at the time of diagnosis in people with Coeliac disease. However, as noted by the company, initial positivity for various anti-glycan antibodies is lost after people with Coeliac disease follow a long-term gluten-free diet.[44] Coeliac disease and IBD can be co-morbid, and studies suggest that people with IBD are at an is an increased risk of Coeliac disease.[45] Therefore, the company recommends against using the IBDX kit without exclusion of diagnosis of Coeliac disease in those who have not followed a gluten-free diet. The EAG's clinical experts fed back that, as the symptoms of CD and Coeliac disease overlap, most people referred for suspicion of CD are likely to be tested for Coeliac disease, which requires a blood test. The EAG's clinical experts commented that the test for risk of course of CD and presence of Coeliac disease could be done simultaneously.

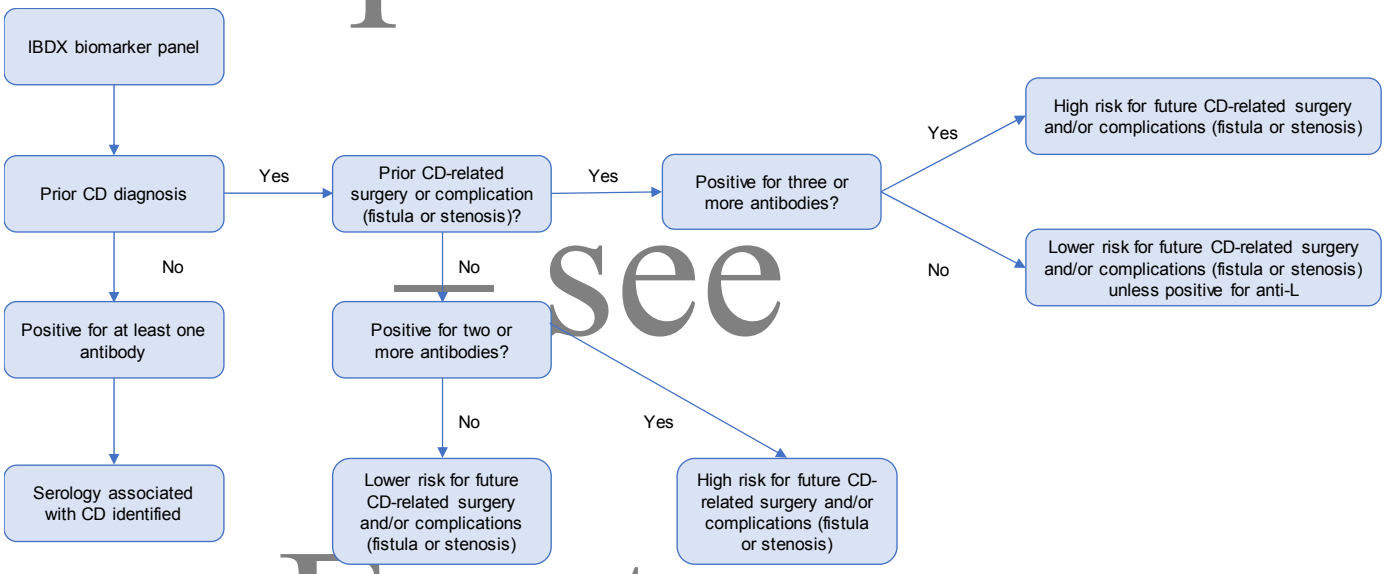
Table 1. Cut-off values for individual IBDX ELISA kits[43]

	gASCA	ACCA	ALCA	AMCA	anti-C	anti-L
Negative	<45	<80	<55	<90	<45	<45
Equivocal[a]	45–50	80–90	55–60	90–100	45–50	45–50
Positive	>50	>90	>60	>100	>50	>50

[a] Repetition of sample assay is recommended.

Abbreviations: ACCA, anti-chitobioside antibodies; ALCA, anti-laminaribioside antibodies; AMCA, anti-mannobioside antibodies; anti-C, anti-chitin antibody; anti-L, anti-laminarin antibody; ELISA, enzyme-linked immunosorbent assay; gASCA, anti-Saccharomyces cerevisiae antibodies.

Figure 2. Overview of interpretation of results from individual IBDX ELISA kits (adapted from instructions provided by Glycominds International[43])



Abbreviations: anti-L, anti-laminarin antibody; CD, Crohn's disease.

1.2.2 PredictSURE-IBD

PredictSURE-IBD™ is proposed for use in adults (16 years or older) with IBD, including CD, who have active disease and are not receiving concomitant glucocorticosteroids, IMs or biological therapies. PredictSURE-IBD could be particularly beneficial for people with:

- newly or recently diagnosed IBD;
- moderate or severe active IBD (people with mild disease are unlikely to have early aggressive treatment with biologics);

- disease that would not require early aggressive treatment with biologics ('top-down' approach) with current standard care in the NHS (e.g., people who do not have fistulising and/or complex perianal CD, or have multiple risk factors).

PredictSURE-IBD facilitates stratification of people with IBD into high and low risk of frequently relapsing course of disease through detection of a gene sequence associated with CD8+ (cluster of differentiation 8) T cell exhaustion.

Gene expression profiling of peripheral blood CD8+ T cells identified a signature gene sequence that was associated with CD8+ T cell exhaustion,[46-48] a state that is reached through the stepwise and progressive loss of T-cell function and that inhibits the immune response.[49] Level of expression of the genes indicating CD8+ T cell exhaustion was found to be linked with course of disease in multiple autoimmune diseases, including IBD.[46-48] People with a CD8+ T cell signature not associated with T cell exhaustion were shown to be at higher risk of a frequently relapsing disease course than those with the signature for T cell exhaustion.[46-48]

The PredictSURE-IBD test determines the presence or absence of the signature gene sequence (15 target genes and 2 control genes;[50] Table 2) indicating CD8+ T cell exhaustion through *in vitro* quantitative reverse transcription-polymerase chain reaction (RT-qPCR) of messenger ribonucleic acid (mRNA) isolated from a whole blood sample (2.5 ml). The blood sample must be taken by a trained professional and stored in a sample tube (PAXgene Blood RNA tube): the vessel for the blood sample is not supplied as a component of the PredictSURE-IBD test kit and must be purchased separately. Isolation of mRNA and subsequent RT-qPCR is carried out in a centralised laboratory (Clinical Genetics Laboratory, Addenbrooke's Treatment Centre, Cambridge University Hospitals NHS Foundation Trust).

Table 2. Informative genes in PredictSURE-IBD optimised qPCR classifier[50]

Gene ID	Gene name
FCRL5	Fc receptor-like 5
GBP5	Guanylate binding protein 5
GZMH	Granzyme H
GZMK	Granzyme K
HP	Haptoglobin
IFI44L	Interferon-induced protein 44-like
IL18RAP	Interleukin 18 receptor accessory protein
LGALSL	Lectin, galactoside-binding-like

LINC01136	Long intergenic non-protein coding ribonucleic acid 1136
LY96	Lymphocyte antigen 96
NUDT7	Nudix (nucleoside diphosphate linked moiety X)-type motif
P2RY14	Purinergic receptor P2Y, G-protein coupled, 14
TRGC2 / TRGJ1	T cell receptor gamma constant 2 / T cell receptor gamma joining 1
TRGV3	T cell receptor gamma variable 3
VTRNA1-1	Vault RNA 43101

In RT-qPCR, because the starting genetic material is RNA rather than deoxyribonucleic acid (DNA), the first step in the process requires transcription of mRNA into complementary DNA (cDNA) using reverse transcriptase (RT). Next, the cDNA acts as the template for qPCR for DNA amplification. qPCR is carried out in a 384-well plate (16 x 24 wells). Given the requirements for quality control of the assay, a maximum of 4 samples can be analysed per plate. Each sample of cDNA is amplified in triplicate, which requires 12 rows of the plate. A quality control RNA (supplied as part of the PredictSURE-IBD kit and run in triplicate [3 rows]) and a no-RNA control (run singularly [1 row]) are tested with each batch of mRNA samples to validate the run. The centralised laboratory uses a Roche LightCycler 480/480 II® platform, which is a standard platform, to carry out RT-PCR. Staff training to process the PredictSURE-IBD kits will not be required at the centralised laboratory as the site is already providing testing services as part of an ongoing study (PROFILE[51]). If required, PredictImmune would support staff training at additional laboratories to facilitate expansion of testing, with training thought to require 2–3 days at each centre.[52]

Results from RT-qPCR are fed into a proprietary algorithm that calculates a continuous risk score, and based on this score, patients are categorised as high- or low-risk of following a frequently relapsing form of IBD. A confidence level associated with the result is also reported and presented as a percentage. Turnaround time for the test is 7–10 days.

1.3 Comparator

As no validated tool or algorithm is available to determine course of Crohn's disease, the relevant comparator is standard clinical care in the NHS.

1.4 Reference standard

As there is no test or algorithm available to determine long-term course of disease or an individual's risk of developing severe course of disease, estimation of prognosis is based on subjective clinical judgement of presenting signs and symptoms, together with potential risk factors for severe course of disease. Thus, there is no reference standard for the tools under evaluation.

1.5 Aim of the assessment

The aim of this diagnostic assessment review is to assess the prognostic test accuracy and clinical and cost-effectiveness of two molecular prognostic tools for IBD in identifying those at a high risk of severe course of CD. The tools assessed in the review reported here are IBDX and PredictSURE-IBD. At the time of writing, no validated test or algorithm is available to stratify people with CD by risk of developing complications of disease. Presence of known risk factors for flare and for complications in CD could influence the treating clinician's management of the condition, but consensus on use of risk factors to determine prognosis of disease has yet to be achieved and treatment can vary. The accuracy, clinical and cost-effectiveness of the tools will be evaluated against standard clinical care in the National Health Service, based on input from clinical advisors, when assessing the likely course of Crohn's disease.

2 METHODS FOR ASSESSING CLINICAL EFFECTIVENESS

A systematic literature review was carried out to evaluate, first, the prognostic test accuracy of the Crohn's disease Prognosis Test (IBDX;[53] Glycominds International) and PredictSURE-IBD (PredictImmune[54]) tools in the identification of those at high risk versus at low risk of developing a severe course of Crohn's disease (CD), and, second, the clinical impact of using the tools in the management of CD.

Methods for the systematic review were in line with those reported in a prespecified protocol that was registered on the international prospective register of systematic reviews (PROSPERO: CRD42019138737[55]). General principles followed were those outlined in the Centre for Reviews and Dissemination (CRD) guidance for conducting reviews in healthcare,[56] the National Institute for Health and Care Excellence's (NICE's) Diagnostics Assessment Programme manual,[57] and the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.[58] The systematic review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for diagnostic test accuracy (DTA) studies. The PRISMA-DTA checklist and PRISMA-DTA for abstracts checklist are presented in Appendix 1 and Appendix 2 of this report, respectively.

2.1 Search strategy

Search strategies for electronic databases were designed with a focus on the target condition of the systematic review (i.e., CD) and the specified prognostic tools (i.e., IBDX and PredictSURE-IBD). Strategies comprised a combination of Medical Subject Headings (MeSH) and free text terms. During the scoping search process, no record was retrieved using the term "PredictSURE-IBD" or any appropriate derivative, and it was noted that terms including tradenames of the prognostic tools must be combined with "or" to avoid omission of known potentially relevant studies. Names for the prognostic tools of interest, and relevant alternative terms, were included in consideration of future updates. No study design filters were applied and all electronic databases were searched from inception through to 14 June 2019. The search strategies applied in electronic databases to retrieve records on studies evaluating prognostic accuracy and the impact of using the tools on the management of CD are available in Appendix 3.

The records retrieved from electronic databases were uploaded to, and deduplicated in, EndNote X7 software. The deduplicated list of records was exported to Rayyan QCRI, which was used to co-ordinate the assessment of titles and abstracts by two independent reviewers. The reference lists of relevant

systematic reviews and eligible studies were hand-searched to identify additional potentially relevant studies.

Data submitted by the manufacturers of the two prognostic tools that are the focus of this assessment were considered for inclusion in the review. Any 'commercial in confidence' data provided by companies, and specified as such, has been highlighted in [redacted] in the assessment report (followed by the company name in parentheses). Any 'academic in confidence' data provided by companies, and specified as such, has been highlighted in [redacted] in the assessment report.

Electronic databases searched for relevant studies were:

- MEDLINE (MEDLINE and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily and Versions; Ovid);
- EMBASE (Ovid);
- the Cochrane Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR).

Clinical trial registers were searched to identify relevant ongoing clinical trials that when completed may have an impact on the results of this review:

- WHO International Clinical Trials Registry Platform;
- ClinicalTrials.gov.

The website of the US Food and Drug Administration (FDA) was also searched to identify unpublished data.

Abstracts from key conference proceedings from the past 2 years, were screened for additional potentially relevant studies. Conferences identified by clinical experts as of importance to the assessment were those organised by:

- British Society of Gastroenterology;
- European Crohn's and Colitis Organisation;
- Digestive Disease Week;

- United European Gastroenterology.

2.2 Eligibility criteria

Eligibility criteria for the inclusion of studies assessing the prognostic test accuracy or clinical impact of the tools that are the focus of this assessment are presented in Table 3.

Considering study design, based on scoping searches, and given that the interventions are prognostic tools, retrieval of relevant randomised controlled trials (RCTs) was deemed to be unlikely. Thus, to ensure that all relevant studies were captured, no limit relating to study design was applied, with the exception that studies must be carried out in humans, and not be an opinion piece (i.e., editorial). Studies analysing the clinical validity (the ability of the test to reliably and accurately identify the biomarkers of interest or determine the risk of developing severe versus non-severe course of CD) or clinical utility (the ability of the test to improve measurable clinical outcomes, and its usefulness and added value to patient management) of the prognostic tool were eligible for inclusion. Studies evaluating analytical validity were included where applicable, where analytical validity denotes the ability of the tool to accurately and reliably measure the biomarker of interest as assessed using laboratory tests on samples that are representative of those with CD. Studies not published in English language were eligible if sufficient relevant data could be extracted from the full-text publication in non-English language, or from an English language abstract.

For the IBDX tool, to be included, a study had to assess all six biomarkers included in the panel:[42]

- ACCA;
- ALCA;
- AMCA;
- gASCA;
- anti-L;
- anti-C.

Table 3. Eligibility criteria for the systematic review of studies evaluating prognostic accuracy or clinical impact of the tools

Aspect of review	**Eligibility criteria**	
Population	Those with active CD, and a diagnosis of disease.	
Prognostic tests (interventions)	IBDX and PredictSURE-IBD	
	Prognostic test accuracy	**Clinical impact**
Comparator	No comparator or comparison of the prognostic tool and clinical judgement versus clinical judgement alone of high risk of following a severe course of CD.	
Reference standard	Not applicable	Standard care in the NHS
Outcomes	Prognostic test accuracy: • sensitivity and specificity; • numbers of true positive, true negative, false positive and false negative test results for predicting course of disease; • diagnostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease); • time to test result; • number of test failures; • number of inconclusive test results.	Outcomes are of interest in the subgroups of those assessed as having high risk versus not being at high risk of following a severe course of CD: • percentage of people for whom early treatment with biologics was offered ('top-down'); • rates and duration of response and remission; • rates and duration of flare-ups and/or relapses; • rates and duration of corticosteroid-free remission; • cumulative corticosteroid 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 (e.g., obstruction, intestinal ulcers, fistula, anal fissure); • composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure); • adverse effects of treatment; • health-related quality of life.
Abbreviations: CD, Crohn's disease; IBDX, Crohn's disease Prognosis Test; NHS, National Health Service.		

2.3 Study selection

Firstly, two reviewers independently assessed the titles and abstracts of studies retrieved from the electronic database searches for potential relevance according to the prespecified eligibility criteria (Table 3). In cases in which consensus could not be achieved, the full texts of potentially relevant studies were ordered. Next, full-text copies of potentially relevant studies were obtained and assessed independently by two reviewers for inclusion against the prespecified eligibility criteria. Any disagreements were resolved by discussion, or a third reviewer was consulted if necessary.

2.4 Data extraction

After creation of a standardised data extraction form (including a pilot process), data were extracted by one reviewer, and independently checked for accuracy by a second reviewer. Discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. Information extracted included details of the study's design and methodology, intervention and comparator tests, reference standard, relevant baseline characteristics of participants (e.g., duration of CD, location of CD, and presence of complications), and outcome measures, including clinical outcome efficacy and any adverse events (Table 3). The companies producing the prognostic tests and the corresponding authors of the studies selected for assessment of test accuracy were, when necessary, contacted for missing data or clarification of the data presented.

2.5 Quality assessment

In a change from the prespecified protocol, taking into account reviewer feedback and a review of the available checklists, the quality of prognostic test accuracy studies was assessed using the QUIPS[59, 60] (Quality of Prognosis Studies in Systematic Reviews) rather than the PROBAST (Prediction model Risk Of Bias ASsessment Tool) tool as originally planned.[61, 62] The quality of clinical effectiveness studies was to be assessed based on their study design: RCTs were to be assessed using the Cochrane Risk of Bias Tool;[63] non-randomised studies were to be assessed using the Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I) tool;[64] and qualitative studies were to be assessed using the Critical Appraisal Skills Programme (CASP) tool.[65] However, all studies identified as relevant to the systematic review were prognostic accuracy studies. All quality appraisal assessments were carried out by one reviewer and verified by another independently.

2.6 Methods of analysis and evidence synthesis

Details of results on accuracy of the prognostic tests and potential impact of their use on clinical outcomes, together with quality assessment for each included study are presented in structured tables and as a narrative summary. The heterogeneity identified across studies associated with clinical (e.g., baseline characteristics and reported outcomes) and methodological (e.g., different study designs and limited reporting of data) characteristics precluded quantitative synthesis of the data. For prognostic accuracy, positive predictive values (PPV), negative predictive values (NPV), sensitivity values and specificity values, with 95% confidence intervals (95% CIs) are presented for each study, where available.

2.6.1 Potential subgroup analyses

Evidence permitting, planned subgroups to be investigated were:

- children versus adults with a diagnosis of CD;
- newly diagnosed CD versus established diagnosis of CD;
- mild versus moderate-severe activity of disease;
- presence versus absence of fistulising or complex perianal disease.

2.6.2 Sensitivity analyses

Planned sensitivity analyses were to include studies deemed to be high risk of bias that were excluded from the primary analyses. Sensitivity analyses stratified by risk of bias were not conducted as lack of sufficient data precluded analysis.

3 RESULTS OF THE REVIEW OF PROGNOSTIC TEST ACCURACY AND CLINICAL IMPACT

The sections that follow discuss the quantity and quality of evidence available, including characteristics and risk of bias of identified studies, retrieved through literature searches to identify data on prognostic accuracy and clinical impact of PredictSURE-IBD and Crohn's disease Prognosis test (IBDX).

3.1 Quantity and quality of the available evidence

3.1.1 Results of the systematic literature search

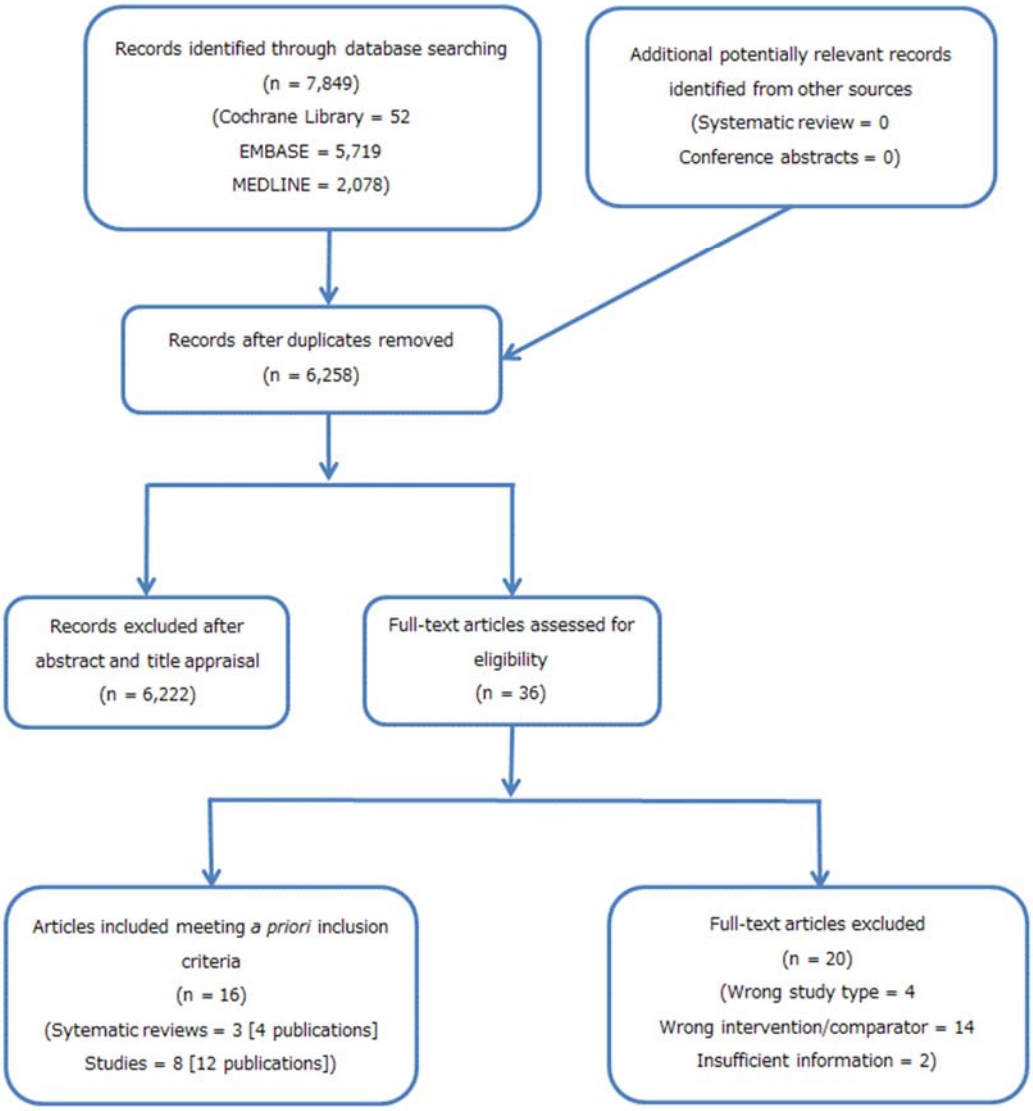
Searches of electronic databases retrieved 6,258 records (post deduplication) that were of possible relevance to the review (Figure 3). Initial screening of titles and abstracts led to the identification of 36 publications for review of full texts. Of the 36 articles evaluated, 16 publications, including systematic reviews, were deemed to be relevant to the review.[38, 50, 66-79] Four records (three full texts[38, 66, 70] and one conference abstract[68]) provided details for three systematic reviews, the reference lists of which were screened for potentially relevant studies. Additionally, documents supplied by the companies marketing the prognostic tools were reviewed.

Limited evidence is available from the included full-text publications on the prognostic accuracy of PredictSURE-IBD, and none is available on prognostic accuracy of IBDX, in identifying those at high risk of following a severe course of Crohn's disease (CD) as determined by measures such as sensitivity and specificity (prognostic outcomes of interest listed in Table 3). Most evidence on the utility of the tools is derived from observational studies that report estimates of risk of experiencing a clinical outcome associated with an aggressive course of CD, for example, need for treatment escalation, development of a complication or surgery. Estimates are presented as increased risk for those categorised, based on test results, as being at higher risk compared with those determined to be at lower risk of following a severe disease course. No study retrieved reported on the clinical impact of use of IBDX or PredictSURE-IBD in terms of influencing the treatments given in the management of active CD.

Authors of two studies[79, 80] were contacted to verify that the kit used in their research was the IBDX tool and not a comparable kit produced by another company. One author confirmed that they had used a kit that was not captured in the scope of this review, and the study was therefore excluded from the review.[80]

Summaries of the studies included in the review are presented by prognostic tool evaluated and key characteristics of studies (Table 4). A list of full-text publications screened but subsequently excluded (with reasons for exclusion) from the review is available in Appendix 4.

Figure 3. PRISMA flow chart



3.1.1.1 Ongoing studies

Searches of prespecified sources, together with information supplied by the companies, identified four ongoing studies of potential relevance to the review, all of which assess use of PredictSURE-IBD.

The PROFILE study is a prospective, multicentre randomised study set in the UK.[51] PROFILE has been designed to compare the clinical efficacy of "top-down" (TD) and "accelerated step-up" (SU) treatment regimens in people with newly diagnosed CD who have first been stratified into subgroups based on risk of following a severe, relapsing course of CD (high versus low) using the PredictSURE-IBD tool. Within the biomarker-stratified groups, people are randomised (1:1) to either TD or accelerated SU treatment. Treatment allocation is open label, but clinicians and patients are masked to subgroup classification. The authors propose that those designated as being at high-risk of severe course of CD will experience a greater benefit of receiving early TD treatment. Conversely, those likely to experience a more indolent course of disease could be managed with the accelerated SU approach and avoid the risk of adverse effects associated with biological therapies. Thus, a goal of the study is to determine whether use of the PredictSURE-IBD tool can facilitate personalised therapy in CD and improve clinical outcomes. The primary outcome is the incidence of sustained surgery and glucocorticosteroid-free remission from the completion of induction treatment through to study completion (48 weeks). Recruitment began in December 2017, with a planned enrolment of 400 people, generating 100 people in each of the four groups.[51] The estimated end date for the trial listed on the ISRCTN (International Standard Randomised Controlled Trials Number) registry is March 2022.[81]

PRECIOUS is a multicentre observational study based in the USA and sponsored by PredictImmune.[82] Set in referral centres and community hospitals, PRECIOUS is designed to assess the efficacy of the PredictSURE-IBD tool in stratifying those newly diagnosed with active inflammatory bowel disease (IBD), including CD, into cohorts at high or low risk of following an aggressive disease course requiring frequent treatment escalations. Patients' blood will be collected at enrolment and be tested with PredictSURE-IBD at a later date. Ideally, people will be treatment naïve. Those enrolled will receive treatment as per local standard of care with a SU or accelerated SU regimen, and will be prospectively followed for 12 months. People enrolled and clinicians will be masked to tests results. With a planned recruitment of 200 people, the estimated end date for the study listed on clinicaltrials.gov is June 2021.[82]

Two additional studies evaluating PredictSURE-IBD were highlighted by PredictImmune in their response to a request for information as part of the Diagnostics Assessment Programme (DAP) process:

- Prospective, masked study stratifying a paediatric cohort with incident IBD (N=80);
- Head-to-head comparison of PredictSURE-IBD with IBDX for stratification of those at higher risk of following a severe course of CD using samples from cohorts previously assessed as part of a study evaluating PredictSURE-IBD.

PredictImmune informed that data from the study evaluating PredictSURE-IBD in a paediatric cohort are, at the time of writing, undergoing analysis and results are likely to be available towards the end of 2019. The EAG notes that only results in children and adolescents with CD will be of relevance to the DAR reported here.

For the head-to-head comparison of PredictSURE-IBD and IBDX, the cohort analysed comprised those with active CD as confirmed by one objective marker (i.e., raised C-reactive protein [CRP], raised calprotectin or endoscopic signs of active disease) in addition to active symptoms. People had been recruited from a single site in the UK for an observational study evaluating PredictSURE-IBD. All enrolled were treated with the accelerated SU regimen in accordance with UK guidelines. Samples for analysis by the two biomarker tests were taken concurrently from the same bleed: PredictSURE-IBD requires whole blood RNA and IBDX uses serum. A conference abstract outlining results of the comparison has been submitted for consideration and, if accepted, will be presented at the Congress of the European Crohn's and Colitis Organisation (ECCO) taking place in February 2020.

3.2 Evidence provided by the companies

3.2.1.1 Glycominds International

Glycominds International provided a list of bibliographic details for the key publications outlining the evidence in support of the IBDX tool. All studies reporting results on the effectiveness of the kit in stratifying those at high risk of following a severe course of CD were retrieved, and subsequently reviewed, by the External Assessment Group (EAG).

3.2.1.2 PredictImmune

PredictImmune provided a list of bibliographic details for several publications relating to PredictSURE-IBD, including references describing the research underpinning the development of the signature gene sequence. All studies flagged by the company were retrieved, and subsequently reviewed, by the EAG.

Additionally, in response to queries from the EAG, PredictImmune supplied anonymised individual patient data (IPD) for results from the cohort that provided results for validation of PredictSURE-IBD, together with data for the head-to-head comparison of PredictSURE-IBD with IBDX. Results provided by PredictImmune for this direct comparison are presented and critiqued in Section 3.3.4.3.

3.3 Assessment of prognostic test accuracy

3.3.1 Characteristics of included studies

All studies informing the evidence base on the prognostic accuracy of the IBDX and PredictSURE-IBD biomarker-stratification tests were observational in design. Key characteristics of included studies are summarised in Table 4, with validated data extraction forms for studies provided in Appendix 5. Twelve publications describing eight studies retrieved from electronic searches were included in the assessment of the prognostic accuracy of the tests, with seven of the studies (11 publications) reporting results on utility of the IBDX kit and one on PredictSURE-IBD in stratifying those at high-risk of a severe course of CD (Table 4). Several studies included a mixed population of CD and ulcerative colitis, and reported results separately for those with CD. Most studies included predominantly adults with CD, with one study (three publications) reporting data for an adolescent or paediatric population. No additional potentially relevant study was identified from hand searching of bibliographies of three systematic reviews.[38, 66, 68, 70]

All included studies assessed outcomes in people reported to have a diagnosis of CD. However, limited reporting was noted across studies relating to the IBDX on stage of diagnosis (newly versus established) at time of test. Baseline characteristics suggest that samples analysed were provided predominantly by people with established CD (Appendix 5). By contrast, most people enrolled in the study on PredictSURE-IBD had received a recent diagnosis of CD.

Prespecified inclusion criteria for the systematic review presented here required that people have active disease (Table 3). Although most of the included studies outlined criteria to be met for a diagnosis of CD, only the study evaluating the PredictSURE-IBD tool required people to have active disease to be eligible for enrolment, and reported how presence of active disease was determined.[50] In retrospect, given the biomarker targets of the two prognostic tests, the reviewers consider that a criteria of active CD is appropriate for inclusion of studies assessing PredictSURE-IBD, but is not essential for studies reporting on IBDX. As outlined in Section 1.2, the PredictSURE-IBD tool detects a gene sequence associated with CD8+ T cell exhaustion that arises from an autoimmune response to active disease, and, therefore, it is appropriate that people have active CD when blood is taken for analysis: it has been reported that, in people with inactive disease after treatment, as determined by endoscopy, level of CD8+ T cells increases to that comparable with those observed in healthy controls.[83] By contrast, the IBDX kit detects serum levels of specific anti-glycan-antibodies, with specified cut-offs for allocation of positive or negative status for each biomarker. Although serum levels of each antibody can change over time, it is purported that status for positivity or negativity for that antibody remains stable throughout the course of disease.[74] Therefore, for IBDX, the reviewers decided to include those studies

not specifying a measure of active disease, if they met all other inclusion criteria and reported assessment of the six biomarkers included in the IBDX panel.

Analyses presented for evaluation of the six biomarkers forming the IBDX kit typically reported the association of positivity for individual biomarkers, or positive status for a higher number of biomarkers, with the increased risk of following a severe course of CD, and not the evaluation of all six biomarkers as a collective.

Considering PredictSURE-IBD, the included study described use of the tool in three cohorts, two training cohorts and one validation cohort.[50] Samples from one training cohort (N=66) were used in biomarker discovery and samples from the second (N=39) were used in whole blood classifier development. Estimates of prognostic accuracy are available for only the validation cohort. Based on IPD data supplied by the company, the reviewers consider the validation cohort together with the second training cohort (N=39) to be the most appropriate data set to inform the evidence base on for economic analysis: discussed in greater detail in 4.2.3.2.

Caveats to interpretation of the results for prognostic accuracy of both tests are discussed in greater detail in the corresponding results sections.

Table 4. Characteristics of studies included in the prognostic test accuracy review

Study	Design, and country	Population	Number eligible for analysis	Duration of disease at time of test	Severity of disease at time of test	Outcomes reported
IBDX						
Harrell 2010[67] (conference abstract)	Unclear, Unclear	People with CD	172	Not reported	Not reported	Association of individual anti-glycan biomarkers with: • Disabling disease course; • Complicated disease behaviour and/or need for surgery.
Paul 2015[69] (full publication)	Cross-sectional, France	People with IBD and a diagnosis for more than1 year	107 with CD	Median 9.4 (IQR 1 to 44) years	Not reported	Differentiating severe from non-severe course of disease
Rieder 2010b[75] (full publication) Related publications [73, 77]	Prospective cohort, Germany	People with IBD, other GI disease and healthy controls	363 with CD	Median 66.8 (IQR 11 to 141) months	Not reported	OR for : • Complication; • CD-related surgery; • Early disease onset. Where analyses based on median number of positive markers: OR reported for median positive markers present 2.0 (1.0 to 3.0).

Rieder 2010c[76] (full publication)	Prospective cohort, Germany	People with CD and no prior complication or surgery	76	Median 10.6 (IQR 1.7 to 52.3) months	Not reported	Time to complication or surgery analysed by number of positive biomarkers (1, 2 or 3)
Rieder 2012[72] (full publication) Related publications [71, 73]	Cross-sectional, Germany	Children (<18 years) with IBD and healthy controls	59 with CD	Median 18.0 (IQR 12.0 to 43.0) months	Not reported	Need for CD-related surgery by number of positive biomarkers (1, 2 or 3)
Seow 2009[78] (full publication)	Cross-sectional, Canada	People with IBD and healthy controls	517 with CD	Median 8.9 (IQR 0.02 to 46.30) years	Not reported	Association of number of positive biomarkers with key prognostic factors for severe course of disease and need for abdominal surgery
Wolfel 2017[79] (conference abstract)	Prospective cohort, Unclear	People with CD who had undergone one surgical resection	118	Not reported	Not reported	Time to repeat surgery
PredictSURE-IBD						
Biasci 2019[50 a] (full publication)	Prospective cohort, UK	People with active CD or UC and not receiving concomitant corticosteroids, immunomodulators or biological therapy	66 with CD (validation cohort)	61 (92.4%) people were newly diagnosed with CD	Not reported	• Sensitivity and specificity for predicting the need for multiple escalations within the first 18 months; • Negative predictive value;

						• Number of treatment escalations required; • Time to treatment escalation.

[a] Additional data were provided by PredictImmune during the Diagnostic Assessment Programme process.

Abbreviations: CD, Crohn's disease; GI, gastrointestinal; IBD, inflammatory bowel disease; IBDX, Crohn's disease Prognosis Test; IQR, interquartile range; OR, odds ratio; UC, ulcerative colitis.

3.3.2 Quality assessment of included studies

Included studies were assessed for risk of bias and applicability using the QUIPS (QUality In Prognosis Studies) tool.[59, 60] A summary of the results of the assessment of risk of bias and generalisability concerns across studies is presented in Table 5. The full critique for each study is presented in Appendix 6.

QUIPS encompasses six domains for assessment of validity and bias of studies evaluating prognosis and factors influencing the course of a condition:[59, 60]

- participation;
- attrition;
- prognostic factor measurement;
- confounding measurement and account;
- outcome measurement;
- analysis and reporting.

Each domain comprises prompting items (between three and seven) for consideration in the overall rating for an item of high, moderate or low risk of bias.[59, 60]

The IBDX and PredictSURE-IBD tools have been designed with the goal of predicting a course of disease based on levels of biomarkers produced in response to presence of CD, with stratification to high or low risk of severe course of disease determined by results of laboratory analysis. The extent to which biomarker levels in blood and serum samples change over time in individual people and what factors influence fluctuations in levels is uncertain. Additionally, as production of the biomarkers assayed is triggered by changes in cellular processes, the effect of physical characteristics that could influence prognosis in CD, for example, smoking status and age, on biomarker levels is unclear. Thus, for the studies informing the evidence on prognostic test accuracy reported here, the External Assessment Group (EAG) considers that the importance of the "confounding measurement and account" domain as a determinant of the risk of bias associated with the studies is also unclear. To reflect the ambiguity around the importance of confounding factors, and to capture uncertainty where limited reporting in the publication precluded an assessment of risk for a particular domain, the EAG adapted the QUIPS tool to include an overall assessment of unclear risk.

About half of the included studies were deemed to have at least one domain at unclear risk of bias (Table 5): for conference abstracts, an unclear rating was predominantly associated with limited reporting of details due to space constraints.

Most studies reporting results for the IBDX tool were determined to be at moderate risk of bias for the population domain as the studies included those with a recent diagnosis and those with an established diagnosis of CD, and, in some studies, those with presence of complicated disease at baseline. Data were not analysed separately for the individual subgroups. Of note, the population of greatest relevance to the economic evaluation is those with a new diagnosis of CD and who have moderate/severe disease activity. The study assessing the prognostic accuracy of PredictSURE-IBD enrolled those with a recent diagnosis of CD but included any level of disease activity at sample assessment, with severity of disease activity determined by endoscopy for some people: severity of disease activity at baseline was not available for all those forming the validation cohort.

Most studies were considered to be at a low risk of bias for attrition and for measurement of prognostic factors because all samples taken were analysed with the relevant tool and results generated as per the company's individual protocols. Additionally, outcome assessment was deemed to be a low risk of bias across many studies as the clinicians were masked to the results of the biomarker assessment.

Table 5. QUIPS assessment of prognostic studies

Study	Participation	Attrition	Measurement of prognostic factor	Outcome assessment	Measurement of confounding factors	Analysis and reporting
IBDX						
Harrell 2010[67] (conference abstract)	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Paul 2015[69] (full publication)	Low	Low	Low	Low	Unclear	Low
Rieder 2010b[75] (full publication)	Moderate	Low	Low	Low	Moderate	Low
Rieder 2010c[76] (full publication)	Low	Low	Low	Low	Moderate	Low
Rieder 2012[72] (full publication)	Moderate	Low	Low	Low	Moderate	Low
Seow 2009[78] (full publication)	Moderate	Low	Low	Low	Moderate	Low
Wolfel 2017[79] (conference abstract)	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear

PredictSURE-IBD						
Biasci 2019[50] [a] (full publication)	Low	Unclear	Low	Low	Unclear	Low

[a] Additional data were provided by PredictImmune during the Diagnostic Assessment Programme process.

Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel disease; IBDX, Crohn's disease Prognosis Test.

3.3.3 Accuracy of prognostic tests

The EAG notes that limited data were available from the included studies on prognostic accuracy of the tools in stratifying risk of severe course of CD in terms of standard measures of test accuracy, for example, sensitivity and specificity. The EAG is unaware of a validated definition for determination of whether a person has followed a severe course of CD, for example, a set number of treatment escalations or development of a complication or need for surgery. Thus, the EAG considers the criterion required for a true positive or false positive for IBDX and PredictSURE-IBD to be unclear. The EAG considers it would be challenging to ascertain an accurate estimate of prognostic accuracy of IBDX and PredictSURE-IBD in stratifying course of CD. Establishing prognostic accuracy of the tools would require carrying out a prospective study that included a group that received only SU treatment after determination of their risk of course of CD with clear prespecified criteria for following a severe course of CD. The ongoing PROFILE RCT randomises people to accelerated SU or TD treatment after determination of high or low risk of following a severe course of CD and so data from the two SU groups will provide additional data to inform estimates of prognostic accuracy. Additionally, no study included in the review prospectively followed people whose treatment was determined by results from IBDX and PredictSURE-IBD: the ongoing PROFILE RCT assesses whether early treatment with TD strategy affords clinical benefit to those categorised as being high risk of severe course of CD and should provide data on clinical impact of use of PredictSURE-IBD.

3.3.3.1 IBDX

No identified study reported on the accuracy of the IBDX kit as a whole (six biomarkers) as per the prespecified prognostic outcomes of interest to this review for stratification of risk of following a severe course of CD (Table 3). One study reported that positivity for the ASCA and AMCA antibodies had the best validity for differentiation of severe from non-severe course of CD, with an AUC of 0.63 and 0.65, respectively. Combination of ASCA and AMCA generated increased precision for differentiation of severe from non-severe course of CD, with an AUC of 0.71.[69]

In their submission to the DAP, Glycominds International reported a sensitivity for IBDX of 78%, and a specificity of 85% to 98% depending on the number of positive biomarkers. Data or details of references to support the reported sensitivity and specificity were not provided in the documentation. None of the studies included by the EAG provided estimates of sensitivity or specificity for the IBDX panel. Additionally, it is unclear whether the reported estimates relate to sensitivity and specificity in the diagnosis of CD, including differentiation of CD from ulcerative colitis, or in the stratification of risk of severe course of CD.

Typical test time for IBDX is reported by Glycominds International to be about 90 minutes, and all samples can be run in parallel.

The instructions on the use of the IBDX kit advise that, in cases of an equivocal test result, the individual biomarker be tested again. Details on the frequency of occurrence of an equivocal result are not available from identified studies.

A longitudinal analysis assessed whether levels of the individual biomarkers fluctuate over time.[74] Between two and seven serum samples were available from each person forming the cohort for analysis. Over a median follow-up of 17.4 months (IQR 8.0 to 31.6 months), the authors noted that, despite marked changes in overall immune response and in levels in individual biomarkers, the status of positivity or negativity for an individual biomarker remained mostly stable over time.

3.3.3.2 PredictSURE-IBD

One publication (Biasci 2019[50]) assessing the PredictSURE-IBD tool was deemed to meet the inclusion criteria for the review.[50] Several related papers were identified and determined not to be relevant because they describe the research underpinning the identification of the signature genetic profile (15 target genes and 2 control genes) that stratifies those with active CD to high or low risk of severe course of disease and not the use of PredictSURE-IBD (full details available in Appendix 4).

The included study enrolled people aged 18 years and over with active CD or ulcerative colitis who were not receiving concomitant glucocorticosteroids, IMs or biological therapy. People were recruited from a specialist IBD clinic before treatment started. Diagnosis of CD or ulcerative colitis was based on standard endoscopic, histological and radiological criteria. Active disease was confirmed by one or more objective marker (raised CRP, raised calprotectin or endoscopic evidence of active disease) in addition to active symptoms and/or signs. People were treated with a conventional SU strategy in accordance with national and international guidelines.

Within the publication, results on stratification to high or low risk of severe course of CD are presented for a training cohort (N=118; CD=66, UC=52) and a validation cohort (N=123; CD=66, UC=57).[50] Additionally, the full text publication refers to a second training cohort (N=39) from which samples were used in development of a whole blood classifier. Results from the training cohort (N=66) used in biomarker discovery were used to finalise the signature gene sequence, which was subsequently applied to analysis of the validation cohort. Two different source cells were used in the process, with mRNA extracted from unseparated peripheral blood mononuclear cells (PBMCs) for the training cohort informing biomarker discovery and from a venous blood sample for the validation cohort, as would be

the case in clinical practice. Both unseparated PBMCs and blood samples were processed for the second training cohort (N=39) but it is unclear from the full publication whether the whole blood samples were analysed using the signature gene sequence identified during biomarker discovery. As part of the DAP, the company clarified that blood samples from the second training cohort were analysed using the finalised gene sequence. Thus, the EAG considers results from the validation cohort and the smaller training cohort to be the most appropriate data set to inform the evidence based on the accuracy of PredictSURE-IBD. However, data on specificity and sensitivity are available for only the validation cohort.

Of the 66 people in the validation cohort, 27 (40.9%) were assigned as high risk of following a severe course of CD (IBDHi) versus 39 (59.1%) categorised as being at low risk (IBDLo). Of the 39 people in the training cohort, 19 (48.7%) and 20 (51.3%) were categorised as IBDHi and IBDLo, respectively. Baseline characteristics for the validation cohort indicate that most people had newly diagnosed CD (61/66 [92.4%]). The EAG notes that level of disease activity at enrolment (mild, moderate, or severe) was not reported and details on proportion of people with complications of CD (e.g., fistulae and perianal disease) at baseline are not available in the full publication but were provided by PredictImmune in their response to a request for information as part of the DAR process (Appendix 5):[50] presence of complications of CD at baseline could indicate an earlier requirement for surgery in the SU algorithm.

Data on number of test failures and number of inconclusive test results were not available.

3.3.3.2.1 Sensitivity and specificity

The study by Biasci and colleagues[50] reports a sensitivity and specificity for predicting the need for multiple escalations within the first 18 months of 72.7% and 73.2%, respectively. The full text publication does not provide a cut off as to how the sensitivity and specificity for multiple escalations were derived. As noted earlier, the EAG is unaware of a validated definition for determination of whether a person has followed a severe course of CD, and, as a consequence, considers the criterion required for a true positive or false positive to be unclear for the prognostic tests assessed in the review.

As part of the DAP process, PredictImmune helpfully provided anonymised IPD for the validation cohort, including the 2 x 2 table for calculation of sensitivity and specificity for multiple escalations at 12 and 18 months (Table 6). PredictImmune applied a cut-off of two or more treatment escalations to categorise people as having followed a more aggressive course of CD. The EAG considers the company's approach reasonable. However, the EAG notes that people in the validation cohort and second training cohort underwent treatments at the discretion of the treating clinician and so a

proportion (29/105; 27.6%) received a therapy other than glucocorticosteroid at entry, including elemental diet, anti-TNF alone or in combination with IM, and IM alone: [redacted]. The EAG recognises that the study is of a more pragmatic design but considers that induction treatment would likely influence the timing and frequency of treatment escalation, and consequently sensitivity and specificity. Moreover, some people included in the calculation of sensitivity and specificity for predicting multiple escalations received surgery as a first treatment escalation (7/66; 10.6%) and continued to be monitored for subsequent treatments, including IMs and biological therapies. Given that RCTs assessing clinical effectiveness of treatment strategies in the management of CD typically report occurrence of CD-related complications (e.g., need for surgery or hospitalisation or development of fistula or stenosis) as a composite clinical outcome or separately, the EAG considers it important to assess time to and occurrence of surgery independently of other treatment escalations to reflect outcomes in other studies, including those assessing the effectiveness of IBDX: the EAG's clinical experts supported the proposal that assessment of CD-related surgery as a separate outcome would be appropriate. Inclusion of people who underwent surgery as a first treatment escalation and received subsequent treatment escalations could influence the accuracy of sensitivity and specificity as assessed by number of treatment escalations. The EAG notes that the sample size for the validation cohort is small (N=66) and, moreover, that not all people in the validation cohort have been included in analyses at 12 or 18 months. Additionally, a proportion of people in the validation cohort received an anti-TNF biologic, with or without an IM (11/66; 16.7%) as their first escalation.[50] The EAG appreciates that the study is pragmatic and likely reflects treatment approaches in clinical practice in the UK but the EAG also considers that analysing those who receive TD or surgery as their first treatment escalation together with those who followed the SU treatment algorithm or were treated at discretion of the treating clinician is unlikely to reflect the true estimate of number of treatment escalations that would occur in the SU or accelerated SU strategy.

Table 6. Data informing the calculation of sensitivity and specificity for PredictSURE-IBD based on predicting need for multiple treatment escalations

PredictSURE-IBD categorisation	<2 treatment escalations	≥2 treatment escalations	Sensitivity	Specificity
Within 12 months				
IBDHi	15	7	77.8%	70.6%
IBDLo	36	2		

Within 18 months				
IBDHi	11	8	72.7%	73.2%
IBDLo	30	3		
Abbreviations: IBD, inflammatory bowel disease.				

3.3.3.2.2 Predictive value

The included study reports a negative predictive value of 90.9% for PredictSURE-IBD of predicting multiple escalations within the first 18 months.[50] Based on the 2x2 table supplied by the company (Table 6), the EAG calculates a positive predictive value of 42.1% for predicting multiple escalations within the first 18 months.

3.3.4 Results for clinical outcomes

The EAG notes that results presented in this section are the risk of experiencing an event for those categorised by the tools as high risk versus low risk of following a severe course of CD and are not related to clinical outcome of treatment decisions based on stratification of risk by IBDX and PredictSURE-IBD.

3.3.4.1 IBDX

Results are reported based on positive status for increasing number of biomarkers, as per the company's recommendations on the interpretation of outputs from the test (please see Figure 2). As noted, all included studies evaluated the full panel of biomarkers comprising the IBDX kit, but there is no single measure of accuracy or clinical outcome for the six biomarkers as a collective.

Clinical and methodological heterogeneity across the identified studies precluded meta-analysis and results are presented in a narrative review.

3.3.4.1.1 Developing a complication

Two studies reported an effect estimate for the risk of experiencing a complication by number of biomarkers testing positive (Table 7).[75, 76] Both studies prospectively followed a cohort of people with CD.

Complicated disease behaviour was defined in both studies as the occurrence of fistulae or stenosis.[75, 76] In one study, 68% of people (249/363) had a complication before or at the time of sample procurement.[75] The second study enrolled people with or without prior complication and prior or no prior CD-related surgery but focussed reporting on those with no prior complication and no CD-related

surgery before or within 20 days of obtaining the sample (N=76).[76] Median follow-up was 59 months for one cohort[75] and 53.7 months for the second.[76]

Median duration of CD was disparate between the two studies, with one study reporting a median of 66.8 months (IQR 11 to 141 months)[75] compared with a much shorter duration of 10.6 months (IQR 1.7 to 52.3 months)[76] in the other: the EAG's clinical experts advised that 10.6 months may be insufficient follow-up to monitor development of a CD-related complication.

In the study including people with complications at baseline, an odds ratio (OR) of 1.5 (95% CI: 1.3 to 1.9; p<0.001; Table 7) was reported for developing a complication compared with not experiencing a complication, with increased risk associated with a positive status for a higher median number of biomarkers.[75] During follow-up an additional 28 people developed a fistula or stenosis, or both.

In people with no prior complication, 20 people experienced a fistula or stenosis, with a higher risk of experiencing a complication noted in those with positive status for at least two or three biomarkers (Table 7), with the risk reaching statistical significance for those testing positive for at least two of the six antibodies (hazard ratio [HR] 2.5; 95% CI: 1.03 to 6.1; p=0.043).[76] The EAG notes the small sample size informing the estimate of risk.

Increasing number of positive antibodies was reported to be significantly associated with complicated disease behaviour and/or surgery (OR 3.3, 95% CI not reported; p=0.0005) for a cohort of people with CD from the USA:[67] results presented in a conference abstract and limited details available. Complicated disease behaviour was defined as intestinal fistula and/or stricture.

One study of a cross-sectional design analysed serum samples from children and adolescents aged 18 years or younger.[71-73] The authors reported results for the younger cohort that were aligned with those derived from an adult cohort, with a higher number of positive serum biomarkers associated with an increased risk of experiencing complicated CD and requiring CD-related surgery (estimates of effect not reported).[72] Additionally, the authors assessed differences in cut-off levels for indicating positivity of biomarkers between the paediatric cohort and adults evaluated in a related study[75] and found lower cut off points to denote positivity for paediatric samples. In a related conference abstract, the authors reported that, in paediatric CD patients, positivity for at least one marker out of the whole panel versus no marker positive was independently associated with fibrostenotic or fistulizing disease behaviour (p=0.036) and ileal disease location (p=0.014).[71] Although the accuracy of the biomarker panel to diagnose and differentiate CD from other gastrointestinal conditions was reported to decrease with increased age at sample procurement, when assessing CD behaviour, the ability of the panel to stratify disease phenotypes remained constant over time.[72]

Table 7. Summary of risk of developing a complication based on number of positive biomarkers

Outcome	N	Population	Result	P value
Complication[75] [a]	Unclear	CD	OR 1.5 (95% CI: 1.3 to 1.9)	<0.001
Complication[76] [b] (subgroup of people experiencing a complication)	20	CD but no prior complication or surgery		
• At least 1 positive marker			HR 1.8 (95% CI: 0.61 to 5.4)	0.29
• At least 2 positive markers			HR 2.5 (95% CI: 1.03 to 6.1)	0.043
• At least 3 positive markers			HR 2.6 (95%CI: 0.92 to 7.2)	0.072

[a] Analyses based on median number of positive markers: OR reported for median positive markers present 2.0 (1.0 to 3.0). OR reported based on median number of positive biomarkers present [2.0 (1.0 to 3.0)] in those developing a complication versus median positive markers present in those not developing a complication [1.0 (0.0 to 2.0)].

[b] Analyses adjusted for age, sex, BMI, disease activity and duration, age at diagnosis and disease location.

Abbreviations: BMI, body mass index; CI, confidence interval; CD, Crohn's disease; HR, hazards ratio; OR, odds ratio.

3.3.4.1.2 Requirement for surgery

Two of the three studies reporting on risk of complication also provided information on the increased likelihood of requiring surgery.[75, 76] A third study of cross-sectional design evaluated serum samples of 517 people with CD and a median duration of disease of 8.9 years (range from 0.02 to 46.30 years).[78]

One study reported an OR of 1.5 (95% CI 1.3 to 1.8; p<0.001; Table 8) for requiring surgery compared with no requirement for surgery, with increased risk associated with a positive status for a higher median number of biomarkers.[75] At the time of sample procurement, 224 people had undergone surgery related to IBD, with an additional 33 people requiring surgery during follow-up.

For the cohort of people who had not undergone surgery at enrolment, 14 people required surgery, with a statistically significantly higher risk for surgery (HR 3.6; 95% CI: 1.2 to 11.0; p=0.023; Table 8).[76] The EAG notes the small sample size informing the analysis, and the large confidence interval accompanying the estimate of risk.

The third study identified a trend towards a larger proportion of people requiring surgery with increasing number of biomarkers testing positive (Table 8).[78] A statistically significant difference across the categories assessed was identified (p<0.0001).

A conference abstract provided results for a cohort of people (N=118) who had undergone one surgical intestinal resection related to CD.[79] Most people evaluated (92%) underwent first surgery due to internal penetrating and/or stricturing disease. Serum samples for analysis with the IBDX kit were taken after surgery. After a median follow-up of 100 months, the authors reported that, when considering the full panel of six biomarkers, neither the quartile sum score nor the number of positive biomarkers combined predicted a shorter time to repeat intestinal surgery. After adjustment for ileal disease location and use of IMs or anti-TNF biologic after first surgery, analysis of individual biomarkers identified that positivity for AMCA (HR 2.6; 95% CI: 1.1 to 5.9; p=0.026) and ALCA (HR 2.3; 95% CI: 1.04 to 5.3; p=0.039) predicted shorter time to second surgery.[79] Another study reported that, of the panel of tested antibodies, only AMCA antibodies tended to be associated with higher risk of CD-related surgery with an OR of 2.1 (95% CI: 0.8 to 5.1; p=0.10) but the association did not reach statistical significance.[69]

Table 8. Summary of need for surgery based on number of positive biomarkers

Outcome	N	Population	Result	P value
Surgery[75] [a]	Unclear	People with CD	OR 1.5 (95% CI: 1.3 to 1.8)	<0.001
Surgery[76] [b] (subgroup of people undergoing surgery)	14	CD but no prior complication or surgery		
• At least 1 positive marker			HR 2.6 (95% CI: 0.58 to 12.0)	0.21
• At least 2 positive markers			HR 3.6 (95% CI: 1.2 to 11.0)	0.023
• At least 3 positive markers			HR 2.8 (95% CI: 0.80 to 9.6)	0.11
Surgery[78] [c] (abdominal)	517	CD		
• 1 positive marker	103		51.64%	<0.0001
• 2 positive markers	130		54.62%	
• 3 positive markers	77		63.64%	
• 4 positive markers	36		57.89%	

• 5 or 6 positive markers	36	76.67%

[a] Analyses based on median number of positive markers: OR reported based on median number of positive biomarkers present [2.0 (1.0 to 3.0)] in those requiring surgery versus median positive markers present in those not requiring surgery [1.0 (0.0 to 2.0)].

[b] Analyses adjusted for age, sex, BMI, disease activity and duration, age at diagnosis and disease location.

[c] Results presented are proportion of people needing surgery by number of positive biomarkers.

Abbreviations: BMI, body mass index; CI, confidence interval; CD, Crohn's disease; HR, hazards ratio; OR, odds ratio.

3.3.4.2 PredictSURE-IBD

3.3.4.2.1 Time to treatment escalation

The full text publication reported that those 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).[50]

The EAG notes that, based on the IPD supplied by PredictImmune, people in the validation cohort underwent treatments at the discretion of the treating clinician and so a proportion (14/66; 21.2%) received a therapy other than glucocorticosteroid at entry.[50] Choice of and time to first treatment escalation is likely to be influenced by response to treatment at study entry, which in turn is likely to be affected by risk of following severe course of CD. The EAG recognises that the study is of a more pragmatic design but considers that, as people within the validation cohort have not followed a standardised algorithm of treatment, that analysis of time to first treatment escalation is subject to a level of bias, the direction of which is unclear.

The EAG analysed IPD provided by PredictImmune for incorporation into the economic model, with a focus on those with a new diagnosis of CD as per the protocol.

3.3.4.3 Comparison of IBDX and PredictSURE-IBD

PredictImmune provided a conference abstract that reports the results of a head-to-head comparison of the IBDX and PredictSURE-IBD tools.[84] The abstract has been submitted for consideration and, if accepted, will be presented at the ECCO Congress taking place in February 2020.

[redacted]

[redacted] [84].

[redacted] [84].

[redacted].

3.3.5 Summary of findings for prognostic test accuracy

3.3.5.1 Sensitivity, specificity and negative predictive value

The evidence base on prognostic accuracy of the IBDX and PredictSURE-IBD tools in identifying those at high risk of following a severe course of CD is limited. No study was identified providing an assessment of the prognostic accuracy of the full panel of six biomarkers for the IBDX, and only one observational study provided results on use of PredictSURE-IBD in stratifying those with a recent diagnosis of CD and disease of any level of activity at the time of sample procurement, with severity of disease activity determined by endoscopy for some people: severity of disease activity at baseline was not available for all those forming the validation cohort.

Use of PredictSURE-IBD was associated with a sensitivity and specificity of 77.8% and 70.6%, respectively, in stratifying by need for multiple treatment escalations within 12 months. Corresponding sensitivity and specificity for multiple escalations within 18 months were 72.7% and 73.2%, respectively. A negative predictive value of 90.9% for PredictSURE-IBD of predicting multiple escalations within the first 18 months was also reported. The EAG notes that the cut-off for multiple escalations applied in the determination of sensitivity and specificity was two treatment escalations, and comprised any type of treatment, including surgery. The EAG is unaware of a validated definition for determination of whether a person has followed a severe course of CD and considers the choice of two escalations to be an arbitrary value. Additionally, the EAG's clinical experts fed back that it would be appropriate to consider escalation to CD-related surgery separately from progression to drug treatment, and also to use development of a complication of CD (fistula or stenosis) as another marker of sensitivity and specificity. The full text publication presenting results for PredictSURE-IBD indicates that those in the validation cohort were treated at the discretion of the treating clinician. IPD data provided by the company indicate that, of those in the validation cohort, 21.2% (14/66) received a therapy other than glucocorticosteroid at entry. Choice of and time to first treatment escalation is likely to be influenced by response to treatment at study entry, which in turn is likely to be affected by risk of following severe course of CD. The EAG recognises that the study is of a more pragmatic design but considers that, as people within the validation cohort have not followed a standardised algorithm of treatment, that induction treatment would likely influence the timing and frequency of subsequent escalations, and consequently sensitivity and specificity. The risk of bias of the study as assessed by the QUIPS tool was determined to be low across most domains. Considering the caveats highlighted by the EAG, together with the small sample size (N=66) informing calculation of prognostic accuracy for PredictSURE-IBD, the EAG considers that the results are potentially unreliable and should be interpreted with caution.

3.3.5.2 Clinical outcomes

Clinical outcomes that could be considered proxies for predicting prognosis are those that are typically associated with following a severe course of CD, including higher risk of developing a complication of CD (fistula or stenosis), of needing CD-related surgery, and a shorter time to and increased frequency of treatment escalations.

Seven studies evaluating the IBDX kit were deemed to be of relevance to the review, all of which were observational in nature: three studies were prospective cohorts and three were of a cross-sectional design. Of those studies reporting estimates of effect, people enrolled in the studies predominantly had an established, rather than a recent, diagnosis of CD. Clinical heterogeneity across studies in terms of various characteristics (prior complication versus no complication, previous IBD-related surgery or no surgery, and unclear whether people had active disease at baseline) was noted, which led to a determination of moderate risk of bias for the population domain based on the QUIPS tool. Two prospective cohort studies reported increased risk of experiencing a complication or of requiring surgery for those testing positive for at least two of the six biomarkers included in the IBDX kit. In addition, some estimates were informed by small sample sizes. Risks of experiencing a complication by positive biomarker status were reported to be:

- OR of 1.5 (95% CI: 1.3 to 1.9; p<0.001; N unclear) based on positivity for a median of two biomarkers;
- HR of 2.5 (95% CI: 1.03 to 6.1; p=0.043; N=20 with no prior complication or surgery) based on positivity for at least two biomarkers;
- HR of 2.6 (95% CI: 0.92 to 7.2; p=0.072; N=20 with no prior complication or surgery) based on positivity for at least three biomarkers.

Considering surgery, three studies reported on increased risk of surgery. One study reported a trend towards a larger proportion of people with CD requiring abdominal surgery with increasing number of positive biomarkers (N=517; p<0.0001 across the groups). Other estimates of higher risk of requiring surgery were:

- OR of 1.5 (95% CI: 1.3 to 1.8; p<0.001; N unclear) based on positivity for a median of two biomarkers;
- HR of 3.6 (95% CI: 1.2 to 11.0; p=0.023; N=14 with no prior complication or surgery) based on positivity for at least two biomarkers;

- HR of 2.8 (95% CI: 0.80 to 9.6; p=0.11; N=14 with no prior complication or surgery) based on positivity for at least three biomarkers.

Estimate of increased risk of treatment escalation by number of positive biomarkers was not available for IBDX.

In a study evaluating IBDX in an adolescent population, results for adolescents and children aligned with those derived from an adult cohort, with a higher number of positive serum biomarkers associated with an increased risk of experiencing complicated CD and requiring CD-related surgery. Research suggests that, although the levels of biomarkers fluctuate over time, the status of positive or negative for an individual biomarker remains constant.

Estimates of increased risk of developing a complication or requirement for surgery were not available for PredictSURE-IBD. The study evaluating PredictSURE-IBD reported that those 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). As noted earlier, based on the IPD supplied by PredictImmune, some in the validation cohort received a therapy other than glucocorticosteroid at entry. The EAG considers that choice of and time to first treatment escalation is likely to be influenced by response to treatment at study entry, which in turn is likely to be affected by risk of following severe course of CD. As people within the validation cohort have not followed a standardised algorithm of treatment, the EAG considers analysis of time to first treatment escalation is subject to a level of bias, the direction of which is unclear. The EAG reiterates that clinical experts fed back that it would be useful to assess CD-related surgery as an independent outcome.

Given the disparity in the clinical outcomes assessed for the IBDX and PredictSURE-IBD, the EAG considers that no conclusions can be drawn on the comparative effectiveness of the two tools in stratifying people by risk of severe course of CD.

4 METHODS FOR ASSESSING COST-EFFECTIVENESS

4.1 Systematic literature review for cost-effectiveness studies

4.1.1 Methods

A systematic literature review (SLR) was undertaken in July 2019 to identify published economic evaluations of the PredictSURE-IBD™ and IBDX tools, as well as economic evaluations of treatments for newly diagnosed patients, with moderate to severe Crohn's disease (CD). The searches were also used to identify potential model parameters in case a *de novo* model was needed. The searches were used to identify resource use and cost data, together with the natural history of CD. Separate searches were carried out for supporting information on utility data.

The following databases were searched for relevant studies:

- Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) (Ovid);
- Embase (Ovid);
- NHS Economic Evaluation Database (NHS EED) (Centre for Reviews and Dissemination, CRD);
- Cochrane Database of Systematic Reviews (CDSR) (Cochrane);
- Cochrane Central Database of Controlled Trials (CENTRAL) (Cochrane);
- Database of Abstracts of Reviews of Effects (DARE) (CRD);
- Health Technology Assessment Database (HTA) (CRD).

Further to the database searches, experts in the field were contacted with a request for details of relevant published and unpublished studies and reference lists of key identified studies were also reviewed for any potentially relevant studies.

The search strategy for existing economic evaluations of prognostic tests combined terms capturing the tests of interest (PredictSURE-IBD™ and IBDX) and the target population (adults who have been newly diagnosed with moderate to severe CD, and who have not been offered biologics under current standard care) with economic and healthcare resource use terms (adapted from the Canadian Agency for Drugs and Technologies in Health's search filter for economic evaluations).

The target population considered in the SLR to identify economic evaluations of treatments for CD and HRQoL evidence (adults with moderate to severe CD) was broader than the population considered in the SLR to identify economic evaluations of prognostic tests to account for the fact that patients' characteristics change along the treatment pathway. The search strategy for existing economic

evaluations of treatments for CD also replaced prognostic tool terms with terms related to corticosteroid, immunomodulator and biologic treatments. The search strategy for HRQoL data was not restricted by prognostic tools or treatments, and combined terms capturing the target population with HRQoL terms (adapted from Arber *et al.* 2017).[85]

Limits were applied to searches to remove animal studies, letters, editorials, comments or case studies. Only conference abstracts published within the last two years were considered for inclusion; it was assumed that any high-quality studies reported in abstract form before that date would have been published in a peer-reviewed journal. Searches were also restricted to studies published in the English language; however, no restriction by setting or geographical location was applied to the search strategy. Full details of the search strategies are presented in Appendix 3.

The titles and abstracts of papers identified through the searches were independently assessed for inclusion by two reviewers using pre-defined eligibility criteria. The inclusion and exclusion criteria for each of the three reviews are outlined in Table 9. The methodological quality of the full economic evaluations identified in the review was assessed using the Drummond checklist.

Table 9. Inclusion and exclusion criteria for the systematic reviews of economic and HRQoL evidence

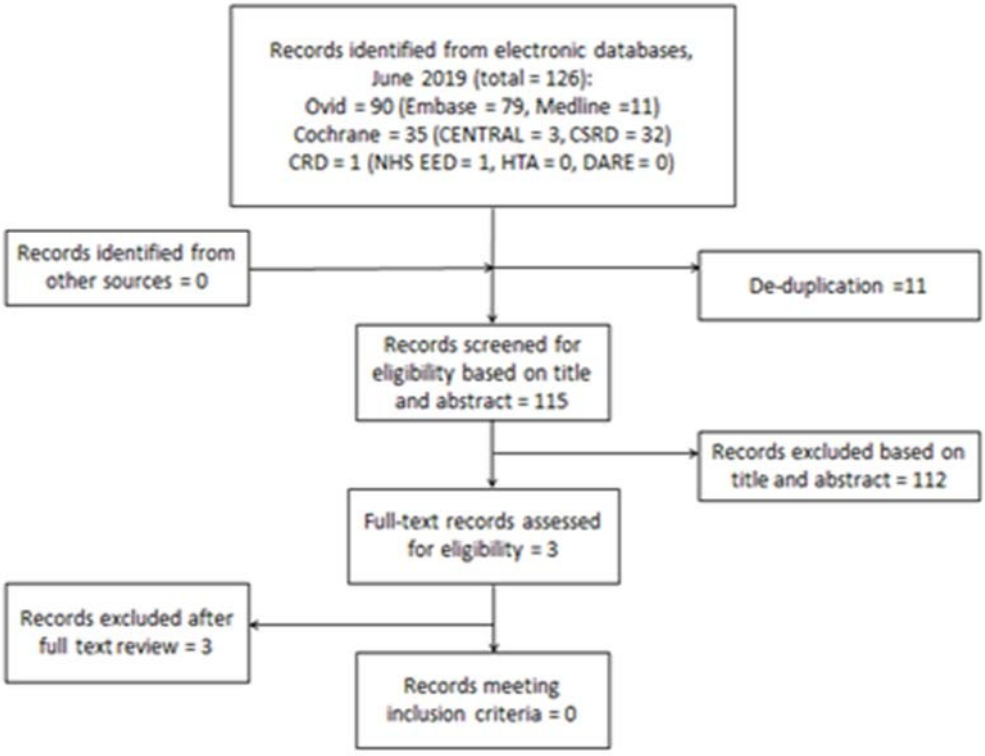
Inclusion criteria: economic evaluations of tests for the identification of those at high risk of developing a severe course of CD
• Prognostic tests according to the scope of the assessment (PredictSURE-IBD™ and CD Prognosis Test (IBDX)); • Study population according to the scope of the assessment (adults aged 16 years and older who have been newly diagnosed with moderate to severe CD, and who have not been offered biologics under current standard care); • Full economic evaluations (cost-utility, cost-effectiveness, cost-benefit or cost-consequence analyses) that assess both costs and outcomes associated with the prognostic tests of interest.
Inclusion criteria: economic evaluations of treatments for CD
• Economic evaluations of treatment strategies for CD, including the "top-down" and "step-up" (standard and accelerated) approaches; however, if insufficient data can be identified on those approaches, economic evaluations of individual treatments will be considered; • Study population included in the conceptual model (adults aged 16 years and older with moderate to severe CD; • Full economic evaluations (cost-utility, cost-effectiveness, cost-benefit or cost-consequence analysis) that assess both costs and outcomes associated with the treatment of interest.
Inclusion criteria: HRQoL of patients with CD
• Studies reporting utility data elicited using a generic or a condition-specific preference-based measure, or vignette and a validated, choice-based technique for valuation (i.e. time trade-off or standard gamble); however, if sufficient EQ-5D data are found during the searches for utility data, the EAG will restrict the data extraction to EQ-5D data; • Studies reporting utility data referring to specific health states associated with the treatment of CD patients in the economic model; • Studies in adults (16 years and above) with moderate to severe CD;

• Primary sources of utility data.
Exclusion criteria: all economic evaluations
• Non-English language; • Abstracts with insufficient methodological details; • Conference papers published 2 years before the search was performed (papers published pre-2017); • Papers published before NICE was formed (1999).
Abbreviations: CD, Crohn's Disease; EAG, External Assessment Group; HRQoL, health-related quality of life.

4.1.1.1 Economic evaluations of prognostic tests

The SLR identified a total of 115 papers after de-duplication and based on title and abstract a total of three papers were identified as potentially relevant and were obtained for full text review. Of the three papers identified for full text review, none were considered relevant for inclusion. Reasons for exclusion are provided in Appendix 4. The results of the process to identify evidence is summarised in Figure 4.

Figure 4. PRISMA diagram of SLR to identify economic evaluations of prognostic tests



4.1.1.2 Economic evaluations of treatments for CD

The SLR identified a total of 2,403 papers after de-duplication and based on title and abstract a total of 80 papers were identified as potentially relevant and were obtained for full text review. Of the 80 papers identified for full text review, 32 were considered relevant for inclusion. Of those, one Italian study

specifically compared the cost-effectiveness of top-down and step-up approaches (Marchetti *et al.* 2013).[86] A second study, in a UK setting, compared nine induction treatment sequences (composed of four treatment lines) (NICE NG129).[30]

The remaining studies compared individual treatment steps. Given the high volume of such studies, data extractions were restricted to UK studies plus the Italian study that compared the top-down with step-up approach. Reasons for paper's exclusion are provided in Appendix 4. The results of the process to identify evidence is summarised in Figure 5.

The type of economic evaluation included in each of the 11 extracted studies was a cost-utility analysis, where the incremental cost-effectiveness ratio (ICER) was expressed as the cost per QALY gained. Of the 11 extracted studies, five were related to NICE guidance, including three NICE Technology Appraisals (Dretzke *et al.* 2011 for TA187, Rafia *et al.* 2016 for TA352 and Hodgson *et al.* 2018 for TA456),[87-89] one NICE clinical guidance (NG 129)[30] and one NICE diagnostics guidance (Freeman *et al.* 2016 for DG 22).[90] For NG129, two economic evaluations were developed, one on treatment sequences to for the induction of remission and a second on treatments for the maintenance of remission.

The most frequent type of decision analytic model used to estimate cost-effectiveness was a Markov model. Three papers also included a decision tree followed by a Markov model to disaggregate the short- and long-term effects (Hodgson 2018 (TA456), Rafia 2016 (TA352) and Bodger 2009).[88, 89, 91] The time horizons in these analyses ranged from 1 to 60 years (lifetime), while the cycle lengths ranged from 2 weeks to 2 months. Decision trees without any Markov component were used to estimate cost-effectiveness over shorter time horizons (30 weeks and 1 year) in the two remaining analyses (NG129 and Saito 2013).[30, 92] A summary of the 11 extracted studies is provided in Table 10 and detailed data extractions can be found in Appendix 5.

Figure 5. PRISMA diagram of SLR to identify economic evaluations of treatments for CD

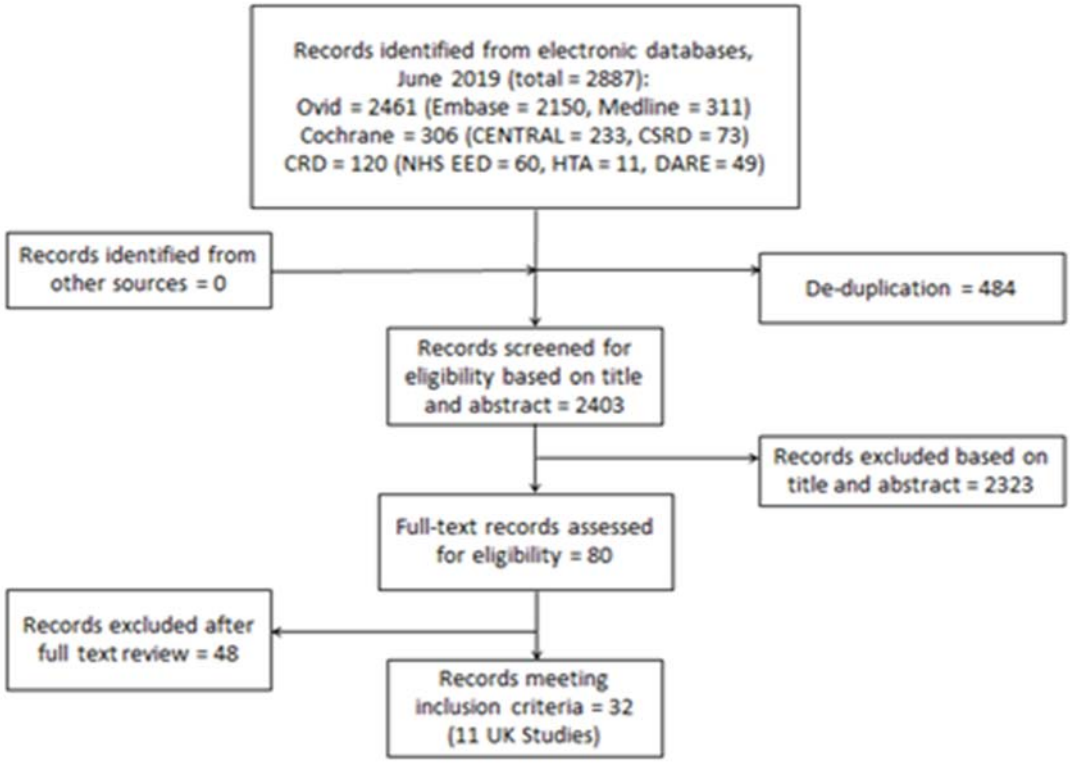


Table 10. Summary of the 11 included economic evaluations

Study	Population	Interventions/ comparators	Model type (cycle length)	Time horizon
Marchetti 2013	Newly diagnosed luminal moderate-to-severe CD	• Top-down: 1st step infliximab plus azathioprine, 2nd step additional infliximab plus azathioprine, 3rd step methylprednisolone plus azathioprine • Step-up: 1st step methylprednisolone, 2nd step methylprednisolone plus azathioprine, 3rd step infliximab plus azathioprine	Markov model (1 month)	5 years
Dretzke 2011 (TA187)	1. Moderate CD that is refractory to conventional treatment 2. Severe CD that is refractory to conventional treatment	• Infliximab induction infusions • Infliximab maintenance infusions • Adalimumab induction infusions • Adalimumab maintenance infusions • Conventional treatment (without TNF-α inhibitors including treatment with aminosalicylates, methotrexate, corticosteroids, azathioprine, metronidazole or surgical intervention)	Markov model (4 weeks)	1 year
Hodgson 2018 (TA456)	Adults with moderate to severe CD in two subpopulations: 1. Anti-TNF alfa failure	1. Ustekinumab compared with conventional care and vedolizumab for ant-TNF alfa failure	Decision tree followed by Markov model (2 weeks)	1 year

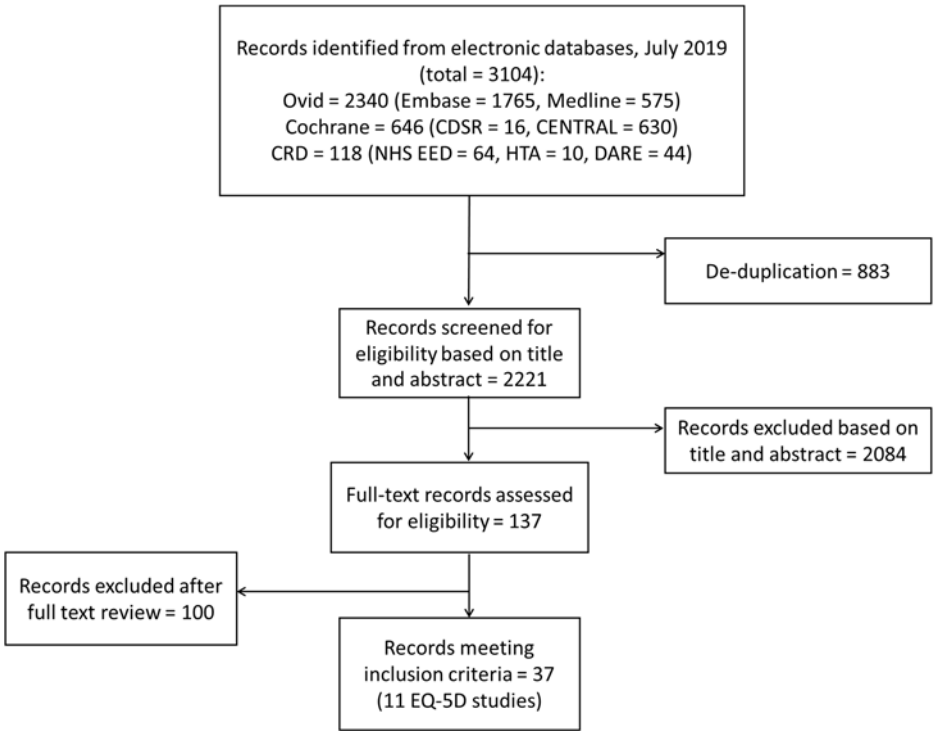
	2. Conventional care failure	2. Ustekinumab was compared with conventional care and adalimumab for conventional care failure		
Rafia 2016 (TA352)	Moderate to severe active disease after failure of initial therapy in 3 subpopulations: 1. The mixed ITT population, which comprised patients who had previously received anti-TNF-a therapy and those who were anti-TNF-a naive 2. Patients who were anti-TNF-a naive only 3. Patients who had previously received anti-TNF-a therapy only	• Vedolizumab induction and maintenance infusion • Conventional nonbiologic therapies (a combination of 5-amino salicylic acids, immunomodulators and corticosteroids)	Decision tree followed by Markov model (8 weeks)	10 years
Mayberry 2013 (NG129)*	1. Acute exacerbation of CD 2. Active CD in medically-induced remission	1. Nine treatment strategies with four treatment lines for acute exacerbations of CD 2. No treatment, azathioprine, mesalazine, olsalazine, budesonide and glucocorticosteroids compared for active CD in medically-induced remission	Decision tree Markov model (2 months)	30 weeks 2 years
Freeman 2016 (DG22)	Moderate to severe active CD treated with infliximab or adalimumab in two subpopulations: 1. Patients responding to treatment 2. Patients who had lost response to treatment	• Monitoring of serum anti-TNF-α compared • No testing	Markov model (4 weeks)	10 years
Saito 2013	Moderate to severe CD refractory to conventional therapies and naive to biologic therapy	• Infliximab induction and maintenance infusions plus azathioprine • Infliximab monotherapy	Decision tree	1 year
Bodger 2009	Moderate to severe active CD	• Infliximab infusions for induction of remission followed by maintenance treatment • Adalimumab injection for induction of remission followed by maintenance treatment • Conventional treatment (5-amino salicylic acids, immunosuppressive agents, corticosteroids, antibiotics, symptomatic therapies, topical therapies and surgery)	Decision tree followed by Markov model (8 weeks)	Lifetime (60 years)
Loftus 2009	1. Severe active CD 2. Moderate to severe active CD	• Adalimumab induction and maintenance therapy injection • Conventional non-biological therapeutics (5-aminosalicylic acid, antibiotics, immunosuppressants and corticosteroids).	No decision analytic model, costs and benefits were attached to estimated rates of hospitalisation	1 year
Lindsay 2008	1. Moderate to severe active luminal Disease	• Infliximab initial infusions and maintenance treatment	Markov model (luminal active	5 years

	2. Fistulising CD	• Conventional treatment, comprising immunomodulators and/or corticosteroids.	CD, 2 to 4 weekly cycles until week 14 then 8-weekly; fistulising active CD, one 14 week cycle and one 16 week cycle, then 24 weekly)	
Clark 2003	1. Chronic active disease resistant to conventional treatment 2. Fistulising CD resistant to conventional treatment	• Infliximab as single and episodic infusions • Placebo	Markov model (2 months)	Lifetime (40 years)
Abbreviations: CD, Crohn's disease, ITT, intention-to-treat; TNF, tumour necrosis factor *economic evaluations reported in the full guideline				

4.1.1.3 HRQoL evidence

The SLR identified a total of 2,221 papers after de-duplication and based on title and abstract a total of 137 papers were identified as potentially relevant and were obtained for full text review. Of the 137 papers identified for full text review, 37 were considered relevant for inclusion and 11 of those reported EQ-5D data. The remaining papers considered generic measure including the SF-36, SF-12, Psychological General Well-Being Index (PGWBI), Cleveland Global Quality of Life (CGQL) and EQ-5D VAS, and disease specific measures including the Crohn's disease activity index (CDAI) and Inflammatory Bowel Disease Questionnaire (IBDQ). Due to the high volume of relevant papers; the availability of EQ-5D data in these papers; and NICE's preference for EQ-5D data, the EAG decided to restrict the data extraction to primary sources of EQ-5D data. Reasons for exclusion of the ordered papers are provided in Appendix 4. The results of the process to identify evidence is summarised in Figure 6.

Figure 6. PRISMA diagram of SLR to identify HRQoL evidence



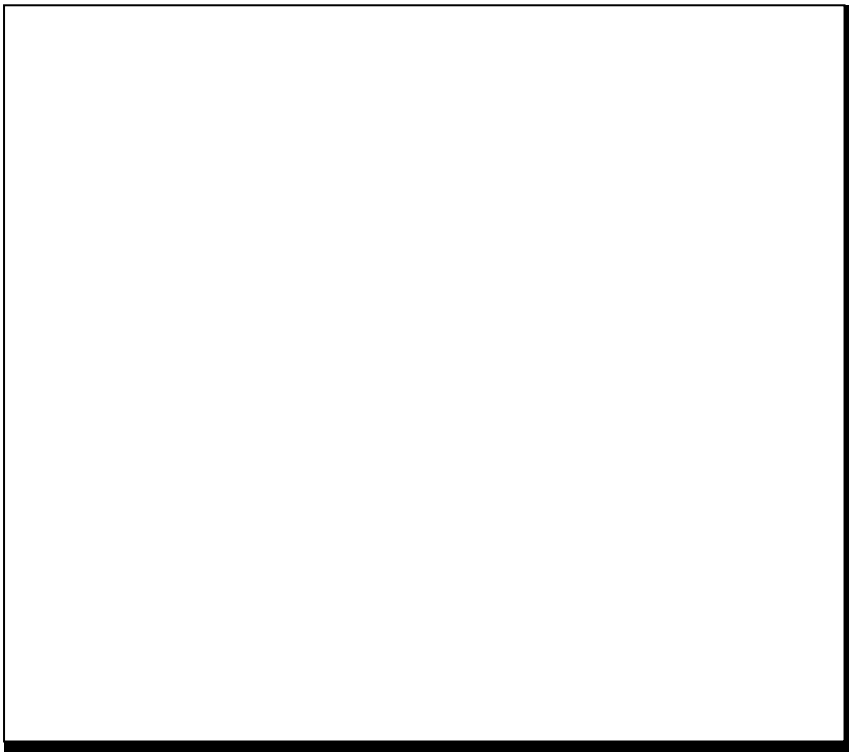
Of the 11 studies that reported EQ-5D data, 10 used the EQ-5D-3L and one of those 10 also collected EQ-5D-5L data. The remaining paper did not specify which levels of the EQ-5D were used. EQ-5D-3L responses were converted into utilities using UK population tariffs in four studies, which were undertaken in Italy (Benedini *et al.* 2012 and Mozzi A *et al.* 2016),[93, 94] Germany (Stark *et al.* 2010)[95] and Hungary (Rencz *et al.* 2019)[96]. However, each of those four studies used different sources of UK population tariffs to value EQ-5D-3L responses. The sources included Dolan *et al.* 1995[97] for Benedini *et al.* 2012,[93] Badia *et al.* 2001[98] for Mozzi *et al.* 2016,[94] Dolan *et al.* 1996[99] for Stark *et al.* 2010[95] and Dolan 1997[100] for Rencz *et al.* 2019.[96]

Six of the 11 studies that reported EQ-5D data were undertaken in Spain and four of those used Spanish population tariffs developed by Badia *et al.* 1999[101] (for Casellas *et al.* 2005a,[102] Casellas *et al.* 2007[103] and Huaman *et al.* 2010)[104] or Rue and Badia 1996[105] (for Casellas *et al.* 2000)[106] to convert EQ-5D-3L responses into utilities. The other two studies undertaken in Spain (Casellas *et al.* 2005b and Saro *et al.* 2017)[107, 108] did not report the sources used to value EQ-5D-3L responses. Finally, one study undertaken in Poland (Holko *et al.* 2016)[109] valued EQ-5D-3L responses using a Polish population tariff developed by Golicki *et al.* 2010.[110] As for the study that collected EQ-5D-5L responses in Hungary (Rencz *et al.* 2019),[96] English tariffs developed by Devlin *et al.* 2018 were employed.[111]

4.1.2 PredictImmune's economic model

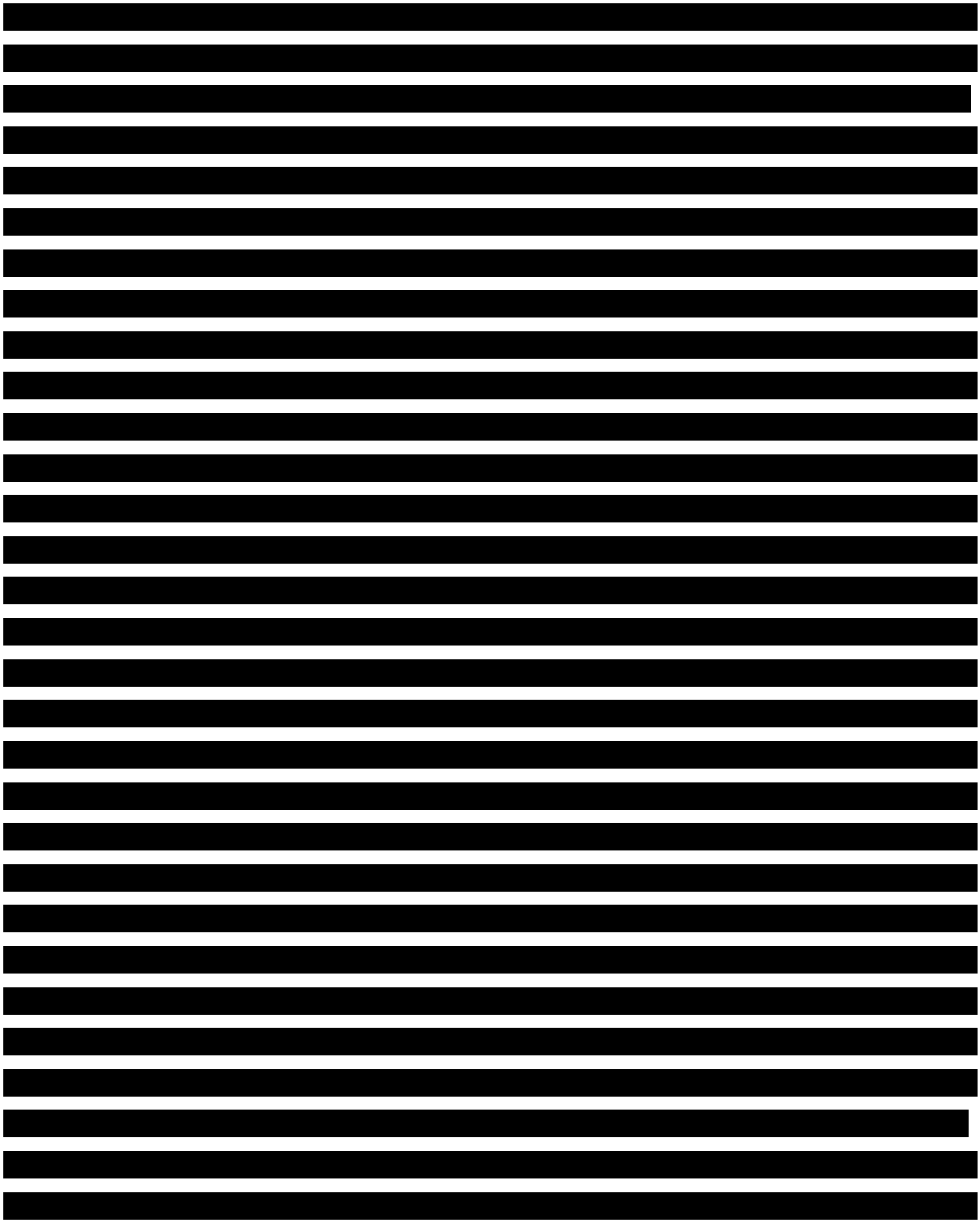
During the diagnostic assessment review subgroup meeting, the EAG became aware of the existence of an economic model built by PredictImmune to assess the cost-effectiveness of PredictSURE IBD™. As a result of a request from the EAG, the company supplied the economic model.

[redacted]



[redacted]

[redacted]



[redacted]

4.2 Development of the health economic model

As reported in Section 3, despite extensive systematic searches of the literature, no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, IBDX® and PredictSURE-IBD™. Furthermore, the EAG considers it would be challenging to ascertain an accurate estimate of prognostic accuracy of the tools in stratifying course of CD.

Therefore, the development of an economic model to accurately assess the cost-effectiveness of IBDX® and PredictSURE-IBD™ was not possible based on the currently available data. Instead, the EAG developed an economic model that provides a structural framework for analysing future available data on prognostic accuracy and to assess the costs and consequences of treating high- and low-risk patients with both TD and SU strategies. Furthermore, the EAG did not find any robust evidence on the effectiveness of the complete TD or SU treatment sequences including no evidence on the effectiveness of these strategies by patients' risk of disease severity.

As the ongoing PROFILE RCT[51] randomises people to accelerated SU or TD treatment after determination of high- or low-risk of following a severe course of CD, the ERG considers that the trial could provide additional data to inform estimates of prognostic accuracy and patients' outcomes stratified by risk and type of treatment received.

As no model found through the SLR met the requirements of the review, the EAG developed its own model. The latter is described in the following sections.

4.2.1 Population

The population included in the economic analysis consists of adults (aged 16 years and older) who have been newly diagnosed with moderate to severe CD, and who have not been offered biologics under current standard care. The population in the economic model is largely based on the Biasci *et al.* population.[50] The paper included a training (38 CD patients) and a validation cohort (66 CD patients);

nonetheless, the published paper did not provide sufficient detail on treatments received by the validation or training cohort. Therefore, the EAG asked PredictImmune to provide additional treatment data for the study cohort in Biasci *et al*. and in response, the company provided the available individual patient-level data (IPD).

The IPD included 88 patients with newly diagnosed CD and a classification of high- or low-risk disease. However, the EAG had to remove patients from the IPD (as explained in detail in Section 4.2.4.1) therefore, the final population included in the model was reduced to 40 patients ([redacted] high-risk patients and [redacted] low-risk patients). The average age in the EAG modelled population was 35 years and 65% of patients were non-smokers, with 25% being smokers and 8% ex-smokers (smoking status missing for 2%). Thirty-three percent of patients were male and 55% female (12% of patients with no information on sex collected). The study did not collect data on patients' weight, therefore the EAG assumed a mean weight of 71.4kg in the model, based on results provided in TA456.

4.2.2 Intervention and comparator

As per the final protocol, the interventions of interest are the IBDX® and the PredictSURE IBD™ tests. Nonetheless, the base case economic model included only the PredictSURE IBD™ test, while a scenario analysis was undertaken to compare the IBD™ test against standard care (SC). Although the EAG considers that there are no robust prognostic accuracy data for either tests, the development of the model was mainly based on the IPD provided by PredictImmune pertaining to the use of PredictSURE IBD™.

The comparator included in the analysis is SC. As there is no test or algorithm available in the UK NHS to determine long-term course of disease or an individual's risk of developing severe course of disease, estimation of prognosis is based on clinical judgement of presenting signs and symptoms, together with potential risk factors for severe course of disease (more details are provided in Section 2.1).

For the purpose of the economic model, the EAG assumed that the PredictSURE IBD™ test (and the IBDX® in the scenario analysis) ultimately categorise patients into high- and low-risk disease categories, so that treatment sequences can be allocated accordingly. The treatment sequences included in the economic model were based on clinical expert opinion provided to the EAG and are intended to describe SC in the UK NHS for the step-up arm (SU); and the accelerated treatment pathway – top-down (TD) arm, not currently recommended in the UK NHS. The clinical experts added that less than 10% of CD patients receive TD therapy in the NHS thus, the EAG assumed that patients in the SC arm of the model can only receive SU therapy. The TD treatment approach is assumed to be received only by high-risk patients who have been tested either with PredictSURE IBD™ or IBDX®.

The two treatment strategies include an induction treatment with prednisolone for 100% of patients in the model. The difference in treatment strategies thereafter are based solely on the fact that the SU strategy includes an additional treatment step with IMs at the beginning of the sequence. The modelled treatment steps include four bundles of different types of therapies: IMs; anti- TNF biologics; and second- and third-line biologics. Clinical expert opinion was used to derive the distribution of treatments comprised in each treatment bundle. The bundles were defined as follows:

1. IM bundle: 80% azathioprine; 10% mercaptopurine; 10% methotrexate;

2. Anti-TNF bundle: 40% infliximab; 60% adalimumab and 30% of all patients get the IM bundle;

3. Second-line biologic bundle: 50% vedolizumab; 50% ustekinumab and 20% of all patients get the IM bundle;

4. Third-line biologic bundle: 50% vedolizumab; 50% ustekinumab and 20% of all patients get the IM bundle (patients getting vedolizumab as second line treatment are assumed to receive ustekinumab as third line treatment and vice-versa).

The order of treatments received in the TD strategy is described in Figure 10, while the treatment sequence assumed for the SU strategy is reported in Figure 11.

Figure 10. Top-down treatment strategy

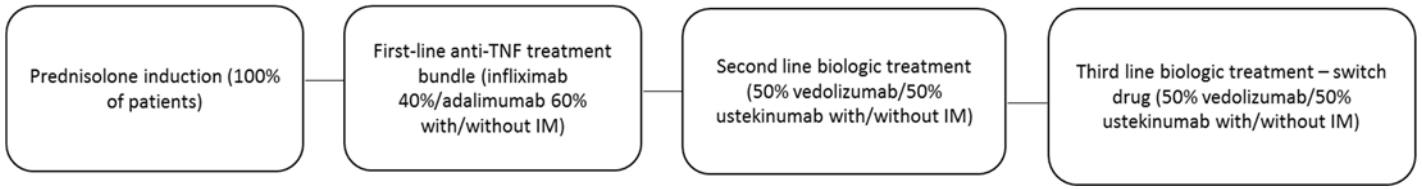
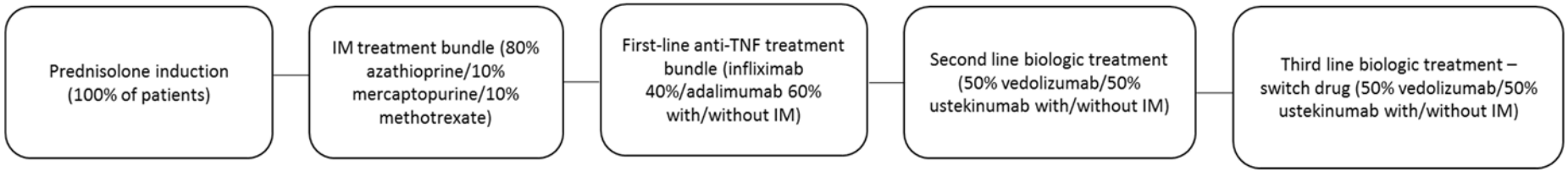


Figure 11. Step-up treatment strategy



4.2.3 Model structure

The EAG adopted a hybrid modelling approach, where a decision tree was developed to allocate patients to a response category after initial induction therapy in either the TD or SU treatment arms. The decision tree is followed by a cohort model, where state membership was estimated through a series of different Markov health states.

Patients enter the decision tree model (Figure 12) after being allocated to the test (with either PredictSURE IBD™ in the base case or IBDX® in as scenario analysis) or no test (SC) arm. In both test and no test arms, patients are categorised as high-risk or low-risk patients, according to test results; or clinical judgment alone, depending on the model arm. Given that patients in the SC arm of the model can only receive the SU treatment approach and that the TD treatment approach is assumed to be received only by high-risk patients, the economic model is ultimately assessing the cost-effectiveness of TD therapy vs SU therapy in high-risk patients. The EAG did not identify any direct evidence on the latter. There is however, an ongoing study (PROFILE[51]) which will provide data on the relative effectiveness of these treatment strategies in high-risk patients. The ERG considers that study should also be able to inform the costs and health consequences of "misdiagnosing" patients as high- and low-risk. The EAG has undertaken a scenario analysis to account for the cost-effectiveness of misdiagnosed cases. The analysis is described in more detail in Section 5.

After being allocated to either the TD or SU treatment strategies, patients are allocated to induction therapy, at the end of which they are classified as responders (an improvement in CDAI score higher than 70) or non-responders (deterioration; no change; or an improvement of less than 70 in CDAI score). Duration of induction therapy differs by class of treatments (i.e., IM, anti-TNF, and second-line biologic). If patients respond to induction therapy, they move to the maintenance cohort model (Figure 13), while non-responders escalate to the next step on their allocated treatment strategy.

Responders to their first induction therapy enter the maintenance cohort model in the remission (CDAI<=150); mild (CDAI 150-220); or moderate to severe (CDAI 220-600) health states. Patients can then move between these states during maintenance therapy, reflecting the different levels of response to maintenance therapy. The probability of patients transitioning between these states is also dependent on the treatment class received.

Non-responders to induction therapy escalate to induction in the next step of their treatment strategy, to which they can become responders or non-responders. Patients receiving their second induction therapy are assessed for response and escalation to the next treatment step, similar to patients receiving their first induction therapy (portrayed by the loop in Figure 12).

Patients in the mild and in the moderate to severe states are at risk of escalating to the next treatment step and death is the absorbing state in the model.

Escalation to next treatment step occurs therefore, due to two reasons in the model: lack of response to induction therapy; or relapse while on maintenance therapy. The former is a default assumption in the model, as 100% of patients who do not respond to induction therapy move to the next step in their treatment strategy. The latter is not explicitly estimated in the economic model, but instead is assumed that time to treatment escalation (taken from Biasci *et al*.[50]) reflects a relapse while on maintenance treatment. This issue is further discussed in Section 4.2.4.1.

The EAG had to estimate surgical events as a stand-alone outcome in the model. This modelling simplification 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 surgeries is estimated, and the associated costs and impact on patients' quality of life is calculated (this is further discussed in Section 4.2.4.4). Patients who receive surgery in the model have an increased probability of dying associated with the procedure.

The economic assessment is taken from the perspective of the NHS and Personal Social Services and both costs and benefits are discounted at 3.5% per annum. Cycle length in the model is 2 weeks, and the time horizon of the model is 65 years (when modelled patients would be 100 years old).

Figure 12. Model for induction treatment

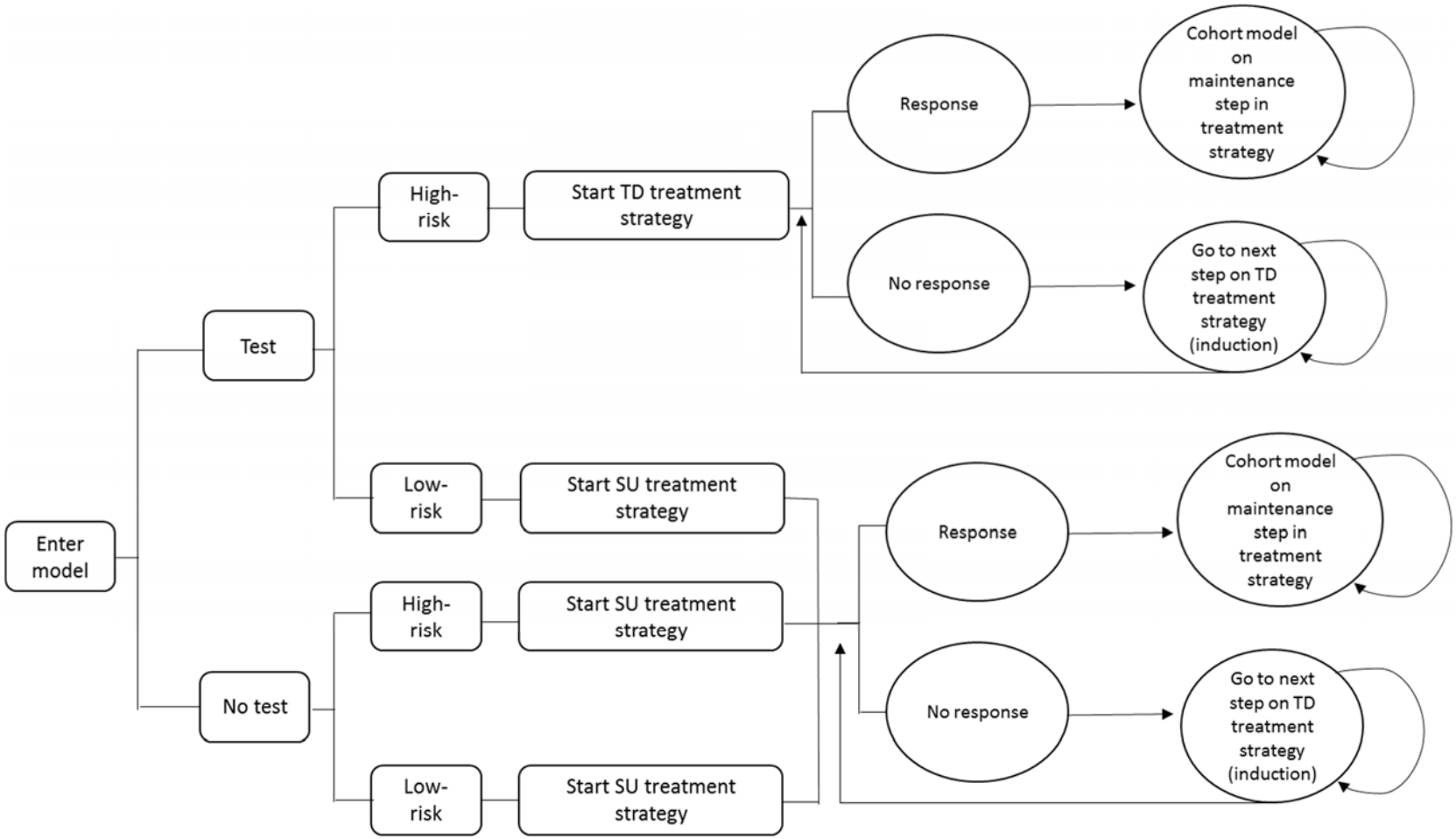
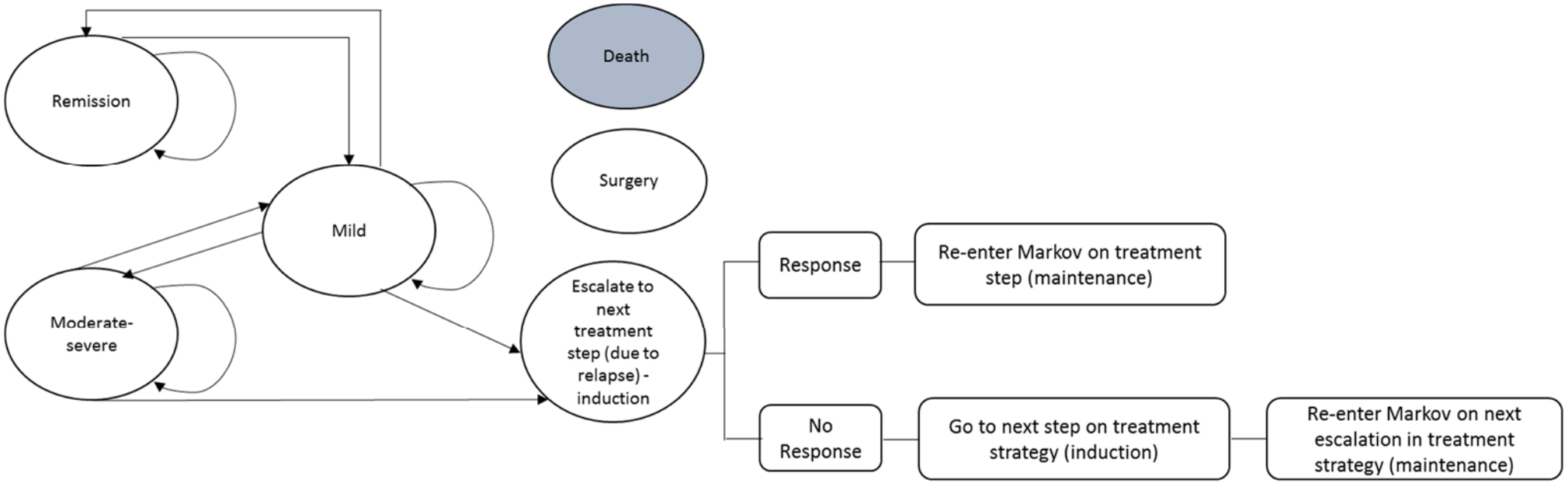


Figure 13. Cohort model for TD and SU maintenance steps



4.2.4 Clinical input parameters

As mentioned in Section 4.2.3, the economic model is ultimately assessing the cost-effectiveness of TD therapy vs SU therapy for high-risk patients. However, the EAG did not identify any direct evidence on the latter thus, the clinical data informing the economic analysis had to be derived from multiple sources. This approach is not ideal and creates a patchwork network of evidence, introducing uncertainty in the economic results. It is anticipated by the EAG that this problem will be (at least partially) overcome when results from the PROFILE trial are available to populate the economic model. The ERG considers that the PROFILE study should also be able to inform the costs and health consequences of "misdiagnosing" patients as high and low risk, therefore allowing the estimation of the cost-effectiveness of under- or over-treating CD's patients in the UK NHS.

The EAG notes that the clinical input parameters in the base case economic model for PredictSURE IBD™ and in the scenario analysis for IBDX® are the same. The only difference in the cost-effectiveness analyses of the two diagnostic tests is the cost of the test.

The EAG found two main sources of evidence that could be used to model time to treatment escalation (TTE) and time to surgery (TTS). Nevertheless, each source could only partially inform the TTE and TTS analyses in the economic model. While the Biasci *et al*.[50] paper could inform TTE and TTS according to high- and low-risk of CD complications (for the SU strategy); the D'Haens *et al.*[34, 35] (and its 10-year follow-up study Hoekman *et al*.)[112] could inform TTE and TTS according to TD and SU treatments (for a population with mixed risk of disease complications).

The Biasci *et al*. [50] study enrolled patients with active CD who were not receiving concomitant corticosteroids, immunomodulators or biological therapy. Forty patients received treatment with a corticosteroid, followed by an IM (out of whom 50% escalated to treatment with an anti-TNF). This treatment strategy was considered to be a good representation of the first three steps in the SU pathway described by the EAG's clinical experts. Biasci *et al*. [50] included TTE outcomes however, did not differentiate outcomes by treatment strategy, but instead by risk of severe disease course. Therefore, the data provided in the study could only potentially inform the difference in TTE and TTS for high- vs low-risk patients receiving SU.

The D'Haens *et al.* [34, 35] study evaluated the clinical efficacy of early immunosuppression compared with conventional therapy. The study consisted on a 2-year open-label randomised trial at 18 centres in Belgium, the Netherlands, and Germany and randomly assigned 133 patients to either early combined immunosuppression or conventional treatment. The study collected outcome data on time to relapse for 62 patients: 20 patients who received conventional therapy; and 42 patients assigned to combined immunosuppression who received three infusions of infliximab (5 mg/kg of bodyweight) at weeks 0, 2,

and 6, with azathioprine. Additional treatment was given with infliximab and, if necessary, corticosteroids, to control disease activity.

Patients assigned to conventional management received corticosteroids, followed, in sequence, by azathioprine and infliximab if needed. If patients responded to treatment with corticosteroids, treatment tapering was initiated. If patients' symptoms worsened during the course of corticosteroid tapering and did not respond to an increase in treatment dose, treatment with azathioprine was initiated (2–2.5 mg/kg per day). Patients who relapsed after withdrawal of corticosteroids were given a second course of corticosteroids in combination with azathioprine. Any patient who remained symptomatic after 16 weeks of azathioprine treatment received an induction course of infliximab (5 mg/kg bodyweight at weeks 0, 2, and 6) and continued antimetabolite treatment.

Therefore, even though the study forms a reasonable evidence base for measuring the relative effectiveness of anti-TNF vs corticosteroid followed by IM and anti-TNF, it does not differentiate outcomes by risk of severe disease course, only by treatment received.[34, 35] Furthermore, the treatment sequences included in the D'Haens *et al.* [34, 35] trial only partially reflect the TD and the SU strategies as described by the clinical experts advising the EAG: the TD and SU strategies in the UK include an initial induction with steroid treatment. In the UK, these clinical strategies are only differentiated after steroid treatment, where TD patients are given treatment with an anti-TNF and SU patients are given an IM treatment. Since the TTE data taken from D'Haens *et al.* [34, 35] was based on time to relapse, the EAG assumed that relapse meant failure on first treatment in both strategies in the study and therefore, time to relapse data was based on the comparison of anti-TNF with corticosteroids. Furthermore, D'Haens *et al.* [34, 35] included a mix of high- and low-risk patients. This means that low-risk patients were overtreated with first-line anti-TNF. The study concluded that TD patients took a longer time to relapse compared to SU patients.

The Hoekman *et al*.[112] study was a retrospective review of medical records of patients included in the D'Haens *et al.*[34, 35] trial, which collected data on hospitalisation, flares, surgery, clinical activity, and other outcomes, for a median follow-up of 10 years. The study concluded that in the long-term, no difference was found in clinical remission rate; endoscopic remission, hospitalisation, surgery or new fistulas. During the follow-up period, the proportion of patients who received an IM was similar across arms (88% SU and 86% TD, p-value=0.76), while the use of anti-TNF was higher in the SU arm compared to the TD arm (73% vs 54%, p-value=0.04). However, the authors explained that the observed lower use of anti-TNF agents during long-term follow-up in TD-treated patients was not directly relevant for current clinical practice, because it was related to the previous practice of episodic anti-TNF treatment with no anti-TNF maintenance.

Given that the EAG did not find any sources of evidence combining CD outcomes differentiated by risk of disease and by treatment received, the EAG had to choose between Biasci *et al.*, [50] which differentiated outcomes by patients' risk of severe disease course and D'Haens *et al.* [34, 35] (and Hoekman *et al.*),[112] which differentiated outcomes by type of treatment strategy received (a proxy for TD vs SU) to form the baseline treatment measure in the model. The EAG chose Biasci *et al.* [50] because it considered that estimating a relative treatment effect of TD vs SU (from D'Haens *et al.*[34, 35] for TTE and from and Hoekman *et al.*[112] for TTS) and applying it to a different population was based on a less flawed assumption than estimating the relative risk of disease to be applied in a different group of patients. Furthermore, given that the purpose of the diagnostic tests is to categorise patients into high- and low-risk disease, the EAG's preference was to prioritise robust evidence for this component of the model. Additionally, the D'Haens *et al.* [34, 35] data did not cover the sequences of treatments included in the TD or the SU approaches as per clinical practice in the UK.

The EAG discusses the TTE and TTS data analysis undertaken using the Biasci *et al*; [50] D'Haens *et al*; [34, 35] and Hoekman *et al*. [112] in the next subsections of the report. However, it caveats the results of the theoretical economic analysis by the lack of robust evidence around the relative clinical effectiveness of TD vs SU strategies for the population defined in the scope.

4.2.4.1 Time to treatment escalation in high- and low-risk patients

Out of the 105 patients included in the Biasci *et al.* [50] IPD provided to the EAG, 88 patients were newly diagnosed with CD (Figure 14). Out of these 88 patients, 75 patients received initial treatment with corticosteroids. The EAG also removed 35 patients from the analysis who never received a subsequent IM after corticosteroids (leaving 40 patients for the TTE analysis - Figure 14). The EAG did not model time to escalation from corticosteroid treatment to IM (SU) or to anti-TNF (TD). This decision was based on the fact that the economic analysis is driven by the impact of giving high-risk patients TD vs SU therapy therefore, considering that 100% of patients in the high-risk group would receive initial treatment with corticosteroids, the impact of treatment would cancel out across the TD high-risk and the SU high-risk arms, as the treatment effect from D'Haens *et al.* [34, 35] was only applied for the IM vs anti-TNF (and subsequent treatment steps) in the model.

Figure 14. Selection of patients from Biasci *et al* for time to escalation analysis.

The TTE data from Biasci *et al.* [50] were used to estimate time to next treatment step in all the SU arms of the economic model (the test and no test arms of the model and the high- and low-risk arms of the no test model). In order to extrapolate TTE data to the model time horizon, the EAG analysed the IPD data to create time to event data for time to first escalation (reported in Figure 15).

The final dataset included [redacted] high-risk and [redacted] low-risk patients, with [redacted] escalation events observed in the high-risk group and [redacted] escalation events in the low-risk group). The EAG censored patients who did not have an escalation event. Time to treatment escalation was statistically significantly different across the high- and low-risk arms (p-value = 0.02). Overall, the EAG notes that both the number of patients and events in the analysis are very small and therefore, the results of the EAG's analysis need to be interpreted with extreme caution.

The EAG had to make some assumptions in its base case analysis in order to use the Biasci *et al*. [50] data. These consist of the following:

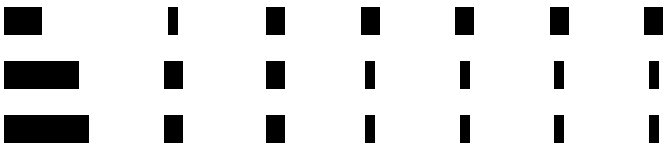
1. Treatment escalations in the model correspond to a relapse to patients' current treatment (or a flare);

2. Patients have the same baseline probability of escalating to the next step in the SU treatment strategy (which is estimated from time to first escalation in Biasci *et al*. [50]) regardless of number of previous escalations.

The EAG acknowledges that these assumptions are a simplification of clinical reality, where time to escalation is likely to depended on number of previous treatments. Nonetheless, given that patients in remission are assumed not to escalate treatment while on maintenance therapy, and that the probability of remission changes according to treatment step, the total number of patients escalating treatment differs by treatment step, across all treatments.

Furthermore, as mentioned in this section, the EAG did not find any sources of evidence containing complete treatment sequences (for the TD and SU strategies), which would have allowed the estimation of TTE by treatment step and response status.

The EAG considered the possibility of splitting the TTE Biasci *et al*. [50] data by first; and second (or more) escalations. However, the number of events in the second (or more) escalations dataset was too small (3 events overall) and therefore the KM data were deemed unreliable. The EAG also considered the possibility of splitting the TTE data according to patients' initial response to treatment (by using a proxy of time to escalation in Biasci *et al*. [50]). However, decided against it given the already very small size of the Biasci *et al*. [50] population for the TTE data available.



The TTE KM data were fitted with an exponential, Weibull, Gompertz, log-logistic, lognormal and generalised gamma models in accordance with guidance from NICE Decision Support Unit (DSU) Technical Support Document (TSD) 14.[113] The fit of each parametric model was compared with the observed KM data and statistical fit was assessed using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The fitted curves were also validated by clinical expert opinion. Given the low number of patients and events across treatment arms, the curves were initially fit dependently for high- and low-risk patients. However, clinical expert opinion provided to the EAG supported the use of different models for high- and low-risk patients, as the clinical expectation is that all high-risk patients will eventually escalate from IM to anti-TNF but that only 65% of low-risk patients

escalate from IM. Amongst the best-fitting curves, the ones that support the clinical predictions are the Gompertz curve for low-risk patients and the lognormal for high-risk patients.

The EAG acknowledges that the DSU advises against fitting different models to same-study arms unless a strong clinical argument exists. The EAG considers that such a clinical argument is present in this case (as supported by clinical expert opinion) and that the nature of the modelled outcome (TTE for different disease severity course) lends to plausibility for the difference in the curves' shape.

According to the AIC and BIC statistics reported in Table 11 and Table 12 , for high- and low-risk patients, respectively, the three best-fitting models to the high-risk KM data from Biasci *et al*. [50] are the lognormal, loglogistic and gamma, while the gamma, exponential and Gompertz are the three best-fitting models for the low-risk group.

Figure 16 shows the fitted curves for high-risk patients along with the TTE KM data, while Figure 17 shows the equivalent curves for low-risk patients. The chosen lognormal (for high-risk patients) and Gompertz (for low-risk patients) curves are presented together in Figure 18.

Table 11. Measure of fit statistics for time to treatment escalation - high-risk patients in Biasci *et al*.

	AIC	BIC
Exponential	150.95	**152.09**
Weibull	152.11	154.38
Gompertz	150.13	152.40
Lognormal	**149.17**	**151.44**
Loglogistic	**149.99**	**152.26**
Gamma	**149.98**	153.39
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion		

Table 12. Measure of fit statistics for time to treatment escalation - low-risk patients in Biasci *et al*.

	AIC	BIC
Exponential	**48.79**	**49.62**
Weibull	50.60	52.27
Gompertz	**49.08**	**50.75**

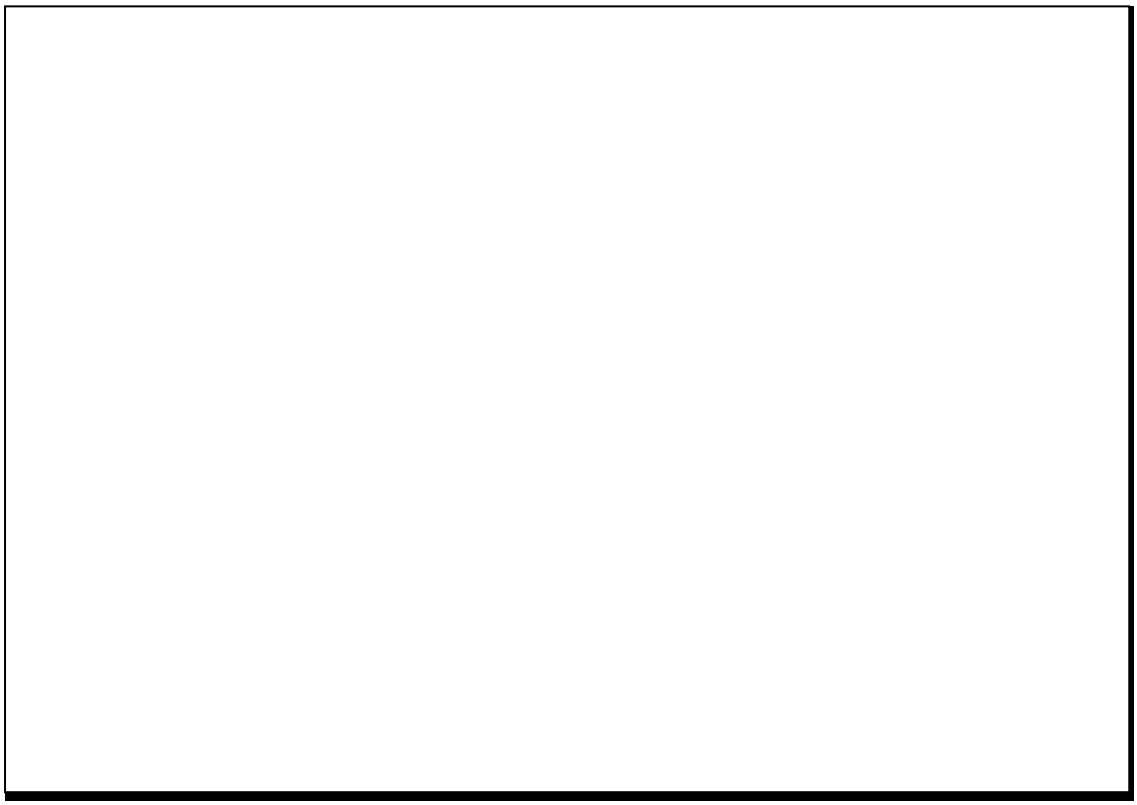
Lognormal	49.47	51.14
Loglogistic	50.17	51.83
Gamma	**46.68**	**49.18**
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion		

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

4.2.4.2 *Effectiveness of top-down vs step-up treatment strategy on time to treatment escalation*

In order to estimate TTE in the high-risk TD strategy arm of the model, the EAG applied a hazard function derived from D'Haens *et al*. to reflect the treatment effect of TD vs SU treatment on escalations.[34, 35]

The EAG had to make some assumptions in their base case analysis in order to use the D'Haens *et al*. [34, 35] data. These consist of the following:

1. Randomisation has resulted in balanced populations of high- and low-risk patients within each treatment group;

2. The relative treatment effect of TD vs SU in a mixed-risk population is the same as the relative treatment effect of TD vs SU in a high-risk population;

3. Time to relapse is a proxy measure for time to next treatment escalation;

4. The effectiveness of the treatment strategies in D'Haens *et al*. [34, 35] is a proxy for the treatment effectiveness of the first step in the TD and SU strategies modelled.

The first regimens in the treatment strategies included in D'Haens *et al*. [34, 35] (corticosteroid vs anti-TNF) are likely to overestimate the relative effectiveness of the modelled first step treatment in the TD strategy (anti-TNF) compared to the first step in the SU strategy (IM). Counterbalancing the direction of this bias, is the fact that the anti-TNF regimen in the study consisted of only three infliximab infusions (at weeks 0, 2 and 6) followed by maintenance monotherapy treatment with azathioprine or methotrexate and additional infliximab infusions only in case of clinical deterioration. As pointed out by the authors of the Hoekman *et al*.[112] study, since the D'Haens *et al.* [34, 35] trial, clinical practice has evolved to continued maintenance treatment with infliximab (in cases of a favourable response to induction treatment), which is consistent with UK's clinical practice and NICE guidelines. Therefore, even though it is not possible to anticipate the overall magnitude or direction of these biases in the data, they work in opposite directions, and so at least partially alleviate the impact of the overall bias in the analysis.

The EAG digitised the time to relapse data KM data in D'Haens *et al*. [34, 35] and used the number of patients at risk provided in the study to simulate the pseudo-individual patient-level data using the Guyot *et al*.[114] method and the algorithm in the *survHE R* package.[115] Subsequently, the EAG fitted a variety of parametric curves to the KM data (reported in Figure 19) using the same process as described in Section 4.2.4.1. The EAG notes that time to relapse was statistically significantly different across the TD and SU arms (p-value = 0.04).

The EAG restricted the modelling of the D'Haens *et al*. [34, 35] data to dependently fitted survival models only. This was to ensure that the relative effect estimated between the two treatment groups was only a scaling factor. Allowing both the scale and shape of the curves to vary would have resulted in implausible estimates of a relative effect, particularly in the probabilistic sensitivity analysis (PSA) where samples of the curves could theoretically cross.

Given that no relapse events took place for the first 14 weeks of the analysis, both KM curves show a plateau from week 0 to week 14 (Figure 19). This made the curve fitting exercise challenging as the shape of the fitted curves was heavily influenced by the plateau.

According to the AIC and BIC statistics reported in Table 13, the three best-fitting models to the time to relapse KM data were the lognormal, log-logistic and the gamma. Figure 20 shows the fitted curves for TD patients along with the time to relapse KM data, while Figure 21 shows the equivalent curves for SU patients. The lognormal provided the second best fit according to AIC and BIC statistics. Given

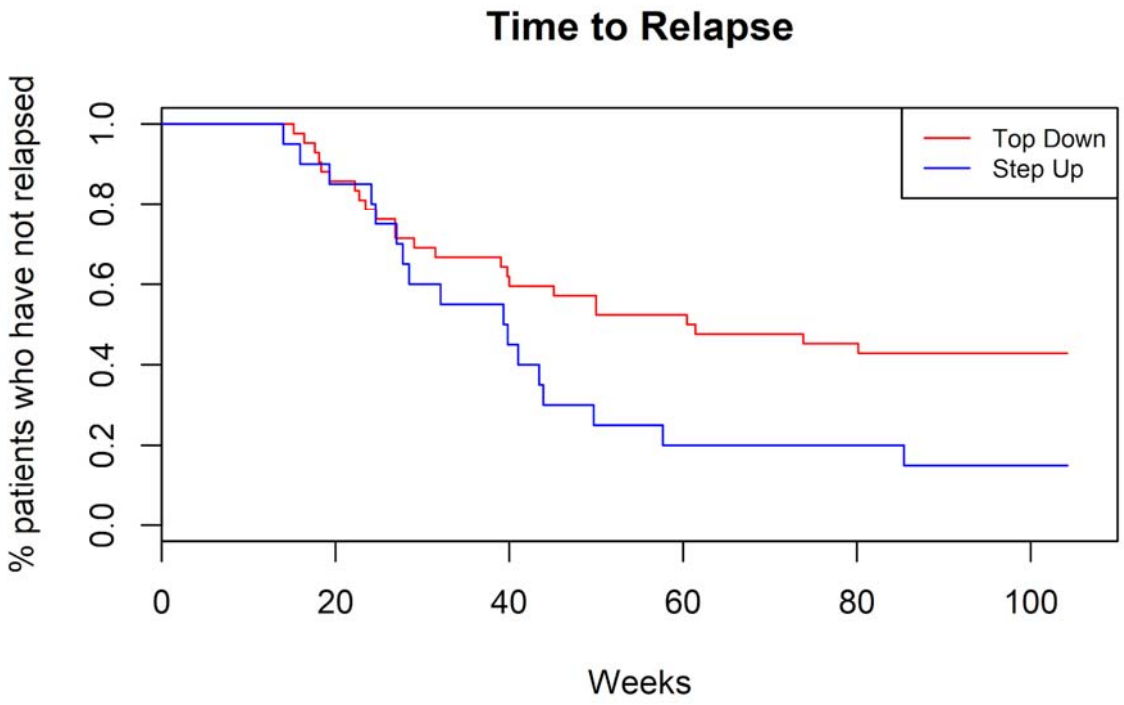
that TTE data were fitted with a lognormal model and that the hazard function derived from D'Haens *et al*. was to be applied to the TTE data, the EAG chose the lognormal curve.

Given that none of the three best-fitting curves provided a great visual fit to the KM data (due to the plateau observed for the initial 14 weeks), the EAG explored the option of truncating the KM data at 12 weeks (Figure 22) in order to fit survival curves.

According to the AIC and BIC statistic reported in Table 14 the three best-fitting models to the truncated KM data were the gamma, lognormal and the Gompertz. The lognormal provided the best fit according to AIC and BIC statistics. Figure 23 shows the fitted curves for TD patients along with the time to relapse KM data, while Figure 24 shows the equivalent curves for SU patients.

Even though the curves fitted to the truncated data provide a better visual fit, the EAG was warry of eliminating 12 weeks of time to relapse data from the analysis. Therefore, the EAG ran the economic analysis with both sets of lognormal curves (i.e. based on the truncated and the original KM data) and concluded that the impact on the final ICER was minimal. Thus, the EAG decided to use the non-truncated lognormal curves in the model (Figure 25).

Figure 19. Time to relapse estimated by the Evidence Assessment Group



Time	0	26	39	65	91	104
Immunosuppression	65	33	28	20	18	7
Conventional care	64	15	10	4	3	3

Table 13. Measure of fit statistics for time to relapse (dependant fit) in D'Haens *et al*.

	AIC	**BIC**
Exponential	326.49	330.75
Weibull	325.43	331.81
Gompertz	328.43	334.81
Lognormal	**315.92**	**322.30**
Loglogistic	**318.59**	**324.97**
Gamma	**299.47**	**307.97**
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion		

Figure 20. Step-up time to relapse curves fitted with Weibull, lognormal and log-logistic

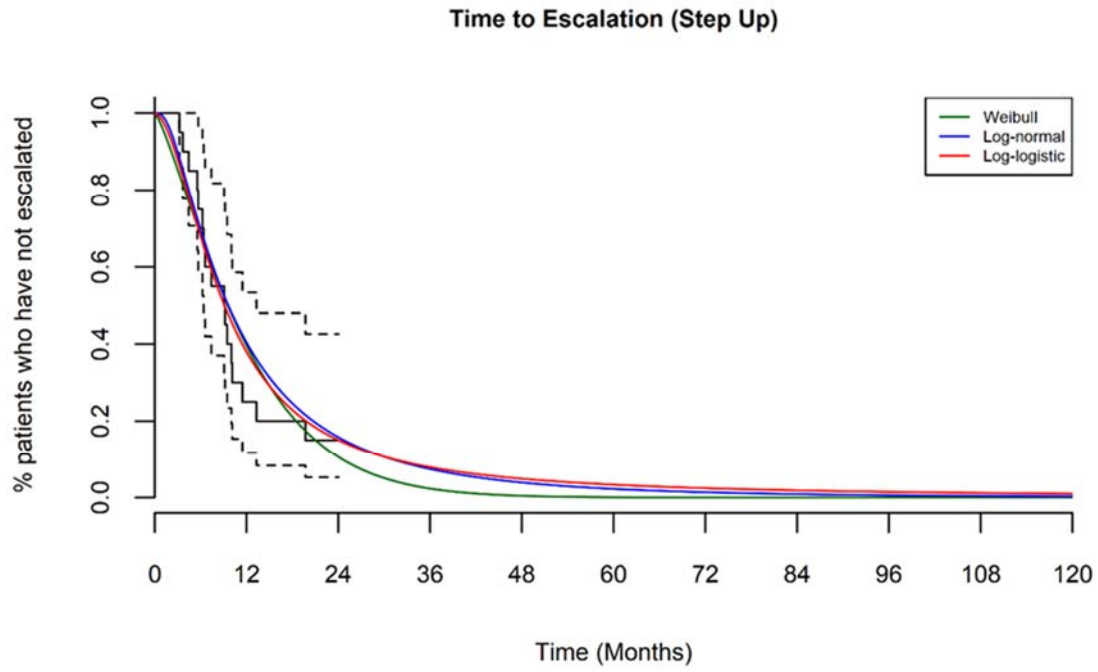


Figure 21. Top-down time to relapse curves fitted with Weibull, lognormal and log-logistic.

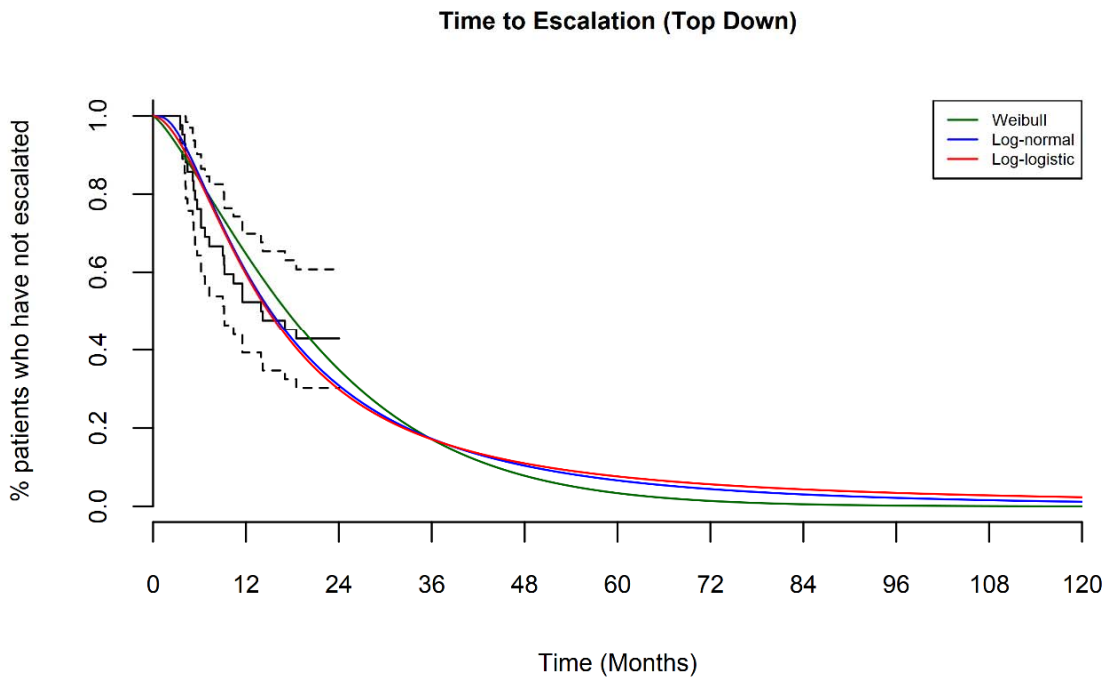


Table 14. Measure of fit statistics for time to relapse (truncated, dependant fit) in D'Haens *et al*.

	AIC	BIC
Exponential	305.93	310.18
Weibull	306.88	313.26
Gompertz	**301.24**	**307.62**
Lognormal	**301.18**	**307.56**
Loglogistic	303.08	**309.46**
Gamma	**301.15**	309.66
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion		

Figure 22. Time to relapse (truncated) estimated by the Evidence Assessment Group

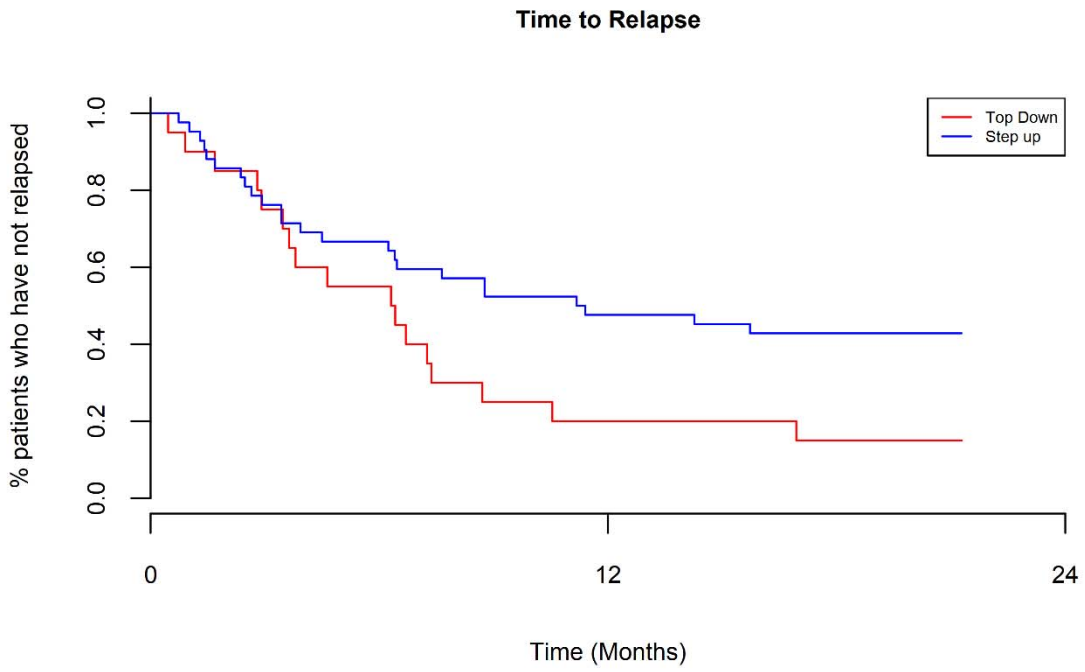


Figure 23. Step-up time to relapse curves fitted with Gompertz, lognormal and gamma (truncated data).

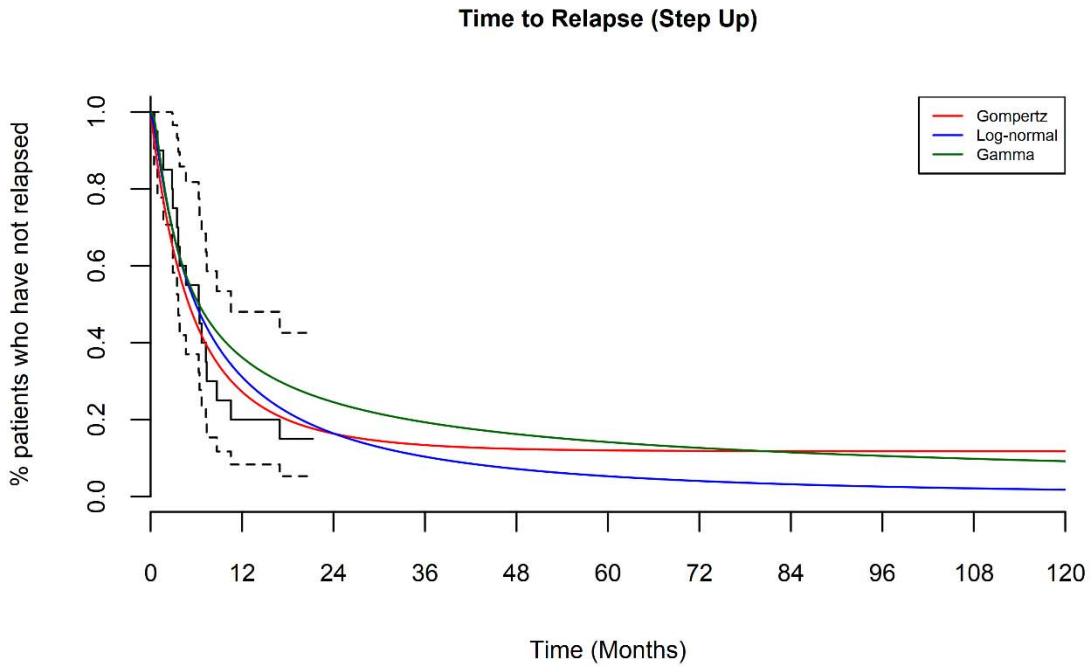


Figure 24. Top-down time to relapse curves fitted with Gompertz, lognormal and gamma (truncated data).

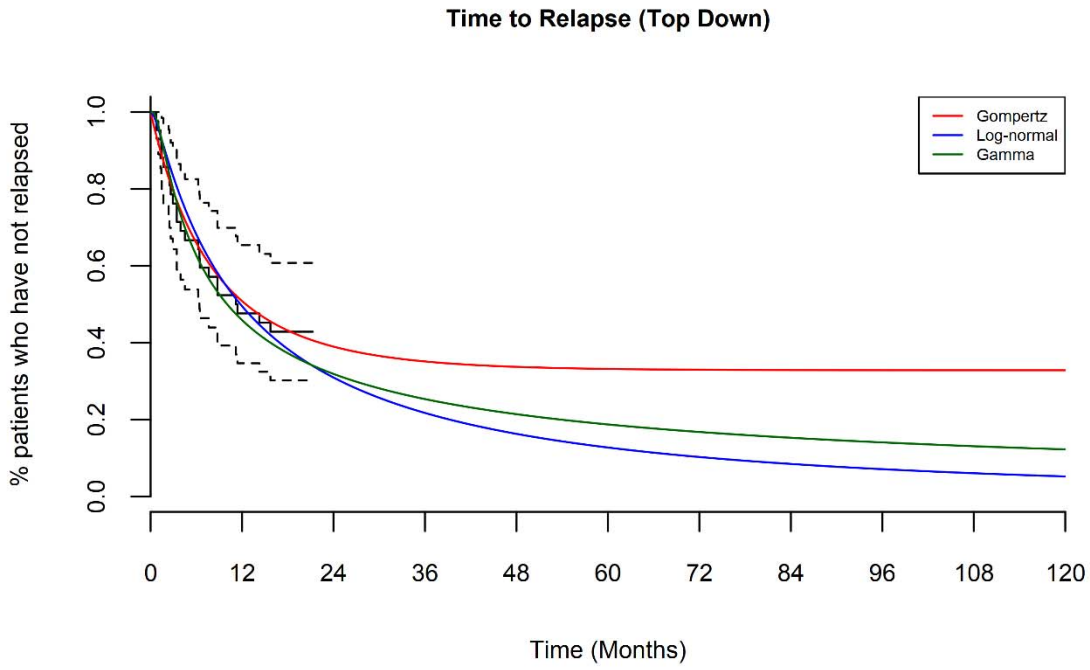
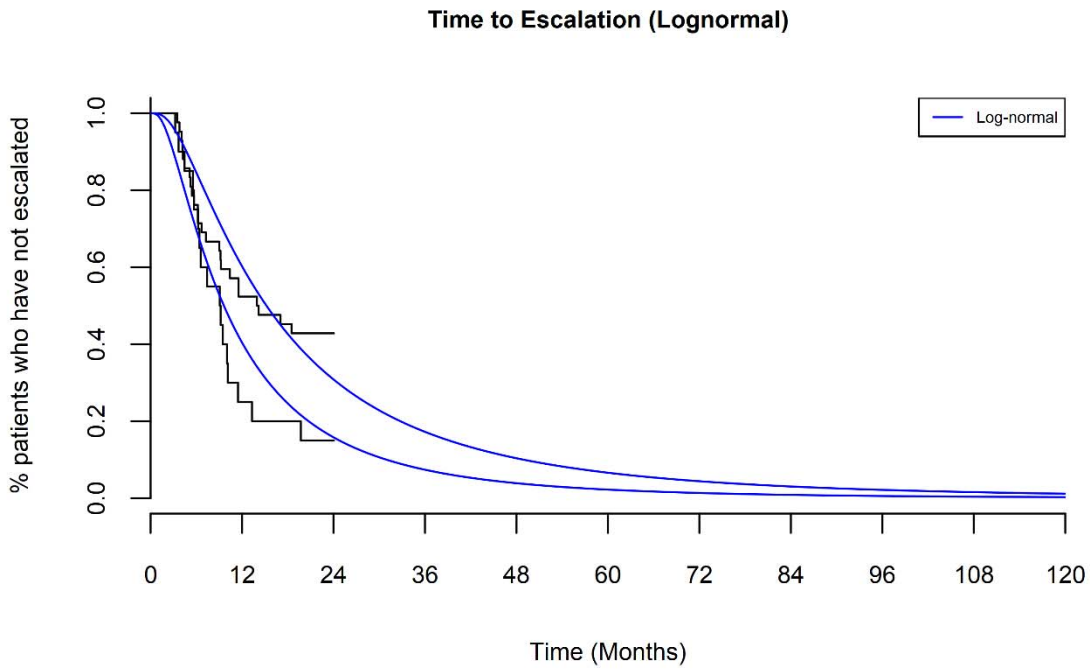


Figure 25. Time to relapse curves fitted with lognormal curves



The EAG used the lognormal fitted curves to estimate a hazard function to apply to the high-risk TD arm of the economic model. The EAG applied the relative hazard function to TTE curves in the first step in the TD strategy (anti-TNF). The TTE associated with the remaining treatment steps in both the TD and the SU arms were assumed to all be the same as TTE for anti-TNF in the TD arm (Figure 26).

The underlying assumption in the EAG's base case approach is that high-risk patients who initiate treatment with IMs (SU arm) escalate treatment quicker than high-risk patients who initiate treatment with anti-TNF (supported by the data presented in D'Haens *et al.*) [34, 35], however, once SU patients initiate treatment with anti-TNF (their second treatment step), they "catch-up" with patients on the TD treatment strategy.

As some high-risk patients who receive SU treatment respond to IM treatment (Section 4.2.4.3), having the additional IM step in the SU strategy is advantageous to patients in the EAG's base case analysis as patients still subsequently receive treatment with biologics, which are assumed to have the same benefit as biologics is the TD arm. Given the paucity of data to substantiate any further benefits in subsequent treatment steps in the TD vs SU approaches, the EAG considered this to be the most conservative modelling approach.

As mentioned in Section 4.2.2 and Section 4.2.4.1, the first treatment step modelled in the TD sequence is anti-TNF, while the first step in the SU strategy is IM treatment. Therefore, there is no modelling of escalation from corticosteroids nor is there any difference captured across TD and SU arms in time to corticosteroid failure and beginning of first treatment.

Assuming that all patients receive steroids but that only patients in the SU strategy would receive a full course of treatment rather than being switched to biologics in the TD strategy as soon as the test results become available was not modelled. Including this step in the model only for SU would add further benefits to the SU strategy as it would allow patients a further chance to respond, as well as reducing the chances of receiving the highly expensive biologic treatments (also considering the very low cost of corticosteroids).

The Specialist Committee Members (SCMs) raised a concern for the potential risk of additional complications associated with the SU strategy given the delay for initiating treatment with biologics. The EAG notes that Hoekman *et al.* concluded that in the long-term (10 year follow up) there was no difference found in complications, such as new fistulas or surgery, across the TD and SU arms. Furthermore, even though not based on comparative evidence, the Biasci *et al*. IPD reported, only very few events that required surgery, and no patients had more than one surgery within their follow up period while receiving a SU strategy.

Therefore, the EAG considers that the SCMs view that early biologics are better than later biologics may apply only to those who do not respond to treatment with IMs. However, removing this step entirely from the model, would mean taking away the benefit for those who do respond to IMs. As well as loosing this benefit, there would also be the addition of highly expensive biologics that are potentially unnecessary for those who would have responded well to IM.

Nonetheless, the EAG has varied these assumptions in a range of scenarios analyses described in Section 5.2. Regarding the measure of treatment effectiveness of TD vs SU in the model, the ERG ran three scenario analyses in the model:

1. High-risk patients on anti-TNF after IM (second step on SU arm) do not do as well as high-risk patients on first-line anti-TNF (first step on TD arm) and thus, the former escalate treatment quicker than the latter. Given that the EAG did not find any data to support this reduction in relative treatment effect, a theoretical estimate of half of the base case relative hazard was assumed (Figure 27);

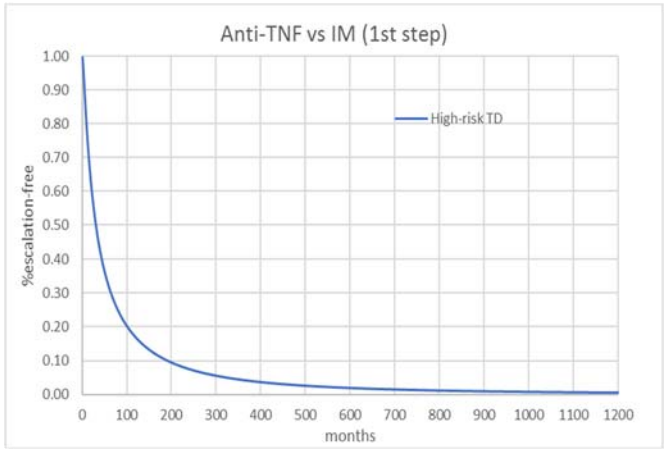
2. Combining scenario 1 with the base case approach, the EAG assumed that high-risk patients on TD only derive a benefit during the first step of the treatment strategy (anti-TNF in TD compared with SU patients on IM treatment), however, once patients have moved on to the second step in both strategies there is no relative benefit for TD vs SU. This scenario differs from the base case as the benefit assumed is the same as that used in scenario 1 (Figure 28);

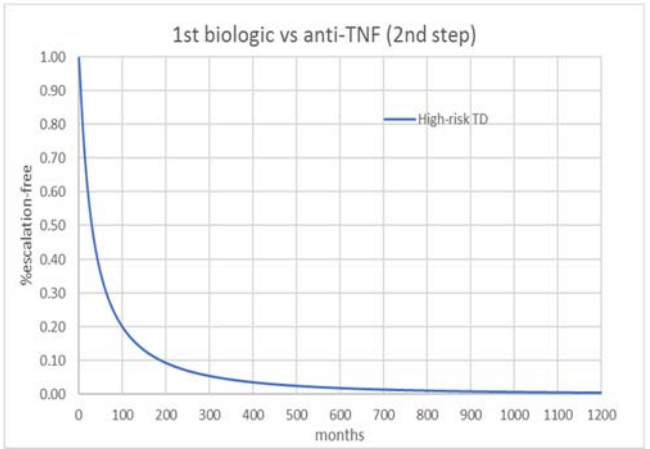
3. Assuming high-risk patients do not respond to treatment with IM, that is, 100% of patients who receive SU do not respond to treatment and therefore escalate to anti-TNF after induction with IMs.

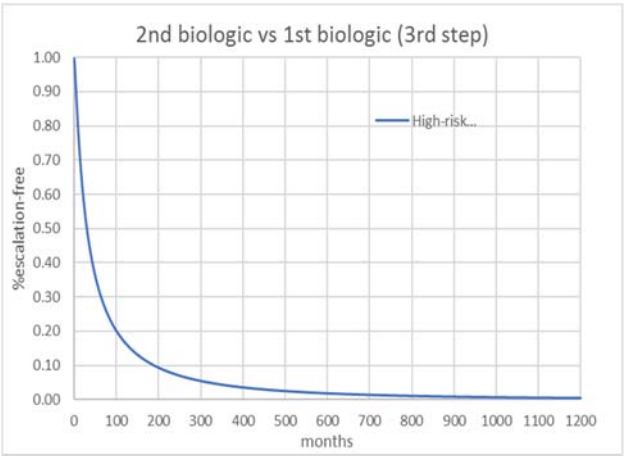
The three TTE curves for high-risk patients used in the base case and in both scenario analyses are reported in Figure 29. Results of these analyses are reposted in Section 5.

Figure 26. Comparison of time to treatment escalation curves across treatment strategies for high-risk patients (base case)

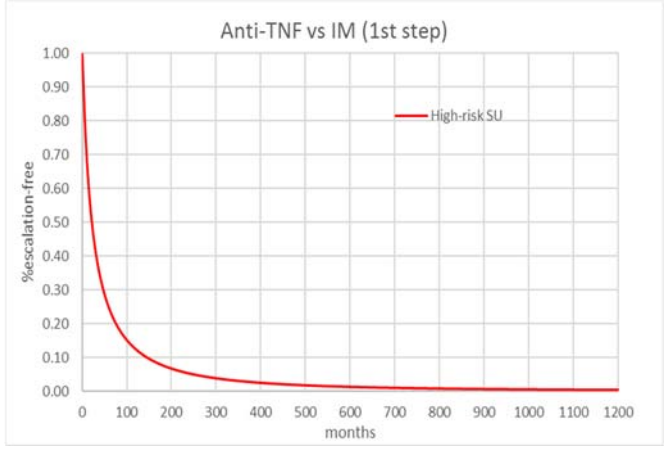
Top-down

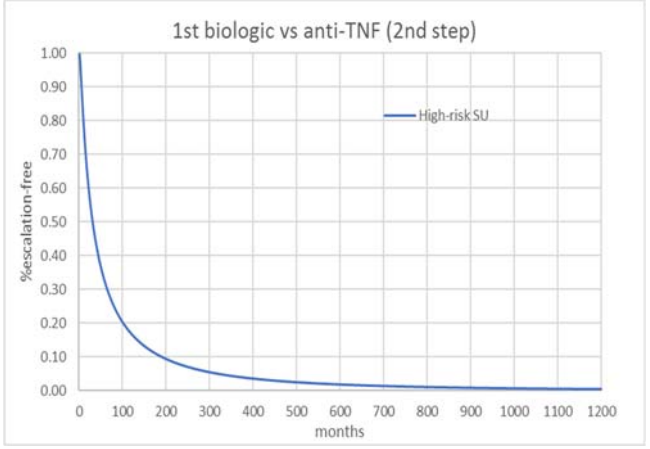






Step-up





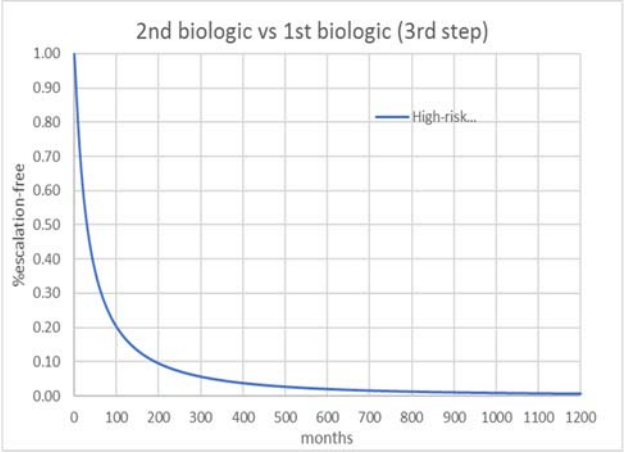
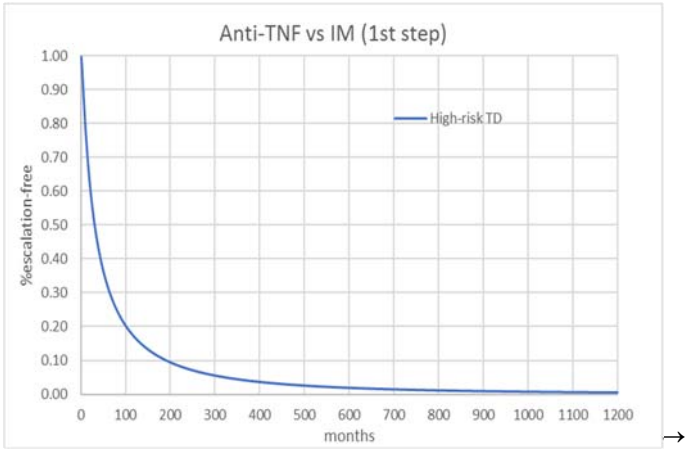
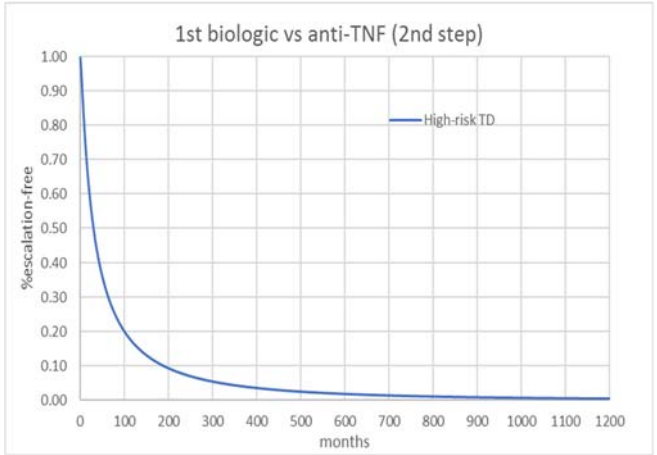
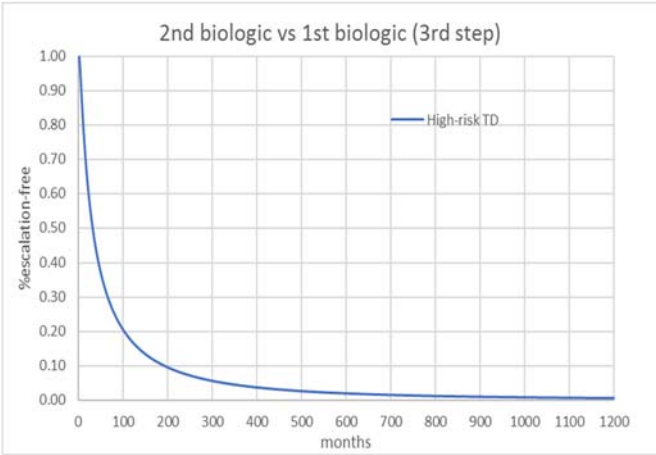


Figure 27. Comparison of time to treatment escalation curves across treatment strategies for high-risk patients (scenario analysis 1 with step-up time to escalation from step 2 estimated with half of the base case relative hazard)

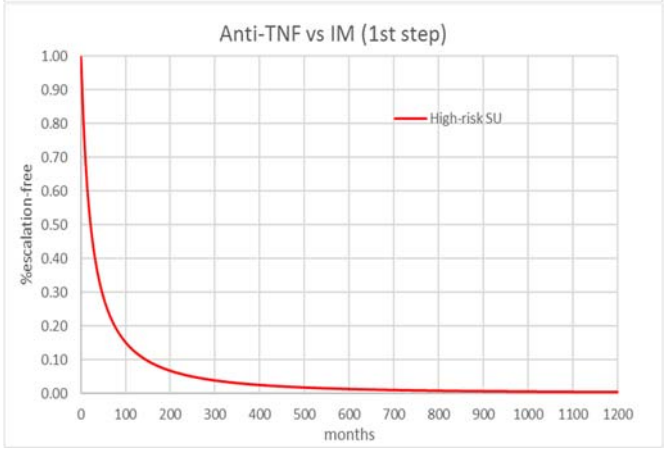
Top-down

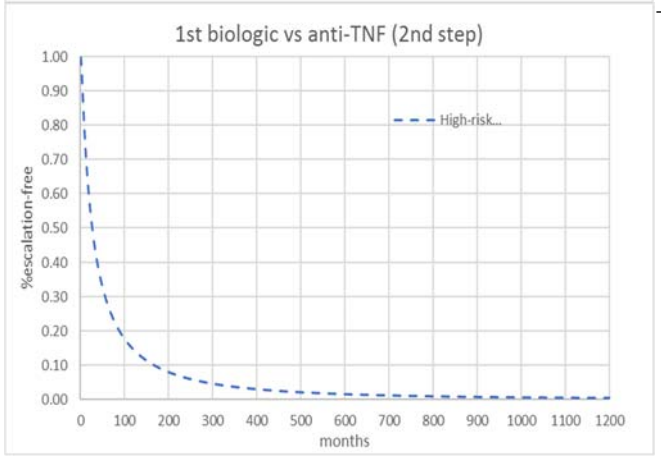






Step-up





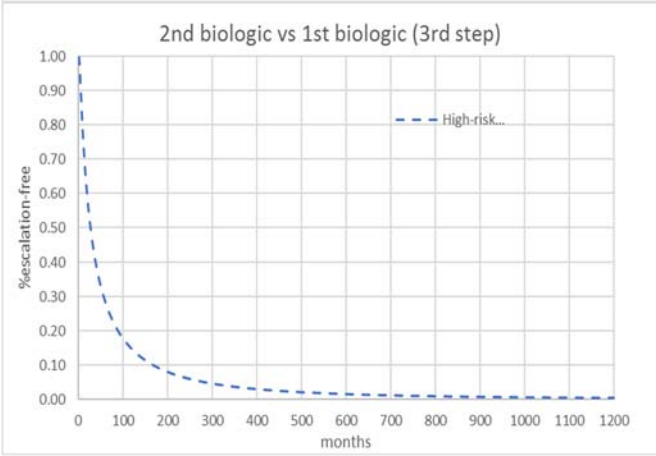
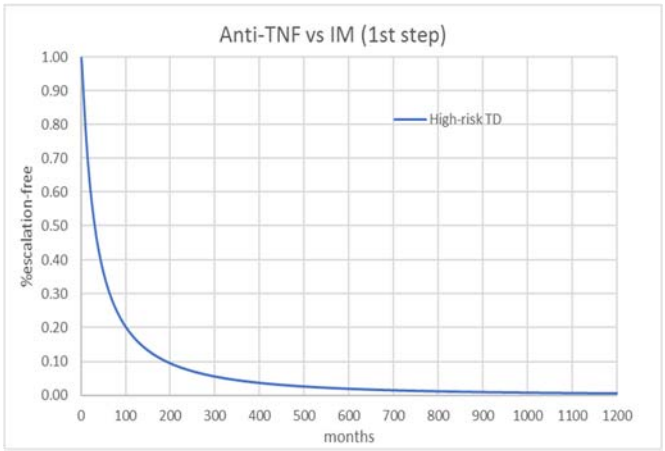
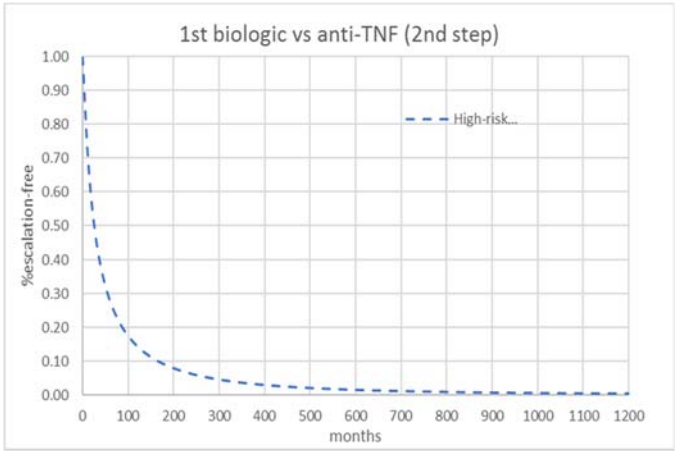
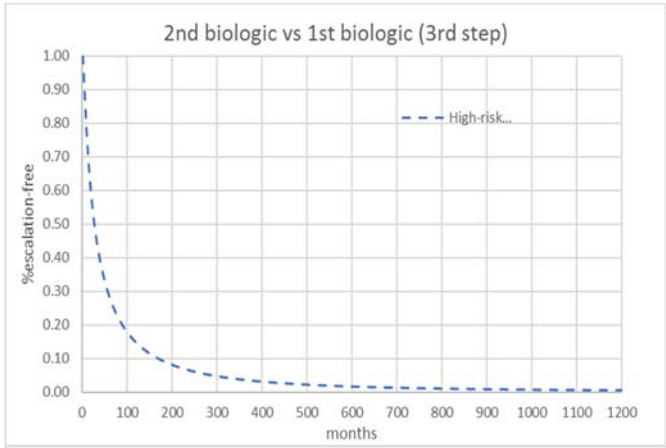


Figure 28. Comparison of time to treatment escalation curves across treatment strategies for high-risk patients (scenario analysis 2 with step-up and top-down time to escalation from step 2 estimated with half of the base case relative hazard)

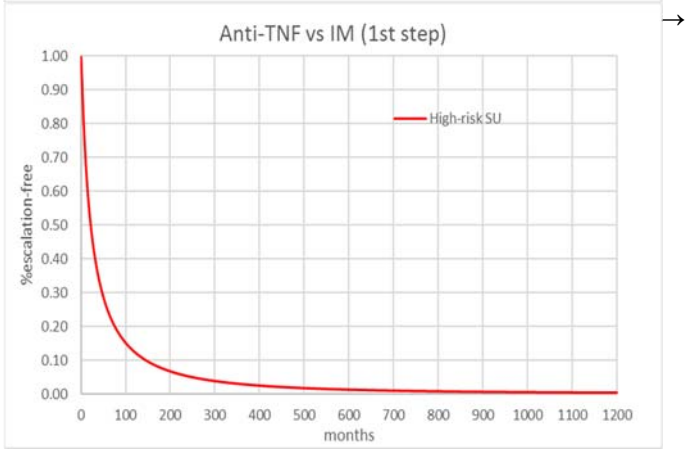
Top-down

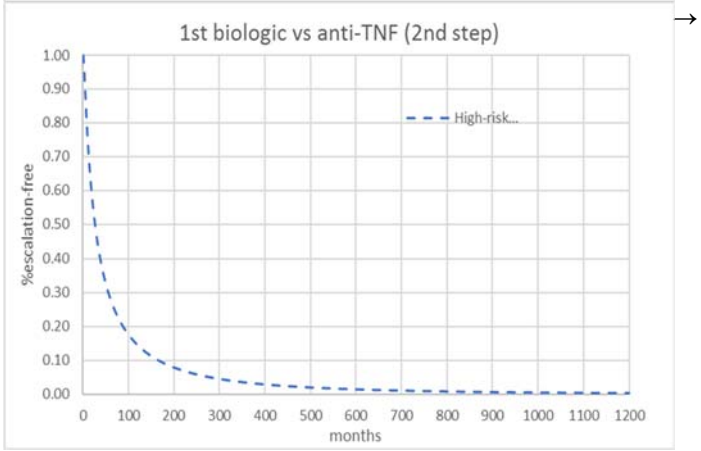






Step-up





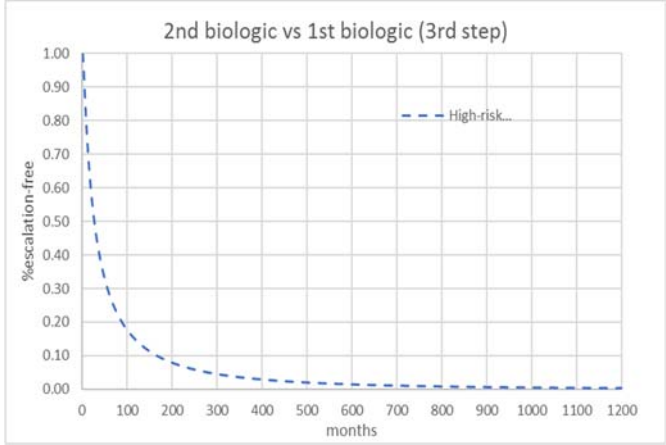
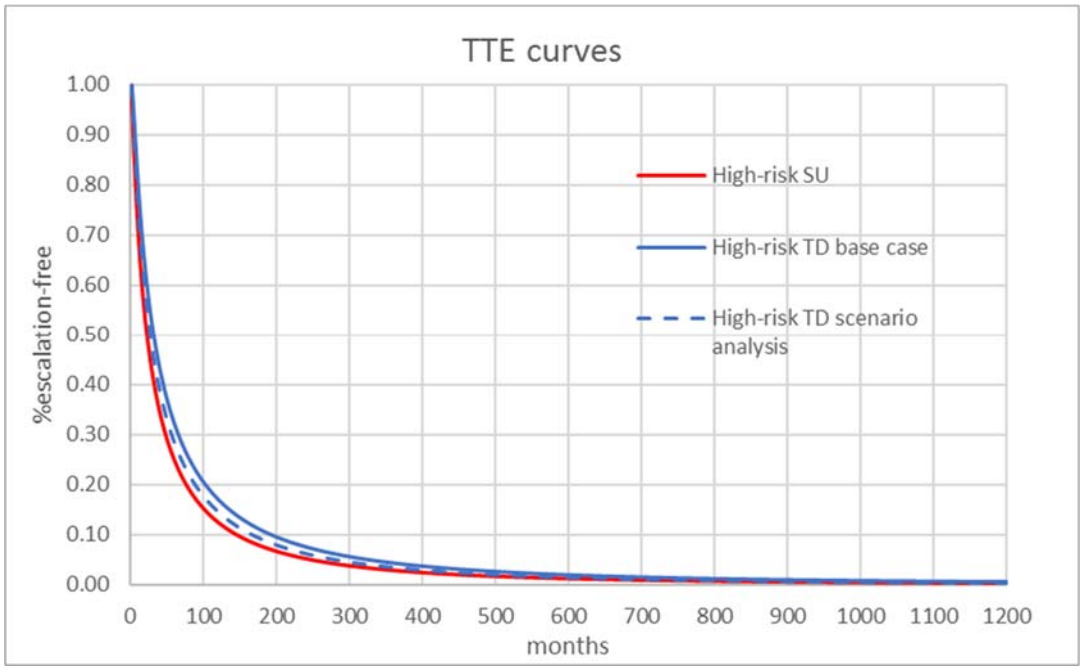


Figure 29. High-risk curves with top-down vs step-up effect



4.2.4.3 Effectiveness of induction and maintenance therapies

To estimate the effectiveness of the different therapies included in the modelled TD and SU strategies, the EAG sought evidence informing the probability of response and remission for the induction and maintenance periods of each treatment step in the respective sequences. The EAG also aimed to identify the proportions of patients expected to be in either a mild or a moderate to severe health state for those who experienced a response.

Initially, advice was sought from clinical experts to verify clinical practice in England relating to administration, scheduling and doses of the SU and TD strategies in the induction of remission of CD, and treatments given to maintain response or remission. Due to time and resource constraints, a pragmatic approach was taken to identify studies with data on clinical outcomes for people receiving induction and maintenance therapies for CD.

A search of electronic databases was carried out to identify systematic reviews of SU or TD treatments for CD. Electronic databases were searched from inception to 14 June 2019 and were:

- MEDLINE (MEDLINE and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily and Versions; Ovid);
- EMBASE (Ovid);

- the Cochrane Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR).

Search strategies for electronic databases included MeSH terms for CD and free text terms for CD and for SU and TD strategies: search strategies are provided in Appendix 3. The searches retrieved 507 records (post deduplication), which were imported into Rayyan QCRI for assessment of titles and abstracts by two independent reviewers. Review of titles and abstracts generated 15 studies for assessment of the full text publication: reasons for exclusion of the 14 studies are provided in Appendix 4. Two reviewers independently identified one SR as the most comprehensive review to be used as a source of studies on SU and TD treatments.[116] The full text publication of all studies listed in Tsui 2018[116] was assessed independently by two reviewers.

For inclusion of a study in the analysis of clinical effectiveness of induction and maintenance strategies, the study should have evaluated therapies at the dose and schedule, or similar, outlined in the licence of the drug for use in the management of CD in England. No IM has marketing authorisation for use in CD in England and, instead, doses reported in the BNF were applied for inclusion of studies.[117] For induction therapy, data should be reported for those with a new or recent diagnosis of CD and with moderate/severe activity of CD at baseline, as per the population of interest to the economic evaluation. For the SU treatment pathway, those moving on to receive second-line biological therapy should have failed treatment with first-line anti-TNF biologic as per NICE guidance.[30]

TA352[118] (vedolizumab) and TA456[119] (ustekinumab) were used as sources to identify studies evaluating clinical effectiveness of non-anti-TNF biological therapies, and also as a supplementary source on anti-TNF therapies as used in SU treatment. The full texts of all studies included in the NMAs presented in TA352 and TA456 were reviewed independently by two reviewers for potential relevance.

One RCT identified by the SLR was deemed relevant to the economic evaluation.[36] The RCT provided results on effectiveness of induction therapy with IM alone for SU treatment and on anti-TNF monotherapy for TD strategy.

Six additional studies included in TA352[118] and TA456[119] were considered to be relevant to inform estimates of clinical effectiveness of induction treatment: two RCTs reported results for anti-TNF biological therapy with or without IM in people naïve to anti-TNF;[120, 121] and four RCTs provided data on ustekinumab or vedolizumab with or without IM as a second-line biological therapy in people who failed treatment with an anti-TNF.[122-125]

Three studies from TA352[118] and TA456[119] informed on maintenance of response or remission for SU treatment: one RCT evaluated anti-TNF biologic with or without IM;[126] and two RCTs assessed ustekinumab or vedolizumab with or without IM.[122, 123]

The goal of the economic evaluation is to compare the cost-effectiveness of the SU and TD treatment pathways rather than determine which therapy within a class of treatments is the most effective at each step. Given the aim of the economic evaluation, and considering the available evidence, a class effect was assumed for each class of treatments (i.e., IM, anti-TNF ± IM, and second-line biologic ± IM) to simplify the complexity of the analyses. Clinical experts fed back that the assumption of a class effect was reasonable. Additionally, the EAG considered using the NMAs reported in TA352[118] (vedolizumab) and TA456[119] (ustekinumab) as potential sources of estimates of clinical effectiveness for anti-TNF and non-anti-TNF biological therapies in the economic model. However, after reviewing the underlying trials as described above, the EAG had concerns around the generalisability of the studies selected (see Appendix 8 for more details). Considering the NMAs reported in TA352 and TA456, given the EAG's assessments of the trials included, the EAG has reservations around the reliability of the results of NMAs for use in the economic model.

Data were extracted from the included studies by one reviewer and validated by a second. Substantial clinical heterogeneity was identified across the studies included for both induction and maintenance analyses given that studies:

- enrolled a mixture of people with a new or recent diagnosis of CD and those with an established diagnosis;
- evaluating treatment with non-anti-TNF biological therapies included people who had failed treatment with more than one anti-TNF (26% to 63% had failed more than one anti-TNF), which does not reflect clinical practice in England where patients not responding to treatment with anti-TNF biologic would move to a different class of biologic rather than receive a second anti-TNF;
- assessing maintenance treatment evaluated different doses and schedules.

Given the anticipated heterogeneity across the studies, a random-effects model was selected for synthesis of data. Data for each treatment bundle were synthesised using single-arm meta-analysis in Comprehensive Meta-Analysis V3.

The pragmatic search for evidence did not provide a complete set of data to be able to estimate transitions between health states for all treatment steps over time. Firstly, no studies provided the

proportional split of patients between the mild and moderate to severe states for those who achieved a response or those who maintained their response. Furthermore, the only treatment step that provided a complete set of response and remission probabilities for both induction and maintenance was the anti-TNF step for the SU pathway. A summary of the required parameter inputs for the model populated by the data extracted from the included studies, where available, is given in Table 15.

Table 15. Clinical outcomes of response and remissions (without levels of response)

Clinical outcomes	Induction		Maintenance	
	Response	Remission	Response	Remission
Top Down				
Biologics	-	-	-	-
Anti-TNF	-	66%	-	-
Step Up				
Biologics	30%	15%	-	28%
Anti-TNF	26%	37%	10%	33%
Immunomodulator	-	26%	-	-

Despite the limitations identified in the NMA in TA352, it proved to be the best available data set to complete the required response and remission outcomes for the economic model. The EAG identified complete data sources in Table 7.3.1.4 of the company's submission for TA352, which provided estimates based on NMAs for induction and maintenance and separated the outcomes by an anti-TNF naïve population and an anti-TNF failure population. The EAG considered it unreliable to combine different data sources for a particular class of treatment, thus the EAG retained only the SU anti-TNF data from its meta-analysis, which was the only complete set of data. The EAG also applied this to the TD anti-TNF but used TA352 data for biologics and immunomodulators.

For the missing SU data, immunomodulator outcomes were informed by the conventional therapy group for the anti-TNF naïve population and biologics were informed by vedolizumab from the anti-TNF naïve population; the latter also being used for TD biologics. For the second-line biologics, the same transitions as the first-line (non-anti-TNF) biologics were assumed to apply for both SU and TD

The combined set of outcomes that the EAG used to estimate transition probabilities is given in Table 16. Note that the response values were re-calculated to not include those in remission, as was the case in Table 7.3.1.4 of the company's submission in TA352.

Table 16. Probability of response (without levels of response) and remission (supplemented with TA352 data)

Clinical outcomes	Induction		Maintenance	
	Response	Remission	Response	Remission
Top Down				
Biologics	32%	13%	2%	28%
Anti-TNF	26%	37%	10%	33%
Step Up				
Biologics	32%	13%	2%	28%
Anti-TNF	26%	37%	10%	33%
Immunomodulator	23%	16%	15%	25%

The next step in estimating transition probabilities was to estimate the proportion of patients who were in the moderate/severe health state or in the mild health state after achieving a response and at the end of the maintenance phase. The EAG did not identify any data in the trials from the SLR, so instead used the values presented in TA352 as an estimate and assumed the same value for both induction and maintenance, given the lack of more robust data sources. The company from TA352 presented a value of 21.2% of patients being in the moderate/severe health state after response for the mixed population. The EAG did not consider there to be sufficient evidence to apply specific values for treatment naïve and treatment failure patients, so applied the value based on the combined patient population for all treatments.

The resulting induction and maintenance vectors for each treatment when this estimate of the mild and moderate/severe split is applied, are given in Table 17 and Table 18, respectively.

Table 17. Estimated induction vectors for step up and top down with levels of response

Clinical outcomes	Induction			
	Remission	Mild	Moderate/Severe	No response
Top Down				
Biologics	13%	25%	7%	55%
Anti-TNF	37%	20%	5%	38%
Step Up				
Biologics	13%	25%	7%	55%
Anti-TNF	37%	20%	5%	38%

Immunomodulator	16%	18%	5%	62%

Table 18. Estimated maintenance vectors for step up and top down with levels of response

Clinical outcomes	Maintenance			
	Remission	Mild	Moderate/Severe	No response
Top Down				
Biologics	28%	1%	0%	70%
Anti-TNF	33%	8%	2%	57%
Step Up				
Biologics	28%	1%	0%	70%
Anti-TNF	33%	8%	2%	57%
Immunomodulator	25%	12%	3%	60%

The economic model developed by the EAG applies transitions for those who are responding to treatment and deals with those who do not respond to treatment or lose response to treatment separately based on TTE data. Therefore, to estimate the transitions for responders (including remission), the data for the three responder states were taken and reweighted to sum to 100%. These data were then used to perform the estimation of transitions.

The *Optim* function from the *Stats* package in R was used to perform the estimation of transitions.[127] This was done in two stages: firstly, to optimise a 52-week transition matrix without constraints; then a second step to estimate 2-weekly transitions with constraints applied to prevent transitions progressing across two health states in one model cycle, e.g. transitions could go from remission to mild or mild to moderate/severe, but not remission straight to moderate/severe. This was based on clinical expert opinion that the latter would not happen in as short a period as two weeks.

The optimisation approach for both steps required an initial transition matrix to be defined with initial values, which were varied by the *Optim* function to minimise a specified objective function. For the first step, the objective function was defined as the sum of the squared difference between the product of the induction vector and the 52-week transition matrix, and the maintenance vector; and for the second step was the sum of the squared differences between the values of the estimated 52-week transition matrix and the 26^{th} power of the estimated 2-week transition matrix.

The initial matrix values applied in the optimisation can have an impact on the resulting transitions derived from the optimisation and some starting values provided poor estimations or even provided

negative probabilities. Therefore, the EAG varied these values until plausible values were generated that produced relatively accurate estimations of the maintenance vectors when the estimated transition matrices were applied to the induction vectors. The initial values were specified as the parameters of beta distributions that were linked to the transition matrix entries to ensure that values were between zero and one. The minimum values of the objective functions and the resulting predicted maintenance outputs are shown in Table 19 as a measure of goodness of fit.

Table 19. Objective function values after minimisation

Annual transitions	First step (annual probabilities)	First step (2-week probabilities)
Top down		
Anti-TNF	7.27e-11	0.0630
1st and 2nd line biologics	3.09e-05	0.0072
Step Up		
Immunomodulator	2.85e-10	0.0034
Anti-TNF	7.27e-11	0.0630
1st and 2nd line biologics	3.09e-05	0.0072

The resulting 2-weekly transition probabilities for each treatment are given in Table 20 and

Table 21 for TD and SU, respectively.

Table 20. Estimated 2-week transition probabilities for top down

Annual transitions	Remission	Mild	Moderate/Severe
Anti-TNF			
Remission	0.9787	0.0213	0.0000
Mild	0.1059	0.8941	0.0000
Moderate/Severe	0.0000	0.0346	0.9654
1st and 2nd line biologics			
Remission	0.9982	0.0018	0.0000
Mild	0.1136	0.8864	0.0001
Moderate/Severe	0.0000	0.0795	0.9205

Table 21. Estimated 2-week transition probabilities for step up

Annual transitions	Remission	Mild	Moderate/Severe
Immunomodulator			
Remission	0.9736	0.0264	0.0000
Mild	0.0616	0.9302	0.0082
Moderate/Severe	0.0000	0.0482	0.9518
Anti-TNF			
Remission	0.9787	0.0213	0.0000
Mild	0.1059	0.8941	0.0000
Moderate/Severe	0.0000	0.0346	0.9654
1st and 2nd line biologics			
Remission	0.9982	0.0018	0.0000
Mild	0.1136	0.8864	0.0001
Moderate/Severe	0.0000	0.0795	0.9205

The EAG also performed a scenario analysis that used only data from TA352 to inform the induction and maintenance vectors. The transition probabilities were re-estimated using this data and these data along with the induction vectors were applied in the model to test the impact on the results. The induction and maintenance vectors for the scenario are given in Table 22 and Table 23, respectively, and the updated transitions for TD and SU are given in Table 24 and Table 25, respectively. The results of the scenario analysis are presented in Section 5.2.

Table 22. Estimated induction vectors for step up and top down with levels of response (TA352 data)

Clinical outcomes	Induction			
	Remission	Mild	Moderate/Severe	No response
Top Down				
Biologics	13%	25%	7%	55%
Anti-TNF	32%	23%	6%	38%
Step Up				
Biologics	13%	25%	7%	55%
Anti-TNF	32%	23%	6%	38%
Immunomodulator	16%	18%	5%	62%

Table 23. Estimated maintenance vectors for step up and top down with levels of response (TA352 data)

Clinical outcomes	Maintenance			
	Remission	Mild	Moderate/Severe	No response
Top Down				
Biologics	28%	1%	0%	70%
Anti-TNF	48%	9%	3%	41%
Step Up				
Biologics	28%	1%	0%	70%
Anti-TNF	48%	9%	3%	41%
Immunomodulator	25%	12%	3%	60%

Table 24. Estimated 2-week transition probabilities for top down (TA352 data)

Annual transitions	Remission	Mild	Moderate/Severe
Anti-TNF			
Remission	0.9691	0.0309	0.0000
Mild	0.1665	0.8335	0.0000
Moderate/Severe	0.0000	0.0548	0.9452
1st and 2nd line biologics			
Remission	0.9982	0.0018	0.0000
Mild	0.1136	0.8864	0.0001
Moderate/Severe	0.0000	0.0795	0.9205

Table 25. Estimated 2-week transition probabilities for step up (TA352 data)

Annual transitions	Remission	Mild	Moderate/Severe
Immunomodulator			
Remission	0.9736	0.0264	0.0000
Mild	0.0616	0.9302	0.0082
Moderate/Severe	0.0000	0.0482	0.9518
Anti-TNF			
Remission	0.9691	0.0309	0.0000
Mild	0.1665	0.8335	0.0000
Moderate/Severe	0.0000	0.0548	0.9452
1st and 2nd line biologics			
Remission	0.9982	0.0018	0.0000
Mild	0.1136	0.8864	0.0001
Moderate/Severe	0.0000	0.0795	0.9205

4.2.4.4 Time to surgery in high- and low-risk patients

The goal of including surgical events in the model was to capture the impact of TD treatment on potentially reducing the need for surgery in high-risk patients. Clinical expert opinion provided to the EAG reflected that CD patients can receive surgery for multiple reasons, including exhausting treatment options or severity of disease (or symptoms) related with developing strictures or perforation of the bowel.

Conversely, the EAG acknowledges that surgery might have a beneficial impact on patients' quality of life as there is a disease "reset" for a period of time after surgery. Even though the EAG has not captured this potential benefit of surgery in the economic analysis, it notes that to do so would benefit the SU strategy, as a higher proportion of patients receive surgery in the SU arm than on the TD arm of the model.

The EAG analysed the IPD available for the 88 patients in the Biasci *et al*.[50] cohort for surgical events and removed one patient who had surgery at study entrance. There was a total of [redacted] surgeries in Biasci *et al*.[50] The EAG began by analysing the data separately by risk of disease complications ([redacted]), however, considered the data insufficiently mature to separate TTS by high- and low-risk groups and thus, pooled the TTS data across both study arms ([redacted]). The implication of this approach is that TTS is the same for high- and low-risk patients, which is unlikely to be an accurate

reflection of clinical reality. Nonetheless, the estimated treatment effect of TD vs SU was applied to the baseline population of high-risk patients on SU treatment, to allow the estimation of the incremental costs and benefits for high-risk patients receiving TD compared to high-risk patients receiving SU.

The limitation of this assumption is that it does not allow the estimation of the impact of misdiagnosis on TTS. However, there are presently no data to allow the estimation of the cost-effectiveness of misdiagnosing patients (as discussed throughout Section 5).

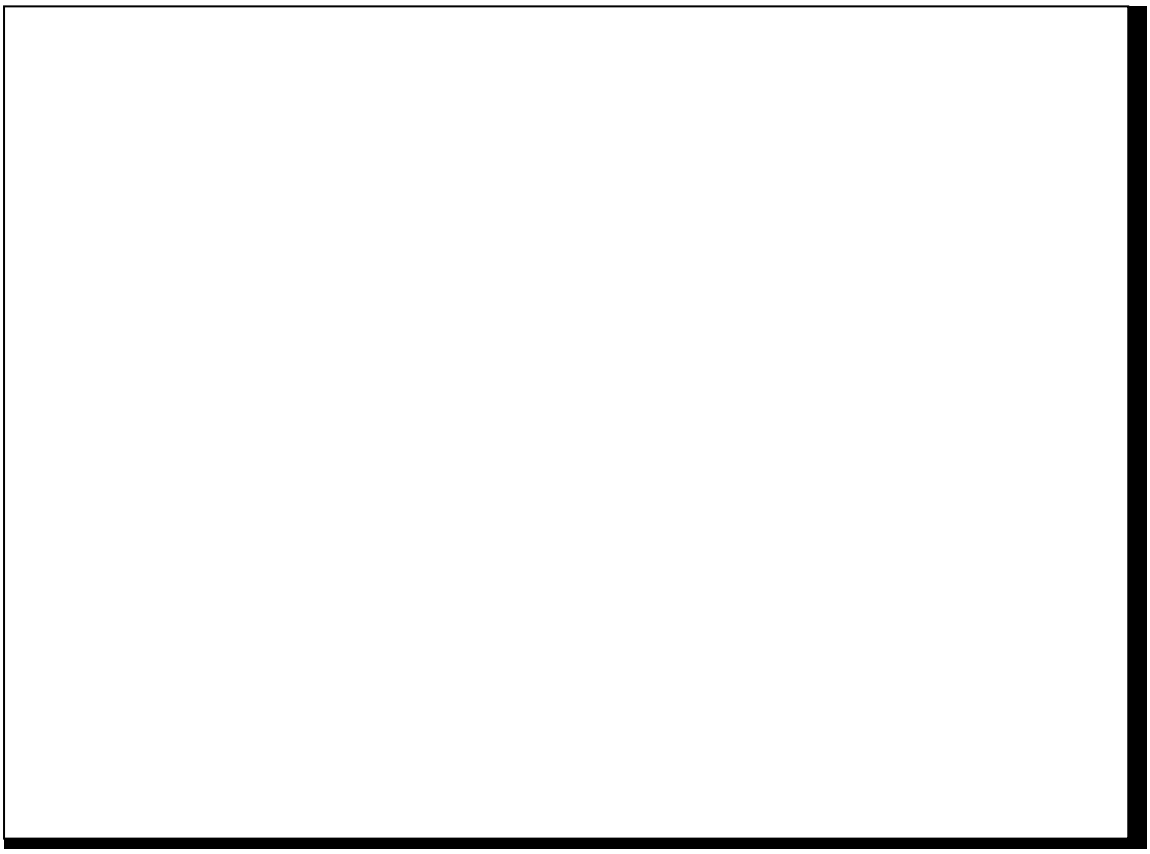
The SLRs of economic evaluations in CD did not produce any data to inform state-specific transition probabilities to/from surgery. Therefore, the EAG had to estimate TTS as a stand-alone outcome in the model. This modelling simplification means that patients do not explicitly leave their health state in a specific cycle to move to the surgery state. Instead, in every model cycle, a proportion of surgeries is estimated, and the associated costs and impact on patients' quality of life is calculated. To avoid double-counting issues, 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. In clinical practice, it is expected that patients might need to change treatment (or receive no treatment for a while) after surgical events, and furthermore, that surgery is dependent on patients' level of response to current treatment. However, the EAG could not find the data to reflect all the possible time-dependent transitions from the different health states in the model.

As a scenario analysis, the EAG allowed a proportion of patients to receive surgery as a final treatment step in the economic model. The results of this analysis are reported in Section 5.

[redacted]

[redacted]

[redacted]



[redacted]

In order to extrapolate TTS data into the model time horizon, the EAG fitted a variety of parametric curves to the KM data, reported in [redacted]. The pooled TTS data were fitted using the same process described in Section 4.2.4.1. Clinical experts were shown the fitted curves and informed the EAG that 50% of CD patients are expected to receive surgery during the first 10 years after initial diagnosis, while 25% of patients would receive surgery in the subsequent 5-year-period. [redacted]. Even though according to the AIC and BIC statistic reported in Table 26, the three best-fitting models are the Gompertz, lognormal and the gamma, given the [redacted], the EAG decided to use the exponential model in the base case analysis.

Table 26. AIC and BIC statistics for pooled data in Biasci *et al*.

	AIC	BIC
Exponential	220.04	**222.50**
Weibull	220.68	225.61
Gompertz	**216.82**	**221.76**
Lognormal	**217.81**	**222.74**
Loglogistic	219.72	224.65
Gamma	**216.25**	223.65
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion		

[redacted]

[redacted]

4.2.4.5 Effectiveness of top-down vs step-up treatment strategy on surgery

In order to estimate TTS in the high-risk TD strategy arm of the model, the EAG applied a hazard function taken from Hoekman *et al*.[112] The study concluded that time to surgery was not statistically significantly different across treatment arms. The authors discussed several potential explanations for

the lack of statistical differences across study outcomes. These included the reasons already discussed in Section 4.2.4.2 regarding the D'Haens *et al.*[35] trial, in addition to the following:

1. The authors mention the relatively early introduction of IM or infliximab in the treatment regime for patients receiving conventional management as a potential factor for underestimating the relative effectiveness of early immunosuppressant therapy (at the start of follow-up, 66% of SU patients had received an IM and 15% had received anti-TNF treatment, compared to 82% and 20% of TD patients, respectively).

   The EAG does not necessarily agree with this point, as the "early" introduction of anti-TNF or IM in the conventional treatment arm of the study could have been a reflection of the poor performance of conventional therapy and thus the need to escalate to anti-TNF treatment faster;

2. The authors also mention the study's potential lack of statistical power. Conversely, the authors also argue that observed statistically significant differences between groups merely reflect type 1 errors due to multiple testing (multiple testing correction was not applied in the study);

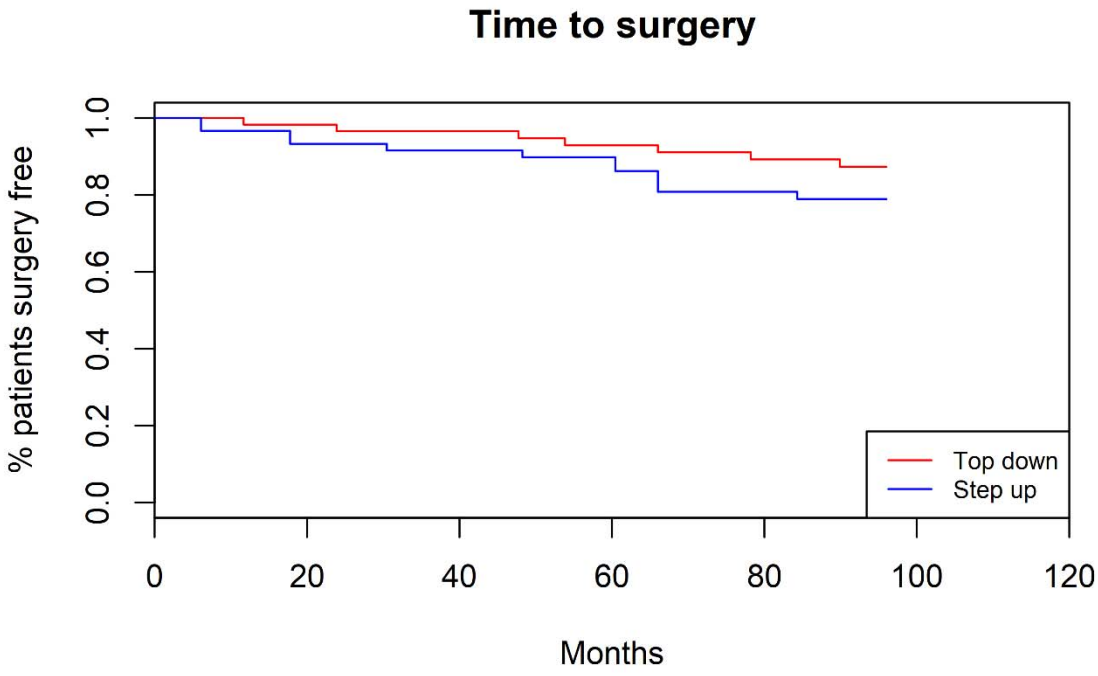
3. Finally, the study reports that the treatment received by patients beyond year 2 (end of the D'Haens *et al.*[35] trial and beginning of the follow-up study by Hoekman *et al.*[112]) was at the discretion of the treating physician. Consequently, patients' outcomes might have be influenced by different treatment strategies at the participating sites. The authors added that subjects in both arms of the trial were subsequently evenly distributed across the participating hospitals, and thus, in theory, equally exposed to the treating physician's preferences.

In conclusion, the EAG cannot be sure if early vs later immunosuppression therapy has an impact on TTS events, as the data demonstrate a non-statistically significant effect. However, given that there are also plausible reasons that could explain an underestimation of the effect (or a lack of statistical power to detect it), the EAG has applied the hazard function taken from Hoekman *et al.*[112] to the TTS in the high-risk TD arm of the model in its base case analysis. As an exploratory analysis, the EAG has assumed that TTS is the same in the TD and the SU arms for high-risk patients. Results of this scenario analysis are reported in Section 6.

The EAG digitised the TTS data KM data in Hoekman *et al.*[112] The study did not provide numbers at risk (except for the total number of patients entering the study). Therefore, the EAG had to manually reconstruct the numbers at risk, by visually analysing the KM data and estimating when (and how many) events happened over time. This task was simplified by the fact that there were no censored events in the TTS data. Subsequently, the EAG used the number of patients at risk to simulate the pseudo-

individual patient-level data using the Guyot *et al*. method and the algorithm in the *survHE R* package.[115] The EAG obtained the KM data (reported in Figure 33) and fitted survival models (dependently due to low number of events across the arms) using the same process as described in Section 4.2.4.1. The EAG notes that TTS was not statistically significantly different across the TD and SU arms in the EAG analysis (p-value = 0.2).

Figure 33. Time to surgery Kaplan-Meier data estimated by the Evidence Assessment Group from Hoekman *et al*.



Time	0	24	48	66	80	90
Immunosuppression	59	57	53	51	49	47
Conventional care	60	57	52	48	43	43

According to the AIC and BIC statistics reported in Table 27 the three best-fitting models are the exponential, lognormal and log-logistic. Figure 34 shows the fitted curves for SU patients along with the time to relapse KM data, while Figure 35 shows the same equivalent curves for TD patients. The EAG chose the exponential model given it was the best fitting model, and for the reasons discussed in Section 4.2.4.4. The EAG used the fitted curves to estimate a hazard function to then be applied to the TTS curve in the high-risk TD arm of the economic model (Figure 37).

Table 27. Measure of fit statistics for time to surgery (dependant fit) in Hoekman *et al*.

	AIC	BIC
Exponential	**278.12**	**283.67**
Weibull	279.84	288.18
Gompertz	280.00	288.34
Lognormal	**279.29**	**287.63**
Loglogistic	**279.72**	**288.05**
Gamma	281.24	292.36
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion		

Figure 34. Step-up time to surgery curves (with confidence intervals) fitted with lognormal, loglogistic and exponential models

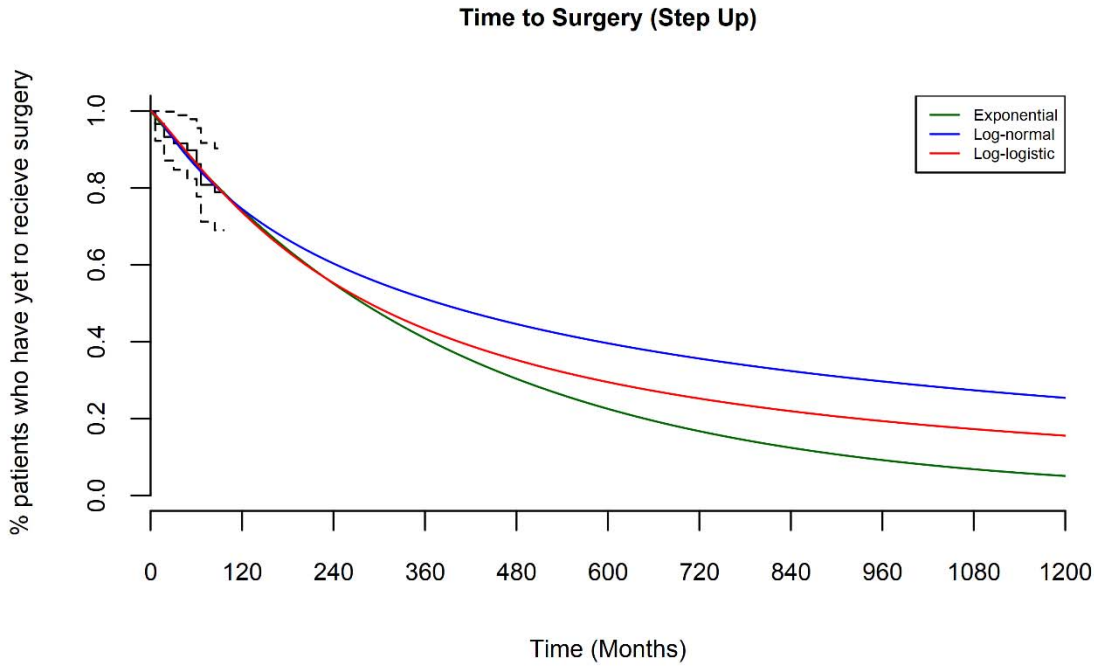


Figure 35. Top-down time to surgery curves (with confidence intervals) fitted with lognormal, loglogistic and exponential models

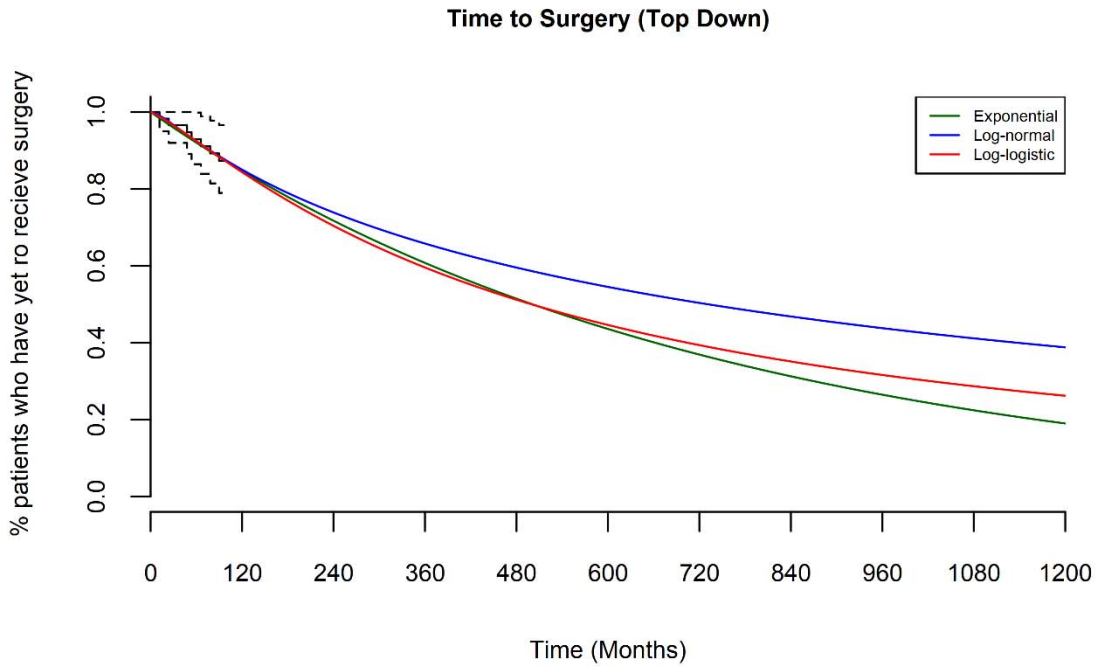


Figure 36. Time to surgery curves fitted with exponential curves

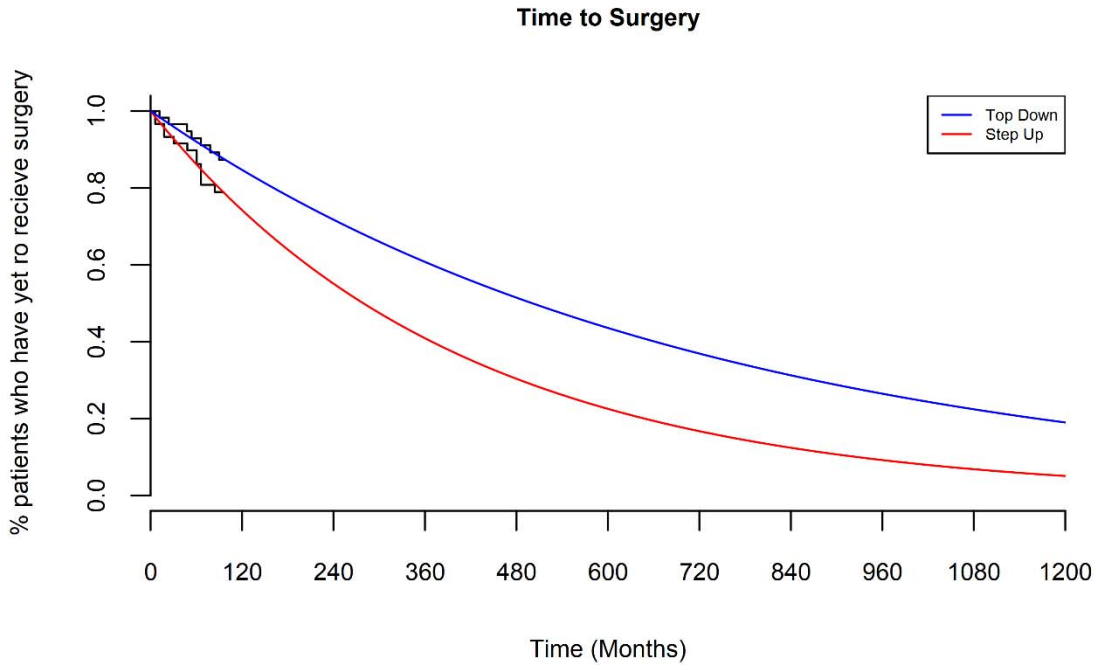
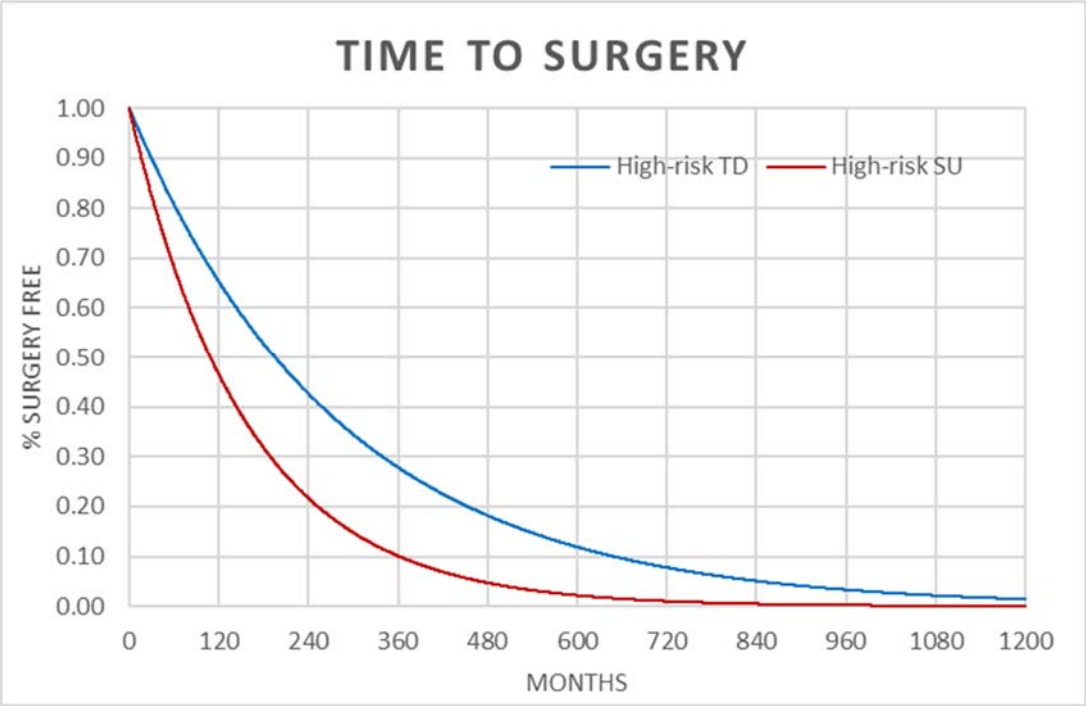


Figure 37. Time to surgery for high-risk patients with top-down vs step-up effect.



4.2.4.6 Mortality

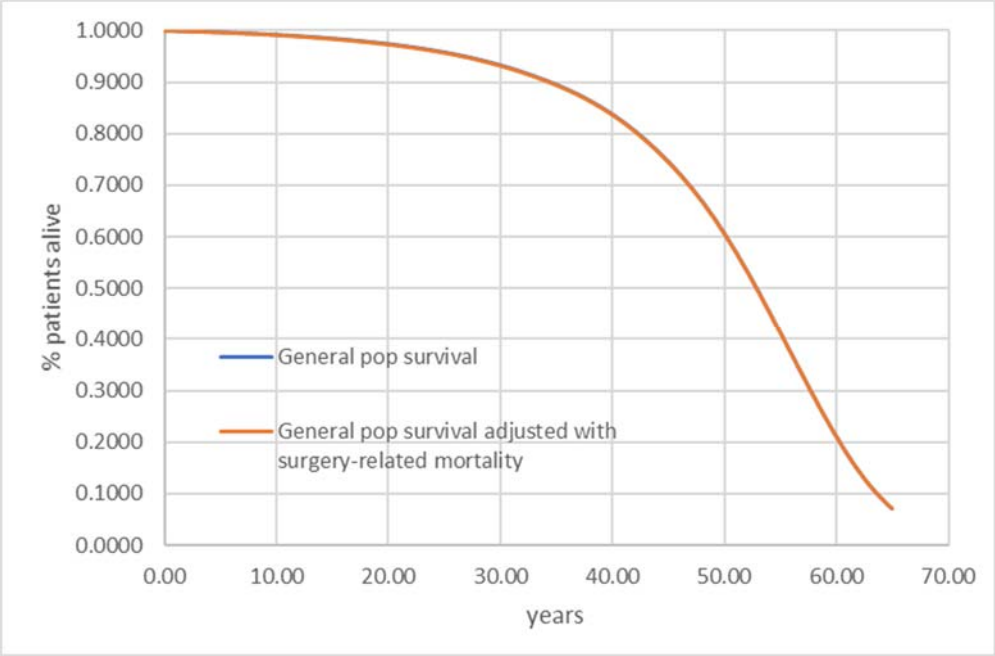
The EAG assumed that CD does not directly impact patients' mortality. Instead, background survival rates matched for gender and age were used to estimate patients' survival in the economic model (ONS Life tables for the UK).[128] The EAG assumed that surgery events were associated with a risk of death, hence, after every surgery in the model, patients have higher probability of dying compared to patients who do not undergo surgery.

In the company's model and in the Marchetti *et al.*[86] study, surgery-related mortality was derived from Silverstein *et al*[129]. (0.0015 increase in the probability of dying per month). The EAG acknowledges that the Silverstein *et al.* is an old study (1999)[129] and so surgery procedures and surgery-related death might have improved since then, however, the EAG did not identify more recent sources to populate this parameter in the model. The study is a 24-year follow-up of a population-based 'inception cohort' of 174 patients with CD in Olmstead County, USA, and provides data on the progress of patients from remission through mild and more severe disease states.

In summary, mortality in the model only differed (albeit very slightly) for high-risk TD vs high-risk SU patients through the difference in TTS outcomes for the two groups. Survival in the model is reported

in Figure 38 for both general population survival and general population adjusted with surgery-related mortality. The impact of the latter on the former is visually negligible, hence the curves overlap.

Figure 38. Survival in the model



4.2.4.7 Adverse events

The EAG decided not to include adverse events (AEs) in the economic analysis. The rational for this decision was twofold: the ERG in TA352 concluded that the exclusion of AEs associated with treatment with biologics (vedolizumab, infliximab and adalimumab) in the model did not have a relevant impact on the final ICER; the aim of the economic model is to asses cost-effectiveness of different treatment sequences for high-risk patients, and not to compare the cost-effectiveness of isolated treatments.

Furthermore, the EAG did not find any evidence on the impact of the long-term use of biologics in the TD vs SU approach on patients' quality of life. However, if AEs were included in the analysis, given that a higher proportion of patients receive biologic treatment in the TD arm, it would negatively impact the outcomes in the TD arm of the model.

4.2.5 Utility values

All utilities were adjusted to account for the age and sex of the modelled population, according to Ara and Brazier 2010.[130]

4.2.5.1 Remission, mild, and moderate to severe health states

The EAG used the two most recent NICE TAs on CD to inform the choice of utility values for the different CDAI states in the model (TA456 and TA456). While TA456 is more recent than TA352, the ERG for TA456 reported that it was, "[…] *unclear why the company did not make use of the utilities used in TA352 which were based on EQ-5D data from GEMINI studies;* […] *The estimated utility values in the GEMINI studies were elicited directly from the EQ5D using pooled data from the GEMINI II and GEMINI III studies and were estimated by health state regardless of study visit or treatment received*". The ERG concluded that the utility values derived from the GEMINI studies were, "*theoretically superior to the values estimated from the mapping algorithm because they are directly elicited. The utility values from GEMINI studies are, however, similar to those used in the company's base-case and therefore it is not expected to impact on estimated QALYs greatly*."

Therefore, the EAG used the utility values accepted in TA352 in the base case analysis and ran a scenario analysis using the TA456 utility values. Both sets of values are reported in Table 28 and results of the scenario analysis are reported in Section 6.

Table 28. Utility values used for remission, mild, moderate/severe health states

Health state	TA352	TA456
Remission	0.820	0.820
Mild disease	0.730	0.700
Moderate to severe	0.570	0.550

4.2.5.2 Surgery disutility

To capture the impact of surgery the EAG used the disutility values reported in Marchetti *et al*.[86] (described in Table 29). The estimates were based on assumptions made by the authors and were also used in the company's model. Marchetti *et al*.[86] assumed that patients undergoing surgery retained 0.5 of their utility estimate for 1 month.

Table 29. Disutilities associated with surgery and switching treatments in the model

Event	Marchetti *et al*. (assumptions)	Disutility estimate
Surgery	0.5 for 1 month	0.04

4.2.6 Costs

The following costs are considered in the model:

- Diagnostic test costs;

- Treatment costs;
- Acute and chronic care costs of CD (including costs of surgery).

All costs considered in the model are valued in 2019 UK pound sterling (£). Where unit costs have been obtained from the published literature before 2019, costs were uplifted using the ONS Consumer Price Inflation Index for Medical Services (DKC3).[131]

4.2.6.1 Diagnostic test costs

Table 30 reports the costs of PredictSURE IBD™ and IBDX®. In order to estimate the cost of IBDX® the EAG had to make some assumptions. The EAG took the midpoint cost in the range provided by the IBDX® company for the cost of the kit ([redacted] to [redacted] using HMRC's exchange rate of USD/GBP 1.2483) and then multiplied the cost by 6 (to reflect the 6 available kits) and divided by 45 (as the full set of tests need to be run twice). This resulted in the estimation of the cost of the test per patient. The EAG then multiplied the cost by [redacted] to account for lab tests and other miscellaneous costs (as suggested by the IBDX® company). Results of the cost-effectiveness analysis using the IBDX® costs are reported in Section 6.

Table 30. Cost of devices

Device name	Unit cost	Source
PredictSURE IBD™	£1,250	Company's reply to request for information
IBDX®	£347 (using HMRC exchange rate USD/GBP 1.2483)	Company's reply to request for information and EAG's assumptions

4.2.6.2 Treatment costs

The treatments included in the model are those described in the TD and SU strategies in Section 4.2.2. The different treatment costs are reported in Table 31. Treatment schedules and doses varied according to induction and maintenance stages (Table 32). As a modelling simplification, the EAG fixed the time on induction (and thus induction costs) by class of treatment in the model. For the base case analysis, the EAG looked at all treatments integrated in the treatment class – for example, for anti-TNF, the EAG looked at duration of induction treatment for adalimumab and infliximab – and chose the maximum induction period (4 weeks with infliximab) to estimate duration of induction with anti-TNF therapy in the model. As a scenario analysis, the EAG used the minimum induction period from the treatment class in the model (in the case of anti-TNF that would be 2 weeks as per the adalimumab schedule). Results of the scenario analysis are reported in Section 6.

The clinical experts advising the EAG consistently reported that treatment with anti-TNF and second-line biologics would be given as long as patients continued to show a response. Therefore, the base case analysis assumed that patients receive treatment with first- and second-line biologics until escalation to next treatment steps occurs. The EAG included two scenario analyses in the model to explore the uncertainty around this assumption and reported the results in Section 5:

1. Assuming that a proportion of patients in remission are cured and therefore stop treatment permanently;

2. Capping the duration of treatment with biologics in the model.

The EAG assumed that

Table 31. Treatment doses and costs

Treatment	Dose per unit (mg)	List price/unit	Source
Ustekinumab	130	£2,147.00	BNF
Vedolizumab	300	£2,050.00	BNF
Infliximab	100	£377,66	BNF
Adalimumab	40	£308.13	BNF (per syringe) based on Hulio (Mylan)
Azathioprine	50	£0.04	BNF (per tablet, 56 tablet pack)
6-mercaptopurine	50	£1.97	BNF (per tablet, 25 tablet pack)
Methotrexate	25	£16.64	BNF (pre-filled pen)
Methotrexate	15	£14.92	BNF (pre-filled pen)
Prednisolone	2.5	£0.04	BNF (per tablet, 28 tablet pack)
IV administration (outpatient)	1	First: £199 Follow-up: £212	NHS Ref. Costs 2017-18 - SB12Z Deliver Simple Parenteral Chemotherapy at First Attendance (outpatient); SB15Z Deliver Subsequent Elements of a Chemotherapy Cycle

Table 32. Induction and maintenance regimens

Treatment	Induction (mg per week, unless stated)	Maintenance (mg per week, unless stated)	Source
Ustekinumab	For body weight up to 56 Kg: 260 mg, 90 mg after 8 weeks For body weight 56 - 85 Kg: 390 mg, 90 mg after 8 weeks For body weight 86 Kg and above: 520 mg, 90 mg after 8 weeks	90 every 8 weeks	Clinical expert opinion
Vedolizumab	Initially 300 mg, then 300 mg after 2 weeks, followed by 300 mg after 4 weeks	300 every 8 weeks	Clinical expert opinion
Infliximab	Initially 5 mg/kg, then 5 mg/kg after 2 weeks, then 5 mg/kg after 4 weeks	5 per kg every 8 weeks	Clinical expert opinion
Adalimumab	Initially 160 mg, then 80 mg after 2 weeks	40 every two weeks	Clinical expert opinion
Azathioprine	2.5 mg/kg	2.5 mg/kg	Clinical expert opinion
6-mercaptopurine	1.25 mg/kg	1.25 mg/kg	Clinical expert opinion
Methotrexate	25	15	Clinical expert opinion
Prednisolone	40 and then taper by 5mg per week – 8 weeks	No maintenance with prednisolone	Clinical expert opinion

4.2.6.3 Acute and chronic care costs of CD

The EAG took the resource use reported in TA352 as a starting basis for discussion with clinical experts. After receiving input from clinical experts, the EAG combined the estimates on resource use by taking the middle point between estimates when clinical opinion was different, or the estimate provided by the experts when there were no discrepancies. The estimates used in the economic analysis are reported in Table 33. Unit costs were sourced from NHS reference costs (2017-2018).[132] A summary of the health care costs by health state is provided in Table 34.

Table 33. Health state costs

	Resource use/year (source: clinical expert opinion)				
	Remission	**Mild**	**Moderate to severe**	**Unit costs**	**Code**
Outpatient					
IBD consultant	0.5	0.75	2.0	First: £165 Follow-up: 132£	NHS Ref. Costs 2017-18. Currency code WF01A/B, service code:301, gastroenterology
Dietician	-	0.38	2.35	£81	NHS Ref. Costs 2017-18. Community Health Services; Currency Code A03 Dietician
Other IBD nurse	0.86	1.82	5.11	£77	NHS Ref. Costs 2017-18. Community Health Services Currency Code N29AF Other Specialist Nursing, Adult, Face to face
Helpline	0.59	1.52	6.09	£33	NHS Ref. Costs 2017-18. Community Health Services Currency Code N29AN Other Specialist Nursing, Adult, Non face to face
Pharmacist	-	0.17	0.63	£8	Assuming 10 minutes of a pharmacist time. Pharmacists cost per hour taken from PSSRU.

Nutritional support	-	-	0.5	£71	NHS Ref. Costs 2017-18. Outpatient attendances; Service Code 654 Dietetics (non-consultant led)
Radiology					
Plain X-ray	-	-	0.94	£30	NHS Ref. Costs 2017-18. DAPF, Direct access plain film
CT scan of abdomen/pelvis	-	-	1.16	£137	NHS Ref. Costs 2017-18. Outpatient, RD28Z, Complex CT scan
MRI scan of abdomen/pelvis	0.25	0.30	0.63	£301	NHS Ref. Costs 2017-18. Outpatient, RD03Z, Magnetic Resonance Imaging Scan requiring extensive patient repositioning
DEXA scan	0.31	0.31	0.31	£71	NHS Ref. Costs 2017-18. Outpatient, RD50Z, Dexa scan
MRI small bowel	-	-	0.5	£205	NHS Ref. Costs 2017-18. Outpatient, RD03Z, Magnetic Resonance Imaging Scan, one area, pre and post contrast
Endoscopies					
Oesphagogastroduodenoscopy	-	-	0.4	£299	NHS Ref. Costs 2017-18. Day case, FZ60Z Diagnostic Endoscopic Upper Gastrointestinal Tract Procedures, 19 years and over
Sigmoidoscopy	0.25	0.35	0.78	£319	NHS Ref. Costs 2017-18. Day case, FZ55Z Diagnostic Flexible Sigmoidoscopy, 19 years and over
Colonoscopy	0.2	0.3	1.23	£517	NHS Ref. Costs 2017-18. Day case, FZ52Z Diagnostic Colonoscopy with Biopsy, 19 years and over
Double balloon enteroscopy	-	-	0.08	£265	NHS Ref. Costs 2017-18. Endoscopies. Currency Code FZ13C Minor Therapeutic or

					Diagnostic, General Abdominal Procedures, 19 years and over
Wireless capsule endoscopy	-	-	0.15	£734	NHS Ref. Costs 2017-18. Endoscopies. Currency Code FZ42A Wireless Capsule Endoscopy, 19 years and over
Hospitalisations	-	-	0.6	£2,773	NHS Ref. Costs 2017-18. Non-elective inpatients (average length of stay 7 days). Currency Code FZ37P Inflammatory Bowel Disease without Interventions, with CC Score 5+

Table 34. 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 EAG matched the type of surgical procedures observed in the Biasci *et al.*[50] IPD to the Healthcare Resource Group (HRGs) 2017/2018 reference costs grouper. The EAG then used the HRG code to cost the specific procedure in the National schedule of reference costs (NHS costs). The resulting costs and average length of stay for the specific procedures underpinning the TTS data used in the model can be found in Table 35, along with the number of occurrences for each surgery observed in the Biasci *et al.* IPD.[50]

In order to estimate surgery costs in the model, the EAG applied a weighted average of the unit costs outlined in Table 35 using data on the number of occurrences of each type of surgery from the Biasci *et al.* IPD.[50] The weighted average cost was calculated by the EAG as £8,813 and this was assumed to apply to the proportion of patients who receive surgery in each model cycle based on the estimated TTS survival curves.

The EAG used the length of stay estimates for each procedure from NHS costs to determine how long surgical patients might be expected to temporarily discontinue pharmacological treatment (with either IM, anti-TNF or biologics) in the model. The weighted length of stay for surgery was estimated to be

12.17 days (Table 35). As this estimate is within a single model cycle of two weeks, the EAG assumed that patients would discontinue treatment for one full cycle when then receive surgery.

The EAG assumed that the risk of surgery was not dependent on the step in the treatment pathway. Therefore, the EAG estimated that the pharmacological treatment costs not incurred in each cycle for the proportion of patients who receive surgery were weighted equally across each of the treatment steps. This was estimated by multiplying the total per-cycle pharmacological treatment costs across all steps by the per-cycle proportion of patients receiving surgery, and then removing these costs from the total per-cycle costs.

Table 35. Costs of surgery

Procedure	HRG code	NHS reference cost description	Cost	Average length of stay (days)	Number of occurrences in Biasci *et al.* IPD
Right hemicolectomy	FF32	Proximal Colon Procedures, 19 years and over, with CC Score 6+	£9,225	10	3
Ileal resection	FF22	Major Small Intestine Procedures, 19 years and over, with CC Score 7+	£10,480	16	11
Defunctioning ileostomy	VA11	Multiple Trauma with Diagnosis Score <=23, with Intervention Score 1-8	£1,907	1	[redacted]
Perianal surgery (percutaneous drain)	FF41	Intermediate Anal Procedures, 19 years and over, with CC Score 3+	£2,469	2	[redacted]
Surgery (enterocutaneous fistula)	FF02	Major Therapeutic Endoscopic, Upper or Lower Gastrointestinal Tract Procedures,	£3,635	4	[redacted]

		19 years and over, with CC Score 3+			
Several perianal operations	FF33	Distal Colon Procedures, 19 years and over, with CC Score 3+	£7,675	6	[redacted]
Weighted average	-	-	**£8,813**	**12.17**	-

4.2.7 Summary of base case model inputs and assumptions

Table 36 reports the key model inputs used in the EAG's base case model and how these were varied in PSA. Table 37 summarises the key assumptions in the EAG's economic analysis, together with the rationale for these and a comparison with the modelling approach adopted in PredictImmune's model.

Table 36. Base case model inputs

Variable	Value/assumption in EAG model	Measurement of uncertainty/ distribution in EAG's model	Source
Model settings			
Time horizon (years)	65	Fixed	Assumption
Discount rate for costs and benefits	3.5%	Fixed	As per NICE guidelines
Days in a cycle	14.00	Fixed	Assumption
Patients' characteristics			
Age (years)	35	Gamma	Biasci *et al* IPD
Patients' weight (Kgs)			
Proportion of males	0.38	Beta	Biasci *et al* IPD
Probability of high-risk disease course	0.58	Beta	Biasci *et al* IPD
Diagnostic test accuracy			
Probability of PredictSURE IBD™ identifying high risk correctly	1.00	Beta	See Section 4
Probability of IBDX™ identifying low risk correctly	1.00	Beta	See Section 4
Treatment bundles			

Proportion on infliximab in anti-TNF biologics bundle	0.40	Beta	Clinical expert opinion
Proportion on adalimumab in anti-TNF biologics bundle	0.60	1- Proportion on infliximab in anti-TNF biologics bundle	Clinical expert opinion
Proportion on vedolizumab in non-anti-TNF biologics bundle	0.50	Beta	Clinical expert opinion
Proportion on ustekinumab in non-anti-TNF biologics bundle	0.50	1- Proportion on vedolizumab in non-anti-TNF biologics bundle	Clinical expert opinion
Proportion on azathioprine in IM bundle	0.80	Beta	Clinical expert opinion
Proportion of 6-mercaptopurine in IM bundle	0.10	Beta	Clinical expert opinion
Proportion of methotrexate in IM bundle	0.10	1- (Proportion of 6-mercaptopurine in IM bundle+ Proportion of methotrexate in IM bundle)	Clinical expert opinion
Proportion of patients receiving IM in anti-TNF bundle	0.30	Gamma	Clinical expert opinion
Proportion of patients receiving IM in non-anti-TNF biologic bundle	0.20	Gamma	Clinical expert opinion
Induction period			
Time spent in induction state with IMs (weeks)	8	Gamma	BNF/clinical expert opinion
Time spent in induction state with anti-TNF (weeks)	4	Gamma	BNF/clinical expert opinion
Time spent in induction state with biologics (weeks)	8	Gamma	BNF/clinical expert opinion
Mortality			
Probability of death following surgery	0.0015	Beta	Marchetti *et al*
Diagnostic test cost			
PredictSURE cost	£1,250	Fixed	Company's reply to request for information

IBDX cost	£347	Uniform	Company's reply to request for information and EAG's assumptions
Health state costs per cycle			
Remission	£17	Gamma	Clinical expert opinion
Mild	£27	Gamma	Clinical expert opinion
Moderate/severe	£122	Gamma	Clinical expert opinion
No response	£122	Gamma	Clinical expert opinion
Surgery	£8,813	Gamma	NHS Ref. Costs 2017-18
Treatment costs			
Induction - Anti TNF	£1,525	Gamma	BNF/clinical expert opinion
Induction - Biologic	£1,545	Gamma	BNF/clinical expert opinion
Induction - Immunomodulator	£4.43	Gamma	BNF/clinical expert opinion
Maintenance - Anti TNF	£536.46	Gamma	BNF/clinical expert opinion
Maintenance - Biologic	£656.47	Gamma	BNF/clinical expert opinion
Maintenance – Immunomodulator	£12.10	Gamma	BNF/clinical expert opinion
IV administration first attendance	£199	Gamma	NHS Ref. Costs 2017-18
IV administration follow-up	£212	Gamma	NHS Ref. Costs 2017-18
Utility			
Remission	0.82	Beta	TA352
Mild	0.73	Beta	TA352
Moderate/severe	0.57	Beta	TA352

No response	0.57	Beta	Assumption

Table 37. Key modelling assumptions

Description	Assumption	Justification	Comparison with PredictSURE model
Structural			
Relative treatment effect for TD vs SU for TTE	The EAG applied the relative hazard function to TTE curves in the first step in the TD strategy (anti-TNF) vs the first step in the SU strategy (IM). The TTE associated with the remaining treatment steps in both the TD and the SU arms were assumed to be the same as TTE for anti-TNF in the TD arm.	The only evidence available for the relative treatment effectiveness of TD vs SU (D'Haens) on time to relapse (and therefore treatment escalation) compares anti-TNF with corticosteroids. The EAG assumed this measure to be a proxy for the relative treatment effect of anti-TNF vs IM. However, the EAG considered that applying a relative treatment effect for TD vs SU across all treatment steps in the model was inappropriate. The underlying assumption in the EAG's base case approach is that high-risk patients who initiate treatment with IMs (SU arm) escalate treatment quicker than high-risk patients who initiate treatment with anti-TNF however, once SU patients initiate treatment with anti-TNF (their second treatment step), they "catch-up" with patients on the TD treatment strategy. Furthermore, Hoekman concluded that in the long-term (10-year follow up) a TD strategy had not proven to alter the natural history of CD's disease.	[redacted]

		Given the paucity of data to substantiate any further benefits in subsequent treatment steps in the TD vs SU approaches, the EAG considered this to be the most conservative modelling approach.	
Surgery modelling	Surgery was modelled with time to event data as a stand-alone health state, with no explicit transitions from/to any other states (except death) in the model.	The SLRs of economic evaluations in CD did not produce any data to inform state-specific transition probabilities to surgery. In every model cycle, a proportion of surgeries is estimated, and the associated costs and impact on patients' quality of life is calculated. To avoid double-counting issues, the EAG applied an adjustment to 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. In clinical practice, it is expected that patients might need to change treatment (or receive no treatment for a while) after surgical events, and furthermore, that surgery is dependent on patients' level of response to current treatment. However, the EAG could not find the data to reflect all the possible time-dependent transitions from the different health states in the model.	[redacted]
Relative treatment effect for TD vs SU for TTS	The EAG applied a relative hazard function to TTS curves.	The Hoekman data do not suggest that there is a statistically significant difference in TTS for TD vs SU therapy. However, given that there are also plausible reasons that could explain an underestimation of	[redacted]

		the effect (or a lack of statistical power to detect it), the EAG has applied the hazard function taken from Hoekman to the TTS in the high-risk TD arm of the model in their base case analysis and as an exploratory analysis, the EAG has assumed that TTS is the same in the TD and the SU arms for high-risk patients.	
Transition between disease severity stages while on maintenance treatment	Patients experience different levels of response to maintenance therapy throughout time.	As discussed in TA352, the DSU has reported the importance of capturing partial response to maintenance treatment (as well as remission, relapse, surgery and post-surgical remission) in CD's modelling approaches.[133] Therefore, the EAG based its model on the Bodger *et al*. structure in order to capture different levels of response.	[redacted]
Flares	The EAG assumed that treatment escalations in the model correspond to a relapse to patients' current treatment (or a severe flare)	The EAG assumed that the Biasci *et al.* TTE data captured flares leading to treatment escalation	[redacted]
TTE modelling			

TTE high-risk	Lognormal curve fitted to IPD KM from Biasci *et al*. (cohort of 40 patients as per Section 4.2.4.1)	The EAG aimed to use time to event data whenever available. Furthermore, the EAG restricted its analysis set to the 40 patients in the Biasci *et al.* IPD who had received treatments representative of the SU strategy in the UK NHS pathway. The EAG only used time to first escalation (IM to anti-TNF) from the Biasci *et al.* IPD as data on further escalations was deemed too incomplete and not robust for analysis.	[redacted]
TTE low-risk	Gompertz curve fitted to IPD KM from Biasci *et al.* (cohort of 40 patients as per Section 4.2.4.1)		[redacted]
TTE TD	Lognormal (dependant fit) curve fitted to IPD KM from D'Haens *et al.*		[redacted]
TTE SU			
TTE high-risk TD	Lognormal curve fitted to IPD KM from Biasci *et al* with hazard function from D'Haens		[redacted]

TTE high-risk SU	Same as TTE high-risk		[redacted]
TTE low-risk SU	Same as TTE low-risk		[redacted]
TTS modelling			
TTS high-risk	Exponential (pooled high- and low-risk curves) fitted to IPD KM from Biasci *et al*	The EAG aimed to use time to event data whenever available. Furthermore, the ERG is TA352 criticised the company in the same appraisal for modelling surgery as a constant probability in the economic analysis.	[redacted]
TTS low-risk			[redacted]
TTS TD	Exponential (dependant fit) fitted to IPD KM from Hoekman *et al.*		[redacted]
TTS SU			[redacted]
TTS high-risk TD	Exponential (pooled high- and low-risk curves) fitted to IPD KM from Biasci *et al* with hazard function from Hoekman *et al.*		[redacted]
TTS high-risk SU	Same as TTS high-risk		[redacted]
TTS low-risk SU	Same as TTS low-risk		[redacted]
Surgery costs			
The EAG assumed that patients would discontinue treatment when they receive surgery.	Patients stop treatment for 1 model cycle (14 days)	The EAG's approach aimed to avoid double-counting surgery and treatment costs. In clinical practice, it is expected that patients might need to change treatment (or receive no treatment for a while) after surgical events, and furthermore, that surgery is dependent on patients' level of response to current treatment. However, the EAG could not find the data to reflect all the possible time-dependent transitions from the different health states in the model. The weighted length of stay for surgery procedures observed in	[redacted]

		Biasci *et al.* was estimated to be 12.17 days. The EAG acknowledges that surgery might have a beneficial impact on patients' quality of life as there is a disease "reset" for a period of time after surgery, when patients might not need any treatment. Even though the EAG has not captured this potential benefit of surgery in the economic analysis, it notes that to do so would benefit the SU strategy, as a higher proportion of patients receive surgery in the SU arm than on the TD arm of the model.	
Biologic costs			
Treatment duration	Treatment given until escalation to next treatment step	The clinical experts advising the EAG consistently reported that treatment with anti-TNF and second-line biologics would be given as long as patients continued to show a response. Therefore, the base case analysis assumed that patients receive treatment with first- and second-line biologics until escalation to next treatment steps occurs. The EAG included two scenario analyses in the model to explore the uncertainty around this assumption and reported the results in Section 5: 1) Assuming that a proportion of patients in remission are cured and therefore stop treatment permanently; 2) Capping the duration of treatment with biologics in the model.	[redacted]

5 COST-EFFECTIVENESS RESULTS

5.1 Base-case deterministic and probabilistic results

Table 38 presents the deterministic base-case incremental cost-effectiveness ratio (ICER) for PredictSURE IBD™ compared with SC. The results show that the TD strategy (via the use of PredictSURE IBD™ in the model) is dominated by SU (via the SC arm of the model), with an additional cost of £9,526 and a QALY loss of 0.06.

Table 38. Base case deterministic cost effectiveness results (discounted)

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard of Care	£199,702	15.76	–	–	–
PredictSURE IBD™	£209,229	15.70	£9,526	-0.06	Dominated
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year.					

The EAG conducted a probabilistic sensitivity analysis (PSA) to assess the impact of the combined uncertainty from all parameters in the model. This was performed by sampling from distributions of the uncertain parameters 10,000 times, to generate the equivalent number of sampled ICERs. The methods for the inclusion of parameter uncertainty are discussed for each parameter type in turn.

There are many sources of uncertainty in the economic model and the key parameters that can have a meaningful impact on the results include the induction vector values to inform the initial cohort distribution across the health states, the transition probability estimates, and the time to escalation survival curves.

The induction vectors and each row of the transition matrices were varied using Dirichlet distributions to ensure that the rows summed to one. These were sampled in R using the Dirichlet function of the MCMCpack[134] package to generate 10,000 samples, which were copied into the economic model and sampled consecutively for each iteration of the PSA.

Each time-to-escalation curve applied in the model was sampled in a similar way by deriving 10,000 samples of each curve, using the *vcov* function of the stats package to estimate covariance matrices for the parameters, which were then used along with the mean parameter estimates in the *mvrnorm* function of the MASS[135] package to generate 10,000 correlated samples for each parameters, which were subsequently used to generate 10,000 survival curves.

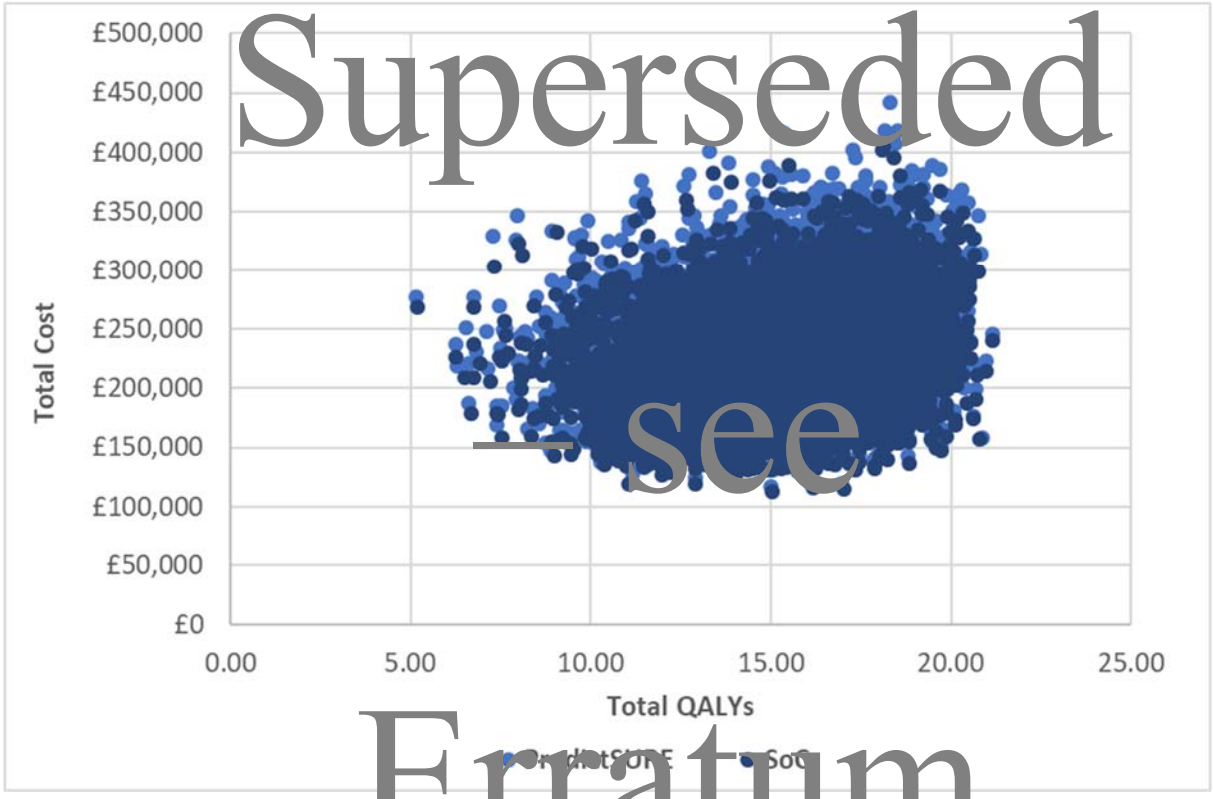
For cost estimates, gamma distributions were applied using 20% of the mean value to estimate standard errors, while for probabilities and utilities, beta distributions were applied; again, with an assumption that the standard errors are 20% of the mean estimate. A summary of the full parameterisations of these estimates varied in the PSA are given in Table 36 and the probabilistic ICER is reported in Table 39. Figure 39 reports the scatterplot showing the spread of results from the individual samples. The incremental costs and QALYs relative to SC are shown in the cost-effectiveness plane in Figure 40, while the cost-effectiveness acceptability curves (CEACs) showing the probability of PredictSURE IBD™ being cost-effective against SC over a range of willingness to pay thresholds, are given in Figure 41.

The probabilistic ICER is dominated against PredictSURE IBD™ and the CEACs show that the diagnostic test has a 0% probability of being cost-effective against SC at the £20,000 – £30,000 ICER threshold used by the National Institute for Health and Care Excellence (NICE).[136] The EAG varied the willingness to pay threshold to assess when the CEACs would begin to converge and at a threshold of £500,000 per QALY gained, the probability of PredictSURE IBD™ being cost-effective was 20% against 80% for the SC arm.

Table 39. Base case probabilistic cost effectiveness results (discounted)

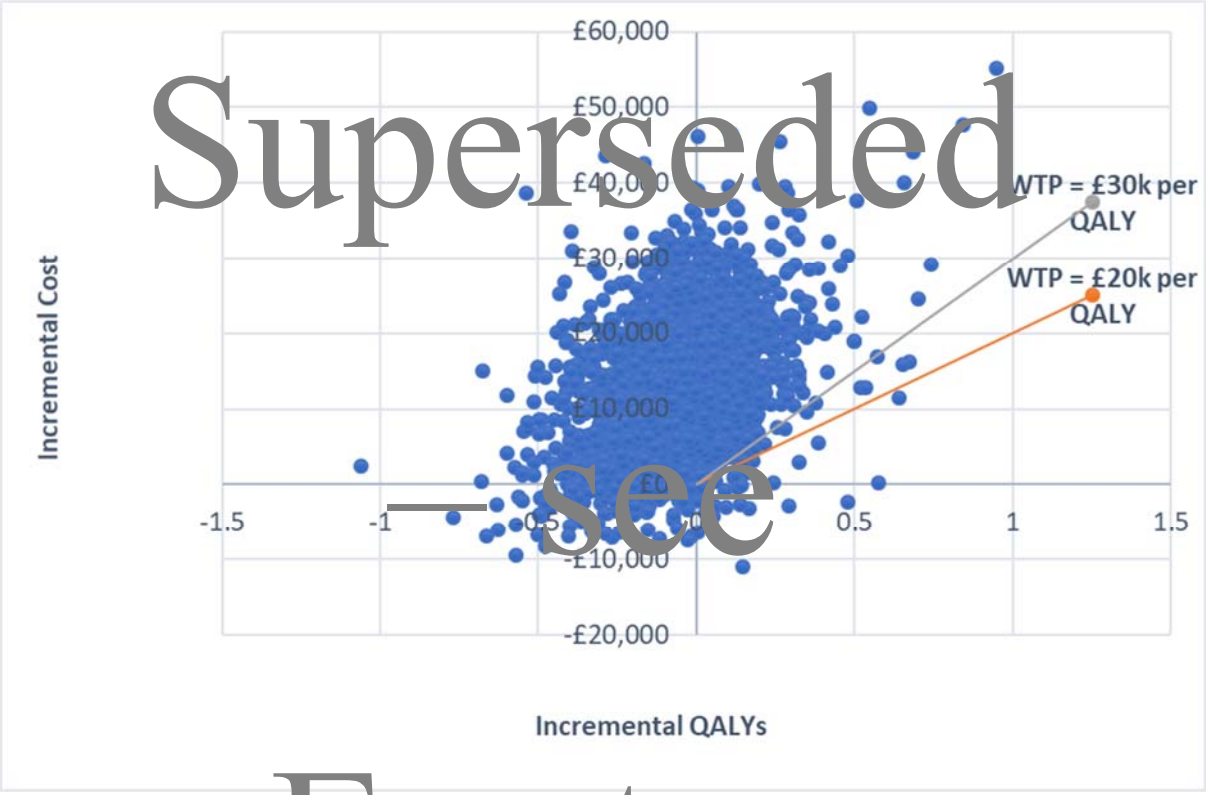
Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard of Care	£225,928	15.65	–	–	–
PredictSURE IBD™	£237,002	15.61	£11,073	-0.03	Dominated
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year.					

Figure 39. Scatterplot of the 10,000 PSA samples of costs and QALYs



Abbreviations in figure: SoC, standard of care.

Figure 40. Cost-effectiveness plane



Abbreviations in figure: WTP, willingness to pay.

Figure 41. Cost-effectiveness acceptability curve

Abbreviations in figure: SoC, standard of care.

5.2 Scenario analyses

The EAG conducted scenario analyses to assess the potential impact of the uncertainty around some of the assumptions made in the model. Results are reported in Table 40.

1. The EAG ran the economic model using the IBDX® cost (reported in Section 4.2.6). The EAG notes that the clinical input parameters in the base case economic model for PredictSURE IBD™ and in the scenario analysis for IBDX® are the same;

2. The EAG used the utility values in TA456 in a scenario analysis;

3. The EAG applied the induction vectors and transition probabilities based on TA352 studies;

4. As an exploratory analysis, the EAG assumed that TTS is the same in the TD and the SU arms for high-risk patients;

5. The EAG removed the age and sex utility adjustments from the economic analysis;

6. As a scenario analysis, the EAG used the minimum induction period from the treatment class in the model to estimate induction costs;

7. The EAG assumed that 100% of high-risk patients who receive SU do not respond to treatment and therefore escalate to anti-TNF after induction with IMs.

All of the scenario analyses undertaken produced dominated ICERs against PredictSURE-IBD™ compared to SC. The only exception was scenario 7, where the EAG assumed that 100% of high-risk patients who receive SU therapy do not respond to IMs (therefore not deriving any benefit from response to this treatment). The ICER for PredictSURE-IBD™ compared to SU changed from dominated (against the diagnostic tool) to £71,294. To note is that the EAG tested the impact of varying the proportion of patients who do not respond to IM treatment in the analysis. When the EAG assumed that 92% of high-risk patients who receive SU therapy do not respond to IMs (therefore not deriving any benefit from response to this treatment), the two strategies (TD and SU) became clinically equivalent.

Table 40. Results of scenario analyses

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Scenario 1: Applying IBDX cost					
Standard of Care	£199,702	15.76	–	–	–
IBDX	£208,326	15.70	£8,623	-0.06	Dominated
Scenario 2: Applying utilities from TA356					
Standard of Care	£199,702	15.47	–	–	–
PredictSURE IBD™	£209,229	15.41	£9,526	-0.06	Dominated
Scenario 3: Applying induction vectors and transition probabilities based on TA352 studies					
Standard of Care	£199,660	15.77	–	–	–
PredictSURE IBD™	£209,208	15.70	£9,548	-0.07	Dominated
Scenario 4: Applying equivalent TTS curves for top down and step up					
Standard of Care	£199,702	15.76	–	–	–
PredictSURE IBD™	£209,796	15.70	£10,093	-0.06	Dominated
Scenario 5: Removing Ara & Brazier utility adjustment					

Standard of Care	£199,702	15.82	–	–	–
PredictSURE IBD™	£209,229	15.75	£9,526	-0.06	Dominated
Scenario 6: Use the minimum induction period from the treatment class to estimate induction costs					
Standard of Care	£192,824	15.73	–	–	–
PredictSURE IBD™	£202,249	15.68	£9,424	-0.05	Dominated
Scenario 7: 100% of high-risk patients who receive SU do not respond to IM treatment					
Standard of Care	£208,077	15.68	–	–	–
PredictSURE IBD™	£209,229	15.70	£1,151	0.02	£71,294
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year; TTS, time-to-surgery.					

Table 41 presents the fully incremental analysis of cost-effectiveness results and demonstrates that out of the diagnostic tools under consideration PredictSURE IBD™ is dominated by IBDX® and both tools are dominated by standard care. However, as discussed throughout the report, despite extensive systematic searches of the literature, no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools and the EAG considers it would be challenging to ascertain an accurate estimate of prognostic accuracy of the tools in stratifying course of CD. Therefore, the only difference in the analysis of cost-effectiveness for the two diagnostic tools is the cost of tests.

Table 41. Base case fully incremental cost effectiveness results (discounted)

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard of Care	£199,702	15.76	–	–	–
IBDX®	£208,326	15.70	£8,623	-0.06	Dominated
PredictSURE IBD™	£209,229	15.70	£903	0	Dominated
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year.					

The EAG also ran a scenario analysis to include price discounts to the cost of anti-TNF and second-line biologic treatments in the analysis. The discounts were applied to the treatment class and a range of discounts was considered: 25%; 50% and 75%. The results of the analysis are reported in Table 42, showing that PredictSURE IBD™ remains dominated by standard of care in all scenarios. Although the increase in the discount of the drugs results in a decreased incremental cost overall, it is not enough to cause the PredictSURE IBD™ group total costs to be lower than the standard of care total costs.

Table 42. Drug price discount scenarios

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Biologic discount: 25%					
Standard of Care	£181,644	15.76	–	–	–
PredictSURE IBD™	£190,507	15.70	£8,863	-0.06	Dominated
Biologic discount: 50%					
Standard of Care	£163,586	15.76	–	–	–
PredictSURE IBD™	£171,785	15.70	£8,199	-0.06	Dominated
Biologic discount: 75%					
Standard of Care	£145,527	15.76	–	–	–
PredictSURE IBD™	£153,063	15.70	£7,536	-0.06	Dominated
Anti-TNF discount: 25%					
Standard of Care	£193,147	15.76	–	–	–
PredictSURE IBD™	£201,974	15.70	£8,827	-0.06	Dominated
Anti-TNF discount: 50%					
Standard of Care	£186,591	15.76	–	–	–
PredictSURE IBD™	£194,719	15.70	£8,128	-0.06	Dominated
Anti-TNF discount: 75%					
Standard of Care	£180,036	15.76	–	–	–
PredictSURE IBD™	£187,464	15.70	£7,428	-0.06	Dominated
Biologic and Anti-TNF discount: 25%					
Standard of Care	£175,083	15.76	–	–	–
PredictSURE IBD™	£183,252	15.70	£8,163	-0.06	Dominated
Biologic and Anti-TNF discount: 50%					
Standard of Care	£150,475	15.76	–	–	–
PredictSURE IBD™	£157,275	15.70	£6,800	-0.06	Dominated
Biologic and Anti-TNF discount: 75%					
Standard of Care	£125,861	15.76	–	–	–
PredictSURE IBD™	£131,298	15.70	£5,438	-0.06	Dominated
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year.					

As discussed throughout the report, and in particular Section 4.2.4.2, the EAG conducted a range of additional analyses to test extreme scenarios around increasing the relative treatment effectiveness of the TD approach while decreasing the relative costs associated with TD. These scenarios are described below, together with the respective results.

5.2.1 Accounting for the cost-effectiveness of misdiagnosed cases

The test accuracy in the base case economic model for PredictSURE IBD™ and in the scenario analysis included in the DAR for IBDX® was the same and assumed to be 100%. This is unlikely to reflect the tests' actual accuracy in clinical practice; however, no robust diagnostic data were found to inform this in the analysis.

There is however, an ongoing study (PROFILE) that will provide data on the relative effectiveness of these treatment strategies in high-risk patients. The ERG considers that study should also be able to inform the costs and health consequences of misdiagnosing patients as high- and low-risk.

In the absence of real data to inform the costs and consequences of misdiagnosing patients according to their risk of disease severity, presently the EAG has undertaken a theoretical scenario analysis. The EAG assumed that both diagnostic tools are 75% accurate and therefore, 25% of CD cases are assumed to be misdiagnosed in the analysis.

The EAG assumed that a proportion of patients incorrectly diagnosed as having a low-risk course of CD (i.e. high-risk patients) who receive SU therapy, do not respond to IMs and thus, move to anti-TNF treatment after induction therapy. Conversely, patients incorrectly diagnosed as being at high-risk (i.e. low-risk patients) who initiate TD treatment, are assumed to enter remission with anti-TNF treatment and do not have the need to escalate treatment any further.

The rationale for the EAG's assumptions is that low-risk patients (misdiagnosed as high-risk) do not need to escalate from anti-TNF to other treatment options (second-line biologics) in the model. Given these are low-risk patients, the EAG assumed that after 2 years of anti-TNF treatment, 100% of low-risk patients would be in a treatment-free remission state. Similarly, a proportion of high-risk patients (identified as low-risk patients) do not respond to IM and so move on to anti-TNF. The proportion of high-risk patients who do not respond to IM was assumed to be the same as in the base case model (62%). The EAG acknowledges that these assumptions are a simplification of clinical reality, however, no robust evidence was found to inform this scenario.

5.2.2 Varying the assumptions around the measure of relative treatment effectiveness for time to treatment escalation

As some high-risk patients who receive SU treatment respond to IM treatment, having the additional IM step in the SU strategy is advantageous to patients in the EAG's base case analysis as patients in the SU still subsequently receive treatment with biologics, which are assumed to have the same benefit as biologics is the TD arm. Given the paucity of data to substantiate any further benefits in subsequent treatment steps in the TD approach, the EAG considered this to be the most conservative modelling approach.

Nonetheless, the ERG varied these assumptions in two scenario analyses. The scenarios are explained below and summarised in Table 43.

a) High-risk patients on anti-TNF after IM (second step on SU arm) do not do as well as high-risk patients on first-line anti-TNF (first step on TD arm) and thus, the former escalate treatment quicker than the latter. Given that the EAG did not find any data to support this reduction in relative treatment effect, a theoretical assumption was made and varied:

    i. Anti-TNFs in the SU approach are assumed to be only half as effective as anti-TNFs in the TD approach;

    ii. Anti-TNFs in the SU approach are assumed to be as effective as anti-TNFs in the TD approach;

This scenario also assumes that the relative benefit in the anti-TNF step of the TD strategy compared to the anti-TNF step in the SU strategy carries through the next treatment steps. Therefore, patients on second line biologic treatment in the TD strategy have a relative benefit comparatively to second line biologic treatment in the SU arm (as do patients on third line biologics). It is also assumed that second and third line biologic treatment is as effective as anti-TNF treatment within the respective TD and SU arms (Figure 27 and Table 43 below).

b) High-risk patients on anti-TNF after IM (second step on SU arm) do not do as well as high-risk patients on first-line anti-TNF (first step on TD arm) and thus, the former escalate treatment quicker than the latter. However, once patients have moved on to second and third line biologics, SU patients "catch up" to TD patients, and there is no further benefit for TD vs SU.

This scenario also assumes, by default, that second and third line biologic treatment is less effective than anti-TNF treatment in the TD arm (Figure 28 and Table 43 below).

Table 43. Summary of exploratory analyses

Steps in the model	Base case	Scenario a	Scenario b
Anti-TNF (TD) vs IM (SU)	Relative benefit for anti-TNF (D'Haens)	Relative benefit for anti-TNF (D'Haens)	Relative benefit for anti-TNF (D'Haens)
Anti-TNF (TD) vs anti-TNF (SU)	No relative benefit	Relative benefit for TD*: ai) Half of D'Haens; aii) Same as D'Haens	Relative benefit for TD*: bi) Half of D'Haens; bii) Same as D'Haens
Second and third line biologic (TD) vs second and third line biologic (SU)	No relative benefit	Relative benefit for TD*: ai) Half of D'Haens; aii) Same as D'Haens	No relative benefit
Second and third line biologic (TD) vs anti-TNF (TD)	No relative benefit	No relative benefit	Relative benefit for anti-TNF* bi) Half of D'Haens; bii) Same as D'Haens
Second and third line biologic (SU) vs anti-TNF (SU)	No relative benefit	No relative benefit	No relative benefit
*Scenarios i and ii consist on two alternative scenarios, where the size of the benefit is varied as indicated.			

5.2.3 Assumptions around treatment discontinuation in the model

a) The EAG assumed that after 2 years in remission with any biologic treatment, a proportion of patients experience mucosal healing and therefore, stop treatment permanently. The EAG used the Marchetti *et al.* paper to inform this scenario. The study reports that after 2 years in remission, 76% of patients in the TD strategy experience mucosal healing, while 40% of patients in the SU arm experience the same outcome (illustrated in scenario 5.3.2 a i).

   The EAG also varied the Marchetti *et al.* assumptions and explored the possibility of TD and SU therapies having the same impact on the 2-year probability of mucosal healing. Therefore, the EAG assumed that both TD and SU arms would experience the same probability (76% in scenario 5.2.3 a ii and 40% in scenario 5.2.3 a iii) of mucosal healing.

The EAG notes that Hoekman *et al.* concluded that in their 10-year follow-up study, "*mucosal healing 2 years after the start of treatment was associated with a reduced use of anti-TNF treatment during long-term follow-up. Other outcomes, however, did not differ significantly between patients with and without mucosal healing 2 years after the start of treatment, which is in contrast to a recent meta-analysis of 12 studies with 673 patients that showed that mucosal healing is associated with an increased likelihood of long-term clinical remission*." Furthermore, Hoekman *et al.* also reported that another study has shown that 2–4 years after randomisation, mucosal healing at week 104 after randomisation, but not treatment allocation, was associated with stable, corticosteroid-free remission (Baert *et al*.).[173]

Therefore, while there is some evidence supporting that 2-year endoscopic mucosal healing is associated with long-term, corticosteroid-free clinical remission, there does not seem to be any evidence supporting that mucosal healing at 2 years differs according to TD or SU treatment. To note is that estimates used in Marchetti *et al.* were taken from another study, which the EAG did not have access to (Baert *et al*.).[173]

b) The company in TA352 assumed that patients discontinued treatment with biologic agents approximately 1 year after maintenance treatment. The ERG in TA352 was concerned that a discontinuation rule may not have been appropriate for patients who are not in remission as the NICE recommendation for infliximab and adalimumab suggests that, "*specialists should discuss the risks and benefits of continued treatment with patients and consider a trial withdrawal from treatment for all patients who are in stable clinical remission. People who continue treatment with infliximab or adalimumab should have their disease reassessed at least every 12 months to determine whether ongoing treatment is still clinically appropriate. People whose disease relapses after treatment is stopped should have the option to start treatment again*". The EAG notes that duration of treatment with biologics in clinical practice remains uncertain. The clinical experts advising the EAG reported that treatment with anti-TNF and second-line biologics would be given as long as patients continue to show a response.

For completeness, the EAG ran an additional scenario analysis assuming that 100% of patients in continuous remission for 12 months with maintenance treatment of any biologic (i.e. anti-TNF, second or third line biologics), discontinue treatment.

5.2.4 Surgery as a final treatment step in the economic model

a) The clinical experts advising the EAG explained that once patients exhaust all the biologic treatments available, they receive surgery. Therefore, the EAG ran a scenario analysis where patients escalating from third line biologic treatment in the model receive surgery. The EAG assumed that surgery had a temporary "curative" effect of 2 years, where patients experience the costs and utility associated with being in the remission state. After 2 years it was assumed that patients revert to the moderate to severe state, where they remain for the rest of the model;

b) To test the sensitivity of the results of the model to assumptions relating to surgery, the EAG ran a scenario analysis excluding surgeries from the model.

5.2.5 Accounting for the cost-effectiveness of misdiagnosed cases and assumptions around treatment discontinuation in the model

The EAG combined a range of scenarios to assess the impact of increasing the effectiveness of the TD strategy while decreasing costs with biologic treatments. Scenario 5.2.5 a, b, and c explored changing the effectiveness of the diagnostic tool (and TD) through the assumptions made for the misdiagnosis scenario.

a) The EAG combined the misdiagnosis scenario 5.2.1 with scenario 5.2.3 a i, where it was assumed that after 2 years in remission, 76% of patients in the TD strategy experience mucosal healing, while 40% of patients in the SU arm experience the same outcome.

b) The EAG also combined scenario 5.2.1 with scenario 5.2.3 a ii, where it was assumed that after 2 years in remission, 76% of patients in both the TD and the SU strategies experience mucosal healing.

c) The EAG also combined scenario 5.2.1 with scenario 5.2.3 a iii, where it was assumed that after 2 years in remission, 40% of patients in both the TD and the SU strategies experience mucosal healing.

5.2.6 Varying the assumptions around the measure of relative treatment effectiveness for time to treatment escalation and assumptions around treatment discontinuation in the model

As with scenario 5.2.5, the EAG explored the impact of combining scenario 5.2.3 (where costs associated with biologics were decreased) with changing the effectiveness of the diagnostic tool (and TD) through the assumptions made for time to treatment discontinuation (TTE) in the model. The EAG used scenario 5.2.2. a ii for all the analyses as this is the scenario that assumes the highest relative effective for TD vs SU in terms of TTE.

a) The EAG combined scenario 5.2.2 a ii with scenario 5.2.3 a i, where it was assumed that after 2 years in remission, 76% of patients in the TD strategy experience mucosal healing, while 40% of patients in the SU arm experience the same outcome.

b) The EAG also combined scenario 5.2.2 a ii with scenario 5.2.3 a ii, where it was assumed that after 2 years in remission, 76% of patients in the TD and the SU strategies experience mucosal healing.

c) The EAG also combined scenario 5.2.2 a ii with scenario 5.2.3 a iii, where it was assumed that after 2 years in remission, 40% of patients in the TD and the SU strategies experience mucosal healing.

5.2.7 Varying the proportion of patients who respond to IM and varying the assumptions around the measure of relative treatment effectiveness for time to treatment escalation

One of the scenario analyses carried out by the EAG assumed that no high-risk patients respond to IMs (therefore not deriving any benefit from response to this treatment). This scenario intended to portray an extreme clinical reality where high-risk patients need treatment with a biologic for a response and its impact on the final ICER. The ICER for PredictSURE-IBD™ compared to SU changed from the EAG's base case of dominated (against the diagnostic tool) to £71,294. To note is that the EAG tested the impact of varying the proportion of patients who do not respond to IM treatment in the analysis and when 92% of patients were assumed not to respond to IM treatment the two strategies (TD and SU) become clinically equivalent.

The EAG combined scenario 5.2.2 a ii with varying the proportion of high-risk patients who receive SU therapy and do not respond to IMs thus, increasing the relative effectiveness of TD and decreasing the effectiveness of SU, both in terms of TTE and the probability of response and remission in the model.

The EAG tested the assumption that 100% of patients do not respond to IM and varied this percentage to assess the impact on the final ICERs.

5.2.8 Varying the proportion of patients who respond to IM; varying the assumptions around the measure of relative treatment effectiveness for time to treatment escalation; and varying treatment discontinuation assumptions

a) The EAG combined scenario 5.2.6 a with varying the proportion of high-risk patients who receive SU therapy and do not respond to IMs (therefore not deriving any benefit from response to this treatment).

b) The EAG combined scenario 5.2.6 b with varying the proportion of high-risk patients who receive SU therapy and do not respond to IMs (therefore not deriving any benefit from response to this treatment).

c) The EAG combined scenario 5.2.6 c with varying the proportion of high-risk patients who receive SU therapy and do not respond to IMs (therefore not deriving any benefit from response to this treatment).

All the scenarios increased the relative effectiveness of TD in terms of TTE and decreased the costs associated biologic treatment (to different amounts). For all scenarios, the EAG tested the assumption that 100% of patients do not respond to IM and varied this percentage to assess the impact on the final ICERs.

5.2.9 Results

Results of the EAG's scenario analyses are reported in Table 43. The majority of the scenarios still produced a dominated ICER, showing that the TD strategy (via the use of PredictSURE IBD™ in the model) is dominated by SU (via the SC arm of the model), with additional costs and a QALY loss.

Scenario 5.2.1 produced an ICER of £59,456 per QALY gained, with PredictSURE IBD™ being more costly than SC but generating a QALY gain of 0.18. Even though this scenario assumes lower test accuracy, the assumed consequences of misdiagnosis produced a QALY gain for the diagnostic tool. This is related to the assumption of allocating low-risk patients (misdiagnosed as high-risk) to the anti-TNF state in the model, without any further need for further escalation. Given that treatment with anti-TNF holds the highest remission rate in the EAG's analysis, and that 62% of high-risk patients (misdiagnosed as low-risk) in the SU arm were assumed to not derive any benefit from treatment with IMs, the results produced positive incremental QALYs for the diagnostic tool (thus, for the TD strategy). The EAG also combined this scenario with reducing the costs associated with TD, through reducing the time spent on biologic treatment (as per scenario 5.2.3) and presents the results in scenario 5.2.5.

Scenario 5.2.3 a i produced an ICER of £48,078 for SC vs PredictSURE IBD™, meaning that the diagnostic tool is less expensive than SC (by £2,978) but also less effective (0.06 QALY loss). This scenario reduced the costs of biologic treatment in the TD arm, by assuming that a higher proportion of patients in the TD arm achieve mucosal healing and thus stop treatment. Even though these patients were "kept" in the remission state, the QALYs generated with this assumption were not enough to produce a QALY gain compared with the benefit patients derive from initial treatment with IMs in SU. The EAG also notes that scenario 5.2.3 a i can also be interpreted as a proxy for a scenario assuming de-escalation from biologic treatment in the TD arm to IMs. This is because the scenario reduced treatment costs (by stopping treatment with biologics) which would be similar to replacing treatment with biologics with IMs in the model due to the low cost of IM treatment.

The other variations of scenario 5.2.3, where the same proportion of patients were assumed to achieve mucosal healing in the TD and SU arms, produced dominated ICERs against the diagnostic tool (and thus TD). The EAG notes that Hoekman *et al.* did not show a difference in mucosal healing for TD vs SU (although it is not clear if the authors investigated the impact that the strategies had on this outcome). Notwithstanding, the authors reported that the rate of mucosal healing reported in another study (Baert *et al.*) had shown that 2–4 years after randomisation treatment allocation was associated with stable, treatment-free remission.[173]

Scenario 5.2.5 a resulted in a dominant ICER for PredictSURE IBD™ (and TD), with the diagnostic tool being associated with less costs and higher QALYs than SC (and SU**).** This scenario combines modelling misdiagnosed cases with reducing the costs associated with TD, therefore generating additional QALYs for the diagnostic tool at a lower cost, given the assumption that a proportion of

patients on TD enter a permanent stage of remission. Given that scenario 5.2.5 a assumes a difference in the rate of treatment discontinuation for biologics (whereby TD patients have a higher probability of discontinuing treatment – due to mucosal healing – than SU patients), this scenario produced the highest cost savings for TD. Scenarios 5.2.5 b and c produced higher ICERs as the relative costs associated with treatment with biologics (and the diagnostic tool) increased, however scenario 5.2.5 b resulted in an ICER of £32,743 per QALY gained, therefore close to the upper threshold (£30,000) typically used in the NICE decision-making process.

Scenario 5.2.6 a, b and c, explored increasing the effectiveness of TD vs SU with respect to time to treatment escalation (TTE), combined with decreasing the treatment costs with biologics. As demonstrated, all scenarios generate a QALY loss for the diagnostic tool compared to SC. When it is assumed that a higher proportion of patients in the TD arm achieve mucosal healing (scenario 5.2.3 a i) than in the SU arm, the diagnostic tool (and TD) becomes cost saving (-£1,332) albeit less effective (-0.02).

Scenario 5.2.7 and scenario 5.2.8 explored increasing the effectiveness of TD vs SU with respect to time to treatment escalation (TTE), combined with decreasing the treatment costs with biologics and with varying the assumption around the rate of response to IM treatment in the SU strategy.

Scenario 5.2.7 shows that when the relative TTE benefit in the anti-TNF step of the TD strategy compared to the IM step in the SU strategy carries through all the next treatment steps in the model (scenario 5.2.2 a ii) and when 100% of SU patients are assumed to not respond to treatment with IM, the ICER amounts to £53,859 per QALY gained. Therefore, even when 100% of high-risk patients do not respond to IMs, the ICER for the diagnostic tool (and TD) compared to SC (and SU) is still above the NICE £30,000 threshold.

Scenario 5.2.8 a shows that when the relative TTE benefit in the anti-TNF step of the TD strategy compared to the IM step in the SU strategy carries through all the next treatment steps in the model (scenario 5.2.2 a ii); when a higher proportion of patients in the TD arm achieves mucosal healing (scenario 5.2.3 a i); and when 100% of SU patients are assumed to not respond to treatment with IM, the final ICER becomes dominant for PredictSURE IBD™ (and TD), with the diagnostic tool being associated with less costs and higher QALYs than SC (and SU). The diagnostic tool remains dominant up to when the assumption around the proportion of high-risk SU patients not responding to IM treatment is decreased from 100% to 70%. To note is that the EAG's base case analysis estimates that 62% of high-risk patients do not respond to initial treatment with IMs.

Scenario 5.2.8 b and c show that when the relative TTE benefit in the anti-TNF step of the TD strategy compared to the IM step in the SU strategy carries through all the next treatment steps in the model (scenario 5.2.2 a ii); when the same proportion of patients in the TD and SU arms achieves mucosal healing (scenario 5.2.3 a ii for 76% and 40%, respectively); and when 100% of SU patients are assumed to not respond to treatment with IM, the final ICERs are £28,724 and £40,630, respectively. Both scenarios generate a QALY gain for the diagnostic tool (and TD) compared to SC (and SU), however the additional costs associated with TD are higher in scenario c (40% of patients in remission stop treatment with biologics in both the TD and SU arms) than in scenario b (76% of patients in remission stop treatment with biologics in both the TD and SU arms).

The EAG has produced plots to demonstrate the impact of reducing the percentage of high-risk patients who do not respond to IM from 100% to zero for scenario 5.8a (where Predict-SURE IBD™ is dominant). The plot in Figure 42 shows the changes in the incremental costs and QALYs on the cost-effectiveness plane and demonstrates the ICER changing from dominant at 100% non-response to IMs, moving into the south-west quadrant (less costly and less effective for TD) at 70%, then becoming dominated from below 53%. Figure 43 shows the resulting final ICERs, and the drastic variation in these at 70% non-response, when the incremental QALYs become negative.

Figure 42. Incremental costs and QALYs as percentage of high risk IM non-responders varies

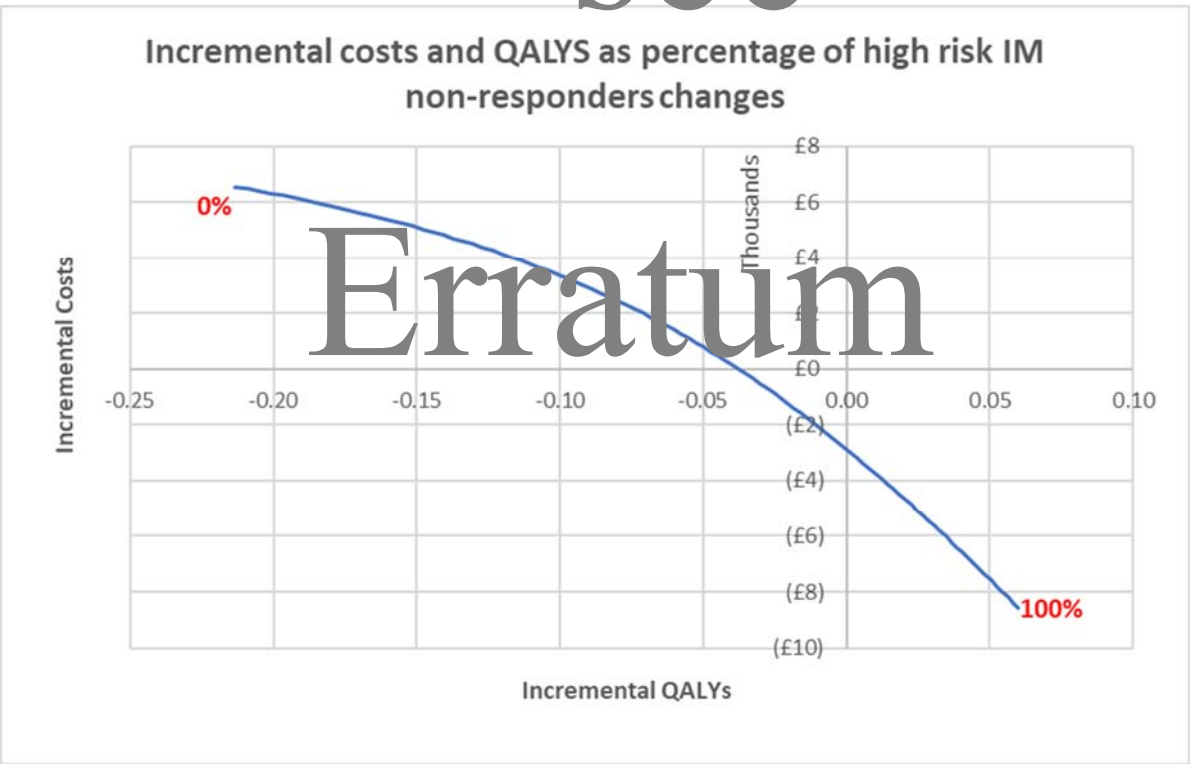
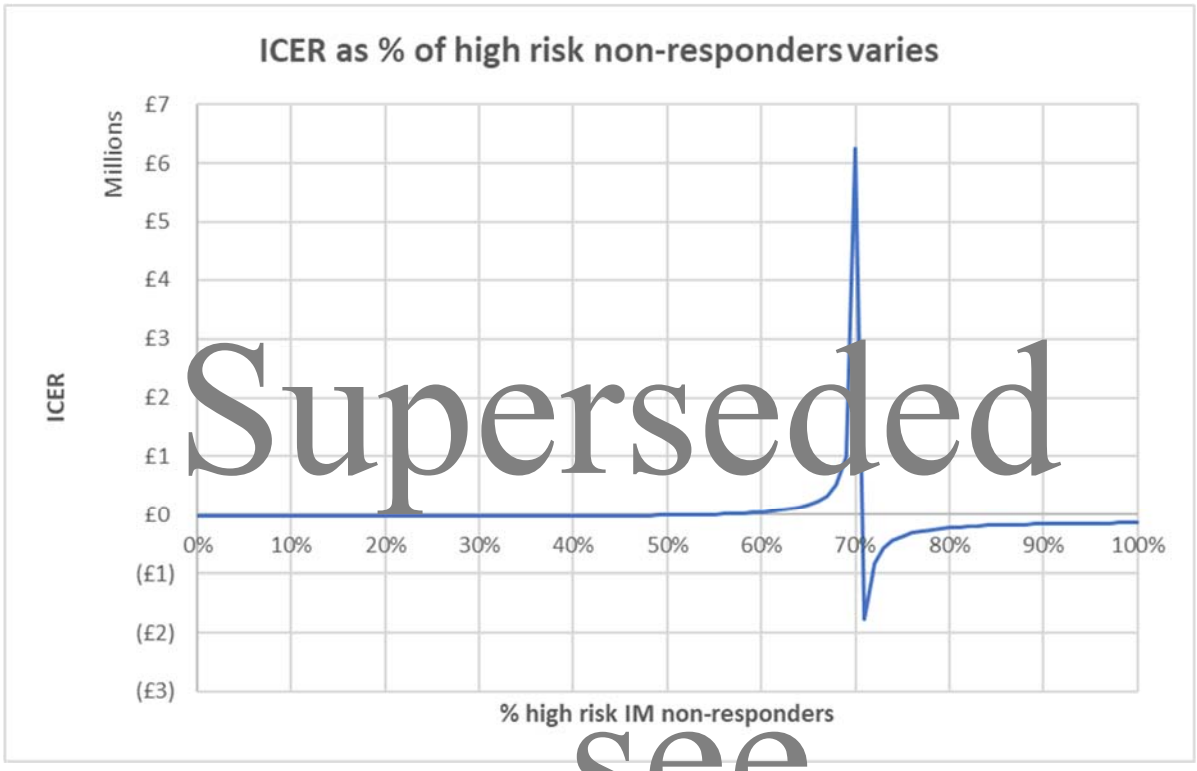


Figure 43. Resulting ICERs as the percentage of high risk IM non-responders varies



5.2.10 Conclusions

1. Estimating the impact of reducing test accuracy was only possible through combining this with an increase in the relative effectiveness of the TD strategy (in order to attribute consequences to misdiagnosing patients). However, changing this alone in the model still produced ICERs above the NICE cost-effectiveness upper threshold of £30,000 (scenario 5.2.1). When this assumption was combined with decreasing the cost associated with biologic treatment (through assuming different rates of mucosal healing leading to remission); the ICER ranged from dominant to £45,397 for PredictSURE IBD™ (and TD) - (scenario 5.2.5 a and c, respectively).

2. By itself, increasing the relative effectiveness of TD on TTE did not have an impact on the dominance of SC over TD (scenario 5.2.2).

3. Assuming that 40% and 76% of patients in remission after 2 years (and 100% of patients in remission after 1 year) on maintenance treatment with anti-TNF, second and third line biologics discontinued treatment in both treatment arms also did not impact the dominance of SC over

TD. Nonetheless, when a higher proportion of patients discontinued treatment with biologics in the TD arm compared with the SD arm, this generated a cost saving for TD, however, still with less QALYs than for SU (scenario 5.2.3).

4. Excluding surgeries from the model did not have an impact on the dominance of SC over TD, and neither did assuming that surgery has a curative effect for 2-years (scenario 5.2.4).

5. Combining the increase in the relative effectiveness of TD on TTE with reducing the costs of biologic treatment did not have an impact on the dominance of SC over TD when the same proportion of patients were assumed to discontinue treatment with biologics in the TD and SU arm. When a higher proportion of patients discontinued treatment with biologics in the TD arm compared with the SD arm, this generated a cost saving for TD, however, with less QALYs than for SU (scenario 5.2.6).

6. Increasing the relative effectiveness of TD on TTE and additionally reducing the effectiveness of SU (through assuming a 0% probability of response to IM treatment for high-risk patients) still generated an ICER above the NICE cost-effectiveness upper threshold of £30,000 (scenario 5.2.7).

7. When the increase in the relative effectiveness of TD on TTE and the additional reduction in the effectiveness of SU are combined with a reduction of time on treatment with biologics, the ICERs for PredictSURE IBD™ and TD drop below the £30,000 per QALY gained threshold with SC (and SU), depending on the assumptions made for the proportion of patients who discontinue treatment with biologics. When the proportion of patients discontinuing treatment with biologics is 76% in the TD arm compared with 40% in the SU arm, the final ICER is dominant for PredictSURE IBD™ against SC, as long as the proportion of high-risk patients who do not respond to initial treatment with IM is 70% (or above).

In conclusion, once the relative effectiveness of TD is artificially increased (through both TTE; probability of response to initial treatment; and the impact it has on low-risk patients) and combined with decreased time on biologic treatment, the ICERs for PredictSURE IBD™ (and TD) compared to SC (and SU) fall below £30,000 - the upper threshold typically used in the decision-making process by NICE. However, the EAG notes that these results need to be interpreted with extreme caution as the assumptions made in these scenarios were designed to test extreme clinical scenarios where TD was assumed to be more effective than SU. Nonetheless, the EAG did not find any evidence to substantiate the benefits modelled in these scenarios, and thus concludes that its base case analysis showing that TD

is dominated by SU remains the most conservative assessment of the relative cost-effectiveness of these treatment strategies.

Table 44. Results of scenario analyses

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Scenario 5.2.1 Misdiagnosis					
Standard of Care	£199,702	15.76	–	–	–
PredictSURE IBD™	£210,345	15.94	£10,643	0.18	£59,456
Scenario 5.2.2 a i - Assuming half of the base case relative effectiveness for TD on TTE for further steps					
Standard of Care	£197,192	15.71	–	–	–
PredictSURE IBD™	£207,827	15.67	£10,635	-0.04	Dominated
Scenario 5.2.2 a ii - Assuming the same as the base case relative effectiveness for TD on TTE for further steps					
Standard of Care	£194,816	15.67	–	–	–
PredictSURE IBD™	£206,458	15.65	£11,642	-0.02	Dominated
Scenario 5.2.2 b i - Assuming half of the base case relative effectiveness for TD on TTE for anti-TNF					
Standard of Care	£197,192	15.71	–	–	–
PredictSURE IBD™	£206,951	15.66	£9,759	-0.05	Dominated
Scenario 5.2.2 b ii - Assuming the same as the base case relative effectiveness for TD on TTE for anti-TNF					
Standard of Care	£194,816	15.67	–	–	–
PredictSURE IBD™	£204,767	15.62	£9,951	-0.05	Dominated
Scenario 5.2.3 a i – Assuming discontinuation of biologic treatment for 76% TD; 40% SU.					
Standard of Care	£179,600	15.76	–	–	–
PredictSURE IBD™	£176,622	15.70	-£2,978	-0.06	£48,078*
Scenario 5.2.3 a ii - Assuming discontinuation of biologic treatment for 76% TD; 76% SU.					
Standard of Care	£161,507	15.76	–	–	–
PredictSURE IBD™	£168,191	15.70	£6,684	-0.06	Dominated
Scenario 5.2.3 a iii - Assuming discontinuation of biologic treatment for 40% TD; 40% SU.					
Standard of Care	£179,600	15.76	–	–	–
PredictSURE IBD™	£187,630	15.70	£8,030	-0.06	Dominated

Scenario 5.2.3 b - Assuming discontinuation of biologic treatment for 100% TD; 100% SU.					
Standard of Care	£149,446	15.76	–	–	–
PredictSURE IBD™	£155,232	15.70	£5,786	-0.06	Dominated
Scenario 5.2.4 a – Assuming surgery as last treatment step					
Standard of Care	£201,736	16.04	–	–	–
PredictSURE IBD™	£211,315	15.99	£9,579	-0.05	Dominated
Scenario 5.2.4 b – Removing surgery from the model					
Standard of Care	£195,598	15.78	–	–	–
PredictSURE IBD™	£205,713	15.71	£10,115	-0.06	Dominated
Scenario 5.2.5 a (Scenario 5.2.1 + Scenario 5.2.3 a i)					
Standard of Care	£179,600	15.76	–	–	–
PredictSURE IBD™	£175,447	15.94	-£4,152	0.18	Dominant
Scenario 5.2.5 b (Scenario 5.2.1 + Scenario 5.2.3 a ii)					
Standard of Care	£161,507	15.76	–	–	–
PredictSURE IBD™	£167,368	15.94	£5,861	0.18	£32,743
Scenario 5.2.5 c (Scenario 5.2.1 + Scenario 5.2.3 a iii)					
Standard of Care	£179,600	15.76	–	–	–
PredictSURE IBD™	£187,726	15.94	£8,126	0.18	£45,397
Scenario 5.2.6 a (Scenario 5.2.2 a ii + Scenario 5.2.3 a					
Standard of Care	£175,575	15.67	–	–	–
PredictSURE IBD™	£174,243	15.65	-£1,332	-0.02	£69,963*
Scenario 5.2.6 b (Scenario 5.2.2 a ii + Scenario 5.2.3 a ii)					
Standard of Care	£158,257	15.67	–	–	–
PredictSURE IBD™	£166,166	15.65	£7,908	-0.02	Dominated
Scenario 5.2.6 c (Scenario 5.2.2 a ii + Scenario 5.2.3 a iii)					
Standard of Care	£175,575	15.67	–	–	–
PredictSURE IBD™	£185,252	15.65	£9,677	-0.02	Dominated
Scenario 5.2.7 (Scenario 5.2.2 a ii + assuming that 100% of SU patients do not respond to IM)					
Standard of Care	£203,229	15.59	–	–	–
PredictSURE IBD™	£206,458	15.65	£3,229	0.06	£53,859
Scenario 5.2.8 a (Scenario 5.2.2 a ii + Scenario 5.2.3 a i + assuming that 100% of SU patients do not respond to IM)					

Standard of Care	£182,816	15.59	–	–	–
PredictSURE IBD™	£174,243	15.65	-£8,573	0.06	**Dominant**
Scenario 5.2. 8 b (Scenario 5.2.2 a ii + Scenario 5.2.3 a ii + assuming that 100% of SU patients do not respond to IM)					
Standard of Care	£164,444	15.59	–	–	–
PredictSURE IBD™	£166,166	15.65	£1,722	0.06	£28,724
Scenario 5.2.8 c (Scenario 5.2.2 a ii + Scenario 5.2.3 a iii + assuming that 100% of SU patients do not respond to IM)					
Standard of Care	£182,816	15.59	–	–	–
PredictSURE IBD™	£185,252	15.65	£2,436	0.06	£40,630
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year; TTS, time-to-surgery. *This ICER is for SC vs PredictSURE IBD™, meaning that the diagnostic tool is cheaper than SC but also less effective.					

5.3 Sensitivity analyses — see

5.3.1 One-way sensitivity analysis

The EAG conducted a number of deterministic one-way sensitivity analyses around the model inputs as described in Table 45. Figure 46 ranks the model key drivers by their impact on the deterministic ICER. The lower and upper bounds of each of the sensitivity analyses were derived from the lower and upper bounds of the 95% confidence intervals of the distributions specified for the PSA. Details of each of the distributions is given in Table 36.

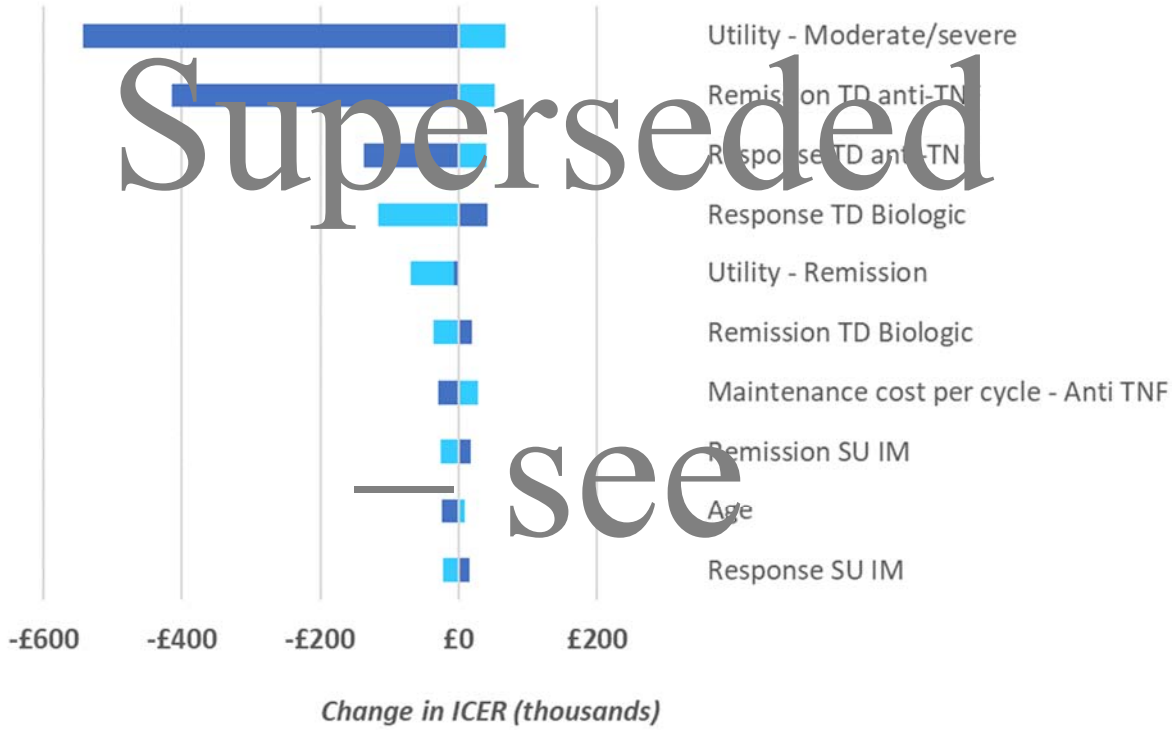
Similar to the results of the scenario analyses undertaken by the EAG, most of the changes in model parameters resulted in dominated ICERs against PredictSURE-IBD™. The only exception (not reported in Figure 46 due to issues of scale of results) is the utility estimate used for the mild state in the economic model. Varying this input led to a variation in the ICER up to £50,000,000 per QALY gained.

Table 45. Inputs and results of OWSAs

Model Parameter	Lower bound	Upper bound	Lower ICER	Upper ICER
Age	21.3	48.7	-£144,151	-£177,040
Crohn's disease expected body weight	43.8	100.2	-£147,326	-£160,540
Proportion of males	0.228	0.5220	-£153,173	-£154,685
Probability of being high risk	0.3496	0.8004	-£164,398	-£149,234
Proportion on infliximab in anti-TNF biologics class	0.2432	0.5568	-£152,977	-£154,889
Proportion on vedolizumab in non-anti-TNF biologics class	0.3040	0.6960	-£151,772	-£156,094
Proportion on azathioprine for immunomodulators	0.4864	1.0000	-£154,466	-£153,593
Proportion of 6-mercaptopurine for immunomodulators	0.0608	0.1392	-£153,981	-£153,910
Proportion of anti-TNF with IM bundle	0.1824	0.4176	-£153,737	-£154,129
Proportion of Biologics with IM bundle	0.1216	0.2784	-£153,850	-£154,016
Response TD Biologic	0.1918	0.4390	-£269,635	-£112,021
Remission TD Biologic	0.0795	0.1821	-£190,416	-£134,160
Response TD anti-TNF	0.1565	0.3583	-£113,632	-£290,304
Remission TD anti-TNF	0.2231	0.5108	-£101,608	-£569,260
Response SU Biologic	0.1918	0.4390	-£153,933	-£153,933
Remission SU Biologic	0.0795	0.1821	-£153,933	-£153,933
Response SU anti-TNF	0.1565	0.3583	-£153,933	-£153,933
Remission SU anti-TNF	0.2231	0.5108	-£153,933	-£153,933

Response SU IM	0.1380	0.3160	-£176,477	-£137,727
Remission SU IM	0.0950	0.2176	-£179,214	-£136,972
Probability of death following surgery	0.0009	0.0021	-£151,742	-£156,183
PredictSURE cost	£760	£1,740	-£146,022	-£161,843
Health state cost - Remission	£10	£23	-£154,200	-£153,766
Health state cost - Mild	£16	£37	-£154,779	-£153,087
Health state cost - Moderate/severe	£74	£170	-£147,332	-£160,534
Health state cost - No response	£74	£170	-£155,409	-£152,457
Induction cost per cycle - Anti TNF	£927	£2,123	-£152,966	-£154,900
Induction cost per cycle - Biologic	£940	£2,151	-£152,007	-£155,859
Induction cost per cycle - Immunomodulator	£3	£6	-£154,013	-£153,853
Maintenance cost per cycle - Anti TNF	£326	£747	-£125,193	-£182,673
Maintenance cost per cycle - Biologic	£399	£914	-£131,595	-£176,271
Maintenance cost per cycle - Immunomodulator	£7	£17	-£155,516	-£152,350
IV administration first attendance	£121	£277	-£153,630	-£154,236
IV administration follow-up	£129	£295	-£135,107	-£172,759
Cost of surgery	£5,359	£12,268	-£157,923	-£149,943
Utility - Remission	0.50	1.00	-£215,791	-£161,004
Utility - Mild	0.44	1.00	-£153,933	-£153,933
Utility - Moderate/severe	0.35	0.79	-£85,522	-£697,276
Disutility for surgery	0.02	0.06	-£155,325	-£152,566

Figure 46. Tornado plot showing OWSAs for PredictSURE IBD™ versus standard care



Abbreviations in figure: ICER, incremental cost effectiveness ratio; IM, immunomodulator; OWSA, one-way sensitivity analysis; SoC, standard of care; SU, step up; TD, top down; TNF, tumour necrosis factor.

6 DISCUSSION

6.1 Statement of principal findings

6.1.1 Prognostic accuracy

Twelve publications describing eight studies were included in the assessment of the prognostic accuracy of the tests. Seven of the studies reported results on utility of the IBDX kit and one study provided data on PredictSURE-IBD in stratifying those at high-risk of a severe course of CD. Limited evidence is available from the included full-text publications on the prognostic accuracy of PredictSURE-IBD, and none is available on prognostic accuracy of IBDX, as determined by measures such as sensitivity and specificity. Most evidence on the utility of the two tools is derived from observational studies that report estimates of risk of experiencing a clinical outcome associated with an aggressive course of CD, for example, need for treatment escalation, development of a complication or surgery. No study retrieved reported on the clinical impact of use of IBDX or PredictSURE-IBD in terms of influencing the treatments given in the management of active CD.

All included studies assessed outcomes in people reported to have a diagnosis of CD. However, limited reporting was noted across studies relating to the IBDX on stage of diagnosis (newly versus established) at time of test. Baseline characteristics suggest that samples analysed were provided predominantly by people with established CD. By contrast, most people enrolled in the study on PredictSURE-IBD had received a recent diagnosis of CD. Although most of the included studies outlined criteria to be met for a diagnosis of CD, only the study evaluating the PredictSURE-IBD tool required people to have active disease to be eligible for enrolment, and reported how presence of active disease was determined. Given the biomarker targets of the two prognostic tests, the reviewers consider that a criteria of active CD is appropriate for inclusion of studies assessing PredictSURE-IBD, but is not essential for studies reporting on IBDX.

Use of PredictSURE-IBD was associated with a sensitivity and specificity of [redacted]% and [redacted]%, respectively, in stratifying by need for multiple treatment escalations within 12 months. Corresponding sensitivity and specificity for multiple escalations within 18 months were 72.7% and 73.2%, respectively. A negative predictive value of 90.9% for PredictSURE-IBD of predicting multiple escalations within the first 18 months was also reported. The cut-off for multiple escalations applied in the determination of sensitivity and specificity was two treatment escalations, and comprised any type of treatment, including surgery.

Seven studies evaluating the IBDX kit were deemed to be of relevance to the review, all of which were observational in nature: three studies were prospective cohorts and three were of a cross-sectional design. Clinical heterogeneity across studies in terms of various characteristics (prior complication versus no complication, previous IBD-related surgery or no surgery, and unclear whether people had active disease at baseline) was noted. Two prospective cohort studies reported increased risk of experiencing a complication or of requiring surgery for those testing positive for at least two of the six biomarkers included in the IBDX kit.

Two studies reported increased risk of experiencing a complication or of requiring surgery for those testing positive for at least two of the six biomarkers included in the IBDX kit. Risks of experiencing a complication by positive biomarker status were reported to be:

- OR of 1.5 (95% CI: 1.3 to 1.9; p<0.001; N unclear) based on positivity for a median of two biomarkers;
- HR of 2.5 (95% CI: 1.03 to 6.1; p=0.043; N=20 with no prior complication or surgery) based on positivity for at least two biomarkers;
- HR of 2.6 (95% CI: 0.92 to 7.2; p=0.072; N=20 with no prior complication or surgery) based on positivity for at least three biomarkers.

Considering surgery, three studies reported on increased risk of surgery. One study reported a trend towards a larger proportion of people with CD requiring abdominal surgery with increasing number of positive biomarkers (N=517; p<0.0001 across the groups). Other estimates of higher risk of requiring surgery were:

- OR of 1.5 (95% CI: 1.3 to 1.8; p<0.001; N unclear) based on positivity for a median of two biomarkers;
- HR of 3.6 (95% CI: 1.2 to 11.0; p=0.023; N=14 with no prior complication or surgery) based on positivity for at least two biomarkers;
- HR of 2.8 (95% CI: 0.80 to 9.6; p=0.11; N=14 with no prior complication or surgery) based on positivity for at least three biomarkers.

The study evaluating PredictSURE-IBD reported that those 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).

6.1.2 Economic

As no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, the development of the economic model sets a structural framework for analysing future available data on prognostic accuracy and assesses the costs and consequences of treating high- and low-risk patients with both TD and SU strategies.

The EAG found two main sources of evidence that could be used to model time to treatment escalation (TTE) and time to surgery (TTS). Nevertheless, each source could only partially inform the TTE and TTS analyses in the model. Therefore, clinical data informing the analysis had to be derived from multiple sources.

One of the key underlying assumptions in the EAG's base case analysis is that high-risk patients who initiate treatment with IMs (SU arm) escalate treatment quicker than high-risk patients who initiate treatment with anti-TNF (supported by the data presented in D'Haens *et al.*). However, once SU patients initiate treatment with an anti-TNF (their second treatment step), they "catch-up" with patients on the TD treatment strategy. As some high-risk patients who receive SU treatment respond to IM treatment, having the additional IM step in the SU strategy is advantageous to patients in the EAG's base case analysis as patients still subsequently receive treatment with biologics, which are assumed to have the same effect as biologics is the TD arm. Given the paucity of data to substantiate any further benefits in subsequent treatment steps in the TD vs SU approaches, the EAG considered this to be the most conservative modelling approach.

The EAG also notes that even though, in theory, a TD approach would suggest a "de-escalation" of treatments, the clinical experts advising the EAG consistently reported that IMs would not be given to patients who are responding well to biologics (instead treatment with biologics would just be continued until loss of response). The experts also explained that after loss of response with first- or second-line biologics, patients would not be given IMs, but instead surgery as a last treatment option. Nonetheless, the EAG undertook a scenario analysis (5.2.3 a i) to explore the impact of de-escalation in the model.

The long-term follow up study Hoekman *et al.* found no difference between SU and TD in 10-year clinical remission rate; endoscopic remission, hospitalisation, surgery or new fistulas. Furthermore, the study concluded that in the long-term a TD strategy had not proven to alter the natural history of CD. However, time to relapse was found statistically significantly different across TD and SU arms in the 2-year analysis of the same data (D'Haens *et al.*).

Hoekman *et al.* concluded that their study was the first to compare the long-term outcomes for newly diagnosed CD patients who received combined immunosuppression vs conventional management. The authors added that early combined immunosuppression may be a preferential strategy given the associated delay in time to relapse. However, the authors noted that the costs and risks of potentially overtreating patients with a potentially 'benign' disease course mean that a TD approach should not be recommended as a universal treatment strategy for all patients with newly diagnosed CD.

The EAG's cost-effectiveness analyses are consistent with the conclusions from Hoekman *et al.* The ICERs indicate that SC (and so SU) dominates use of both diagnostic tools (and so TD) even when assuming the tests are 100% accurate. In the base case results, the incremental analysis of cost-effectiveness demonstrates that the TD strategy (via the use of PredictSURE IBD™ in the model) is dominated by SU (via the SC arm of the model), with an additional cost of £9,526 and a QALY loss of 0.06.

In order to mitigate some of the concerns raised by the specialist committee members (SCMs), the EAG conducted a range of analyses to test extreme scenarios around increasing the relative treatment effectiveness of the TD approach while decreasing the relative costs associated with TD. The EAG concluded that:

1. Estimating the impact of reducing test accuracy was only possible through combining this with an increase in the relative effectiveness of the TD strategy (in order to attribute consequences to misdiagnosing patients). However, changing this alone in the model still produced ICERs above the NICE cost-effectiveness upper threshold of £30,000. When this assumption was combined with decreasing the costs associated with biologic treatment (through assuming different rates of mucosal healing leading to remission), the ICER ranged from dominant to £45,397 for PredictSURE IBD™ (and TD);

2. By itself, increasing the relative effectiveness of TD on TTE did not have an impact on the dominance of SC over TD;

3. Assuming that 40% and 76% of patients in remission after 2 years (and 100% of patients in remission after 1 year) on maintenance treatment with anti-TNF, second and third line biologics discontinued treatment in both treatment arms also did not impact the dominance of SC over TD. Nonetheless, when a higher proportion of patients discontinued treatment with biologics in the TD arm compared with the SD arm, this generated a cost saving for TD, however, still with less QALYs than for SU;

4. Excluding surgeries from the model did not have an impact on the dominance of SC over TD, and neither did assuming that surgery has a curative effect at 2-years;

5. Combining the increase in the relative effectiveness of TD on TTE with reducing the costs of biologic treatment did not have an impact on the dominance of SC over TD when the same proportion of patients were assumed to discontinue treatment with biologics in the TD and SU arm. When a higher proportion of patients discontinued treatment with biologics in the TD arm compared with the SU arm, this generated a cost saving for TD, however with less QALYs than for SU;

6. Increasing the relative effectiveness of TD on TTE and additionally reducing the effectiveness of SU (through assuming a 0% probability of response to IM treatment from high-risk patients) still generated an ICER above the NICE cost-effectiveness upper threshold of £30,000;

When the increase in the relative effectiveness of TD on TTE and the additional reduction in the effectiveness of SU are combined with a reduction of time on treatment with biologics, the ICERs for PredictSURE IBD™ (and TD) can become cost-effective compared with SC (and SU), depending on the assumptions made for the proportion of patients who discontinue treatment with biologics. When the proportion of patients discontinuing treatment with biologics is 76% in the TD arm compared with 40% in the SU arm, the final ICER is dominant for PredictSURE IBD™ against SC, as long as the proportion of high-risk patients who do not respond to initial treatment with IM is 70% (or above).

6.2 Strengths and limitations of the analysis

6.2.1 Clinical

Despite extensive systematic searches of the literature, no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, IBDX and PredictSURE-IBD. In terms of sensitivity and specificity as estimates of prognostic accuracy, the External Assessment Group (EAG) is unaware of a validated definition for determination of whether a person has followed a severe course of CD, and, thus, considers the criterion required for a true positive or false positive for IBDX and PredictSURE-IBD to be unclear. The EAG considers it would be challenging to ascertain an accurate estimate of prognostic accuracy of the tools in stratifying course of CD and to do so would require carrying out a prospective study that included a group that received only "step-up" (SU) treatment after determination of risk of course of CD. The ongoing PROFILE RCT randomises people to accelerated SU or TD treatment after determination of high or low risk of following a severe course of CD and so will provide additional data to inform estimates of prognostic accuracy

One study reporting on the sensitivity and specificity of PredictSURE-IBD was identified. The EAG has reservations about the generalisability of the estimates. To determine sensitivity and specificity, the authors of the study applied a cut off of two or more treatment escalations as denoting a high risk of severe course of CD, with surgery included as treatment escalation. The EAG considers the choice of two escalations to be an arbitrary value. Additionally, the EAG's clinical experts fed back that it would be appropriate to consider escalation to CD-related surgery separately from progression to drug treatment, and also to use development of a complication of CD (fistula or stenosis) as an alternative marker of sensitivity and specificity.

Studies informing the evidence around effectiveness of the tools predominantly estimated increased risk of experiencing a clinical outcome for those designated as high risk versus low risk of following a severe course of CD. Clinical outcomes that could be considered proxies for predicting prognosis of CD are developing a complication (fistula or stenosis), needing CD-related surgery, and a shorter time to and increased frequency of treatment escalations.

For IBDX, estimates were available for increased risk of developing a complication and for need for surgery for those classified as high risk of following a severe disease course, but not for time to treatment escalation. Conversely, estimates were available for PredictSURE-IBD for time to treatment escalation but not for risk of developing a complication or need for surgery. Given the disparity in the clinical outcomes assessed for IBDX and PredictSURE-IBD, the EAG considers that no conclusions can be drawn on the comparative effectiveness of the two tools in stratifying people by risk of severe course of CD.

Another limitation of the identified evidence base is that no study included in the review prospectively followed people whose treatment was determined by results from IBDX and PredictSURE-IBD. The ongoing PROFILE RCT assesses whether early treatment with TD strategy affords clinical benefit to those categorised as being high risk of severe course of CD. However, given that people are first stratified to high or low risk using PredictSURE-IBD and subsequently randomised to SU or TD treatment, the EAG considers that there is potential for misdiagnosis of people who are truly low risk but categorised as high risk to go undetected. However, analysis of those randomised to accelerated SU after determination of high or low risk of following a severe course of CD will provide additional data to inform estimates of prognostic accuracy.

6.2.2 Economic

The EAG's model offers methodological advantages when compared to the PredictSURE-IBD™ model. The main strength of the economic analysis is that it captures partial response to maintenance

treatment (as well as remission, relapse, surgery and post-surgical remission). The analysis also uses time to event data (TTE and TTS) more extensively than previous models. Furthermore, the EAG has conducted a series of scenario analyses exploring structural and parameter uncertainty in the economic model. The EAG also conducted a series of scenarios testing extreme clinical assumptions around the potential benefit of TD compared to SU in order to mitigate the concerns raised by the SMCs,

However, clinical data informing the analysis had to be derived from multiple sources. This approach is not ideal and creates a patchwork network of evidence, introducing uncertainty in the economic results. It is nonetheless, anticipated by the EAG that this problem will be potentially overcome when results from the PROFILE trial are available to populate the economic model.

The test accuracy in the base case economic model for PredictSURE IBD™ and in the scenario analysis for IBDX® is the same and assumed to be 100%. The only difference in the cost-effectiveness analyses of the two diagnostic tests is the cost of the test. This is unlikely to reflect the tests' actual accuracy in clinical practice; however no robust diagnostic data was found to inform this in the analysis.

The potential benefits of TD treatment for high-risk patients are dependent on two questions which remain answered: 1) do some high-risk patients derive a benefit from receiving IM treatment before moving to biologic treatment? 2) do SU high-risk patients have the same benefits as TD high-risk patients once they initiate the TD treatment pathway (i.e. treatment with anti-TNF). In the EAG's model, the potential disadvantage of waiting to initiate treatment with anti-TNF was only based on the increased risk of surgery in the SU arm, however, the negative impact of surgery in the analysis was not enough to offset the advantages of initial treatment with IM for SU patients.

Finally, the EAG acknowledges that AEs, specifically relating to long-term use of biologics, and the potential benefits associated with surgery were not included in the economic analysis. However, if AEs were included in the analysis, given that a higher proportion of patients receive biologic treatment in the TD arm, this would negatively impact the outcomes in the TD arm of the model compared to the SU arm. Similarly, even though the EAG has not captured the potential benefit of surgery in the economic analysis, it notes that to do so would benefit the SU strategy, as a higher proportion of patients receive surgery in the SU arm than in the TD arm of the model. Therefore, including AEs and the benefits of surgery in the analysis would not change the conclusions likely to be drawn from the current results.

7 CONCLUSIONS

7.1 Clinical effectiveness

Despite extensive systematic searches of the literature, no robust evidence was identified on the prognostic accuracy of the biomarker-stratification tools, IBDX and PredictSURE-IBD. In terms of sensitivity and specificity as estimates of prognostic accuracy, the External Assessment Group (EAG) is unaware of a validated definition for determination of whether a person has followed a severe course of CD, and, thus, considers the criterion required for a true positive or false positive for IBDX and PredictSURE-IBD to be unclear. The EAG considers it would be challenging to ascertain an accurate estimate of prognostic accuracy of the tools in stratifying course of CD as to do so would require carrying out a prospective study that included a group that received only "step-up" (SU) treatment after determination of risk of course of CD. The ongoing PROFILE RCT randomised people to accelerated SU or TD treatment after determination of high or low risk of following a severe course of CD and so will provide additional data to inform estimates of prognostic accuracy

Estimates of risk of experiencing a clinical outcome associated with severe course of CD were not available for comparable outcomes for IBDX and PredictSURE-IBD. Given the disparity in the outcomes assessed for IBDX and PredictSURE-IBD, the EAG considers that no conclusions can be drawn on the comparative effectiveness of the two tools in stratifying people by risk of severe course of CD.

7.2 Cost-effectiveness

Given the lack of robust evidence on the prognostic accuracy of the biomarker-stratification tools, the development of the economic model to assess the cost-effectiveness of IBDX® and PredictSURE-IBD™ consisted mainly of a theoretical exercise. The EAG anticipates that the economic model developed provides a structural framework for analysing future available data on prognostic accuracy and to assess the costs and consequences of treating high- and low-risk patients with both TD and SU strategies.

The economic model is ultimately assessing the cost-effectiveness of TD therapy compared to SU therapy for high-risk patients. However, the EAG did not identify any robust evidence on the latter thus, the clinical data informing the economic analysis had to be derived from multiple sources. This approach is not ideal and creates a patchwork network of evidence, introducing uncertainty in the economic results.

One of the key underlying assumption in the EAG's base case analysis is that high-risk patients who initiate treatment with IMs (SU arm) escalate treatment quicker than high-risk patients who initiate treatment with anti-TNF (supported by the data presented in D'Haens *et al*.[35]). However, once SU patients initiate treatment with an anti-TNF (their second treatment step), they "catch-up" with patients on the TD treatment strategy. As some high-risk patients who receive SU treatment respond to IM treatment, having the additional IM step in the SU strategy is advantageous to patients in the EAG's base case analysis as patients still subsequently receive treatment with biologics, which are assumed to have the same effect as biologics is the TD arm. Given the paucity of data to substantiate any further benefits in subsequent treatment steps in the TD vs SU approaches, the EAG considered this to be the most conservative modelling approach.

The EAG also notes that even though, in theory, a TD approach would suggest a "de-escalation" of treatments, the clinical experts advising the EAG consistently reported that IMs would not be given to patients who are responding well to biologics (instead treatment with biologics would just be continued until loss of response). The experts also explained that after loss of response with first- or second-line biologics, patients would not be given IMs, but instead surgery as a last treatment resource.

The long-term follow up study Hoekman *et al.*[112] found no difference in 10-year clinical remission rate; endoscopic remission, hospitalisation, surgery or new fistulas. Furthermore, the study concluded that in the long-term a TD strategy had not proven to alter the natural history of CD's disease. However, time to relapse was found statistically significantly different across TD and SU arms in the 2-year analysis of the data (D'Haens *et al*.)[35].

Hoekman *et al.*[112] concluded that their study was the first to compare the long-term outcomes for newly diagnosed CD patients who received combined immunosuppression vs conventional management. The authors added that early combined immunosuppression may be a preferential strategy given the associated delay in time to relapse. However, the authors noted that the costs and risks of potentially overtreating patients with a potentially 'benign' disease course mean that a TD approach should not be recommended as a universal treatment strategy for all patients with newly diagnosed CD.

The EAG's analysis has shown that too much uncertainty remains around the potential benefit of TD treatment for high-risk patients. The cost-effectiveness of a TD strategy compared with a SU strategy in high-risk patients is highly dependent on two unanswered questions: 1) do some high-risk patients derive a benefit from receiving IM treatment before moving to biologic treatment? 2) do SU high-risk patients have the same benefits as TD high-risk patients once they initiate the TD treatment pathway (i.e. treatment with anti-TNF)? In the EAG's model, the potential disadvantage of waiting to initiate treatment with anti-TNF was only based on the increased risk for surgeries in the SU arm, however, the

negative impact of surgeries in the analysis was not enough to offset the advantages on initial treatment with IM for SU patients.

For the reasons discussed above, most of the EAG's ICERs have shown that SC (and SU) dominates both diagnostic tools (and TD).In order to mitigate some of the concerns raised by the SMCs, the EAG conducted a range of analyses to test extreme scenarios around increasing the relative treatment effectiveness of the TD approach while decreasing the relative costs associated with TD. The EAG concluded that once the relative effectiveness of TD is artificially increased (through both TTE and probability of response to initial treatment) and combined with decreasing time on biologic treatment, the ICERs for PredictSURE IBD™ (and TD) compared with SC (and SU) are below £30,000. However, the EAG notes that these results need to be interpreted with extreme caution as the assumptions made in these scenarios were designed to test extreme clinical scenarios where TD was assumed to be more effective than SU. The EAG did not find any evidence to substantiate the benefits modelled in these scenarios, and thus concludes that its base case analysis showing that TD is dominated by SU remains the most conservative assessment of the relative cost-effectiveness of these treatment strategies.

Finally, the EAG acknowledges that AEs and the potential benefits associated with surgery were not included in the economic analysis. However, if AEs were included in the analysis, given that a higher proportion of patients receive biologic treatment in the TD arm, this would negatively impact the outcomes in the TD arm of the model. Similarly, even though the EAG has not captured the potential benefit of surgery in the economic analysis, it notes that to do so would benefit the SU strategy, as a higher proportion of patients receive surgery in the SU arm than on the TD arm of the model. Therefore, including AEs and the benefits of surgery in the analysis would contribute further for the dominance of SC over PredictSURE-IBD™.

7.3 Suggested research priorities

A high quality clinical trial that directly compares IBDX and PredictSURE-IBD would facilitate capturing robust data for sensitivity and specificity of the tools. The EAG considers it would be important to prespecify trial parameters, for example, eligible population, assessment of disease activity and severity at baseline, criteria for treatment escalation, and treatment algorithm. In addition, clinical experts would likely need to be consulted to determine which outcome would be the most appropriate measure for prognostic accuracy, for example, time to treatment escalation, development of a complication or need for surgery. An economic evaluation based on the results of the PROFILE RCT would also be warranted.

8 REFERENCES

1. Abraham C, Cho JH. Inflammatory bowel disease. *N Engl J Med* 2009; **361**: 2066-78.

2. Hendrickson BA, Gokhale R, Cho JH. Clinical aspects and pathophysiology of inflammatory bowel disease. *Clin Microbiol Rev* 2002; **15**: 79-94.

3. Park JH, Peyrin-Biroulet L, Eisenhut M, Shin JI. IBD immunopathogenesis: A comprehensive review of inflammatory molecules. *Autoimmun Rev* 2017; **16**: 416-26.

4. Crohn's and Colitis UK. Crohn's disease. 2019. Available from: https://www.crohnsandcolitis.org.uk/about-inflammatory-bowel-disease/crohns-disease. Date accessed: 24 July 2019.

5. Johns Hopkins Medicine. Crohn's disease: introduction. 2001. Available from: https://www.hopkinsmedicine.org/gastroenterology_hepatology/_pdfs/small_large_intestine/crohns_disease.pdf. Date accessed: 5 Aug 2019.

6. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology* 2011; **140**: 1785-94.

7. York Teaching Hospital. Crohn's disease. 2019. Available from: https://www.yorkhospitals.nhs.uk/our-services/a-z-of-services/inflammatory-bowel-disease-at-york-hospital/crohns-disease/. Date accessed: 24 July 2019.

8. Ananthakrishnan AN. Environmental triggers for inflammatory bowel disease. *Curr Gastroenterol Rep* 2013; **15**: 302.

9. Feagins LA, Iqbal R, Spechler SJ. Case-control study of factors that trigger inflammatory bowel disease flares. *World J Gastroenterol* 2014; **20**: 4329-34.

10. Langholz E. Current trends in inflammatory bowel disease: the natural history. *Therap Adv Gastroenterol* 2010; **3**: 77-86.

11. Loftus EV, Jr., Schoenfeld P, Sandborn WJ. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. *Aliment Pharmacol Ther* 2002; **16**: 51-60.

12. Munkholm P, Langholz E, Davidsen M, Binder V. Disease activity courses in a regional cohort of Crohn's disease patients. *Scand J Gastroenterol* 1995; **30**: 699-706.

13. Lapidus A, Bernell O, Hellers G, Lofberg R. Clinical course of colorectal Crohn's disease: a 35-year follow-up study of 507 patients. *Gastroenterology* 1998; **114**: 1151-60.

14. Regueiro M, Hashash JA. Overview of the medical management of mild (low risk) Crohn disease in adults. 2018. Available from: https://www.uptodate.com/contents/overview-of-the-medical-management-of-mild-low-risk-crohn-disease-in-adults. Date accessed: 18 Apr 2019.

15. Gomollon F, Dignass A, Annese V, Tilg H, Van Assche G, Lindsay JO, et al. 3rd European Evidence-based Consensus on the Diagnosis and Management of Crohn's Disease 2016: Part 1: Diagnosis and Medical Management. *J Crohns Colitis* 2017; **11**: 3-25.

16. Duricova D, Burisch J, Jess T, Gower-Rousseau C, Lakatos PL, EpiCom E. Age-related differences in presentation and course of inflammatory bowel disease: an update on the population-based literature. *J Crohns Colitis* 2014; **8**: 1351-61.

17. Hovde O, Moum BA. Epidemiology and clinical course of Crohn's disease: results from observational studies. *World J Gastroenterol* 2012; **18**: 1723-31.

18. Garcia Rodriguez LA, Gonzalez-Perez A, Johansson S, Wallander MA. Risk factors for inflammatory bowel disease in the general population. *Aliment Pharmacol Ther* 2005; **22**: 309-15.

19. Rubin GP, Hungin AP, Kelly PJ, Ling J. Inflammatory bowel disease: epidemiology and management in an English general practice population. *Aliment Pharmacol Ther* 2000; **14**: 1553-9.

20. Floyd DN, Langham S, Severac HC, Levesque BG. The economic and quality-of-life burden of Crohn's disease in Europe and the United States, 2000 to 2013: a systematic review. *Dig Dis Sci* 2015; **60**: 299-312.

21. Ghosh N, Premchand P. A UK cost of care model for inflammatory bowel disease. *Frontline Gastroenterol* 2015; **6**: 169-74.

22. Shivananda S, Hordijk ML, Pena AS, Mayberry JF. Crohn's disease: risk of recurrence and reoperation in a defined population. *Gut* 1989; **30**: 990-5.

23. Maaser C, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annese V, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019; **13**: 144-64.

24. Sturm A, Maaser C, Calabrese E, Annese V, Fiorino G, Kucharzik T, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 2: IBD scores and general principles and technical aspects. *J Crohns Colitis* 2019; **13**: 273-84.

25. Peyrin-Biroulet L, Panes J, Sandborn WJ, Vermeire S, Danese S, Feagan BG, et al. Defining Disease Severity in Inflammatory Bowel Diseases: Current and Future Directions. *Clin Gastroenterol Hepatol* 2016; **14**: 348-54 e17.

26. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998; **43**: 29-32.

27. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980; **1**: 514.

28. Vermeire S, Schreiber S, Sandborn WJ, Dubois C, Rutgeerts P. Correlation between the Crohn's disease activity and Harvey-Bradshaw indices in assessing Crohn's disease severity. *Clin Gastroenterol Hepatol* 2010; **8**: 357-63.

29. Iskandar HN, Ciorba MA. Biomarkers in inflammatory bowel disease: current practices and recent advances. *Transl Res* 2012; **159**: 313-25.

30. National Institute for Health and Care Excellence. Crohn’s disease: management (NG129). 2019. Available from: https://www.nice.org.uk/guidance/ng129/chapter/Recommendations#inducing-remission-in-crohns-disease. Date accessed: 5 Aug 2019.

31. Rogler G. Top-down or step-up treatment in Crohn's disease? *Dig Dis* 2013; **31**: 83-90.

32. Tsui JJ, Huynh HQ. Is top-down therapy a more effective alternative to conventional step-up therapy for Crohn's disease? *Ann Gastroenterol* 2018; **31**: 413-24.

33. MP. S, GY. M, S. D, P. K, L. R, Jr. LE, et al. De-escalating medical therapy in Crohn's disease patients who are in deep remission: A RAND appropriateness panel. *GastroHep* 2019; **1**: 108-17.

34. Colombel JF, Sandborn WJ, Reinisch W, Mantzaris GJ, Kornbluth A, Rachmilewitz D, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med* 2010; **362**: 1383-95.

35. D'Haens G, Baert F, van Assche G, Caenepeel P, Vergauwe P, Tuynman H, et al. Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. *Lancet* 2008; **371**: 660-7.

36. Fan R, Zhong J, Wang ZT, Li SY, Zhou J, Tang YH. Evaluation of "top-down" treatment of early Crohn's disease by double balloon enteroscopy. *World J Gastroenterol* 2014; **20**: 14479-87.

37. Khanna R, Bressler B, Levesque BG, Zou G, Stitt LW, Greenberg GR, et al. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* 2015; **386**: 1825-34.

38. Kaul A, Hutfless S, Liu L, Bayless TM, Marohn MR, Li X. Serum anti-glycan antibody biomarkers for inflammatory bowel disease diagnosis and progression: a systematic review and meta-analysis. *Inflamm Bowel Dis* 2012; **18**: 1872-84.

39. Segal AW. Making sense of the cause of Crohn's - a new look at an old disease. *F1000Res* 2016; **5**: 2510.

40. Mitsuyama K, Niwa M, Takedatsu H, Yamasaki H, Kuwaki K, Yoshioka S, et al. Antibody markers in the diagnosis of inflammatory bowel disease. *World J Gastroenterol* 2016; **22**: 1304-10.

41. Kamm F, Strauch U, Degenhardt F, Lopez R, Kunst C, Rogler G, et al. Serum anti-glycan-antibodies in relatives of patients with inflammatory bowel disease. *PLoS ONE* 2018; **13**: e0194222.

42. Glycominds International. Crohn's Disease: Prognostic Serological Marker Profile. 2016. Available from: http://ibdx.net/assets/img/flyers/cd_prognosis.pdf. Date accessed: 08 Aug 2019.

43. Glycominds International. IBDX® ALCA IgG ELISA Kit. 2019. Available from: http://www.ibdx.net/assets/img/product/alca_insert.pdf. Date accessed: 08 Aug 2019.

44. Papp M, Foldi I, Altorjay I, Palyu E, Udvardy M, Tumpek J, et al. Anti-microbial antibodies in celiac disease: trick or treat? *World J Gastroenterol* 2009; **15**: 3891-900.

45. Roman AR, Munoz F. Comorbidity in inflammatory bowel disease. *World J Gastroenterol* 2011; **17**: 2723-33.

46. Lee JC, Lyons PA, McKinney EF, Sowerby JM, Carr EJ, Bredin F, et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest* 2011; **121**: 4170-9.

47. McKinney EF, Lee JC, Jayne DR, Lyons PA, Smith KG. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature* 2015; **523**: 612-6.

48. McKinney EF, Lyons PA, Carr EJ, Hollis JL, Jayne DR, Willcocks LC, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nat Med* 2010; **16**: 586-91, 1p following 91.

49. Yi JS, Cox MA, Zajac AJ. T-cell exhaustion: characteristics, causes and conversion. *Immunology* 2010; **129**: 474-81.

50. Biasci D, Lee JC, Noor NM, Pombal DR, Hou M, Lewis N, et al. A blood-based prognostic biomarker in IBD. *Gut* 2019.

51. Parkes M, Noor NM, Dowling F, Leung H, Bond S, Whitehead L, et al. PRedicting Outcomes For Crohn's dIsease using a moLecular biomarkEr (PROFILE): protocol for a multicentre, randomised, biomarker-stratified trial. *BMJ Open* 2018; **8**: e026767.

52. National Institute for Health and Care Excellence. PredictSURE-IBD and IBDX to guide personalised treatment of Crohn's disease. Final scope. 2019. Available from: https://www.nice.org.uk/guidance/gid-dg10029/documents/final-scope. Date accessed: 06 Sept 2019.

53. Glycominds International. Non-Invasive Solutions to Gastrointestinal Diseases Detection. 2019. Available from: https://www.glycominds.com/. Date accessed: 20 Aug 2019.

54. PredictImmune. PredictImmune. 2019. Available from: https://www.predictimmune.com/. Date accessed: 20 Aug 2019.

55. Barton S, Edwards SJ, Bacelar M, Karner C, P. C, S. S. PredictSURE-IBD and IBDX to guide personalised treatment of Crohn's disease in adults. 2019. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=138737. Date accessed: 20 Aug 2019.

56. Centre for Reviews and Dissemination (CRD). CRD's guidance for undertaking reviews in healthcare. 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Date accessed: 20 Aug 2019.

57. National Institute for Health and Care Excellence. Diagnostics Assessment Programme manual. 2011. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf. Date accessed: 20 Aug 2019.

58. Cochrane Methods. Screening and Diagnostic Tests. Handbook for DTA reviews. 2018. Available from: https://methods.cochrane.org/sdt/handbook-dta-reviews. Date accessed: 20 Aug 2019.

59. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; **144**: 427-37.

60. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013; **158**: 280-6.

61. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019; **170**: W1-W33.

62. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019; **170**: 51-8.

63. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions v.5.1.0. 2011. Available from: http://handbook-5-1.cochrane.org/. Date accessed: 18 Apr 2019.

64. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016; **355**: i4919.

65. Critical Appraisal Skills Programme. CASP Checklist: 10 questions to help you make sense of a Qualitative research. 2018. Available from: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist.pdf. Date accessed: 18 Apr 2019.

66. Bonneau J, Dumestre-Perard C, Rinaudo-Gaujous M, Genin C, Sparrow M, Roblin X, et al. Systematic review: new serological markers (anti-glycan, anti-GP2, anti-GM-CSF Ab) in the prediction of IBD patient outcomes. *Autoimmun Rev* 2015; **14**: 231-45.

67. Harrell L, Weyer G, Yarden J, Dotan N, Hanauer S, editors. T1289. Anti-glycan antibodies are associated with disabling disease course and complicated disease behavior in patients with Crohn's disease.2010: Gastroenterology.

68. Kaul A HS, Liu L, Bayless TM, Marohn MR, Li X., editor Serum anti-glycan antibody biomarkers for inflammatory bowel disease diagnosis and progression: A meta-analysis.2011: Gastroenterology.

69. Paul S, Boschetti G, Rinaudo-Gaujous M, Moreau A, Del Tedesco E, Bonneau J, et al. Association of Anti-glycan Antibodies and Inflammatory Bowel Disease Course. *J Crohns Colitis* 2015; **9**: 445-51.

70. Prideaux L, De Cruz P, Ng SC, Kamm MA. Serological antibodies in inflammatory bowel disease: a systematic review. *Inflamm Bowel Dis* 2012; **18**: 1340-55.

71. Rieder F, Hahn P, Finsterhoelzl L, Dirmeier A, Cai H, Shen B, et al., editors. Serum anti-glycan antibodies can contribute to differential diagnosis and disease stratification of pediatric Crohn's disease patients.2010: Gastroenterology.

72. Rieder F, Hahn P, Finsterhoelzl L, Schleder S, Wolf A, Dirmeier A, et al. Clinical utility of anti-glycan antibodies in pediatric Crohn's disease in comparison with an adult cohort. *Inflamm Bowel Dis* 2012; **18**: 1221-31.

73. Rieder F, Hahn P, Finsterholzl L, Dirmeier A, Shen B, Rogler G, et al., editors. Clinical utility of anti-glycan antibodies in pediatric Crohn's disease in comparison with an adult cohort.2011: J Crohn's Colitis.

74. Rieder F, Lopez R, Franke A, Wolf A, Schleder S, Dirmeier A, et al. Characterization of changes in serum anti-glycan antibodies in Crohn's disease--a longitudinal analysis. *PLoS ONE* 2011; **6**: e18172.

75. Rieder F, Schleder S, Wolf A, Dirmeier A, Strauch U, Obermeier F, et al. Association of the novel serologic anti-glycan antibodies anti-laminarin and anti-chitin with complicated Crohn's disease behavior. *Inflamm Bowel Dis* 2010; **16**: 263-74.

76. Rieder F, Schleder S, Wolf A, Dirmeier A, Strauch U, Obermeier F, et al. Serum anti-glycan antibodies predict complicated Crohn's disease behavior: a cohort study. *Inflamm Bowel Dis* 2010; **16**: 1367-75.

77. Rieder F, Schleder S, Wolf A, Schirbel A, Franke A, Dirmeier A, et al., editors. Characterization of changes of serum anti-glycan antibodies in individual patients over time in inflammatory bowel disease (IBD).2010: Gastroenterology.

78. Seow CH, Stempak JM, Xu W, Lan H, Griffiths AM, Greenberg GR, et al. Novel anti-glycan antibodies related to inflammatory bowel disease diagnosis and phenotype. *Am J Gastroenterol* 2009; **104**: 1426-34.

79. Wolfel G, Lopez R, Hosl J, Kunst C, Martina M, Rieder F, editors. The anti-glycan antibodies amca and alca are associated with shorter time to surgical recurrence in Crohn's disease (CD).2017: Gastroenterology.

80. Halder SL, Stempak JM, Sharaf A, Xu W, Greenberg GR, Steinhart H, et al., editors. Biomarkers associated with progressive behaviour in Crohn's disease.2010: Gastroenterology.

81. International Standard Randomised Controlled Trials Number Registry. PROFILE - personalised medicine in Crohn's disease. 2019. Available from: http://www.isrctn.com/ISRCTN11808228. Date accessed: 9 Oct 2019.

82. ClinicalTrials.gov. The PRECIOUS Study: Predicting Crohn's & ColitIs Outcomes in the United States. 2019. Available from: https://clinicaltrials.gov/ct2/show/NCT03952364. Date accessed: 02 Sept 2019.

83. Smids C, Horjus Talabur Horje CS, Drylewicz J, Roosenboom B, Groenen MJM, van Koolwijk E, et al. Intestinal T Cell Profiling in Inflammatory Bowel Disease: Linking T Cell Subsets to Disease Activity and Disease Course. *J Crohns Colitis* 2018; **12**: 465-75.

84. Lyons PA, Noor NM, Lee JC, McKinney EF, Parkes M, KGC. S. Anti-glycan antibody seropositivity at diagnosis does not predict future disease course in patients with Crohn's disease. Submitted to United European Gastroenterology conference2019.

85. Arber M, Garcia S, Veale T, Edwards M, Shaw A, Glanville JM. Performance of Ovid Medline Search Filters to Identify Health State Utility Studies. *Int J Technol Assess Health Care* 2017; **33**: 472-80.

86. Marchetti M, Liberato NL, Di Sabatino A, Corazza GR. Cost-effectiveness analysis of top-down versus step-up strategies in patients with newly diagnosed active luminal Crohn's disease. *European Journal of Health Economics* 2013; **14**: 853-61.

87. Dretzke J, Edliln R, Round J, Connock M, Hulme C, Czeczot J, et al. A systematic review and economic evaluation of the use of tumour necrosis factor-alpha (TNF-a) inhibitors, adalimumab and infliximab, for Crohn's disease. *Health Technology Assessment* 2011; **15**: 1-250.

88. Rafia R, Scope A, Harnan S, Stevens JW, Stevenson M, Lobo A. Vedolizumab for Treating Moderately to Severely Active Crohn's Disease After Prior Therapy: An Evidence Review Group Perspective of a NICE Single Technology Appraisal. *PharmacoEconomics* 2016; **34**: 1241-53.

89. Hodgson R, Walton M, Biswas M, Mebrahtu T, Woolacott N. Ustekinumab for Treating Moderately to Severely Active Crohn's Disease after Prior Therapy: An Evidence Review Group Perspective of a NICE Single Technology Appraisal. *PharmacoEconomics* 2018; **36**: 387-98.

90. Freeman K, Connock M, Auguste P, Taylor-Phillips S, Mistry H, Shyangdan D, et al. Clinical effectiveness and cost-effectiveness of use of therapeutic monitoring of tumour necrosis factor alpha (TNF-alpha) inhibitors [LISA-TRACKER enzyme-linked immunosorbent assay (ELISA) kits, TNF-

alpha-Blocker ELISA kits and Promonitor ELISA kits] versus standard care in patients with Crohn's disease: Systematic reviews and economic modelling. *Health Technology Assessment* 2016; **20**.

91. Bodger K, Kikuchi T, Hughes D. Cost-effectiveness of biological therapy for Crohn's disease: Markov cohort analyses incorporating United Kingdom patient-level cost data. *Alimentary pharmacology & therapeutics* 2009; **30**: 265-74.

92. Saito S, Shimizu U, Nan Z, Mandai N, Yokoyama J, Terajima K, et al. Economic impact of combination therapy with infliximab plus azathioprine for drug-refractory Crohn's disease: A cost-effectiveness analysis. *Journal of Crohn's and Colitis* 2013; **7**: 167-74.

93. Benedini V, Caporaso N, Corazza GR, Rossi Z, Fornaciari G, Cottone M, et al. Burden of Crohn's disease: Economics and quality of life aspects in Italy. *ClinicoEconomics and Outcomes Research* 2012; **4**: 209-18.

94. Mozzi A, Meregaglia M, Lazzaro C, Tornatore V, Belfiglio M, Fattore G. A comparison of EuroQol 5-dimension health-related utilities using Italian, UK, and US preference weights in a patient sample. *ClinicoEconomics and Outcomes Research* 2016; **8**: 267-74.

95. Stark RG, Reitmeir P, Leidl R, Konig HH. Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. *Inflammatory Bowel Diseases* 2010; **16**: 42-51.

96. Rencz F, Lakatos PL, Gulacsi L, Brodszky V, Kurti Z, Lovas S, et al. Validity of the EQ-5D-5L and EQ-5D-3L in patients with Crohn's disease. *Quality of Life Research* 2019; **28**: 141-52.

97. Dolan P, Gudex C, Kind P ea. A Social Tariff for EuroQol: Results from a UK General Population Survey, Discussion Paper 138. *New York, NY: Centre of Health Economics, University of York* 1995.

98. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making* 2001; **21**: 7-16.

99. Dolan P, Gudex C, Kind P ea. The time trade-off method: results from a general population study. *Health Econ* 1996; **5**: 141-54.

100. Dolan P. Modeling valuations for EuroQol health states. *Medical Care* 1997; **35**: 1095-108.

101. Badia X, Roset M, Montserrat S, Herdman M, Segura A. [The Spanish version of EuroQol: a description and its applications. European Quality of Life scale]. *Med Clin (Barc)* 1999; **112 Suppl 1**: 79-85.

102. Casellas F, Vivancos JL, Sampedro M, Malagelada JR. Relevance of the phenotypic characteristics of crohn's disease in patient perception of health-related quality of life. *American Journal of Gastroenterology* 2005; **100**: 2737-42.

103. Casellas F, Rodrigo L, Nino P, Pantiga C, Riestra S, Malagelada JR. Sustained improvement of health-related quality of life in Crohn's disease patients treated with infliximab and azathioprine for 4 years. *Inflammatory Bowel Diseases* 2007; **13**: 1395-400.

104. Huaman JW, Casellas F, Borruel N, Pelaez A, Torrejon A, Castells I, et al. Cutoff values of the Inflammatory Bowel Disease Questionnaire to predict a normal health related quality of life. *Journal of Crohn's and Colitis* 2010; **4**: 637-41.

105. Rue M, Badia X. The Spanish EuroQol tariff: Results from the Catalan health survey based on self-rated health. In: Badia X, Herdman M, Segura A, eds. EuroQol, plenary meeting. Barcelona: Institut Universitari de Salut Publiea de Catalunya. 1996: 77-98.

106. Casellas F, Lopez-Vivancos J, Badia X, Vilaseca J, Malagelada JR. Impact of surgery for Crohn's disease on health-related quality of life. *American Journal of Gastroenterology* 2000; **95**: 177-82.

107. Casellas F, Arenas JI, Baudet JS, Fabregas S, Garcia N, Gelabert J, et al. Impairment of health-related quality of life in patients with inflammatory bowel disease: A spanish multicenter study. *Inflammatory Bowel Diseases* 2005; **11**: 488-96.

108. Saro C, Ceballos D, Munoz F, De-La-Coba C, Aguilar MD, Lazaro P, et al. Clinical status, quality of life, and work productivity in Crohn's disease patients after one year of treatment with adalimumab. *Revista Espanola de Enfermedades Digestivas* 2017; **109**: 122-9.

109. Holko P, Kawalec P, Mossakowska M, Pilc A. Health-Related Quality of Life Impairment and Indirect Cost of Crohn's Disease: A Self-Report Study in Poland. *PLoS ONE* 2016; **11**: e0168586.

110. Golicki D, Jakubczyk M, Niewada M, Wrona W, JJ. B. Valuation of EQ-5D health states in Poland: first TTO based social value set in Central and Eastern Europe. *Value Health* 2010; **13**: 289-97.

111. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ* 2018; **27**: 7-22.

112. Hoekman D, Stibbe J, Baert F, Caenepeel P, Vergauwe P, De VM, et al. Long-term Outcome of Early Combined Immunosuppression Versus Conventional Management in Newly Diagnosed Crohn's Disease. 2018. Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01656067/full. Date accessed.

113. Latimer N. NICE DSU Technical Support Document 14. Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2013. Available from: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf. Date accessed: November 2019.

114. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 2012; **12**: 9.

115. Baio G. survHE: Survival Analysis in Health Economic

Evaluation. R package version 1.0.65. https://CRAN.R-project.org/package=survHE. 2018.

116. Tsui JJ, Huynh HQ. Is top-down therapy a more effective alternative to conventional step-up therapy for crohn's disease? *Annals of Gastroenterology* 2018; **31**: 413-24.

117. British National Formulary. BNF. 2019. Available from: https://bnf.nice.org.uk/. Date accessed: 30 Oct 2019.

118. National Institute for Health and Care Excellence (NICE). Vedolizumab for treating moderately to severely active Crohn's disease after prior therapy. Technology appraisal guidance [TA352]. 2015. Available from: https://www.nice.org.uk/guidance/ta352. Date accessed: July 2019.

119. National Institute for Health and Care Excellence (NICE). Ustekinumab for moderately to severely active Crohn's disease after previous treatment. Technology appraisal guidance [TA456]. 2017. Available from: https://www.nice.org.uk/guidance/ta456. Date accessed: July 2019.

120. Hanauer SB, Sandborn WJ, Rutgeerts P, Fedorak RN, Lukas M, MacIntosh D, et al. Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn's disease: the CLASSIC-I trial. *Gastroenterology* 2006; **130**: 323-33; quiz 591.

121. Watanabe M, Hibi T, Lomax KG, Paulson SK, Chao J, Alam MS, et al. Adalimumab for the induction and maintenance of clinical remission in Japanese patients with Crohn's disease. *J Crohns Colitis* 2012; **6**: 160-73.

122. Feagan BG, Sandborn WJ, Gasink C, Jacobstein D, Lang Y, Friedman JR, et al. Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. *N Engl J Med* 2016; **375**: 1946-60.

123. Sandborn WJ, Feagan BG, Rutgeerts P, Hanauer S, Colombel JF, Sands BE, et al. Vedolizumab as induction and maintenance therapy for Crohn's disease. *N Engl J Med* 2013; **369**: 711-21.

124. Sandborn WJ, Gasink C, Gao LL, Blank MA, Johanns J, Guzzo C, et al. Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N Engl J Med* 2012; **367**: 1519-28.

125. Sands BE, Feagan BG, Rutgeerts P, Colombel JF, Sandborn WJ, Sy R, et al. Effects of vedolizumab induction therapy for patients with Crohn's disease in whom tumor necrosis factor antagonist treatment failed. *Gastroenterology* 2014; **147**: 618-27 e3.

126. Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JF, et al. Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet* 2002; **359**: 1541-9.

127. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

128. Office for National Statistics (ONS). National life tables: England and Wales. 2015-17. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandandwalesreferencetables. Date accessed: Oct 2019.

129. Silverstein MD, Loftus EV, Jr., Sandborn WJ, Tremaine WJ, Feagan BG, Nietert PJ, et al. Clinical course and costs of care for Crohn's disease: Markov model analysis of a population-based cohort. *Gastroenterology* 1999; **117**: 49-57.

130. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010; **13**: 509-18.

131. Office of National Statistics (ONS). Consumer Price Inflation Index for Medical Services (DKC3). 2018. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/dkc3/mm23. Date accessed: 12 Dec 2018.

132. Department of Health and Social Care (NHS Improvement). NHS National Schedule of reference costs 2017 to 2018. Accessed November 2019.

133. Decision Support Unit NICE. Use of tumour necrosis factor alpha (TNF a) inhibitors (adalimumab and infliximab) for Crohn's disease (2009). Report by the Decision Support Unit NICE 2009.

134. Andrew D. Martin KMQ, Jong Hee Park.,. MCMCpack:

Markov Chain Monte Carlo in R. Journal of Statistical Software.

42(9): 1-21. URL http://www.jstatsoft.org/v42/i09/. 2011.

135. Venables WN, Ripley BD. *Modern Applied Statistics with S. Fourth Edition*. New York: Springer; 2002.

136. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. 2013. Available from: https://www.nice.org.uk/process/pmg9/chapter/foreword. Date accessed: 18 December 2018.

137. Beaven SW, Abreu MT. Biomarkers in inflammatory bowel disease. *Curr Opin Gastroenterol* 2004; **20**: 318-27.

138. Dotan I, Fishman S, Dgani Y, Schwartz M, Karban A, Lerner A, et al. Antibodies against laminaribioside and chitobioside are novel serologic markers in Crohn's disease. *Gastroenterology* 2006; **131**: 366-78.

139. Dotan I, Meringer H, Lerner A, Naftali T, Yaron A, Reif S, et al., editors. Anti-glycan antibodies are significantly increased in Crohn's disease patients and their first degree relatives.2010: Gastroenterology.

140. Fengming Y, Jianbing W. Biomarkers of inflammatory bowel disease. *Dis Markers* 2014; **2014**: 710915.

141. Gasparetto M, Nayak K, Lee C, Kraiczy J, Payne F, Torrente F, et al. Investigating CD8+ T-cell gene expression signatures as potential prognostic biomarkers in paediatric inflammatory bowel disease. *J Ped Gastroenterol Nutrition* 2018; **66**: 61-2.

142. Koutroubakis IE, Drygiannakis D, Tsirogianni A, Oustamanolakis P, Karmiris K, Papamichael K, et al. Antiglycan antibodies in Greek patients with inflammatory bowel disease. *Dig Dis Sci* 2011; **56**: 845-52.

143. Lee JC, Lyons P, Parkes M, Smith KG, editors. A CD8 T cell gene expression signature predicts disease behaviour in inflammatory bowel disease.2011: Gut.

144. Lee JC, Lyons P, McKinney E, Carr E, Rayner T, Parkes M, et al., editors. Gene expression profiling in CD8 T cells predicts disease course in Crohn's disease and ulcerative colitis.2011: J Crohn's Colitis.

145. Lee JC. Predicting the course of IBD: light at the end of the tunnel? *Dig Dis* 2012; **30 Suppl 1**: 95-9.

146. Lee J, Biasci D, Noor N, McKinney E, Ahmad T, Lewis N, et al. PROFILE trial: Predicting outcomes for Crohn's disease using a molecular biomarker. *Journal of Crohn's and Colitis* 2017; **11**: S55.

147. Lee JC, Biasci D, Noor NM, McKinney EF, Ahmad T, Lewis NR, et al. Profile trial: Predicting outcomes for Crohn's disease using a molecular biomarker. *United European Gastroenterology Journal* 2017; **5**: A98.

148. Lee JC, Biasci D, Noor NM, McKinney EF, Ahmad T, Lewis NR, et al. Profile trial: Predicting outcomes for Crohn's disease using a molecular biomarker. *Gut* 2017; **66**: A22-A3.

149. Lyons PA, editor A blood-based prognostic biomarker in inflammatory bowel disease; towards personalised medicine in IBD.2019: Gastroenterology.

150. Papp M, Lakatos PL. Serological studies in inflammatory bowel disease: how important are they? *Curr Opin Gastroenterol* 2014; **30**: 359-64.

151. Rieder F, Schleder S, Wolf A, Schirbel A, Dirmeier A, Obermeier F, et al., editors. Level changes of serum anti-glycan antibodies in individual Crohn's disease patients over time.2011: J Crohn's Colitis.

152. Rieder F, Franke A, Lopez R, Huebenthal M, Dirmeier A, Wang MH, et al., editors. Genetic risk profiling alone or in combination with serum anti-microbial antibodies for the stratification of complicated crohn's disease courses.2013: Gastroenterology.

153. Ryan JD, Silverberg MS, Xu W, Graff LA, Targownik LE, Walker JR, et al. Predicting complicated Crohn's disease and surgery: Phenotypes, genetics, serology and psychological characteristics of a population-based cohort. *Alimentary Pharmacol Therapeutics* 2013; **38**: 274-83.

154. Simondi D, Mengozzi G, Betteto S, Bonardi R, Ghignone RP, Fagoonee S, et al. Antiglycan antibodies as serological markers in the differential diagnosis of inflammatory bowel disease. *Inflamm Bowel Dis* 2008; **14**: 645-51.

155. Chau NA, Romero D, Toolsie P, Jimmy J, Bleidt BA. Assessment of biological agents in the treatment regimen of moderate to severe Crohn's disease. *Pharmacotherapy* 2015; **35**: e299-e300.

156. Colombel JF, Ungaro R, Aggarwal S, Topaloglu O, Skup M, Lee WJ. Efficacy and safety of early biologic treatment of Crohn's disease in adult and paediatric patients: A systematic review. *Journal of Crohn's and Colitis* 2018; **12**: S461.

157. Fan R, Zhong J, Wang ZT, Li SY, Zhou J, Tang YH. Evaluation of "top-down" treatment of early Crohn's disease by double balloon enteroscopy. *World Journal of Gastroenterology* 2014; **20**: 14479-87.

158. Hirschmann S, Neurath MF. Top-down approach to biological therapy of Crohn's disease. *Expert Opinion on Biological Therapy* 2017; **17**: 285-93.

159. Hommes DW. Step-Up Versus Top-Down Therapy in the Treatment of Crohn's Disease. *Gastroenterology & hepatology* 2006; **2**: 546-7.

160. Hutfless S, Lau B, D. y, Wilson LM, Lazarev M, Bass EB. *Pharmacological Management of Crohn's Disease: Future Research Needs. Identification of Future Research Needs From Comparative Effectiveness Review No. 131*2014.

161. Katz JA. Postoperative endoscopic surveillance in Crohn's disease: bottom up or top down? *Gastrointestinal Endoscopy* 2007; **66**: 541-3.

162. Kuznar W, Writer M. Step-up therapy program for anti-inflammatory biologic agents does not increase cost nor adversely affect patient outcomes. *American Health and Drug Benefits* 2013; **6**.

163. Meier J, Sturm A. Top-down versus step-up: New strategies in the treatment of Crohn's disease. *Zeitschrift fur Gastroenterologie* 2009; **47**: 240-2.

164. Parkes M, Noor NM, Dowling F, Leung H, Bond S, Whitehead L, et al. PRedicting Outcomes for Crohn's dIsease using a moLecular biomarkEr (PROFILE): Protocol for a multicentre, randomised, biomarker-stratified trial. *BMJ Open* 2018; **8**: e026767.

165. Peyrin-Biroulet L, Bigard MA, Malesci A, Danese S. Step-Up and Top-Down Approaches to the Treatment of Crohn's Disease: Early May Already Be Too Late. *Gastroenterology* 2008; **135**: 1420-2.

166. Sucong L, Baili C, Yinglian X, Kang C, Yao H, Zhirong Z, et al. Endoscopic and clinical follow-up of step-up and top-down Infliximab therapy in Crohn's disease. *Journal of Gastroenterology and Hepatology* 2013; **28**: 369.

167. Xiao Y, Chen B, He Y, Gao X, Huang M, Hu P, et al. The clinical and endoscopic efficacy of step-up and top-down infliximab therapy in Crohn's disease. 2012. Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-00914623/full. Date accessed: 2012.

168. Colombel JF, Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Panaccione R, et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: the CHARM trial. *Gastroenterology* 2007; **132**: 52-65.

169. Sandborn WJ, Hanauer SB, Rutgeerts P, Fedorak RN, Lukas M, MacIntosh DG, et al. Adalimumab for maintenance treatment of Crohn's disease: results of the CLASSIC II trial. *Gut* 2007; **56**: 1232-9.

170. Rutgeerts P, Van Assche G, Sandborn WJ, Wolf DC, Geboes K, Colombel JF, et al. Adalimumab induces and maintains mucosal healing in patients with Crohn's disease: data from the EXTEND trial. *Gastroenterology* 2012; **142**: 1102-11 e2.

171. Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Colombel JF, Panaccione R, et al. Adalimumab induction therapy for Crohn disease previously treated with infliximab: a randomized trial. *Ann Intern Med* 2007; **146**: 829-38.

172. Targan SR, Hanauer SB, van Deventer SJ, Mayer L, Present DH, Braakman T, et al. A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. *N Engl J Med* 1997; **337**: 1029-35.

173. Baert, F.J., Moortgat, L., Van Assche, G.A., Caenepeel, P., Vergauwe, P.L., De Vos, M., Stokkers, P.C., Hommes, D.W., Vermeire, S., Ritgeerts, P.J., Feagan, B.G., D'Haens, G.: Mucosal healing predicts sustained clinical remission in early Crohn's disease. Gastroenterology 134 (suppl. 1), A640 2008

9 APPENDICES

9.1 Appendix 1. PRISMA DTA checklist

Section/topic	#	PRISMA-DTA Checklist item	Reported on page #
TITLE/ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	i
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	v–vi
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1–15
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	10–15
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	15
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	16
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	18–20

Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	16–17
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	173–176
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	21
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	21
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	18–20
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	21
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	18–20
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	22
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			

Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	23–28, 192–193
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	23–31, 199–250
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	32–35, 299–329
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	38–39
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	N/A

DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	48–50, 142
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	32–33, 36, 145–146
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	148
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	vi

Abbreviations: DTA, diagnostic test accuracy; FN, false negative; FP, false negative; ROC, receiver operator characteristic; TN, true negative; TP, true positive.

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

9.2 Appendix 2. PRISMA DTA for abstracts checklist

Section/topic	#	PRISMA-DTA for Abstracts Checklist item	Reported on page #
TITLE and PURPOSE			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic accuracy (DTA) studies.	v
Objectives	2	Indication the research question, including components such as participants, index test, and target conditions.	v
METHODS			
Eligibility criteria	3	Include study characteristics used as criteria for eligibility.	v–vi
Information sources	4	List the key databases searched and the search dates.	v
Risk of bias and applicability	5	Indicate the methods of assessing risk of bias and applicability.	v
Synthesis of results	A1	Indicate the methods for the data synthesis.	N/A
RESULTS			
Included studies	6	Indicate the number and type of included studies and the participants and relevant characteristics of the studies (including the reference standard).	v
Synthesis of results	7	Include the results for the analysis of diagnostic accuracy, preferably indicating the number of studies and participants. Describe test accuracy including variability; if meta-analysis was done, include summary results and confidence intervals.	v
DISCUSSION			
Strengths and limitations	9	Provide a brief summary of the strengths and limitations of the evidence.	v–vi
Interpretation	10	Provide a general interpretation of the results and the important implications.	v–vi
OTHER			

Funding	11	Indicate the primary source of funding for the review.	vi
Registration	12	Provide the registration number and the registry name.	vi

Abbreviations: DTA, diagnostic test accuracy.

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

9.3 Appendix 3. Search strategies

9.3.1 Prognostic accuracy and clinical impact

Database searched: Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily and Versions(R) **Database searched from inception through to 14 June 2019**		
#	**Terms**	**Hits**
1	Crohn Disease/	37169
2	Crohn*.mp	53162
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	42992
4	Inflammatory bowel diseases/	20151
5	IBD.mp.	22462
6	Inflammatory bowel disease*.mp.	48138
7	or/1-6	84595
8	CD8-Positive T-Lymphocytes/	34782
9	CD8+ T cells.mp.	34207
10	CD8 Antigens/	8666
11	CD8 antigens.mp.	8793
12	CD8*.mp.	107288
13	CD 8*.mp.	1186
14	T-Lymphocytes, Regulatory/	29558
15	Regulatory t cells.mp.	20593
16	(PredictSure or PredictImmune).mp.	0
17	or/8-16	139456
18	Antibodies/	97179
19	antibod*.mp.	1121351
20	glycan.mp.	15625
21	(antichitobioside carbohydrate antibod* or ACCA or chitobioside).mp.	383
22	(antilaminaribioside carbohydrate antibod* or ALCA or laminaribioside).mp.	291
23	(antimannobioside carbohydrate antibod* or AMCA or mannobioside).mp.	322
24	(anti-Saccharomyces cerevisiae antibod* or ASCA or gASCA or mannan).mp.	4629
25	(anti-laminarin carbohydrate antibod* or anti-L or laminarin).mp.	1458
26	(neutrophil elastase degraded elastin or EL-NE).mp.	7
27	glycominds.mp.	4

28	(Crohn's disease prognosis test or IBDX).mp.	1
29	or/18-19	1121351
30	or/20-28	22376
31	29 and 30	4882
32	7 and 17	1667
33	7 and 31	421
34	32 or 33	2078

Database searched: EMBASE **Database searched from inception through to 14 June 2019**		
#	**Terms**	**Hits**
1	Exp Crohn Disease/	83531
2	Crohn*.mp	94568
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	68633
4	Exp Inflammatory bowel disease/	134801
5	IBD.mp.	46227
6	Inflammatory bowel disease*.mp.	79562
7	or/1-6	168160
8	Exp CD8+ T Lymphocyte/	58540
9	CD8+ T cells.mp.	51161
10	Exp CD8 Antigen/	69304
11	CD8 antigen*.mp.	69501
12	CD8*.mp.	177082
13	CD 8*.mp.	1822
14	exp regulatory T lymphocyte/	60131
15	Regulatory t cell*.mp.	38135
16	(PredictSure or PredictImmune).mp.	0
17	or/8-16	226662
18	Exp Antibody/	1137041
19	antibod*.mp.	1355710
20	Exp Glycan/	12642
21	glycan.mp.	23624
22	(antichitobioside carbohydrate antibod* or ACCA or chitobioside).mp.	439

23	(antilaminaribioside carbohydrate antibod* or ALCA or laminaribioside).mp.	442
24	(antimannobioside carbohydrate antibod* or AMCA or mannobioside).mp.	438
25	(anti-Saccharomyces cerevisiae antibod* or ASCA or gASCA or mannan).mp.	8703
26	(anti-laminarin carbohydrate antibod* or anti-L or laminarin).mp.	1603
27	(neutrophil elastase degraded elastin or EL-NE).mp.	21
28	glycominds.mp.	28
29	(Crohn's disease prognosis test or IBDX).mp.	11
30	or/18-19	1620205
31	or/20-29	34509
32	30 and 31	8156
33	7 and 17	4850
34	7 and 32	885
35	33 or 34	5719

Database searched: Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR)

Database searched from inception through to 14 June 2019

#	Terms	Hits
1	Crohn:ti,ab,kw	4482
2	MeSH: [Inflammatory bowel diseases] explode all trees	2889
3	IBD:ti,ab,kw	1738
4	“Inflammatory bowel disease”:ti,ab,kw	2650
5	#1 or #2 or #3 or #4	7295
6	MeSH: [CD8-Positive T-Lymphocytes] explode all trees	641
7	MeSH: [CD8 Antigens] explode all trees	71
8	“CD8 Positive T-Lymphocytes”:ti,ab,kw	503
9	“CD8”:ti,ab,kw	4211
10	MeSH: [T-Lymphocytes, Regulatory] explode all trees	274
11	“Regulatory t cells”:ti,ab,kw	620
12	PredictSure” or “PredictImmune:ti,ab,kw	0
13	#6 or #7 or #8 or #9 or #10 or #11 or #12	4910
14	MeSH: [Antibodies] explode all trees	21533
15	antibod*:ti,ab,kw	38301
16	MeSH: [Polysaccharides] explode all trees	15185

17	glycan:ti,ab,kw	144
18	“antichitobioside carbohydrate antibod*” or ACCA or chitobioside:ti,ab,kw	18
19	“antilaminaribioside carbohydrate antibod*” or ALCA or laminaribioside:ti,ab,kw	6
20	“antimannobioside carbohydrate antibod*” or AMCA or mannobioside:ti,ab,kw	37
21	“anti-Saccharomyces cerevisiae antibod*” or ASCA or gASCA or mannan:ti,ab,kw	139
22	“anti-laminarin carbohydrate antibod*” or anti-L or laminarin:ti,ab,kw	1694
23	“neutrophil elastase degraded elastin” or EL-NE:ti,ab,kw	12
24	glycominds:ti,ab,kw	0
25	“Crohn's disease prognosis test” or IBDX:ti,ab,kw	0
26	#14 or #15	44459
27	#16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25	17149
28	#26 and #27	914
29	#5 and #13	38
30	#5 and #28	14
31	#29 or #30	52

9.3.2 Economic evaluations

Embase <1996 to 2019 Week 22> **Date of search 07/06/2019**		
#	**Terms**	**Hits**
1	Crohn disease/	69892
2	Crohn*.mp.	81047
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	58806
4	inflammatory bowel disease/	26886
5	IBD.mp.	45145
6	Inflammatory bowel disease*.mp.	73947
7	or/1-6	126526
8	CD8-Positive T-Lymphocyte/	58180
9	CD8+ T cells.mp.	48481
10	CD8 Antigen/	62068
11	(CD8 antigen* or CD 8 antigen*).mp.	62126
12	CD8*.mp.	161485
13	CD 8*.mp.	1596

14	regulatory T lymphocyte/	58360
15	Regulatory t cells.mp.	32962
16	(PredictSure or PredictImmune).mp.	0
17	or/8-16	208245
18	antibody/	154965
19	antibod*.mp.	962533
20	glycan.mp.	21832
21	(Antichitobioside carbohydrate antibod* or ACCA or chitobioside).mp.	405
22	(antilaminaribioside carbohydrate antibod* or ALCA or laminaribioside).mp.	348
23	(antimannobioside carbohydrate antibod* or AMCA or mannobioside).mp.	313
24	(anti-Saccharomyces cerevisiae antibod* or ASCA or gASCA or mannan).mp.	7087
25	(anti-laminarin carbohydrate antibod* or anti-L or laminarin).mp.	1201
26	(neutrophil elastase degraded elastin or EL-NE).mp.	20
27	glycominds.mp.	28
28	(Crohn's disease prognosis test or IBDX).mp.	11
29	18 or 19	962533
30	or/20-28	30472
31	29 and 30	7021
32	7 and 31	839
33	7 and 17	4336
34	32 or 33	5160
35	prednisolone/	91187
36	prednisone/	119609
37	cortisone/	6431
38	methylprednisolone/	72061
39	hydrocortisone/	80947
40	(corticosteroid or prednisolone or prednisone or methylprednisolone or hydrocortisone or budesonide).ti,ab.	133365
41	mesalamine/	14425
42	6-mercaptopurine/	16507
43	sulfasalazine/	18757
44	(mesalamine or sulfasalazine or "5-aminosalicylic*" or "5-aminosalicylate*" or "5-asa" or 5aminosalicylic* or 5aminosalicyclate* or 5asa or pentasa or mesalazine or mesalamine or	11513

	asacol or sulfasalazine* or salazopyrin* or salazosulfapyridine* or asulfidine* or azulfadine* or azulfidine*).ti,ab.	
45	azathioprine/	67700
46	methotrexate/	130108
47	(immunosuppressant or immunomodulator or mercaptopurine or methotrexate or amethopterin or Otrexup or Rasuvo or Rheumatrex or Trexall or Maxtrex or Nordimet or Zlatal or Methofill or Metoject or Jylamvo or azathioprine or Imuran or Azapress or thiopurine).ti,ab.	83647
48	((biologic or biologics or tumour necrosis factor alpha or TNF) adj2 inhibitor*).ti,ab.	7815
49	(infliximab or Remicade or Remsima or Inflectra or Zessly or Flixabi or adalimumab or Humira or Imraldi or Amgevita or Hulio or vedolizumab or Entyvio or ustekinumab or Stelara).ti,ab.	34039
50	top-down.mp.	14798
51	top down.mp.	14798
52	step-up.mp.	3588
53	step up.mp.	3588
54	or/35-50	548227
55	7 and 54	35020
56	economics/	153443
57	"health care cost"/ or "cost"/ or "drug cost"/ or "hospital running cost"/ or "hospital cost"/ or "hospitalization cost"/ or "nursing cost"/	267442
58	"cost utility analysis"/ or "cost benefit analysis"/ or "cost effectiveness analysis"/	195011
59	exp health economics/	672653
60	budget/	23702
61	budget*.ti,ab,kw.	29555
62	(economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ti,kw.	214226
63	(economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ab. /freq=2	330337
64	(cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kw.	191330
65	(value adj2 (money or monetary)).ti,ab,kw.	2780
66	statistical model/	149074
67	economic model*.ab,kw.	4233
68	probability/	85320
69	Monte Carlo method/	35702

70	monte carlo.ti,ab,kw.	41897
71	decision theory/	1327
72	"decision tree"/	10402
73	Markov chain/ or hidden Markov model/	6939
74	markov.ti,ab,kw.	25713
75	(decision* adj2 (tree* or analy* or model*)).ti,ab,kw.	28032
76	or/56-75	1280505
77	7 and 76	6583
78	34 and 76	84
79	54 and 77	2391
80	letter.pt.	772802
81	editorial.pt.	512984
82	note.pt.	648565
83	80 or 81 or 82	1934351
84	78 not 83	84
85	79 not 83	2235
86	limit 84 to human	79
87	limit 85 to human	2150

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) <1946 to June 04, 2019>

Date of search 07/06/2019

#	Terms	Hits
1	Crohn Disease/	37141
2	Crohn*.mp.	53094
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	42937
4	Inflammatory bowel diseases/	20114
5	IBD.mp.	22412
6	Inflammatory bowel disease*.mp.	48052
7	or/1-6	84477
8	CD8-Positive T-Lymphocytes/	34762
9	CD8+ T cells.mp.	34163

10	CD8 Antigens/	8663
11	(CD8 antigen* or CD 8 antigen*).mp.	8934
12	CD8*.mp.	107197
13	CD 8*.mp.	1184
14	T-Lymphocytes, Regulatory/	29536
15	Regulatory t cells.mp.	20560
16	(PredictSure or PredictImmune).mp.	0
17	or/8-16	139329
18	Antibodies/	97155
19	antibod*.mp.	1120813
20	glycan.mp. (15595)	15595
21	(Antichitobioside carbohydrate antibod* or ACCA or chitobioside).mp.	381
22	(antilaminaribioside carbohydrate antibod* or ALCA or laminaribioside).mp.	291
23	(antimannobioside carbohydrate antibod* or AMCA or mannobioside).mp.	321
24	(anti-Saccharomyces cerevisiae antibod* or ASCA or gASCA or mannan).mp.	4624
25	(anti-laminarin carbohydrate antibod* or anti-L or laminarin).mp.	1457
26	(neutrophil elastase degraded elastin or EL-NE).mp.	7
27	glycominds.mp.	4
28	(Crohn's disease prognosis test or IBDX).mp.	1
29	18 or 19	1120813
30	or/20-28	22337
31	29 and 30	4877
32	7 and 17	1667
33	7 and 31	421
34	32 or 33	2078
35	prednisolone/	32048
36	prednisone/	38452
37	cortisone/	19548
38	methylprednisolone/	18301
39	hydrocortisone/	70258
40	(corticosteroid or prednisolone or prednisone or methylprednisolone or hydrocortisone or budesonide).ti,ab.	120312
41	mesalamine/	3345

42	sulfasalazine/	4014
43	(mesalamine or sulfasalazine or "5-aminosalicylic*" or "5-aminosalicylate*" or "5-asa" or 5aminosalicylic* or 5aminosalicyclate* or 5asa or pentasa or mesalazine or mesalamine or asacol or sulfasalazine* or salazopyrin* or salazosulfapyridine* or asulfidine* or azulfadine* or azulfidine*).ti,ab.	7655
44	6-mercaptopurine/	6144
45	azathioprine/	14345
46	methotrexate/	36645
47	(immunosuppressant or immunomodulator or mercaptopurine or methotrexate or amethopterin or Otrexup or Rasuvo or Rheumatrex or Trexall or Maxtrex or Nordimet or Zlatal or Methofill or Metoject or Jylamvo or azathioprine or Imuran or Azapress or thiopurine).ti,ab.	67759
48	((biologic or biologics or tumour necrosis factor alpha or TNF) adj2 inhibitor*).ti,ab.	4154
49	(infliximab or Remicade or Remsima or Inflectra or Zessly or Flixabi or adalimumab or Humira or Imraldi or Amgevita or Hulio or vedolizumab or Entyvio or ustekinumab or Stelara).ti,ab.	15559
50	top-down.mp.	13641
51	top down.mp.	13641
52	step-up.mp.	2492
53	step up.mp.	2492
54	or/35-53	344592
55	7 and 54	13382
56	exp Economics, Nursing/ or exp Economics, Pharmaceutical/ or exp Economics, Medical/ or exp Economics, Hospital/	43836
57	Economics/	27043
58	exp "Costs and Cost Analysis"/	225181
59	exp "Fees and Charges"/	29739
60	exp Budgets/	13513
61	(economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ti,kf.	213256
62	(economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ab. /freq=2	263695
63	(cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf.	147684
64	(value adj2 (money or monetary)).ti,ab,kf.	2192
65	exp Models, Economic/	14158

66	economic model*.ab,kf.	3037
67	markov chains/	13431
68	markov.ti,ab,kf.	20502
69	monte carlo method/	26768
70	monte carlo.ti,ab,kf.	45800
71	exp Decision Trees/ or exp Decision Theory/	11468
72	(decision* adj2 (tree* or analy* or model*)).ti,ab,kf.	21223
73	or/56-72	665817
74	7 and 73	1542
75	34 and 73	13
76	54 and 74	412
77	letter.pt.	1029243
78	editorial.pt.	492302
79	historical article.pt.	351921
80	comment.pt.	777323
81	case reports.pt.	2023130
82	or/77-81	3878097
83	75 not 82	13
84	76 not 82	387
85	limit 83 to humans	11
86	limit 84 to humans	311

Cochrane **Date of search 07/06/2019**		
#	**Terms**	**Hits**
#1	MeSH descriptor: [Crohn Disease] this term only	1406
#2	Crohn*	5032
#3	Crohn* near/2 (disease or syndrome)	4482
#4	regional enteritis	45
#5	MeSH descriptor: [Inflammatory Bowel Diseases] this term only	436
#6	inflammatory bowel disease* or IBD	7938
#7	#1 or #2 or #3 or #4 or #5 or #6	10448

#8	MeSH descriptor: [CD8-Positive T-Lymphocytes] this term only	502
#9	MeSH descriptor: [CD8 Antigens] this term only	71
#10	CD8 antigens or " CD8 antigen*" or CD8* or "CD 8*"	4571
#11	Regulatory t cells	654
#12	PredictSure or PredictImmune	0
#13	#8 or #9 or #10 or #11 or #12	5076
#14	MeSH descriptor: [Antibodies] this term only	1353
#15	antibod*	39288
#16	glycan	146
#17	antichitobioside carbohydrate antibod* or ACCA or chitobioside	18
#18	antilaminaribioside carbohydrate antibod* or ALCA or laminaribioside	6
#19	antimannobioside carbohydrate antibod* or AMCA or mannobioside	37
#20	anti-Saccharomyces cerevisiae antibod* or ASCA or gASCA or mannan	208
#21	anti-laminarin carbohydrate antibod* or anti-L or laminarin	1695
#22	neutrophil elastase degraded elastin or EL-NE	386
#23	glycominds	0
#24	IBDX	0
#25	crohn* disease prognos* test	50
#26	#14 or #15	39288
#27	#16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25	2404
#28	#26 and #27	347
#29	#7 and #13	65
#30	#7 and #28	44
#31	#29 or #30	105
#32	MeSH descriptor: [Prednisolone] this term only	2693
#33	MeSH descriptor: [Prednisone] this term only	3666
#34	MeSH descriptor: [Cortisone] this term only	126
#35	MeSH descriptor: [Methylprednisolone] this term only	2371
#36	MeSH descriptor: [Hydrocortisone] this term only	5665
#37	(corticosteroid or prednisolone or prednisone or methylprednisolone or hydrocortisone or budesonide):ti,ab	27102
#38	MeSH descriptor: [Mesalamine] this term only	509
#39	MeSH descriptor: [Sulfasalazine] this term only	448

#40	(mesalamine or sulfasalazine or "5-aminosalicylic*" or "5-aminosalicylate*" or "5-asa" or 5aminosalicylic* or 5aminosalicyclate* or 5asa or pentasa or mesalazine or mesalamine or asacol or sulfasalazine* or salazopyrin* or salazosulfapyridine* or asulfidine* or azulfadine* or azulfidine*):ti,ab	2216
#41	MeSH descriptor: [Mercaptopurine] this term only	252
#42	MeSH descriptor: [Azathioprine] this term only	1166
#43	MeSH descriptor: [Methotrexate] this term only	3755
#44	(immunosuppressant or immunomodulator or mercaptopurine or methotrexate or amethopterin or Otrexup or Rasuvo or Rheumatrex or Trexall or Maxtrex or Nordimet or Zlatal or Methofill or Metoject or Jylamvo or azathioprine or Imuran or Azapress or thiopurine):ti,ab	12481
#45	(biologic or biologics or tumour necrosis factor alpha or TNF near/2 inhibitor*):ti,ab	9180
#46	(infliximab or Remicade or Remsima or Inflectra or Zessly or Flixabi or adalimumab or Humira or Imraldi or Amgevita or Hulio or vedolizumab or Entyvio or ustekinumab or Stelara):ti,ab	4817
#47	top-down	483
#48	top down	483
#49	step-up	731
#50	step up	731
#51	#32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 and #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50	55540
#52	#7 and #51	3092
#53	MeSH descriptor: [Economics] 1 tree(s) exploded	11682
#54	economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed	88865
#55	value near/2 (money or monetary)	351
#56	cost* near/2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcome)	33230
#57	MeSH descriptor: [Models, Economic] this term only	222
#58	MeSH descriptor: [Decision Trees] this term only	157
#59	MeSH descriptor: [Probability] this term only	3092
#60	markov or "monte carlo" or "economic model"	2283
#61	decision* near/2 (tree* or analy* or model*)	2352
#62	#53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61	94635
#63	#7 and #62	900
#64	#31 and #62	35
#65	#52 and #62	306

Centre for Reviews and Dissemination (CRD) Date of search 07/06/2019		
#	**Terms**	**Hits**
1	MeSH DESCRIPTOR Crohn Disease EXPLODE ALL TREES	220
2	(crohn*)	374
3	(Crohn* NEAR2 (disease or syndrome) OR regional enteritis)	356
4	MeSH DESCRIPTOR Inflammatory Bowel Diseases EXPLODE ALL TREES	456
5	(inflammatory bowel disease*)	285
6	(IBD)	80
7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	640
8	MeSH DESCRIPTOR CD8-Positive T-Lymphocytes EXPLODE ALL TREES	2
9	MeSH DESCRIPTOR Antigens, CD8 EXPLODE ALL TREES	0
10	(CD8 antigen* OR CD 8 antigen* OR CD8* OR CD 8*)	17
11	MeSH DESCRIPTOR T-Lymphocytes, Regulatory EXPLODE ALL TREES	2
12	(Regulatory t cells)	1
13	(PredictSure OR PredictImmune)	0
14	#8 OR #9 OR #10 OR #11 OR #12 OR #13	19
15	MeSH DESCRIPTOR Antibodies EXPLODE ALL TREES	2097
16	(antibod*)	2285
17	(glycan)	1
18	(antichitobioside carbohydrate antibod* OR ACCA OR chitobioside)	0
19	(antilaminaribioside carbohydrate antibod* OR ALCA OR laminaribioside)	0
20	(antimannobioside carbohydrate antibod* OR AMCA OR mannobioside)	0
21	(anti-Saccharomyces cerevisiae antibod* OR ASCA OR gASCA OR mannan)	11
22	(anti-laminarin carbohydrate antibod* OR anti-L OR laminarin)	0
23	(neutrophil elastase degraded elastin OR EL-NE)	0
24	(glycominds)	0
25	(prognos* test OR IBDX)	5
26	#15 OR #16	2578
27	#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25	17
28	#26 AND #27	7
29	#7 AND #14	1

30	#7 AND #28	4
31	#29 OR #30	5
32	(* in DARE)	45418
33	(* in NHSEED)	17613
34	(* in HTA)	17351
35	#31 AND #32	3
36	#31 AND #33	1
37	#31 AND #34	1
38	MeSH DESCRIPTOR Economics EXPLODE ALL TREES	17657
39	MeSH DESCRIPTOR Economics, Hospital EXPLODE ALL TREES	1247
40	MeSH DESCRIPTOR Economics, Medical EXPLODE ALL TREES	53
41	MeSH DESCRIPTOR Economics, Nursing EXPLODE ALL TREES	9
42	MeSH DESCRIPTOR Economics, Pharmaceutical EXPLODE ALL TREES	199
43	MeSH DESCRIPTOR Budgets EXPLODE ALL TREES	54
44	MeSH DESCRIPTOR Markov Chains EXPLODE ALL TREES	2056
45	MeSH DESCRIPTOR Decision Trees EXPLODE ALL TREES	864
46	MeSH DESCRIPTOR Decision Theory EXPLODE ALL TREES	873
47	MeSH DESCRIPTOR Probability EXPLODE ALL TREES	10018
48	MeSH DESCRIPTOR Costs and Cost Analysis EXPLODE ALL TREES	17164
49	(economic* OR cost OR costs OR costly OR costing OR price OR prices OR pricing OR pharmacoeconomic* OR pharmaco-economic* OR expenditure OR expenditures OR expense OR expenses OR financial OR finance OR finances OR financed)	26785
50	(cost* NEAR2 (effective* OR utilit* OR benefit* OR minimi* OR analy* OR outcome OR outcomes))	20789
51	(Markov OR economic model OR monte carlo)	3877
52	(decision* NEAR2 (tree* OR analy* OR model*))	3590
53	#38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52	32706
54	#7 AND #53	260
55	#32 AND #54	108
56	#33 AND #54	121
57	#34 AND #54	31
58	MeSH DESCRIPTOR Prednisolone EXPLODE ALL TREES	97
59	MeSH DESCRIPTOR Prednisone EXPLODE ALL TREES	86

60	MeSH DESCRIPTOR Cortisone EXPLODE ALL TREES	1
61	MeSH DESCRIPTOR Methylprednisolone EXPLODE ALL TREES	43
62	MeSH DESCRIPTOR Hydrocortisone EXPLODE ALL TREES	25
63	(corticosteroid OR prednisolone OR prednisone OR methylprednisolone OR hydrocortisone OR budesonide)	1068
64	MeSH DESCRIPTOR Mesalamine EXPLODE ALL TREES	46
65	MeSH DESCRIPTOR 6-Mercaptopurine EXPLODE ALL TREES	80
66	MeSH DESCRIPTOR Sulfasalazine EXPLODE ALL TREES	18
67	(mesalamine OR sulfasalazine OR "5-aminosalicylic*" OR "5-aminosalicylate*" OR "5-asa" OR 5aminosalicylic* OR 5aminosalicyclate* OR 5asa OR pentasa OR mesalazine OR mesalamine OR asacol OR sulfasalazine* OR salazopyrin* OR salazosulfapyridine* OR asulfidine* OR azulfadine* OR azulfidine*)	142
68	MeSH DESCRIPTOR Azathioprine EXPLODE ALL TREES	73
69	MeSH DESCRIPTOR Methotrexate EXPLODE ALL TREES	176
70	(immunosuppressant OR immunomodulator OR mercaptopurine OR methotrexate OR amethopterin OR Otrexup OR Rasuvo OR Rheumatrex OR Trexall OR Maxtrex OR Nordimet OR Zlatal OR Methofill OR Metoject OR Jylamvo OR azathioprine OR Imuran OR Azapress OR thiopurine)	644
71	(biologic OR biologics OR tumour necrosis factor alpha OR TNF NEAR2 inhibitor*)	308
72	(infliximab OR Remicade OR Remsima OR Inflectra OR Zessly OR Flixabi OR adalimumab OR Humira OR Imraldi OR Amgevita OR Hulio OR vedolizumab OR Entyvio OR ustekinumab OR Stelara)	438
73	(top-down)	37
74	(top down)	37
75	(step-up)	28
76	(step up)	28
77	#58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67 OR #68 OR #69 OR #70 OR #71 OR #72 OR #73 OR #74 OR #75 OR #76	2147
78	#54 AND #77	120
79	#55 AND #77	49
80	#56 AND #77	60
81	#57 AND #77	11

9.3.3 Health-related quality of life

Embase <1974 to 2019 July 09> **Date of search 25/07/2019**		
#	**Terms**	**Hits**
1	Crohn disease/	83044
2	Crohn*.mp.	95973
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	69843
4	inflammatory bowel disease/	27738
5	IBD.mp.	47446
6	Inflammatory bowel disease*.mp.	80993
7	or/1-6	145131
8	exp quality adjusted life year/	24118
9	(qaly$ or qald$ or qale$ or qtime$).ti,ab,kw.	18449
10	(quality adjusted or adjusted life year$).ti,ab,kw.	22397
11	disability adjusted life.ti,ab,kw.	3630
12	daly$1.ti,ab,kw.	3588
13	((index adj3 wellbeing) or (quality adj3 wellbeing) or qwb).ti,ab,kw.	1051
14	(multiattribute$ or multi attribute$).ti,ab,kw.	1079
15	(utility adj3 (score$1 or scoring or valu$ or measur$ or evaluat$ or scale$1 or instrument$1 or weight or weights or weighting or information or data or unit or units or health$ or life or estimat$ or elicit$ or disease$ or mean or cost$ or expenditure$1 or gain or gains or loss or losses or lost or analysis or index$ or indices or overall or reported or calculat$ or range$ or increment$ or state or states or status)).ti,ab,kw.	47492
16	utility.ab. /freq=2	24677
17	utilities.ti,ab,kw.	10761
18	disutili$.ti,ab,kw.	845
19	(HSUV or HSUVs).ti,ab,kw.	100
20	health$1 year$1 equivalent$1.ti,ab,kw.	44
21	(hye or hyes).ti,ab,kw.	130
22	(hui or hui1 or hui2 or hui3).ti,ab,kw.	2108
23	(illness state$1 or health state$1).ti,ab,kw.	10420
24	(euro qual or euro qual5d or euro qol5d or eq-5d or eq5-d or eq5d or euroqual or euroqol or euroqual5d or euroqol5d).ti,ab,kw.	18050
25	(short form$ or shortform$).ti,ab,kw.	41624

26	(sf36$ or sf 36$ or sf thirtysix or sf thirty six).ti,ab,kw.	35765
27	(sf6 or sf 6 or sf6d or sf 6d or sf six or sfsix or sf8 or sf 8 or sf eight or sfeight).ti,ab,kw.	4185
28	(sf12 or sf 12 or sf twelve or sftwelve).ti,ab,kw.	7524
29	(sf16 or sf 16 or sf sixteen or sfsixteen).ti,ab,kw.	47
30	(sf20 or sf 20 or sf twenty or sftwenty).ti,ab,kw.	328
31	(15D or 15-D or 15 dimension).ti,ab,kw.	6267
32	(standard gamble$ or sg).ti,ab,kw.	14405
33	(time trade off$1 or time tradeoff$1 or tto or timetradeoff $1).ti,ab,kw.	2609
34	or/8-33	197191
35	7 and 34	1781
36	letter.pt.	1074956
37	editorial.pt.	623536
38	note.pt.	761313
39	36 or 37 or 38	2459805
40	35 not 39	1765

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) <1946 to July 08, 2019>

Date of search 10/07/2019

#	Terms	Hits
1	Crohn Disease/	37301
2	Crohn*.mp.	53410
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	43196
4	inflammatory bowel diseases/	20297
5	IBD.mp.	22666
6	Inflammatory bowel disease*.mp.	48463
7	or/1-6	85061
8	exp Quality-Adjusted Life Years/	11153
9	"Value of Life"/	5652
10	(qaly$ or qald$ or qale$ or qtime$).ti,ab,kf.	9753
11	(quality adjusted or adjusted life year$).ti,ab,kf.	15185
12	disability adjusted life.ti,ab,kf.	2961

13	daly$1.ti,ab,kf.	2647
14	((index adj3 wellbeing) or (quality adj3 wellbeing) or qwb).ti,ab,kf.	654
15	(multiattribute$ or multi attribute$).ti,ab,kf.	835
16	(utility adj3 (score$1 or scoring or valu$ or measur$ or evaluat$ or scale$1 or instrument$1 or weight or weights or weighting or information or data or unit or units or health$ or life or estimat$ or elicit$ or disease$ or mean or cost$ or expenditure$1 or gain or gains or loss or losses or lost or analysis or index$ or indices or overall or reported or calculat$ or range$ or increment$ or state or states or status)).ti,ab,kf.	30914
17	utility.ab. /freq=2	15905
18	utilities.ti,ab,kf.	6591
19	disutili$.ti,ab,kf.	431
20	(HSUV or HSUVs).ti,ab,kf.	60
21	health$1 year$1 equivalent$1.ti,ab,kf.	40
22	(hye or hyes).ti,ab,kf.	66
23	(hui or hui1 or hui2 or hui3).ti,ab,kf.	1392
24	(illness state$1 or health state$1).ti,ab,kf.	6021
25	(euro qual or euro qual5d or euro qol5d or eq-5d or eq5-d or eq5d or euroqual or euroqol or euroqual5d or euroqol5d).ti,ab,kf.	9651
26	(short form$ or shortform$).ti,ab,kf.	30596
27	(sf36$ or sf 36$ or sf thirtysix or sf thirty six).ti,ab,kf.	20791
28	(sf6 or sf 6 or sf6d or sf 6d or sf six or sfsix or sf8 or sf 8 or sf eight or sfeight).ti,ab,kf.	3059
29	(sf12 or sf 12 or sf twelve or sftwelve).ti,ab,kf.	4303
30	(sf16 or sf 16 or sf sixteen or sfsixteen).ti,ab,kf.	27
31	(sf20 or sf 20 or sf twenty or sftwenty).ti,ab,kf.	330
32	(15D or 15-D or 15 dimension).ti,ab,kf.	4982
33	(standard gamble$ or sg).ti,ab,kf.	9836
34	(time trade off$1 or time tradeoff$1 or tto or timetradeoff$1).ti,ab,kf.	1800
35	or/8-34	133401
36	7 and 35	714
37	letter.pt.	1033774
38	editorial.pt.	495816
39	historical article.pt.	352651
40	comment.pt.	785307
41	case reports.pt.	2030156

42	or/37-41	3896584
43	36 not 42	701
44	limit 43 to humans	575

Cochrane **Date of search 09/07/2019**		
#	**Terms**	**Hits**
#1	MeSH descriptor: [Crohn Disease] this term only	1410
#2	Crohn*	5065
#3	Crohn* near/2 (disease or syndrome)	4514
#4	regional enteritis :ti,ab,kw	29
#5	MeSH descriptor: [Inflammatory Bowel Diseases] this term only	2911
#6	inflammatory bowel disease* or IBD	7989
#7	#1 or #2 or #3 or #4 or #5 or #6	10940
#8	MeSH descriptor: [Quality-Adjusted Life Years] explode all trees	1125
#9	("quality-adjusted life year*" or QALY or QALYs or "quality adjusted life year*"):ti,ab,kw	4129
#10	("quality near/2 life" or QOL):ti,ab,kw	16848
#11	("disability-adjusted life year*" or DALY or DALYs or "disability adjusted life years*"):ti,ab,kw	214
#12	(HRQL or HRQOL):ti,ab,kw	5525
#13	(sf36 or sf-36 or "sf 36" or "short form 36" or "shortform 36" or "sf thirtysix" or "sf thirty six" or "shortform thirtysix" or "shortform thirty six" or "short form thirty six" or "short form thirtysix" or "short form thirty six"):ti,ab,kw	11476
#14	(sf6 or "sf 6" or "sf-6" or "short form 6" or "shortform 6" or "sf six" or sfsix or "shortform six" or "short form six"):ti,ab,kw	179
#15	(sf6d or "sf 6d" or "sf-6d" or "short form 6d" or "shortform 6d" or "sf six dimension" or "short form six dimension"):ti,ab,kw	285
#16	(sf12 or "sf 12" or sf-12 or "short form 12" or "shortform 12" or "sf twelve" or sftwelve or "shortform twelve" or "short form twelve"):ti,ab,kw	2333
#17	(sf16 or "sf 16" or "sf-16" or "short form 16" or "shortform 16" or "sf sixteen" or sfsixteen or "shortform sixteen" or "short form sixteen"):ti,ab,kw	5
#18	(sf20 or "sf 20" or sf-20 or "short form 20" or "shortform 20" or "sf twenty" or sftwenty or "shortform twenty" or "short form twenty"):ti,ab,kw	86
#19	(euroqol or "euro qol" or eq5d or "eq 5d" or eq-5d):ti,ab,kw	7392
#20	(hye or hyes or "health* year* equivalent*"):ti,ab,kw	9

#21	("standard gamble" or SG):ti,ab,kw	1266
#22	((quality near/3 wellbeing index) or QWB):ti,ab,kw	226
#23	("time trade off" or "time tradeoff" or TTO or "time trade-off"):ti,ab,kw	540
#24	(utility near/3 value):ti,ab,kw	119
#25	disutil*:ti,ab,kw	70
#26	("health utilities index" or HUI):ti,ab,kw	264
#27	{OR #8-#26}	40751
#28	#7 and #27	646

Centre for Reviews and Dissemination (CRD) **Date of search 09/07/2019**		
#	**Terms**	**Hits**
#1	MeSH DESCRIPTOR Crohn Disease EXPLODE ALL TREES	220
#2	(crohn*)	374
#3	(Crohn* NEAR2 (disease or syndrome) OR regional enteritis)	356
#4	MeSH DESCRIPTOR Inflammatory Bowel Diseases EXPLODE ALL TREES	456
#5	(inflammatory bowel disease*)	285
#6	(IBD)	80
#7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	640
#8	MeSH DESCRIPTOR Quality-Adjusted Life Years EXPLODE ALL TREES	3547
#9	(quality-adjusted life year OR QALY or QALYs OR quality adjusted life year*)	5531
#10	(quality NEAR2 life OR QOL)	11586
#11	(disability-adjusted life year* OR DALY OR DALYs OR disability adjusted life years*)	226
#12	(HRQL OR HRQOL)	198
#13	(sf36 OR sf-36 OR sf 36 OR short form 36 OR shortform 36 OR sf thirtysix OR sf thirty six OR shortform thirtysix OR shortform thirty six OR short form thirty six OR short form thirtysix OR short form thirty six)	409
#14	(sf6 OR sf 6 OR sf-6 OR short form 6 OR shortform 6 OR sf six OR sfsix OR shortform six or short form six)	6
#15	(sf6d OR sf 6d OR sf-6d OR short form 6d OR shortform 6d OR sf six dimension OR short form six dimension)	57
#16	(sf12 OR sf 12 or sf-12 OR short form 12 or shortform 12 OR sf twelve OR sftwelve OR shortform twelve or short form twelve)	60

#17	(sf16 OR sf 16 OR sf-16 OR short form 16 OR shortform 16 OR sf sixteen OR sfsixteen OR shortform sixteen OR short form sixteen)	1
#18	(sf20 OR sf 20 OR sf-20 OR short form 20 OR shortform 20 OR sf twenty OR sftwenty OR shortform twenty OR short form twenty)	4
#19	(euroqol OR euro qol OR eq5d OR eq 5d OR eq-5d)	790
#20	(hye OR hyes OR health* year* equivalent*)	11
#21	(standard gamble OR SG)	455
#22	(quality NEAR3 wellbeing index OR QWB)	18
#23	(time trade off OR time tradeoff or TTO or time trade-off)	375
#24	(utility NEAR3 value)	151
#25	(disutil*)	184
#26	(health utilities index OR HUI)	201
#27	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26	12314
#28	#7 AND #27	118

9.4 Appendix 4. Excluded studies

9.4.1 Prognostic accuracy and clinical impact

Study	Reason for exclusion
Beaven 2004[137]	Not systematic review.
Dotan 2006[138]	Incorrect biomarker panel (assesses only 4 of the 6 biomarker panel forming the IBDX kit).
Dotan 2010[139]	Incorrect biomarker panel (assesses only 4 of the 6 biomarker panel forming the IBDX kit).
Fengming 2014[140]	Not systematic review.
Gasparetto 2018[141]	Conference abstract with insufficient information reported to include in review.
Halder 2010[80]	Kit used to assess biomarker panel not in the scope of this review.
Koutroubakis 2011[142]	Incorrect biomarker panel (assesses only 4 of the 6 biomarker panel forming the IBDX kit).
Lee 2011a[143]	Related to PredictSURE-IBD: describes development of tool rather than prognostic accuracy or clinical impact.
Lee 2011b[144]	Related to PredictSURE-IBD: describes development of tool rather than prognostic accuracy or clinical impact.

Lee 2011c[46]	Related to PredictSURE-IBD: describes development of tool rather than prognostic accuracy or clinical impact.
Lee 2012[145]	Related to PredictSURE-IBD: describes development of tool rather than prognostic accuracy or clinical impact.
Lee 2017a[146]	Related to PredictSURE-IBD: describes development of tool and results that pre-date the full publication.
Lee 2017b[147]	Related to PredictSURE-IBD: describes development of tool and results that pre-date the full publication.
Lee 2017c[148]	Related to PredictSURE-IBD: describes development of tool and results that pre-date the full publication.
Lyons 2019[149]	Related to PredictSURE-IBD: editorial that describes development of tool rather than prognostic accuracy or clinical impact.
Papp 2014[150]	Not systematic review.
Rieder 2011[151]	Conference abstract with insufficient information reported to include in review.
Rieder 2013[152]	Incorrect intervention and not question of interest to this review: assesses link between gene profiling and biomarker panel.
Ryan 2013[153]	Kit used to assess biomarker panel not in the scope of this review.
Simondi 2008[154]	Incorrect biomarker panel (assesses only 4 of the 6 biomarker panel forming the IBDX kit).

9.4.2 Economic evaluations

Study (48)	**Reason for exclusion**
Aliyev et al. 2018	Conference abstract with insufficient detail
Ananthakrishnan et al. 2013	Irrelevant comparison (mucosal healing v clinical response to escalate dose)
Arhan et al. 2018	Conference abstract with insufficient detail
Azzabi et al. 2017	Conference abstract with insufficient detail
Baji et al. 2016	Conference abstract with insufficient detail
Beilman et al. 2017 (S137)	Conference abstract with insufficient detail
Beilman et al. 2017 (S447)	Conference abstract with insufficient detail
Di Sabatino et al. 2011	Conference abstract published before the specified cut-off date (2016)
Ghosh et al. 2015	Not a full economic evaluation
Hansson-Hedblom et al. 2017	Conference abstract with insufficient detail

Jean et al. 2018	Not a full economic evaluation (systematic review of existing evidence)
Jewell et al. 2005	Not available
Koelewijn et al. 2006	Not a full economic evaluation (systematic review of existing evidence)
Lee et al. 2012	Conference abstract with insufficient detail
Lindsay et al. 2013	Not a full economic evaluation
Marchetti et al. 2014	Not available
Marshall et al. 2002	Not a full economic evaluation
Marshall 2002	Not a full economic evaluation (subjective review of existing evidence)
Mlcoch et al.	Conference abstract with insufficient detail
Mobinizadeh et al. 2012	Non-English language
Noble et al. 1998	Outdated clinical practice
Ntr 2005	Study protocol
Ob et al. 2018	Conference abstract with insufficient detail
Panaccione et al. 2017	Conference abstract with insufficient detail
Panaccione et al. 2018	Conference abstract with insufficient detail
Pillai et al. 2017	Not a full economic evaluation (systematic review of existing evidence)
Priest et al. 2006	Irrelevant population (IBD not limited to Crohn's disease)
Rencz et al. 2017	Not available
Robson et al. 2018	Conference abstract with insufficient detail
Rosim et al. 2017	Conference abstract with insufficient detail
Rudakova 2012	Non-English language
Saro et al. 2015	Methods and results unable to inform conceptual model
Schneider et al. 2017	Conference abstract with insufficient detail
Schneider et al. 2017	Conference abstract with insufficient detail
Schneider et al. 2016	Conference abstract with insufficient detail
Scott et al. 2017	Conference abstract with insufficient detail
Scott et al. 2013	Not a full economic evaluation
Shah et al. 2016	Conference abstract with insufficient detail
Siegel et al. 2006	Not a full economic evaluation

Sprakes et al. 2010	Not a full economic evaluation
Steenholdt et al. 2014	Methods and results unable to inform conceptual model
Swaminath et al. 2013	Not a full economic evaluation Irrelevant comparison and outcomes related to tuberculosis screening
Tang et al. 2013	Not a full economic evaluation (systematic review of existing evidence)
Trallori et al. 1997	Outdated clinical practice
Tsui et al. 2018	Not a full economic evaluation
Wilson and Lucas 2018	Conference abstract with insufficient detail
Winter et al. 2004	Population unclear (types of IBD not reported) Irrelevant comparison and outcomes related to genotype screening
Zaboli et al. 2017	Conference abstract with insufficient detail

Economic evaluations of tests for the identification of those at high risk of developing a severe course of Crohn's disease

Study (3)	**Reason for exclusion**
Odes et al. 2007	Not a full economic evaluation
Teml et al. 2003	Not available
Spizzo et al. 2017	Conference abstract with insufficient detail

9.4.3 Health-related quality of life (HRQoL) evidence

Study	**Reason for exclusion**
Araki et al 2009	Not available
Barreiro-de Acosta et al 2012	Conference abstract with insufficient detail
Bastida et al 2011	Conference abstract with insufficient detail
Baumgart et al 2015	Conference abstract with insufficient detail
Bernklev et al 2006	Utility data not relevant to model health states
Bernklev et al 2002	Subgroup data irrelevant
Bokemeyer et al 2014 Adalimumab…	Conference abstract with insufficient detail

Bokemeyer et al 2014 TNF-alpha…	Conference abstract with insufficient detail
Bokemeyer et al 2019	Conference abstract with insufficient detail
Bracher et al 2019	Conference abstract with insufficient detail
Buxton et al 2007	Not primary source of data
Cappello et al 2012 Personality traits…	Conference abstract with insufficient detail
Cappello et al 2012 Personality profile…	Conference abstract with insufficient detail
Cappello et al 2013	Conference abstract with insufficient detail
Casellas et al 2012 Mucosal…	Not available
Casellas et al 2017	Conference abstract with insufficient detail
Casellas et al 2003	Not available
Casellas et al 2012 Restoration…	Conference abstract with insufficient detail
Ceballos et al 2013	Conference abstract with insufficient detail
Chiarini et al 2017	Non-English Language
Chrobak-Bien et al 2017	Non-English Language
Cicchetti et al 2013	Conference abstract with insufficient detail
Cohen et al 2014	Subgroup data irrelevant
Colombel et al 2009	Not available
Colombel et al 2013	Insufficient detail from the abstract.
Coteur et al 2009	Utility data not relevant to model health states
Danese et al 2019	Conference abstract with insufficient detail
Friger et al 2014	Conference abstract with insufficient detail
Fritzel et al 2018	Conference abstract with insufficient detail
Geccherle et al 2015	Conference abstract with insufficient detail
Geccherle et al 2013	Conference abstract with insufficient detail
Ghazi et al 2010	Conference abstract with insufficient detail

Ghosh et al 2019	Subgroup data irrelevant
Ghosh et al 2013	Conference abstract with insufficient detail
Gratzer et al 2013	Conference abstract with insufficient detail
Greenberg et al 2015	Conference abstract with insufficient detail
Grochenig et al 2017	Not available
Hashimoto et al 1999	Non-English language
Hibi et al 2010	Conference abstract with insufficient detail
Hotokezaka et al 2010	Conference abstract with insufficient detail
Huang et al 2015	Not available
Hummel et al 2011	Not available
Huppertz-Hauss et al 2015	Conference abstract with insufficient detail
Huppertz-Hauss et al 2016	Conference abstract with insufficient detail
Iglesias et al 2009	Conference abstract with insufficient detail
Kane et al 2012	Conference abstract with insufficient detail
Kiran et al 2011	No utility data available for different health states
Kniazev et al 2011	Non-English language
Knowles et al 2018	No utility data available for different health states
Larsson et al 2008	No subgroup utility data for CD patients
Lazzaro et al 2014	Conference abstract with insufficient detail
Liu et al 2018	No subgroup utility data for CD patients
Loftus et al 2009	Conference abstract with insufficient detail
Longworth et al 2018	Conference abstract with insufficient detail
Manuela et al 2013	Not available.
Mnif et al 2010	No subgroup utility data for CD patients
Mostafa et al 2017	Not available.
Munoz et al 2013	Conference abstract with insufficient detail
Novacek et al 2011	Conference abstract with insufficient detail
Ormerod et al 2012	Conference abstract with insufficient detail
Panaccione et al 2018	Conference abstract with insufficient detail

Panaccione et al 2009	Conference abstract with insufficient detail
Petryszyn et al 2016	Conference abstract with insufficient detail
Petryszyn et al 2015	Conference abstract with insufficient detail
Reinshagen et al 2013	Conference abstract with insufficient detail
Renczet al 2018	Conference abstract with insufficient detail
Romberg-Camps et al 2010	Not available
Sandborn. et al 2011	Conference abstract with insufficient detail
Sandborn et al 2012	Conference abstract with insufficient detail
Sandborn et al 2015	Conference abstract with insufficient detail
Sands et al 2016	Conference abstract with insufficient detail
Schwartz et al, 2015	Conference abstract with insufficient detail
Schwartz et al 2016	Conference abstract with insufficient detail
Sherman et al 2014	No subgroup utility data for CD patients
Stjernman et al 2006	Not available
Szepes et al 2012	Conference abstract with insufficient detail
Taxonera et al 2016	Conference abstract with insufficient detail
Timmer et al 2009	Utility values not relevant for the model
Toruner et al 2017	Conference abstract with insufficient detail
Toya et al 2015	Conference abstract with insufficient detail
Vardi et al 2015	Conference abstract with insufficient detail
Vermeire et al 2017	Conference abstract with insufficient detail
Wang et al 2016	Conference abstract with insufficient detail
Wang et al 2013	Not available
Worbes-Cerezo et al 2017	Conference abstract with insufficient detail
Wright et al 2014	Conference abstract with insufficient detail
Xu et al 2014	Utility data not relevant to model health states
Yarlas A et al 2016	Conference abstract with insufficient detail
Yazdanpanah et al 1997	Insufficient detail on utility values (presented graphically)
Zakharash et al 2007	Non-English language

9.5 Appendix 5. Data extraction tables

9.5.1 Prognostic accuracy and clinical impact

9.5.1.1 IBDX

Item	Details
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Harrell 2010
Study details (journal, year, volume, page range)	*Gastroenterology*, 2010, 138, (5), S529
Type of report (full paper/only abstract/conference abstract)	Conference abstract
Section 2: Study information	
Location and number of sites	US patient cohort.
Study sponsor	Not reported.
Conflicts of interest	Not reported.
Patient enrolment (how patients were enrolled, and date to date of enrolment)	Tertiary IBD centre.
Study design (e.g., RCT, cross-over RCT)	Unclear.
Study duration (including any period of follow-up)	Unclear.
Inclusion criteria	People with CD. Disabling disease was defined as at least one of the following criteria: a) >2 courses of steroids and/or dependence on steroids; b) surgery for an inflammatory mass, intra-abdominal abscess or intestinal fistula; c) >1 surgical bowel resection within 5 years for intestinal stricture or clinical bowel obstruction; d) perianal abscess or fistula requiring surgery; e) intestinal cancer. Complicated disease behaviour defined as intestinal fistula and/or stricture.

Exclusion criteria	Not reported.
Subgroups evaluated	Not reported.
Definition of response	Not reported.
Definition of remission	Not reported.
Test	**IBDX**
Number in study, N	172 124 (72.5%) of patients had disabling disease based on definition, 113 (66.1%) of patients had complicated disease behaviour and/or need for surgery. Unclear whether number of people in each category is at baseline or during follow up.
Withdrawals, n (%)	Not reported.
Details of follow-up for development of severe course of CD	Not reported.
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Not reported.
Baseline patient characteristics	**IBDX**
Mean age, (with SD/SE if given), years (range)	Not reported.
Sex (M/F), n (%)	Not reported.
Ethnicity, n (%)	Not reported.
Disease duration (months)	Not reported.
Disease location	Not reported.
• L1 (ileum)	Not reported.
• L2 (colon)	Not reported.
• L3 (ileum + colon)	Not reported.
• L4 (upper digestive tract)	Not reported.
Predictors of disabling CD	Not reported.
• Age <40 y	Not reported.

• Corticosteroid use	Not reported.
• Perianal lesions	Not reported.
Smoking	Not reported.
• Current	Not reported.
• Former	Not reported.
• Never	Not reported.
Perianal examination	Not reported.
• No lesion	Not reported.
• Skin tags	Not reported.
• Fissure or ulcer	Not reported.
• Simple fistula	Not reported.
• Complex fistula	Not reported.
CDAI score	Not reported.
HBI score	Not reported.
IBDQ	Not reported.
Haemoglobin concentration (g/dL)	Not reported.
C-reactive protein concentration (mg/L)	Not reported.
Albumin (g/L)	Not reported.
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy	Not reported.
Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported.
Time to test result	Not reported.
Number of test failures	Not reported.
Number of inconclusive test results	Not reported.

Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported.
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported.
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Need for surgery.
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Complicated disease behaviour.
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Data reported on associations between titres of individual markers and disabling disease.

Adverse effects of treatment	Not reported.			
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported.			
Section 4: Data extraction form				
Outcome	**IBDX**			
Prognostic test accuracy				
	N	N		
Clinical outcomes				
Dichotomous outcomes				
	N	N		
Disabling disease course	124	Increased titres of anti-L and ACCA were associated with disabling disease course, (p=0.0006 and p=0.0367, respectively).		
Complicated disease behaviour and/or need for surgery	113	Elevated titres of ACCA, gASCA IgG, gASCA IgA, Anti-C and Anti-L were all significantly associated with complicated disease behaviour and/or need for surgery (p=0.0022, <0.0001, <0.0001, 0.0003, and 0.0002, respectively).		
Complicated disease behaviour and/or surgery	113	Increasing number of positive antibodies was significantly associated with complicated disease behaviour and/or surgery (OR 3.3, 95% CI not reported; p=0.0005).		
Continuous outcomes				
	Time-frame	Mean	SD/SE	N
Time to event outcomes				
	Time-frame	HR	95% CI	p value
Section 5: Additional comments				
Additional comments				

Further information that could be requested from authors	Are the results of this study published in full elsewhere?
Abbreviations: CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; HRQoL, health-related quality of life; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; OR, odds ratio; RCT, randomised controlled trial; SD, standard deviation; SE, standard error.	

Item	**Details**
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Paul 2015
Study details (journal, year, volume, page range)	*J Crohn's Colitis*, 2015, 9, (6), 445–451
Type of report (full paper/only abstract/conference abstract)	Full paper
Section 2: Study information	
Location and number of sites	Department of Gastroenterology of Saint-Etienne's University Hospital, France.
Study sponsor	None
Conflicts of interest	None for most authors. One author had received fees from Merck for lecturing and acting as a consultant.
Patient enrolment (how patients were enrolled, and date to date of enrolment)	Consecutive in- and out-patients with IBD were enrolled between September 2009 and October 2010.
Study design (e.g., RCT, cross-over RCT)	Cross-sectional study.
Study duration (including any period of follow-up)	Not applicable
Inclusion criteria	Confirmed diagnosis of IBD for more than 1 year were enrolled. The diagnosis of IBD was assessed on common endoscopic and histological evidences of the disease, according to the Lennard-Jones criteria. For CD, disease activity was assessed using the CDAI.
Exclusion criteria	Those with indeterminate colitis.
Subgroups evaluated	Those with severe CD.
Definition of response	Not reported.
Definition of remission	Not reported.

Test	IBDX
Number in study, N	195 (107 with CD and 88 with UC)
Withdrawals, n (%)	Not applicable.
Details of follow-up for development of severe course of CD	68 people with CD met criteria for severe disease, defined as: • Uncontrolled active disease requiring adjunction of biotherapy with anti-TNF after failure with conventional immunosuppressant • Two or more IBD-related previous surgeries or bowel resection longer than 70 cm, • Concomitant active perianal disease with complex fistulas or spread bowel disease (spread bowel disease was defined regarding a length of small bowel higher than 50 cm with a damage of diffuse jujenal injury).
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Not applicable.
Baseline patient characteristics	**IBDX**
Median age (interquartile range), years	At inclusion: 41.3 (20 to 101) At diagnosis: 32.9 (9 to 84)
Sex (M/F), n (%)	Male: 50/107 (46.7%); Female: 57/107 (53.3%)
Ethnicity, n (%)	Not reported.
Median disease duration at inclusion (interquartile range), years	9.4 (1 to 44)
Disease location, n (%)	
• L1 (ileum)	14 (19%)
• L2 (colon)	12 (15%)
• L3 (ileum + colon)	52 (66%)
• L4 (upper digestive tract)	Not reported.
Disease behaviour, n (%)	
• Non-stricturing and non-penetrating (B1)	61 (58.7%)
• Stricturing (B2)	28 (26.9%)

• Penetrating (B3)	15 (14.4%)
• Perianal disease	22 (21.2%)
Predictors of disabling CD	
• Age <40 y	Not reported.
• Corticosteroid use	First-line: 73 (68.2%) Concomitant medications: 86 (80.4%) (steroid dependent: 74 [69.2%]; refractory: 18 [16.8%])
• Perianal lesions	Not reported: perianal disease is reported.
Smoking	
• Current	42.1%
• Former	Not reported.
• Never	Not reported.
Perianal examination	
• No lesion	Not reported.
• Skin tags	Not reported.
• Fissure or ulcer	Not reported.
• Simple fistula	Not reported.
• Complex fistula	Not reported.
CDAI score	Not reported.
HBI score	Not reported.
IBDQ	Not reported.
Haemoglobin concentration (g/dL)	Not reported.
C-reactive protein concentration (mg/L)	Not reported.
Albumin (g/L)	Not reported.
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy	Not reported.

Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported, only data for individual antibodies or combination of a few reported.
Time to test result	Not reported.
Number of test failures	Not reported.
Number of inconclusive test results	Not reported.
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported.
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported.
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure)	Not reported.

by subgroup of severe versus non-severe course of disease		
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported.	
Adverse effects of treatment	Not reported.	
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported.	
Section 4: Data extraction form		
Outcome	**IBDX**	
Prognostic test accuracy		
	n	N
Positive for at least one anti-glycan Antibody	70 (65.4%)	107
Differentiating severe from non-severe course of disease	As reported	The authors report that, using manufacturers' cut-offs, they found that a significant number of patients with severe CD had positive ASCA and anti-glycan AMCA antibodies (n=32 and n = 35, respectively) when compared with the number of patients who had none of the predefined criteria of severity in their past medical history disease (n=8 and n=8, p=0.011 and p=0.003, respectively). In CD, ASCA and AMCA antibodies had the best validity for association with a non-severe course, stable from severe disease, with an AUC of 0.63 for ASCA and 0.65 for AMCA antibodies. On combining both antibodies, the authors obtained a better diagnostic precision with an AUC of 0.71.
Differentiating severe from non-severe course of disease	Unclear	A titre of antibody higher than the threshold of 63 U/ml for ASCA, 50.7 U/ml for ACCA and 77.4 U/ml for AMCA antibodies was significantly associated with severe disease (OR 3.5: 95% CI; 1.37 to 9.2, p=0.0009; OR 2.8: 95% CI; 1.12 to 7.39, p=0.027; OR 4.3: 95% CI; 1.69 to 11.03, p=0.0022, respectively, for ASCA, ACCA, and AMCA antibodies). ASCA, ACCA, and AMCA antibodies were the best serological markers to diagnose severe CD with an AUC of 0.68 for ASCA, 0.66 for ACCA, and 0.68 for AMCA antibodies. The combination of the three antibodies [ASCA, ACCA, and AMCA]

		improved the diagnosis value associated with severe CD course with an AUC of 0.79.
Clinical outcomes		
Dichotomous outcomes		
	n	N
CD-related surgery	NR	Among all the panel of tested antibodies, only AMCA antibodies tended to be associated with higher risk of CD-related surgery with an odds of 2.1 (95% CI 0.8 to 5.1) but the level of significance was not reached (p=0.10).

Continuous outcomes				
	Time-frame	Mean	SD/SE	N

Time to event outcomes				
	Time-frame	HR	95% CI	p value

Section 5: Additional comments	
Additional comments	Clinical data analyses and serological assessments were performed in a blinded manner without knowledge of the patients' diagnosis and medical history. Study included a healthy control group [N=45] who were age- and gender-matched to IBD group. Consecutive healthy volunteers were selected from blood donors. Other than glucocorticosteroids, other first-line medications were anti-TNF therapy (1 [0.95%]), immunosuppressants (8 [7.5%]), and 5-ASA (25 [23.4%]). Other concomitant medications were: anti-TNF therapy (19 [17.8%]), immunosuppressants (50 [46.7%]), IBD-related surgery (41 [38.3%]).
Further information that could be requested from authors	Are any other outcomes available?

Abbreviations: ACCA, anti-chitobioside antibodies; AMCA, anti-mannobioside antibodies; ASA, aminosalicylate; ASCA, anti-*Saccharomyces cerevisiae* antibodies; AUC, are under curve; CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; HRQoL, health-related quality of life; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; OR, odds ratio; RCT, randomised controlled trial; SD, standard deviation; SE, standard error; TNF, tumour necrosis factor.

Item	Details
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Rieder 2010b
Study details (journal, year, volume, page range)	*Inflamm Bowel Dis*, 2010, 16, 263–274. Related paper: Rieder 2011a *PLoS ONE*, 2011, 6, (5), e18172 (presents data on subgroup of people reported in Rieder 2010b). Related paper: Rieder 2010d *Gastroenterology*, 2010, 138, (5), S522 (conference abstract for Rieder 2011a).
Type of report (full paper/only abstract/conference abstract)	Full paper.
Section 2: Study information	
Location and number of sites	IBD Centre of the Department of Internal Medicine I, University of Regensburg, Germany.
Study sponsor	Supported by the German Ministry of Education and Research and the Competence network “Inflammatory Bowel Disease”.
Conflicts of interest	Four authors were employees of Glycominds Ltd.
Patient enrolment (how patients were enrolled, and date to date of enrolment)	All CD in- and out-patients seen at the clinic between 2000 and 2006 were asked to participate in the serum repository.
Study design (e.g., RCT, cross-over RCT)	Prospective cohort.
Study duration (including any period of follow-up)	Not applicable.
Inclusion criteria	Diagnosis of CD was made based on clinical, radiographic, endoscopic, and histopathologic criteria.
Exclusion criteria	Indeterminate colitis.
Subgroups evaluated	Not reported.
Definition of response	Not reported.
Definition of remission	Not reported.
Test	**IBDX**
Number in study, N	363 people with CD (890 serum samples, multiple samples were available for 209 people) and a control group that comprised 461 people, 130 with ulcerative colitis,

	74 with other gastrointestinal diseases, and 257 controls with no IBD, no gastrointestinal disorder, or no apparent disease. Longitudinal analysis of a subgroup of people (859 serum samples of 253 people with IBD [207 CD, 46 ulcerative colitis] reported in Rieder 2011a.[74]
Withdrawals, n (%)	None.
Details of follow-up for development of severe course of CD	Not applicable.
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Time of follow-up was defined as months between CD diagnosis and occurrence of first complication or IBD-related surgery, or 1 April 2007 if no complication or surgery occurred.
Baseline patient characteristics	**IBDX**
Mean age (SD), years	At study: 35.7 (12.2) At diagnosis: 28.3 (12.0)
Sex (M/F), n (%)	Male: 170 (46.8%) Female: 193 (53.2%)
Ethnicity, n (%)	Not reported.
Median disease duration at inclusion (interquartile range), months	66.8 (11 to 141)
Disease location (non-exclusive), n (%)	
• Jejunum, prox. Ileum	40 (11.2%)
• Ileocecal	109 (30.0)
• Colon w/o cecum	54 (14.9%)
• Ileum and colon	199 (55.0%)
• Upper digestive tract)	40 (11.1%)
• Rectum	131 (36.3%)
Disease behaviour, n (%)	
• Inflammatory	87 (24.0%)
• Stricturing	90 (24.8%)

• Fistulae	186 (51.2%)
Predictors of disabling CD	
• Age <40 y	Not reported.
• Corticosteroid use	Not reported.
• Perianal lesions	Not reported.
Smoking	
• Current	Not reported.
• Former	Not reported.
• Never	Not reported.
Perianal examination	
• No lesion	Not reported.
• Skin tags	Not reported.
• Fissure or ulcer	Not reported.
• Simple fistula	Not reported.
• Complex fistula	Not reported.
CDAI score	Not reported.
HBI score	Not reported.
IBDQ	Not reported.
Haemoglobin concentration (g/dL)	Not reported.
C-reactive protein concentration (mg/L)	Not reported.
Albumin (g/L)	Not reported.
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy	Not reported; data reported only for CD versus no CD, or versus individual control groups.

Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported.
Time to test result	Not reported.
Number of test failures	Not reported.
Number of inconclusive test results	Not reported.
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported.
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported.
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure)	Complicated disease behaviour was defined as the occurrence of fistulae or stenosis before or during follow-up (249 [68.6%] people had a complication before or at the time of sample procurement).

by subgroup of severe versus non-severe course of disease	Internal penetrating disease was retrospectively distinguished from perianal penetrating disease. Disease behavioural phenotype was assigned based on the Vienna classification.	
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported.	
Adverse effects of treatment	Not reported.	
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported.	
Section 4: Data extraction form		
Outcome	**IBDX**	
Prognostic test accuracy		
	n	N
Number of positive markers (%)		363
• 0	22.6	
• 1	24.2	
• 2	25.9	
• 3	10.5	
• 4	10.7	
• 5	3.9	
• 6	2.2	
Number of positive markers	The authors reported that positivity for an increasing number of antibodies raised the likelihood for the occurrence of complicated disease behaviour and IBD-related surgery.	
Clinical outcomes		
Dichotomous outcomes		
	n	N

Analyses based on median number of positive markers: OR reported for median positive markers present 2.0 (1.0 to 3.0) versus median positive markers absent 1.0 (0.0 to 2.0)		
OR for complication (95% CI; p value)	Unclear	1.5 (1.3 to 1.9; p<0.001)
OR for CD-related surgery (95% CI; p value)	Unclear	1.5 (1.3 to 1.8; p<0.001)
OR for early disease onset (95% CI; p value)	Unclear	1.4 (1.1 to 1.8; p=0.003)
Complications	Unclear	In univariate analysis, anti-L had the strongest association.
Complications	Unclear	In multivariate analysis gASCA, AMCA and anti-L were all independently associated with complications.
CD-related surgery	Unclear	In univariate analysis, anti-L and anti-C had the strongest associations.
CD-related surgery	Unclear	In multivariate analysis gASCA, AMCA, anti-C and anti-L were all independently associated with surgery.

Results from longitudinal analysis[74]

Analysis based on 75 people with CD for whom at least four serum samples were available during follow-up.

Data presented below are for the validity of markers for association with disease phenotypes over time in individual patients and values are median (25th percentile, 75th percentile).

	Ileum involvement (n=67)	**No ileum involvement (n=8)**	**p value**
First sample			
Sum of quartiles	16.0 (13.5 to 19.5)	10.0 (8.0 to 14.0)	<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)	<0.001
Second sample			
Sum of quartiles	15.5 (12.0 to 19.0)	10.0 (8.0 to 14.0)	<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)	<0.001
Third sample			
Sum of quartiles	15.0 (12.0 to 19.0)	10.0 (8.0 to 14.0)	<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)	<0.001
Fourth sample			
Sum of quartiles	15.0 (12.0 to 19.0)	10.0 (8.0 to 14.0)	<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)	<0.001
	Complication	**No complication**	**p value**

	(n=65)	**(n=10)**		
First sample				
Sum of quartiles	17.0 (13.0 to 20.0)	10.0 (8.0 to 14.0)		<0.001
Number of positive markers	2.0 (1.0 to 4.0)	0.0 (0.0 to 1.0)		<0.001
Second sample				
Sum of quartiles	16.0 (12.0 to 19.0)	10.0 (8.0 to 13.0)		<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)		<0.001
Third sample				
Sum of quartiles	15.0 (12.0 to 19.0)	9.0 (7.0 to 13.0)		<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)		<0.001
Fourth sample				
Sum of quartiles	15.0 (12.0 to 19.0)	10.0 (8.0 to 14.0)		<0.001
Number of positive markers	2.0 (1.0 to 3.0)	0.0 (0.0 to 1.0)		<0.001
	Surgery (n=60)	**No surgery (n=15)**		**p value**
First sample				
Sum of quartiles	17.5 (14.0 to 21.0)	14.0 (10.0 to 16.0)		0.003
Number of positive markers	2.0 (1.0 to 4.0)	1.0 (0.0 to 2.0)		0.002
Second sample				
Sum of quartiles	16.5 (13.0 to 19.5)	12.0 (12.0 to 15.0)		0.004
Number of positive markers	2.0 (1.0 to 3.0)	1.0 (0.0 to 2.0)		0.003
Third sample				
Sum of quartiles	16.5 (13.0 to 19.0)	11.0 (8.0 to 15.0)		0.004
Number of positive markers	2.0 (1.0 to 3.0)	1.0 (0.0 to 2.0)		0.006
Fourth sample				
Sum of quartiles	15.0 (13.0 to 20.0)	13.0 (10.0 to 16.0)		0.02
Number of positive markers	2.0 (1.0 to 3.0)	1.0 (0.0 to 2.0)		0.025
Continuous outcomes				
	Time-frame	Mean	SD/SE	N
Time to event outcomes				

	Time-frame	HR	95% CI	p value
Section 5: Additional comments				
	For people with multiple visits, the serum samples of the earliest timepoint were used for cross-sectional analysis. After a median follow-up (after sample procurement) of 59 months, 277 people had experienced a complication. 196 people with complications had more than one complication during their disease course. 116 people with complicated disease behaviour had both fistulae and stenoses. Twenty percent of the sera were collected within 6-months of CD diagnosis and 30% within 1-year. At time of first sample, 154 patients (74.4%) had undergone CD-related surgery. CD-associated abdominal surgery was separated from CD-related perianal surgery (e.g., perianal abscess drainage or perianal fistula treatment). 257 (70.8%) people required IBD-related surgery (224 at the time of sample procurement and 33 during follow-up). Samples were analysed in a blinded manner. Marked changes in the levels of anti-glycan antibodies were noted over time (median time between procurement of samples was 6.2 months (interquartile range 3.5 to 12.2 months). However, stability in marker status (positive or negative for a particular antibody) was stable — between 74.2% and 89.0% of people with CD had same antibody status in subsequent samples as in their initial sample. **Longitudinal analysis** The occurrence of first time complicated CD behaviour, first time CD-related surgery, active disease or the start of immunosuppressive medication did not influence the overall immune response or the levels of the individual markers. Independent of the time of sample procurement over the follow-up period the overall immune response remained to be associated with ileal involvement, complicated CD behaviour or CD-related surgery. The authors reported that the strengths of the study are the prospective follow-up design, blinded data abstraction, availability of up to 11 samples per patient and a longitudinal analysis paired with detailed clinical information. Limitations were reported as the cohort was derived from a single university hospital, the first serum sample per patient was not in all cases taken close to diagnosis (corrected for in statistical analysis), follow-up samples were taken at arbitrary visits to the hospital and not in a fixed relation to certain events such as complications or surgery. Using this method, patients with a more severe disease course could be selected out, as only they have to come to a referral centre for multiple treatments. The length of time in which immunosuppressive medication was taken by the patients before sample procurement was unknown.			

	The authors concluded, "*Considering the follow-up time of our study we cannot recommend serial measurements, when considering the marker status and claim to use the overall immune response (QSS) with caution for disease stratification, due to strong fluctuations in a subgroup of CD subjects*". Two people with CD did not have any CRP measures during follow-up.
Further information that could be requested from authors	
Abbreviations: AMCA, anti-mannobioside antibodies; anti-C, anti-chitin antibody; anti-L, anti-laminarin antibody; ASCA, anti-*Saccharomyces cerevisiae* antibodies; CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; CRP, C-reactive protein; HRQoL, health-related quality of life; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; RCT, randomised controlled trial; SD, standard deviation; SE, standard error.	

Item	**Details**
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Rieder 2010c
Study details (journal, year, volume, page range)	*Inflamm Bowel Dis*, 2010, 16, (8), 1367–1375.
Type of report (full paper/only abstract/conference abstract)	Full paper.
Section 2: Study information	
Location and number of sites	IBD Centre of the Department of Internal Medicine I, University of Regensburg, Germany.
Study sponsor	Funding came from the German Ministry of Education and Research (BMBF) via the Kompetenznetz ''Chronisch entzu ¨ndliche Darmerkrankungen''.
Conflicts of interest	Four authors are employees of Glycominds Ltd.
Patient enrolment (how patients were enrolled, and date to date of enrolment)	All CD in- and out-patients seen at the clinic between 2000 and 2006 were considered for participation in the study. Clinical data collected included: • age at diagnosis; • body mass index (BMI); • gender; • date of sample procurement;

	• date and type of first complication and surgery; • disease location.		
Study design (e.g., RCT, cross-over RCT)	Prospective cohort.		
Study duration (including any period of follow-up)	Unclear. In July 2008 , all patient charts and the database were reviewed and clinical data updated.		
Inclusion criteria	Diagnosis of CD was made based on clinical, radiographic, endoscopic, and histopathologic criteria. People with CD were included if they did not have a complication of CD (defined as fistula or stenosis), and had not undergone CD-related surgery before or within 20 days of sample procurement.		
Exclusion criteria	Excluded if time to sample procurement to event was ≤20 days or if they had a pure inflammatory disease course with a follow-up of less than 3 years.		
Subgroups evaluated	People with no previous complication and no prior surgery; People progressing to first complication after previous CD-related surgery; People progressing to first surgery after previous complication.		
Definition of response	Not reported.		
Definition of remission	Not reported.		
Test	**IBDX**		
Number in study, N	76 with no prior complication or surgery, 33 with prior surgery or complication and progression to new event and 33 with prior surgery or complication and no new event (149 patients in cohort, 7 excluded for not meeting inclusion criteria, 17 with prior surgery but no complication and 49 with prior complication but no surgery)		
Withdrawals, n (%)	None.		
Details of follow-up for development of severe course of CD	People were followed for development of an event, which was defined as complication or surgery.		
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Median follow up was 53.7 months in those with no prior complication or surgery.		
Baseline patient characteristics	**IBDX**		
	Without prior complication or surgery **(N=76)**	**With prior surgery or complication and progression to new event**	**With prior surgery or complication and no new event** **(N=33)**

		(N=33)	
Mean age (SD), years At study entry At diagnosis	 33.3 (12.5) 30.4 (12.1)	 32.3 (11.9) 26.8 (11.2)	 35.5 (11.6) 28.0 (11.2)
Sex (M/F), n (%) Female	 41 (54.0%)	 18 (54.6)	 21 (63.6)
Ethnicity, n (%)	Not reported.		
Median disease duration at time of first sample (interquartile range), months	10.6 (1.7 to 52.3)	39.1 (11.7 to 110.9)	72.7 (16.4 to 122.4)
Disease location (non-exclusive), n (%)			
• Jejunum, prox. ileum	2 (2.7%)	4 (12.1)	1 (3.0)
• Ileal involvement	64 (84.2%)	27 (81.8)	29 (87.9)
• Colon w/o cecum	12 (15.8%)	6 (18.2)	4 (12.1)
• Upper digestive tract)	15 (19.7%)	3 (9.7)	5 (15.6)
• Rectum	12 (16.0%)	11 (33.3)	11 (34.4)
Predictors of disabling CD			
• Age <40 y	Not reported.		
• Corticosteroid use	Not reported.		
• Perianal lesions	Not reported.		
Smoking			
• Current	Not reported.		
• Former	Not reported.		
• Never	Not reported.		
Perianal examination			
• No lesion	Not reported.		
• Skin tags	Not reported.		

• Fissure or ulcer	Not reported.
• Simple fistula	Not reported.
• Complex fistula	Not reported.
CDAI score	Not reported.
HBI score	Not reported.
IBDQ	Not reported.
Haemoglobin concentration (g/dL)	Not reported.
C-reactive protein concentration (mg/L)	Not reported.
Albumin (g/L)	Not reported.
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy	Not reported.
Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported.
Time to test result	Not reported.
Number of test failures	Not reported.
Number of inconclusive test results	Not reported.
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported.
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported.

Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Time to complication. Complicated disease behaviour was defined as occurrence of fistula or stenoses.	
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported.	
Adverse effects of treatment	Not reported.	
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported.	
Section 4: Data extraction form		
Outcome	**IBDX**	
Prognostic test accuracy		
	n	N
At least 1 positive marker • Without prior complication or surgery	53	76
• With prior surgery or complication and no new event	31	33

• With prior surgery or complication and progression to new event	27	33
At least 2 positive markers		
• Without prior complication or surgery	28	76
• With prior surgery or complication and no new event	19	33
• With prior surgery or complication and progression to new event	17	33
At least 3 positive markers		
• Without prior complication or surgery	11	76
• With prior surgery or complication and no new event	5	33
• With prior surgery or complication and progression to new event	11	33

Clinical outcomes

Dichotomous outcomes

	N	N

Continuous outcomes

	Time-frame	Mean	SD/SE	N

Time to event outcomes

	Time-frame	HR	95% CI	p value

Unadjusted analyses for those with no prior complication or surgery

Time to an event (complication or surgery) (23 people experienced an event)

At least 1 positive marker		2.3	0.78 to 6.8	0.13
At least 2 positive markers		2.9	1.3 to 6.7	0.011
At least 3 positive markers		3.4	1.3 to 8.7	0.01
Time to a complication (20 people experienced a complication)				
At least 1 positive marker		2.0	0.65 to 5.9	0.23
At least 2 positive markers		2.7	1.1 to 6.5	0.028
At least 3 positive markers		3.0	1.07 to 8.2	0.036
Time to surgery (14 people underwent surgery)				
At least 1 positive marker		2.6	0.58 to 11.9	0.21
At least 2 positive markers		3.6	1.2 to 10.7	0.023
At least 3 positive markers		2.7	0.81 to 9.1	0.1
Adjusted analyses for those with no prior complication or surgery (adjusted for age, sex, BMI, disease activity and duration, age at diagnosis and disease location)				
Time to an event (complication or surgery) (23 people experienced an event)				
At least 1 positive marker		2.2	0.74 to 6.5	0.16
At least 2 positive markers		2.8	1.2 to 6.4	0.016
At least 3 positive markers		3.1	1.2 to 8.1	0.019
Time to a complication (20 people experienced a complication)				
At least 1 positive marker		1.8	0.61 to 5.4	0.29
At least 2 positive markers		2.5	1.03 to 6.1	0.043
At least 3 positive markers		2.6	0.92 to 7.2	0.072
Time to surgery (14 people underwent surgery)				
At least 1 positive marker		2.6	0.58 to 12.0	0.21
At least 2 positive markers		3.6	1.2 to 11.0	0.023
At least 3 positive markers		2.8	0.80 to 9.6	0.11

Section 5: Additional comments	
Additional comments	Disease activity was determined by the treating physician and patients were grouped according to active versus non-active disease. Authors retrospectively distinguished internal penetrating from perianal fistulising disease based on the Montreal Classification. Samples were considered positive at cut off ELISA units of: • gASCA: 35; • ACCA: 70; • ALCA: 60; • AMCA: 100; • anti-L: 120; • anti-C: 50. Of the 76 people evaluated, 26.3% experienced a complication during follow-up. Median time to complication was 11.6 months. Median time to surgery was 11.6 months. Authors reported that CD patients positive for at least 2 antibodies showed a more severe disease course.
Further information that could be requested from authors	
Abbreviations: ACCA, anti-chitobioside antibodies; ALCA, anti-laminaribioside antibodies; AMCA, anti-mannobioside antibodies; ASCA, anti-Saccharomyces cerevisiae antibodies; anti-C, anti-chitin antibody; anti-L, anti-laminarin antibody; AUC, are under curve; CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; HRQoL, health-related quality of life; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; RCT, randomised controlled trial; SD, standard deviation; SE, standard error.	

Item	**Details**
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Rieder 2012
Study details (journal, year, volume, page range)	*Inflamm Bowel Dis*, 2012, 18, (7), 1221–1231. Related paper: Rieder 2010a *Gastroenterology*, 2010, 138, (5), S301–S302. Related paper: Rieder 2011b *J Crohns Colitis*, 2011, 5, (1), S48 (conference abstract).

Type of report (full paper/only abstract/conference abstract)	Full paper.
Section 2: Study information	
Location and number of sites	IBD Centre of the Children's Hospital St Hedwig, University of Regensburg, Germany.
Study sponsor	Funding came from the German Ministry of Education and Research (BMBF) via the Kompetenznetz ''Chronisch entzu ¨ndliche Darmerkrankungen'', the DFG Excellence Cluster Inflammation at Interfaces and by the Deutsche Forschungsgemeinschaft (DFG).
Conflicts of interest	Not reported
Patient enrolment (how patients were enrolled, and date to date of enrolment)	All paediatric CD in- and out-patients seen at the clinic in 2008 were asked to participate in the serum repository.
Study design (e.g., RCT, cross-over RCT)	Cross-sectional analysis
Study duration (including any period of follow-up)	Not applicable
Inclusion criteria	Diagnosis of CD and age below 18 years. Diagnosis of CD was made based on clinical, radiographic, endoscopic and histopathologic criteria. Clinical data collected included: • age at and time of diagnosis; • BMI; • gender; • date of sample procurement; • type of complication (fistula versus stenosis); • disease location; • extraintestinal manifestations; • steroid response; • family history.
Exclusion criteria	No exclusion criteria within the CD population.
Subgroups evaluated	None
Definition of response	Not reported
Definition of remission	Not reported

Test	**IBDX**
Number in study, N	59 children
Withdrawals (please specify reasons for withdrawal, including loss to follow-up; use different rows for different reasons), n (%)	Not applicable.
Details of follow-up for development of severe course of CD	Not applicable.
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Not applicable.
Baseline patient characteristics	**IBDX (N=59)**
Mean age (SD), months	At study: 151.9 (42.8) At diagnosis: 124.4 (39.4)
Sex (M/F), n (%)	Male: 36 (61.0) Female: 23 (39.0)
Ethnicity, n (%)	Not reported
Median disease duration (IQR), months	18.0 (12.0 to 43.0)
Disease location (non-exclusive), n (%)	
• L1 (ileum)	Terminal ileum: 40 (67.8)
• L2 (colon)	Colonic involvement: 56 (94.9)
• L3 (ileum + colon)	Not clear
• L4 (upper digestive tract)	42 (71.2)
Disease behaviour, Vienna classification	
• Inflammatory	36 (61.0)
• Stricture	7 (11.9)
• Fistulae	16 (27.1)

CD-related surgery	20 (33.9)
Predictors of disabling CD	
• Age <40 y	Not reported
• Corticosteroid use	Not reported
• Perianal lesions	Not reported
Smoking	
• Current	Not reported
• Former	Not reported
• Never	Not reported
Perianal examination	
• No lesion	Not reported
• Skin tags	Not reported
• Fissure or ulcer	Not reported
• Simple fistula	Not reported
• Complex fistula	Not reported
CDAI score	Not reported
HBI score	Not reported
IBDQ	Not reported
Haemoglobin concentration (g/dL)	Not reported
C-reactive protein concentration (mg/L)	Not reported
Albumin (g/L)	Not reported
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy (e.g., sensitivity and specificity, and/ or if raw data are available, the numbers of true positive, true negative, false positive	Not reported for whole panel. Specificity and sensitivity are reported for the individual biomarkers in comparison to patients with no CD, UC and healthy controls.

and false negative test results for predicting course of disease)	
Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported
Time to test result	Not reported
Number of test failures	Not reported
Number of inconclusive test results	Not reported
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Steroid response was determined by the treating clinician (one of the authors) based on clinical characteristics included in paediatric CDAI without documenting a specific change in paediatric CDAI score.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Assessed need for CD-related surgery at any time during the disease course until the end of follow-up.
Rates of and time to serious complication (e.g., obstruction,	Complicated disease behaviour in CD was defined as the occurrence of fistulae or stenoses at any time during the disease course until the end of follow-up.

intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Fistulising disease included internal and perianal penetrating fistulae. A stenosis was defined endoscopically or radiologically.
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported
Adverse effects of treatment	Not reported
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported

Section 4: Data extraction form			
Outcome	**IBDX**		
Prognostic test accuracy			
	N	N	
Number of positive markers			
• 0	13	59	
• 1	16	59	
• 2	13	59	
• 3	8	59	
• 4	7	59	
• 5	2	59	
• 6	0	59	
Clinical outcomes			
Dichotomous outcomes			
	n (%)	N	OR (95% CI; p-value)
Complication, n (%)	23 (39.0)	23	Complication vs. no complication
At least 1 marker positive	20 (87.0)	23	2.6 (0.62 to 10.6; 0.18)

At least 2 markers positive At least 3 markers positive	16 (69.6) 12 (52.2)	23 23	3.6 (1.2 to 10.9; 0.022) 6.8 (1.9 to 23.6; 0.002)
Complications		The authors reported that ALCA and Anti-L had the strongest association with complications. ACCA and Anti-C were not associated with complicated disease behaviour.	
CD-related surgery, n (%) At least 1 marker positive At least 2 markers positive At least 3 markers positive	20 (33.9) 19 (95.0) 13 (65.0) 9 (45.0)	59 20 20 20	CD-related surgery vs. no CD-related surgery 8.4 (1.01 to 70.5; 0.043) 2.4 (0.79 to 7.3; 0.12) 3.2 (0.98 to 10.3; 0.049)
CD-related surgery		The authors reported that ALCA was associated with CD-related surgery.	

Continuous outcomes				
	Time-frame	Mean	SD/SE	N

Time to event outcomes				
	Time-frame	HR	95% CI	p value

Section 5: Additional comments	
Additional comments	Early disease onset was defined as an age of onset of <10 years. Included a control group of 72 people (27 with ulcerative colitis and 45 controls [no IBD]). Additionally, compared the paediatric cohort against an adult cohort derived from the cohort described in 2010b (355 adults with CD, 129 adults with ulcerative colitis, and 244 controls). Authors found lower cut off points for paediatric samples compared with adults. No clinically relevant association of marker expression and extraintestinal manifestations, response to corticosteroids, IBD family history, CRP levels or disease activity could be detected. Compared with the adult cohort, authors report that the ability of the panel to discriminate disease phenotypes remained constant over the different age groups,

	with the exception of AMCA and anti-C, which showed a decrease accuracy with increased age.
	After correcting for age, sex, BMI, disease duration and disease location, respectively, multivariate logistic regression analysis of the CD subjects using each marker separately identified independent variables for marker positivity of gASCA IgG (p=0.015), gASCA IgA (p=0.012) and AMCA (P=0.005) for complicated disease behaviour.[a]
	In paediatric CD patients, positivity for at least one marker out of the whole panel versus no marker positive was independently associated with fibrostenotic or fistulizing disease behaviour (p=0.036) and ileal disease location (p=0.014).[a]
Further information that could be requested from authors	

[a] Taken from Rieder 2010a.

Abbreviations: ACCA, anti-chitobioside antibodies; ASCA, anti-*Saccharomyces cerevisiae* antibodies; BMI, body mass index; CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; HRQoL, health-related quality of life; RCT, randomised controlled trial; SD, standard deviation; SE, standard error.

Item	**Details**
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Seow 2009
Study details (journal, year, volume, page range)	*Am J Gastro*, 2009, 104, (6), 1426–1434.
Type of report (full paper/only abstract/conference abstract)	Full paper.
Section 2: Study information	
Location and number of sites	Mount Sinai Hospital and the Hospital for Sick Children, both hospitals located in Toronto, Canada.
Study sponsor	Funding came from various organisations and fellowships, including Glycominds Limited.
Conflicts of interest	One of the authors is a stockholder and employee of the company, and another had received partial research funding from the company.

Patient enrolment (how patients were enrolled, and date to date of enrolment)	People were recruited from the hospitals above between 2002 and 2006.
Study design (e.g., RCT, cross-over RCT)	Cross-sectional analysis
Study duration (including any period of follow-up)	Not applicable.
Inclusion criteria	People with IBD. Clinical diagnosis of CD was used as the gold standard (further details not available). People were independently classified using serology. People were designated as having CD if their serum tested positive for at least one of gASCA IgG, gASCA IgA, ACCA, ALCA or AMCA. If none of the markers was positive but the person was seropositive for atypical pANCA, they were designated as having ulcerative colitis.
Exclusion criteria	Not reported.
Subgroups evaluated	Not reported.
Definition of response	Not reported.
Definition of remission	Not reported.
Test	**IBDX**
Number in study, N	818 with IBD (517 with CD and 301 with ulcerative colitis; 97 healthy controls were also included).
Withdrawals, n (%)	Not applicable.
Details of follow-up for development of severe course of CD	Not applicable.
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Not applicable.
Baseline patient characteristics	**IBDX**
Median age (range), years	At enrolment: 33 (2 to 76) At diagnosis (interquartile range): 19 (13 to 26)
Sex (M/F), n (%)	Male: 253 (48.9%) Female: 264 (51.1%)
Ethnicity, n (%)	Not reported.
Median disease duration (range), years	8.9 (0.02 to 46.30)

Disease location	
• L1 (ileum)	161
• L2 (colon)	93
• L3 (ileum + colon)	254
• L4 (upper digestive tract)	105
Disease behaviour based on Montreal classification	
• B1	259
• B2	133
• B3	123
• Perianal	138
Predictors of disabling CD	
• Age <40 y	Not reported.
• Corticosteroid use	Not reported.
• Perianal lesions	Not reported.
Smoking, n (%)	
• At time of diagnosis	134 (25.9%)
• Former	112 (21.7%)
• Never	Unclear
Perianal examination	
• No lesion	Not reported.
• Skin tags	Not reported.
• Fissure or ulcer	Not reported.
• Simple fistula	Not reported.

• Complex fistula	Not reported.
CDAI score	Not reported.
HBI score	Not reported.
IBDQ	Not reported.
Haemoglobin concentration (g/dL)	Not reported.
C-reactive protein concentration (mg/L)	Not reported.
Albumin (g/L)	Not reported.
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy	Not reported.
Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported, only data on numbers of positive antibodies in relation to different characteristics were reported.
Time to test result	Not reported.
Number of test failures	Not reported.
Number of inconclusive test results	Not reported.
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported.

Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported.	
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Not reported.	
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported.	
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported.	
Adverse effects of treatment	Not reported.	
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported.	
Section 4: Data extraction form		
Outcome	**IBDX**	
Prognostic test accuracy		
	n (%)	N

Positive for at least one of the glycan markers	378 (73%)	517				
Phenotype	Number of positive antibodies					
	1 (n=103)	2 (n=130)	3 (n=77)	4 (n=38)	≥5 (n=30)	p value
Mean age at diagnosis (SD), years	23.5 (12.3)	19.6 (8.6)	20.2 (8.5)	18.0 (7.1)	19.8 (8.0)	0.0004
Mean duration of disease, years	10.0	12.1	13.0	12.7	13.4	0.005
Penetrating CD (B3) (%)	24.27	25.38	31.17	36.84	46.67	<0.0001
Perianal disease (%)	21.36	28.46	24.68	36.84	53.33	0.0005
Ileocolonic disease (%)	48.54	56.15	41.56	65.79	73.33	0.0002
Abdominal surgery (%)	51.64	54.62	63.64	57.89	76.67	<0.0001
Clinical outcomes						
Dichotomous outcomes						
	n	N				
Continuous outcomes						
	Time-frame	Mean	SD/SE	N		
Time to event outcomes						
	Time-frame	HR	95% CI	p value		
Section 5: Additional comments						
Additional comments	Date of serum collection was used as the reference date to calculate the age of the patient, and phenotype was assigned at that time using the Montreal classification with no knowledge of the results of serological testing. Samples were considered positive at cut off ELISA units of: • gASCA IgG: 50; • gASCA IgA: 50; • ACCA: 90; • ALCA: 60;					

	• AMCA: 100; • anti-L: 60; • anti-C: 90. 206 people (29.8) were aged 16 years or younger at diagnosis. 271 people had any type of CD surgery (abdominal or perianal). Authors carried out a quartile sum analysis, with score in range of 7 to 28 (increasing score represents higher antibody titres). Authors reported that the higher the score the less likely a person was to have non-stricturing, non-penetrating disease, and the more likely they were to disease of penetrating behaviour and were more likely to require abdominal surgery, with odds ratios of: • non-stricturing, non-penetrating disease: OR 0.88 (95% CI: 0.85 to 0.92; p<0.0001); • penetrating behaviour: OR1.12 (95% CI: 1.07 to 1.18; p<0.0001); • requiring abdominal surgery: OR 1.16 (95% CI: 1.12 to 1.21; p<0.0001).
Further information that could be requested from authors	Can you confirm that Glycominds Ltd provided the kits for assay of anti-L and anti-C?
Abbreviations: ACCA, anti-chitobioside antibodies; ALCA, anti-laminaribioside antibodies; AMCA, anti-mannobioside antibodies; ASCA, anti-Saccharomyces cerevisiae antibodies; anti-C, anti-chitin antibody; anti-L, anti-laminarin antibody; AUC, are under curve; CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; HRQoL, health-related quality of life; RCT, randomised controlled trial; SD, standard deviation; SE, standard error.	

Item	**Details**
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Wolfel 2017
Study details (journal, year, volume, page range)	*Gastroenterology*, 2017, 152, (5), S605
Type of report (full paper/only abstract/conference abstract)	Conference abstract
Section 2: Study information	
Location and number of sites	Tertiary referral centre: location unclear.
Study sponsor	Not reported.
Conflicts of interest	Not reported.

Patient enrolment (how patients were enrolled, and date to date of enrolment)	Not reported.
Study design (e.g., RCT, cross-over RCT)	Prospective cohort
Study duration (including any period of follow-up)	Not reported
Inclusion criteria	People with CD who had undergone one surgical resection. Use of Glycominds tool was confirmed with authors.
Exclusion criteria	Not reported.
Subgroups evaluated	Not reported.
Definition of response	Not reported.
Definition of remission	Not reported.
Treatment	**IBDX**
Number in study, N	118
Withdrawals, n (%)	Not reported.
Details of follow-up for development of severe course of CD	Not reported.
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	Median follow-up after first surgery was 100 months.
Baseline patient characteristics	**IBDX**
Mean age (SD), years	37.7 (not reported)
Sex (M/F), n (%)	Male: 56 (47.5) Female: 62 (52.5)
Ethnicity, n (%)	Not reported.
Disease duration (months)	Not reported.
Disease location	
• L1 (ileum)	89% had ileal disease location before surgery
• L2 (colon)	Not reported.
• L3 (ileum + colon)	Not reported.

• L4 (upper digestive tract)	Not reported.
Predictors of disabling CD	
• Age <40 y	Not reported.
• Corticosteroid use	Not reported.
• Perianal lesions	Not reported.
Smoking	
• Current	Not reported.
• Former	Not reported.
• Never	Not reported.
Perianal examination	
• No lesion	Not reported.
• Skin tags	Not reported.
• Fissure or ulcer	Not reported.
• Simple fistula	Not reported.
• Complex fistula	Not reported.
CDAI score	Not reported.
HBI score	Not reported.
IBDQ	Not reported.
Haemoglobin concentration (g/dL)	Not reported.
C-reactive protein concentration (mg/L)	Not reported.
Albumin (g/L)	Not reported.
Section 3: Outcomes	
Outcome	**Definition**
Prognostic accuracy	Not reported.
Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported.

Time to test result	Not reported.
Number of test failures	Not reported.
Number of inconclusive test results	Not reported.
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported.
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported.
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported.
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported.
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Time to surgical recurrence.
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported.
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal	Not reported.

ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease				
Adverse effects of treatment	Not reported.			
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported.			
Section 4: Data extraction form				
Outcome	**IBDX**			
Time to event outcomes				
	Time-frame	HR	95% CI	p value
Time to surgery			After adjustment for ileal disease location and the use of immunosuppressants or anti-TNF after the first surgery, only positivity for AMCA (HR 2.6: 95% CI; 1.1 to 5.9; p=0.026) and ALCA (HR 2.3: 95% CI 1.04 to 5.3; p=0.039) predicted a shorter time to surgical recurrence. The authors state “Considering the entire antibody panel, neither the quartile sum score nor the number of positive markers combined predicted a shorter time to surgery suggesting a specific effect of AMCA and ALCA”.	
Section 5: Additional comments				
Additional comments	92% of people underwent first surgery due to internal penetrating and/or stricturing disease. 22% of patients underwent re-surgery within a median time after sample procurement of 98.5 months with 73.1% of resections being located in the small bowel and 92% of re-surgery being due to fistulae or strictures. The median interval between sample procurement and first surgery was 14.7 months.			
Further information that could be requested from authors	Are the results of this study published in full elsewhere?			

Abbreviations: ACCA, anti-chitobioside antibodies; ALCA, anti-laminaribioside antibodies; AMCA, anti-mannobioside antibodies; ASCA, anti-Saccharomyces cerevisiae antibodies; anti-C, anti-chitin antibody; anti-L, anti-laminarin antibody; CD, Crohn’s disease; CDAI, Crohn’s disease Activity Index; CI, confidence interval; HR, hazard ratio; HRQoL, health-related quality of life; IBDQ, Inflammatory Bowel Disease Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; RCT, randomised controlled trial; TNF, tumour necrosis factor; SD, standard deviation; SE, standard error.

9.5.1.2 PredictSURE-IBD

Item	Details
Section 1: Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Biasci 2019
Study details (journal, year, volume, page range)	*Gut*, 2019, (68), 1386–1395
Type of report (full paper/only abstract/conference abstract)	Full paper
Section 2: Study information	
Location and number of sites	Training cohort (N=66): Addenbrooke's Hospital, Cambridge, UK. Validation cohort (N=66): Four UK teaching hospitals in Cambridge, Nottingham, Exeter and London.
Study sponsor	Study was funded by grants from the Wellcome Trust, Crohn's and Colitis UK, Medical Research Council, and the Cambridge NIHR Biomedical Research Centre. Analytical validation experiments were funded by PredictImmune. One contributor was supported by Marie Curie.
Conflicts of interest	DB, JCL, EM, PAL and KGCS are coinventors on a patent covering the method of assessing prognosis in IBD. EM, PAK and KGCS are cofounders and consultants for PredictImmune. JCL is a consultant for PredictImmune.
Patient enrolment (how patients were enrolled, and date to date of enrolment)	Two different time periods for patient recruitment are reported in the full publication: (i) between 2008 and 2014 (training cohort); and (ii) between 2009 and 2017 (validation cohort).
Study design (e.g., RCT, cross-over RCT)	Prospective cohort study.
Trial duration (including any period of follow-up)	Not reported.
Inclusion criteria	Patients with active CD and UC, who were not receiving concomitant corticosteroids, immunomodulators or biological therapy, were recruited from a specialist IBD clinic before their treatment commenced. A stable dose of topical or oral 5-ASA was permitted if patients had been diagnosed previously. All patients were diagnosed with CD or UC based on standard endoscopic, histological and radiological criteria.

	To be enrolled, patients had to have active disease confirmed by one or more objective marker (raised CRP, raised calprotectin or endoscopic evidence of active disease) in addition to active symptoms and/or signs. Disease activity was assessed by considering symptoms, clinical signs, blood tests (CRP, haemoglobin and albumin), stool markers (calprotectin) and endoscopic assessment where indicated. All people were aged 18 years or older.			
Exclusion criteria	None reported.			
Subgroups evaluated	Within CD and UC, based on phenotype data, those identified as IBD1 (high risk of severe course of disease) and those classified as IBD2 (low risk of severe course of disease). Detailed phenotype data were collected prospectively.			
Definition of response	Not defined.			
Definition of remission	Not defined.			
Treatment	**PredictSURE-IBD**			
Number in study, N	Training cohort: 118 (CD=66, UC=52) Validation cohort: 123 (CD=66, UC=57)			
Withdrawals (please specify reasons for withdrawal, including loss to follow-up; use different rows for different reasons), n (%)	Not reported.			
Details of follow-up for development of severe course of CD	Time to first treatment escalation and number of treatment escalations were used as indicators for severe course of disease.			
Duration/length of follow-up for development of severe course of CD (mean, with SD/SE if given. If no mean presented, median values with ranges)	For those with CD in training cohort: Median follow up in IBD1 group = 4.9 years (interquartile range of 3.6 to 7.4 years) Median follow up in IBD2 group = 5.3 years (interquartile range of 4.3 to 8.3 years). For those with CD in validation cohort: Median follow up in IBDHi group = 1.6 years (interquartile range of 1.0 to 3.7 years) Median follow up in IBDLo group = 2.4 years (interquartile range of 1.8 to 3.8 years).			
Baseline patient characteristics	**PredictSURE-IBD**			
	Training cohort CD8T cell only (N=66)		**Validation cohort[a] (N=66)**	
	CD IBD1 (n=33)	**CD IBD2 (n=33)**	**CD IBDHi (n=27)**	**CD IBDLo (n=39)**

Age, years (interquartile range)	30.3 (25.3 to 36.1)	30.3 (23.2 to 38.7)	[redacted]	[redacted]
Sex (M), n (%)	14 (42.4%)	13 (39.4%)		
Ethnicity, n (%)	Not reported			
Newly diagnosed, n (%)	27 (81.8%)	24 (72.7%)	[redacted]	[redacted]
Disease duration (years)	0.0 (0.0 to 0.0)	0.0 (0.0 to 0.0)		
Disease location, n (%)				
• L1 (ileum)	9 (27.3%)	13 (39.4%)	[redacted]	[redacted]
• L2 (colonic)	11 (33.3%)	9 (27.3%)	[redacted]	[redacted]
• L3 (ileocolonic)	13 (39.4%)	11 (33.3%)	[redacted]	[redacted]
• L4 (upper digestive tract)	2 (6.1%)	3 (9.1%)	[redacted]	[redacted]
Predictors of disabling CD				
• Age <40 y	Not reported			
• Corticosteroid use	Not reported			
• Perianal lesions, n (%)	6 (18.2%)	3 (9.1%)	[redacted]	[redacted]
Smoking, n (%)				
• Current	10 (28.6%)	12 (33.3%)	[redacted]	[redacted]
• Former	Not reported			
• Never	Not reported			
Perianal examination				
• No lesion	Not reported		[redacted]	[redacted]
• Skin tags	Not reported			
• Fissure or ulcer	Not reported			
• Simple fistula	Not reported			
• Complex fistula	Not reported			
CDAI score	Not reported			
HBI score	Not reported			

IBDQ	Not reported			
Median haemoglobin concentration (g/L) (interquartile range)	12.5 (11.7 to 13.3)	13.1 (11.8 to 13.6)	[redacted]	[redacted]
Median C-reactive protein concentration (mg/L) (interquartile range)	26 (16 to 39)	25 (10 to 59)	[redacted]	[redacted]
Median albumin (g/L) (interquartile range)	35 (32 to 37)	37 (34 to 39)	[redacted]	[redacted]
Section 3: Outcomes				
Outcome	**Definition**			
Prognostic accuracy	Not reported			
Prognostic yield (number of diagnoses of severe versus non-severe course of Crohn's disease)	Not reported			
Time to test result	Not reported			
Number of test failures	Not reported			
Number of inconclusive test results	Not reported			
Percentage of people for whom early treatment with biologics was offered ('top-down') by subgroup of severe versus non-severe course of disease	Not reported			
Rates and duration of response and remission by subgroup of severe versus non-severe course of disease	Not reported			
Rates and duration of flare-ups and/or relapses by subgroup of severe versus non-severe course of disease	Not reported			
Rates and duration of corticosteroid-free remission by subgroup of severe versus non-severe course of disease	Not reported			
Cumulative corticosteroid exposure by subgroup of severe versus non-severe course of disease	Not reported			
Measures of mucosal healing by subgroup of severe versus non-severe course of disease	Not reported			

Rates of and time to treatment escalation by subgroup of severe versus non-severe course of disease	Maximum medical therapy during prospective follow-up split by immunomodulator, anti-TNF alpha, second-line biologic (ustekinumab or vedolizumab) and biologic therapy (excluding those who received biologic therapy due to immunomodulator intolerance)	
Rates of and time to hospitalisation by subgroup of severe versus non-severe course of disease	Not reported	
Rates of and time to surgical intervention by subgroup of severe versus non-severe course of disease	Number of people requiring any type of surgery	
Rates of and time to serious complication (e.g., obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported	
Composite outcomes formed of hospitalisation, surgery or serious complication (obstruction, intestinal ulcers, fistula, anal fissure) by subgroup of severe versus non-severe course of disease	Not reported	
Adverse effects of treatment	Not reported	
Health-related quality of life by subgroup of severe versus non-severe course of disease	Not reported	
Section 4: Data extraction form		
Outcome	**PredictSURE-IBD**	
Prognostic test accuracy for validation cohort		
	n	N
Sensitivity for predicting the need for multiple escalations within the first 18 months	66	72.7%
Specificity for predicting the need for multiple escalations within the first 18 months	66	73.2%

Negative predictive value for predicting multiple escalations within the first 18 months	66	90.9%
Clinical outcomes		
Dichotomous outcomes for training cohort		
	n	N
Risk of requiring immunomodulator as maximum medical therapy during prospective follow-up for IBD1 versus IBD2	66	12/33 (36.4%) with IBD1 versus 11/33 (33.3%) with IBD2 RR 1.09 (95% CI not reported)
Risk of requiring anti-TNF alpha as maximum medical therapy during prospective follow-up for IBD1 versus IBD2b	66	11/33 (33.3%) with IBD1 versus 6/33 (18.2%) with IBD2 RR 1.83 (95% CI not reported)
Risk of requiring second line biologic (vedolizumab or ustekinumab) as maximum medical therapy during prospective follow-up for IBD1 versus IBD2	66	2/33 (6.1%) with IBD1 versus 1/33 (3.0%) with IBD2 RR 2.0 (95% CI not reported)
Risk of requiring escalation to biologic therapy (excluding those who received biologic therapy due to immunomodulator intolerance) as maximum medical therapy during prospective follow-up for IBD1 versus IBD2	66	12/33 (36.4%) with IBD1 versus 4/33 (12.1%) with IBD2 RR 3.0 (95% CI not reported)
Risk of surgery (all procedures) c for IBD1 versus IBD2	66	10/33 (30.3%) with IBD1 versus 7/33 (21.2%) with IBD2 RR 1.43 (95% CI not reported) Difference reported to be non-significant. All people who required a panproctocolectomy were in the IBD1 subgroup
Risk of surgery (panproctocolectomy) for IBD1 versus IBD2	66	3/33 (9.1.3%) with IBD1 versus 0/33 (0.0%) with IBD2 RR not reported (95% CI not reported)
Continuous outcomes		

	Time-frame	Mean	SD/SE	N

Time to event outcomes				
	Time-frame	HR	95% CI	p value
Time to first treatment escalation for IBDHi versus IBDLo (validation cohort [N=66])		2.65	1.32 to 5.34	0.006
Time to first treatment escalation for IBD1 versus IBD2 (Training cohort [N=66])		NR	NR	0.016

Section 5: Additional comments	
Additional comments	Clinicians were blinded to the biomarker results at diagnosis, and to gene expression analyses. It is reported that most people (86/118) were recruited at the time of diagnosis. There were 67 treatment escalations in total across the CD validation cohort (N = 66). Neither clinical parameters (any two of: steroid requirement, age <40 years and perianal disease) nor severe endoscopic features (including deep and extensive ulceration in at least one colonic segment) were able to predict the need for early treatment escalation. Additional outcome reported of risk of not requiring any medical therapy 8/33 (24.2%) with IBD1 versus 15/33 (45.5%) with IBD2 RR 0.53 (95% CI not reported) All patients were assessed and treated at the discretion of their gastroenterologists in accordance with national and international guidelines, rather than following a formal protocol. As patients were recruited before induction therapy, it is not known how the biomarker would perform if treatment had already been started.
Further information that could be requested from authors	Please clarify: • For those with CD, the number of people who were newly diagnosed by IBD1 and IBD2 subgroup.

[a] Data provided by the company on request.

[b] One person categorised as IBD1 and three people classified as IBD2 received anti-TNFα due to intolerance to immunomodulators rather than refractory disease.

[c] Includes any single abdominal operation for treatment of Crohn's disease (e.g. ileocaecal resection or panproctocolectomy) or multiple operations for perianal disease.

Abbreviations: ASA, aminosalicylic acid; CD, Crohn's disease; CDAI, Crohn's disease Activity Index; CI, confidence interval; CRP, C reactive protein; HR, hazard ratio; IBD, inflammatory bowel disease; IBDQ, Inflammatory Bowel Disease

Questionnaire score; n, number of patients with the outcome; N, number of patients assessed; HRQoL, health-related quality of life; RCT, randomised controlled trial; SD, standard deviation; SE, standard error; TNF, tumour necrosis factor; UC, ulcerative colitis.

9.5.2 Economic evaluations

Population, intervention and comparator	Perspective, discounting, cost year and model structure	Clinical effectiveness	Resource and cost use sources	Utility sources and elicitation methods	Results
Marchetti 2013					
Population Newly diagnosed luminal moderate-to-severe CD **Intervention, top-down** 1st step infliximab plus AZA, 2nd step additional infliximab plus AZA, 3rd step methylprednisolone plus AZA **Comparator, step-up** 1st step methylprednisolone, 2nd step methylprednisolone plus AZA, 3rd step infliximab plus AZA.	**Perspective:** third party payer (the Italian healthcare system) **Discounting:** Costs and QALYs 3.5% per year **Cost year:** NR **DAM:** Probabilistic Markov model using 10,000 simulation **Health states:** 1st step (all patients enter the model in this health state), 2nd step, 3rd step and death. Patients enter the 2nd step and 3rd step health states because of	**Top-down** • The monthly transition probability from step 1 to step 2 was calculated according to the rate of patients requiring additional infliximab and to the relapse-free survival curve in D'Haens *et al.* 2008 (0.07 in the first year and 0.01 in subsequent years) • The monthly probability of incurring a mild flare needing additional infliximab for patients already in step 2 was calculated based on the rate of infliximab use during the 2nd year (15% at any 8-week interval equates to a monthly probability of 0.08). • After 24 months/cycles, patients were categorised according to the presence or absence of mucosal healing to estimate their clinical outcome in the following 3 years. The monthly probability of incurring a mild flare was obtained from Baert *et al.* 2008 (0.013 if mucosal healing and 0.027 otherwise) • The step 2 to step 3 transition probability was calculated according to the rate of patients receiving CSs (0.005 per month)	Follow-up resource form Jewell *et al.* 2005. Inpatient resource use obtained from D'Haens *et al*. 2008 and Marchal *et al.* 2004. Unit costs for services obtained from DRG hospital cost codes in Italy. Drug costs obtained from local hospitals in Italy	**Sources and elicitation methods** CDAI scores converted to utility scores using the published equation in Buxton *et al.* 2017 CDAI scores obtained from D'Haens *et al.* 2008 and Casellas *et al.* 2005 **Utility values** Top-down 1st step heath state: 1st month 0.52 2nd month 0.67 3rd month and thereafter 0.82	Discounted results **Total QALYs:** top-down, 3.9; step-up 3.7 **Total costs:** top-down €14,631; step-up, €15,404 **ICER:** top-down dominates step-up **Results of SA** Top-down continued to dominate step-up when the discount rate, cost of surgery, rate of surgery and rate of relapse were varied in OWSA. Top-down no longer dominated step-up when the cost of infliximab was increased to more than €666 (baseline €512) per

	clinical worsening occurring in the previous step. **Cycle length:** 1 month **Time horizon:** 5 years (60 cycles)	**Step-up** • The monthly transition probability from step 1 to step 2 was calculated according to the rate of use of thiopurines in D'Haens *et al.* 2008 (0.09 in the first year and 0.01 subsequent years) • The step 2 to step 3 transition probability was calculated according to the rate of use of infliximab (the cumulative rate from week 26 to week 72 was 75% and equates to a monthly probability of 0.12) • The rate of flares in step 3 was assumed to be the same as for step 2 of top-down. • After 24 months/cycles, patients were categorised according to the presence or absence of mucosal healing, as described for top-down **Surgery** • The monthly probability of undergoing surgery (bowel resection) was derived from D'Haens *et al.* 2008 (0.0038 in the top-down arm and 0.0052 in the step-up arm) • The probability of undergoing a severe flare requiring surgical treatment was assumed constant • Surgery-related mortality was derived from Silverstein *et al.* 1999 (0.0015 per month) **Other** • The probability of hospitalisation due to severe AEs in patients treated with infliximab was taken from Faraawi *et al.* 2008 (0.018 per infliximab infusion) • The probability of death was calculated from Italian life tables and no treatment-related mortality was assumed		Step-up 1st step health state: 1st month 0.52 2nd month 0.62 3rd month 0.72 4th month and thereafter 0.82 2nd step and 3rd step health states 0.82 **Disutility values** Assumed symptom worsening requiring step transition incurs a disutility of 0.2 for 3 months. Assumed surgery incurs a disutility of 0.5 for 1 month. Assumed clinical worsening incurs a disutility of 0.2 per month for 3 months.	100 mg vial or and when the time horizon was reduced to less than 4 years (baseline 5 years)

Freeman 2016					
Population Patients with moderate to severe active CD treated with infliximab or adalimumab. Two populations were considered: 1)Patients responding to treatment 2) Patients who had lost response to treatment. **Intervention** Monitoring of serum anti-TNF-α (IFX or ADA) and/or of anti-drug antibody levels using test assays with a test–treatment algorithm. Two test strategies were assessed: (1) concurrent and reflex	**Perspective** NHS **Discounting** Costs discounted at 3.5%. Benefits NR. **Cost year** 2013/14 **DAM** Two separate Markov models were built to reflect the two populations Three separate decision tree structures were made for the responder's model to reflect: 1) concurrent testing 2) no testing 3) reflexing testing. **Health states**: **Responder** Maintenance	**Transition Probabilities; sources** **Standard care** 1. IFX maintenance to LOR 0.008075 Juillerat et al 2015 2. IFX maintenance to LOR after IFX escalation. 0.017415 Ma et al. 2014 3. ADA after IFX, failure to LOR 0.058553 Sandborn et al 2007 and Karmiris et al 2009 **All** 4. Time to surgery 0.002591 Nguyen et al. 2011 5. Time to recurrent surgery 0.003122 Nguyen et al 2011 6. Time to post-surgical relapse on no therapy 0.049792 Gordon et al. 2014 7. Time to post-surgical relapse on immunosuppressant. 0.029714 Gordon et al 2014 8. Time to post-surgical relapse on anti-TNF-alpha 0.020784 Baert et al 2014 **Intervention arm: test algorithm strategy** 9. Time to post-surgical relapse on anti-TNF alpha and immunosuppressant As 8 above due to lack of data. 10. IFX maintenance to LOR (dose-escalation group) As 1 above Juillerat et al. 2015 11. IFX maintenance to LOR (dose-unchanged group) As 1 above Juillerat et al. 2015 12. IFX maintenance to LOR (dose-decreased	The resource use and costs were those directly incurred by the NHS. The cost of reagents for monitoring trough concentration of anti-TNFs and of antibody measuring kits, treatment for CD and laparoscopic ileocolic resection were all included in the analysis. Resource use and costs associated with occupying all health states (apart from dead) were also included. Resource use and costs came from the following sources: NICE, BNF, NHS Reference Costs and expert opinion.	**Utility values** Health state; utility; source Responder; 0.77; Velayos et al 2013 Loss of response; 0.62; Gregg et al., 1997 Regain response; 0.77; assumption Surgery; 0.60; Marchetti et al. 2014 Post-surgery; 0.86; Velayo et al. 2013 **Sources and elicitation methods** Utility values reported in Velayos et al. 2013 were obtained from the study undertaken by Gregor et al. 1997 who compared various elicitation techniques (SG, TTO and VAS) in 180 consecutive CD	***Testing strategy; Mean cost per strategy; Difference in costs; Total QALYs; Incremental*** ***QALYs; ICER*** **Base case** Reflex testing; 138,700; –; 6.2761; –; – Concurrent testing; 139,800; 1100; 6.2637; –0.0124; Dominated No testing; 150,500; 11,800; 6.5084; 0.2323; 50,800 **Annual testing in responder model** Concurrent testing 114,000; –; 6.2201; –; – Reflex testing 114,100; 100; 6.2281; 0.0080; 12,500 No testing 150,500; 36,400; 6.5084; 0.2803; 129,900 **Annual testing in LOR model** Concurrent testing 106,900;

testing of drugs and (2) antibodies to the drugs (i.e. simultaneous or sequential drug and antibody testing). Concurrent testing yields four possible outcomes: drug+/antibody–, drug+/antibody+, drug–/antibody+ or drug–/antibody–. Reflex testing (antibody testing if drug tests are negative) yields three outcomes: drug+, drug–/antibody– or drug–/antibody+. **Comparator**-Standard care (no testing/ therapeutic practice)	treatment when the patient has supportable active symptoms of abdominal pain, diarrhoea, rectal bleeding or weight loss. **Loss Of Response (LOR)** Recurrence of active symptoms while on treatment with maintenance regimen, after having responded to treatment. **LOR (no anti-TNF-α)** Recurrence of active symptoms having discontinued anti-TNF-α treatment with maintenance regimen, but receiving best supportive care. **Regain response** Maintenance	group) As 1 above Juillerat et al. 2015 13. Regained response on ADA to LOR (group 1, IFX negative/antibodies to IFX positive) As 3 above 14. Regained response on intensified IFX to LOR (group 2, IFX negative/antibodies to IFX negative) As 2 above 15. Regained response on un-prescribed treatment for LOR (group 3 or 4 IFX positive/antibodies to IFX positive or negative) to LOR 0.086173 Rutgeerts et al. 2005		patients. These authors suggested that the SG reflected the true value for health states related to patients with CD, and these values may be the most appropriate for an economic analysis.	–; 6.1406; –; – Reflex testing 108,100; 1200; 6.1532; 0.0126; 95,200 No testing 215,800; 107,700; 6.4961; 0.3429; 314,100 ***Other SA ICERs (gully incremental results)*** **1-year time horizon in responder model** No testing –; Concurrent testing Dominated; Reflex testing Dominated **1-year time horizon in LOR model** Concurrent testing -; Reflex testing £111,100; No testing £170,500 **One-off testing at 3 months followed by yearly retesting** Reflex testing –; Concurrent testing Dominated; No testing £132,800 **One-off testing at 3**

	treatment when the patient has no active symptoms having previously lost response. **Post-Surgery** Medication/no medication after inpatient surgical procedure **Dead**. **Cycle length** 4 weeks **Time horizon** 10 years with a hypothetical cohort of patients aged 30 years				**months and one retest for those who regained response** Concurrent testing -; Reflex testing £74,100; No testing £176,300 **One-off testing at 3 months and no retesting for responders/regained response** Concurrent testing -; Reflex testing £66,700; No testing £176,700 **In the LOR model: 3-monthly testing for patients with LOR; no testing for patients who have regained response** Concurrent testing -; Reflex testing £84,700; No testing £354,500 **No regain of response following best supportive care (responders)** Reflex testing –; Concurrent testing Dominated; No testing £86,600 **No regain of response following best supportive**

					care (LOR) Concurrent testing -; Reflex testing £107,900; No testing £158,500
Saito 2013					
Population Moderate to severe CD refractory to conventional therapies and naive to biologic therapy **Intervention** Infliximab induction and maintenance infusions plus azathioprine **Comparator** Infliximab monotherapy	**Perspective** NHS **Discounting** NR **Cost year** 2008 **DAM** Decision tree **Health states** N/A **Cycle length** N/A **Time Horizon** 1 year	The probability of clinical efficacy, probability of therapy discontinuation owing to AEs or lymphoema risk, were derived from published data. **Treatment efficacy, Probability, Source** **Clinical response rate at week 12** IFX monotherapy, 0.735, Miheller et al Combination therapy with infliximab and AZA ,0.882, Miheller et al **Maintenance remission rate** IFX monotherapy sustained remission at 1 year, 0.309, Hanauer et al 2002, Colombel et al 2010 IFX monotherapy and sustained response at 1 year, 0.487, Hanauer et al 2002, Colombel et al 2010 IFX monotherapy and loss of response, 0.513, Hanauer et al 2002, Colombel et al 2010 **Combination therapy with IFX plus AZA** Sustained remission at one year, 0.446, Colombel et al 2010, Lemann et al 2006	The drug costs of infliximab and AZA were extracted from UK sources (Buchanan et al 2011) Annual care costs were obtained from Sprakes et al 2010 who assessed the care costs of CD patients for the 12 months before and after infliximab therapy by looking at NHS reference costs. These annual costs included inpatient admissions, day case admissions for infliximab infusions, outpatient visits, surgical procedures,	**Sources and elicitation methods** Gregor et al 1997 used a SG approach to define a utility score with the CDAI. With non-responding active disease or lymphoma complicated by CD a utility of 0.4 was given to the non-responding active state based on a consultation with a panel of UK gastroenterologists reported by Lindsay et al, 2008 and assumed that the lymphoma state decreased utility scores by 0.15 following Lewis et al	**Total Costs:** IFX monotherapy £6979.68 Combination + AZA £8573.04 **Total QALYS**: IFX monotherapy 0.064 Combination plus AZA 0.668 **ICER** £24,917 **Results of SA** The OWSA demonstrated that ICERS remain in the £17,147-£45,564 /QALY range and that QoL utilities for nonresponding active disease had the highest impact on the ICER (£45,564 /QALY over infliximab monotherapy).

		Sustained response at one year, 0.705, Colombel et al 2010, Lemann et al. 2006 Loss of response, 0.295, Colombel et al, 2010 Lemann et al. 2006 **Adverse effect associated with IFX** Discontinue infliximab due to serious adverse effect, 0.111, Hamzaoglu et al 2010. Death due to serious adverse effect of infliximab, 0.004, Siegel et al 2012 Discontinue AZA due to adverse effect, 0.089, Pearson el al 1995 Death from lymphoma in first year, 0.297, Lewis et al 2000 **Non-biologic therapy** Remission, 0.068, Odes et al 2010 Post-surgery remission, 0.15, Odes et al 2010 Improvement to mild level of disease, 0.201, Odes et al 2010 Remain drug refractory, 0.711, Odes et al 2010 Death related to CD, 0.005, Odes et al. 2010	endoscopic procedures, radiological investigations blood tests and the cost of all prescribed medications. No evidence was found for the cost associated with lymphoma complicated by CD. Annual cost of £4908.43 was assumed based on study of illness costs in Germany. (Reis A et al 2006). Due to the perspective productivity costs were omitted.	2000. **Utility values** Post-surgery, 0.86 Mild disease, 0.77 Non-responding active disease, 0.4 Lymphoma, 0.25 Death, 0	The impact of changes in lymphoma risk was less significant, with ICERS for the lower and upper risk rates being £24,849 and £25,026 /QALY respectively. With a combination therapy of infliximab and AZA if a cost of £30,000/QALY is deemed acceptable then the probability of combination therapy with infliximab and AZA being less than or equal to this value is 0.750. Results of two-way SA that evaluated thresholds for simultaneous change in lymphoma risk and maintenance remission rate: At a relative risk of lymphoma greater than 129 the combination therapy dominated the infliximab monotherapy when the maintenance remission rate

					at 1 year was 44.6%.
Bodger 2009					
Population Moderate to severe active CD. **Intervention** Addition to standard care of infliximab 5mg/kg intravenous infusions at weeks 0,2,6 for the induction of remission; then 8 weeks for maintenance of clinical remission. Or adalimumab 80mg subcutaneously at week 0, 40mg at week 2 for induction of remission, then 40mg on alternate weeks for maintenance of	**Perspective** NHS **Discounting** costs and QALYs 3.5% per year **Cost year** 2006/7 **DAM** Decision tree followed by Markov model. In the decision tree patients were on infliximab for 10 weeks and adalimumab for 12 weeks, and then entered the Markov cycle. **Health states** A reduced 5 state model was derived from the original 8	Response was assessed at 2 weeks (Infliximab) or 4 weeks (adalimumab). Non responders enter the Markov model for standard care in the non-response state at weeks 10 (infliximab) or week 12 (adalimumab). in the Silverstein et al model, transition probabilities were adjusted proportionally to reflect the changing probabilities of death, but specific details not provided in the original publication. Jess et al reported the survival and cause specific mortality for patients with CD in the same patient registry. A Cox proportional hazard regression model which described age to be independently associated with mortality (HR 1.6 per 5-year increment). The transition probabilities in the reduced model were adjusted accordingly for mortality. Simulations of the reduced Markov model revealed some differences from the original data and corrections were applied to the probabilities associated with the full response and partial response states to result in a reduced matrix that reproduced the data reported by Silverstein et al. Analysis were conducted in Tree Age-Pro.	For the direct costs relating to the hospital care of 160 patients these were based on healthcare resource use associated with in-patient and out-patient services, investigations, medications and surgery. None of the patients received biologic therapy. Costs were classified according the states defined by Silverstein et al. After applying the 5.6% annualized rate of inflation costs were recalculated to match	**Sources and Elicitation Methods** For the health states of full, partial, and non-responsive to treatment, the estimated EQ5D utility score was calculated from the midpoint CDAI scores based on the algorithm developed by Buxton et al, where EQ5D= 0.9168-0.0012*CDAI. This is based on data from 905 patients with moderate to severe CD included in the ENACT 1 trial and 2 trials of natalizumab.	**Mean costs; Mean QALYs; ICER vs. standard care** Standard care £43 490; 14.209; – Infliximab – 1 year £50 330; 14.568; £19 050 Infliximab – 2 years £58 230; 14.901; £21 300 Adalimumab – 1 year £46 730; 14.682; £7190 Adalimumab – 2 years £53 090; 15.156; £10 310 **Results of SA** Impact on the results were made by varying key parameters such as duration of treatment with infliximab or adalimumab, from 1 year to lifetime,

remission. **Comparator** Standard care- assumed to include medical management (5 aminosalicylic acid derivatives, immunosuppressive agents, CSs, antibiotics, topical therapy and surgery)	state model described by Silverstein et al. The states in the reduced model were defined according to CDAI scores to match data from clinical trials. **1st State**- Original health state of medical remission and post-surgical remission were combined to define a new full response state where CDAI scores less than 150. **2nd State**- The original mild and severe disease states were combined to form a partial responsive state between (CDAI between 150 to 220). **3rd Health State**- The original severe	Two trials were used for the analysis: ACCENT I for infliximab and CHARM for adalimumab. Efficacy data were not reported for patients who failed to respond to initial doses of therapy as defined in the ACCENT I trial for infliximab and CHARM study for adalimumab. Transition probabilities not reported however possibility of transition probability matrices mentioned in supporting online document. Non responders to infliximab and adalimumab were assumed in the model to experience the same prognosis as those receiving standard care and entered the reduced (Silverstein) reduced transition probability matrix representing standard care from weeks 10 and 12 respectively. To achieve close replication of the results in the ACCENT I and CHARM, the transition probabilities of remaining in the states of full, partial, and non-response were multiplied by fixed values, determined from simulations and manual adjustments. Since CD is a lifelong condition that affects patient survival, the base case was a life time horizon that was adopted to reflect differences between treatment strategies. In the base case analysis treatment with	the four alive health states; and with the exception of surgery divided by three to calculate 8-week costs. For surgery it was assumed that all costs would be incurred in the full response state. Each health state was assigned a mean cost. Unit costs of the biologic agents were taken from the BNF Costs of infliximab were based on patients' body weights. An administration cost of £168 per infusion (day case hospital attendance) was included and the costs associated with wastage were	Data from patients at multiple time points in the two trials were pooled. A moderate but statistically significant relationship was observed between CDAI and EQ-5D with 29% of variability between in EQ-5D scores explained by CDAI. For each cycle in the surgical state, patients were assumed to experience 2 weeks at an equivalent state of health as non-responders and 6 weeks at an equivalent state of health as full responders. **Mean 8-week utility score for each health state:** Full response 0.128 Partial response 0.106 Non response 0.065 Surgery 0.112	relative risk of surgery with the biologics compared with standard care from 0.1 to 0.9, discount rate for costs and QALYS, from 0% to 6%, and the time horizon of analysis to coincide with duration of treatment (from 1 year to lifetime).

	disease states (drug dependent and drug refractory) were combined to form a non-responsive state (CDAI score between 220 and 600). **4th and 5th State**- Surgical and Death state. **Cycle length**: Markov, 8-weekly **Time Horizon-** 60 years (lifetime horizon with up to 4 years continuous therapy)	biologic agents, for the maintenance of clinical remission was modelled to continue for 1-2 years. The extrapolation from 1-year trial data to 2 years was based on assumption of continuation of treatment effect according to transition probability matrices developed for year 1. After period of treatment ex-responders were assumed to revert to standard care and follow the reduced Silverstein et al model. Assumed treatment suspended during periods of surgery.	calculated based on the vials would not be shared between patients.		
Loftus 2009					
Population Severe active CD and moderate to severe active CD.	**Perspective** NHS **Discounting** 3.5% for costs and QALYS. **Cost year** 2006	CD-related hospitalisations were collected from two RCTs: CHARM and CLASSIC 1. A Poisson regression model was used to estimate the number of hospitalisations per patients over the 56-week timeframe:	UK costs taken from Bassi et al. 2004 who quantified direct medical costs by reviewing patient abstracts for 172	**Sources and elicitation methods** Gregor et al measured HRQoL in 180 Canadian patients with	***Severe CD*** **Total Costs** Adalimumab £10,882 Conventional non-biologic

Intervention Adalimumab induction and maintenance therapy **Comparator** Conventional non-biological therapeutics including azathioprine, 6-mercaptopurine, methotrexate, 5-aminosalicylic acid, sulfasalazine, mesalamine and/or CSs.	**DAM** No DAM. Rate of hospitalisations were estimated from a regression analysis. Costs and benefits were attached to those rates of hospitalisation to estimate cost-effectiveness. **Health States** Efficacy was captured by mapping patients' CD Activity Index over time into four disease states: remission (<150), moderate (150 to 300), severe (300 to 450) and very severe (450 plus). **Cycle length** N/A **Time horizon** 1	Adalimumab treated patients- first year 0.230 Adalimumab treated patients- after week 56 0.264 Non-biologic treated patients- first year 0.708 Non-biologic treated patients- after week 56 0.661	patients with CD over a six-month period in 2000-2001. Other direct medical cost estimates were derived from Bassi et al 2004 who presented coefficients from a general linear regression model with a log-link function estimating the relationship between CD-related costs, disease severity, and other factors.	CD using the SG. **Utility Values** Remission 0.859 Moderate 0.795 Severe 0.693 Very severe 0.433	£8,992 **Total QALYS**: Adalimumab 0.8516 Conventional non-biologic 0.7339 **ICER:** £16,064 ***Moderate to severe CD*** **Total Costs:** Adalimumab £9,696 Conventional non-biologic £6,649 **Total QALYS**: Adalimumab 0.8647 Conventional non-biologic 0.7743 **ICER:** £33,731 **Results of SA** The results of the univariate SA indicated that the model results for patients with moderate to severe CD were more sensitive to the one-way sensitivity tests than the

	year				results for severe only patients. In the lifetime model adalimumab patients were expected to continue on the drug for a median time of 7.6 years. This resulted in ICERs of £6550 per patient with severe CD and £17,873 per patient with moderate to severe CD. *Moderate to severe CD; severe CD* LOCF for patients who dropped out, non-responsive or missing values: £57,571; £34,230 Utility values -/+ 10%: £37,479 /£30,664; £14,603/ £17,848 Hospitalisation costs +/- 10%: £19,430 /£48,032; adalimumab dominant /£33,061 CDAI disease-based costs: £42,570; £24,575

					Adalimumab induction regimen 160mg/80mg instead of 80mg/40mg: £45,604; £25,177
Lindsay 2008					
Population Adults with CD with and without fistulae. This includes patients who have active luminal CD. I**ntervention** Infliximab initial infusions and maintenance treatment **Comparator** Standard care, comprising immunomodulators and/or CSs.	**Perspective** NHS **Discounting** Costs and benefits 3.5% **Cost year** 2005/6 **DAM** One markov model was constructed to stimulate the progression of adult CD patients with and without fistulae during treatment with infliximab (5mg/kg). A second Markov model was produced for active luminal CD.	Results from the published, randomized placebo-controlled induction studies were used to estimate the initial response to infliximab, whereas infliximab efficacy was in maintaining this remission was obtained from the ACCENT I and ACCENT 2 maintenance trials. All patients in ACCENT I received infliximab at week 0 and all patients in ACCENT II received the full induction dose of infliximab (week 0, 2 and 6. Transition probabilities for subsequent model studies were derived from patients randomized as responders at week 2 in the ACCENT I trial. Transition probabilities for the induction phase (week 0–14) in the fistulizing CD analysis were estimated using the data from the pre-sent study,27 and subsequent transitions were estimated using patient level data from responders to the induction regimen in ACCENT II.19 The transition probabilities for the first 54 weeks were derived from the patient level data in ACCENT trials as explained above. The transitions observed in the last assessment cycle in ACCENT trials (weeks 46–54) were then used to extrapolate the analysis up to 5 years.	The total cost associated with infliximab treatment was broken down into acquisition costs (£419.73 per 100 mg vial) and administration costs (£96.00 per infusion). The estimated cost of surgical interventions used was £5277 on the basis of data published in the NHSRC. This was an average cost taking into account elective and non-elective admissions with or without complications or comorbidities. The cost of hospitalisation and other	**Sources and elicitation methods** Casellas et al. 2005 estimated health state preferences of Spanish CD patients used the EQ-5D. The EQ-5D responses of these patients were then converted into utilities using UK tariffs. **Utility Values** Remission 0.83 Active 0.55 Surgery 0.73 Post-surgery remission 0.67 Post-surgery complications 0.50 Remission + fistula	***Severe active luminal CD*** **Total Costs** Infliximab £31499 Standard care £26627 **Total QALYS** Infliximab 2.145 Standard care 1.959 **ICER** £26128 ***Fistulizing CD*** **Total Costs:** Infliximab £37,488 Standard care £31,490 **ICER** £29,752 **Results of SA**

	Health states: **Active Luminal CD** For patients receiving treatment in the model the disease severity was characterised by two discrete on treatment health states: “remission” (CDAI <=150) and treatment response but not achieving remission CDAI>150) often referred as active health state or moved to a different health state. All patients started in the active health state and remained in this health state for the first model cycle. At the end of the first and each subsequent model cycle, patients either remained in the active health	Transition probabilities for surgery and post-surgical states were obtained from the following published literature. (Jess T et al 2006, D’Haens et al 2004, Wolters FL et al 2006, Makowiec F et al 1995. The probability of a post-surgery complication was based on a cohort control study which compared the post-surgical complications between groups of patients treated with or without infliximab before surgery.	assessments was adapted from the Jewell study by adjusting the reported resource use with published estimates from NHSRC. The Jewell study did not estimate the cost associated with post-surgery health states. Cost associated with post-surgery remission was assumed to be equivalent to medical remission and the cost of post-surgery complications £1460.40/cycle was calculated based on the resource use by a hypothetical patient in these health states as estimated by UK gastroenterologists.	0.68 Active and fistula closure 0.55 Active and fistula 0.4	Results remained in the range of £23752 to £38848 at 5 years for active luminal CD and £27047 to £44206 at 5 years for fistulizing CD. Due to the weight-based dosing of infliximab, patient weight had the most impact on the ICER with the ICER increasing to £38848 in luminal CD and £44206 in fistulizing CD for an 80kg patient. The change in health state preferences had the most significant impact on improving the ICERs with a 10 percent increase in utilities resulting in an ICER of £23752 in luminal CD and £27047 in fistulizing CD. Results from Probability SA shows that health state preferences had a significant impact on the ICERS. A proportion of PSA simulations resulted in

	state, or moved to a different health state. Patients responding to treatment and achieving a CDAI <= 150 moved to the remission and remained on treatment. Patients demonstrating a clinical response but not achieving remission remained in the active state and continued to receive treatment. Patients classified as non-responders or patients discontinuing treatment stopped treatment and moved to the nonresponding active state. Once patients had failed treatment and				negative incremental QALY gains for patients treated with infliximab. However clinical practice and trial data suggests QoL gain with the use of infliximab.

	moved to this 'off-treatment' state, they could not restart infliximab treatment and return to the 'on-treatment' states. However, these non-responders and treatment failures were followed throughout the model to capture their costs and outcomes. Patients in both the 'on-treatment' and 'off-treatment' health states could transition to surgery. In the subsequent model cycle, patients undergoing surgery could either undergo repeat surgery due to immediate post-surgery complications and				

	remain in the surgery state or move to a posts-surgery health state (post-surgery remission or post-surgery complications). Patients in post-surgery remission could continue in the same health state, enter surgery to undergo repeat surgery, suffer from a complication to enter post-surgery com- plications or have recurrence of CD to enter the nonresponding active state. In the absence of any evidence to estimate the effectiveness of re-treatment with infliximab for infliximab failures, it				

	was assumed that patients with a recurrence of their CD would not be offered re-treatment with infliximab. Similarly, patients experiencing post-surgery complications could continue in the same health state, enter surgery to undergo repeat surgery, respond to the treatment for their complications and enter post-surgery remission, or have recurrence of CD and enter the nonresponding active state. **Fistulising CD** Similar model to that used in the active luminal CD patients except that the 'on-treatment' health states of				

	remission and active were further classified as with or without fistulae. This resulted in four 'on-treatment' health states. The remaining health states and the patient progression within these health states were identical to the active luminal CD model. **Cycle length**- Luminal active CD- Weeks 0–2, 2–6, 6–10, 10–14 and every 8 weeks thereafter. Fistulizing active CD- Weeks 0–14, 14–30 and every 24 weeks thereafter. **Time horizon** 5 years				
Clark 2003					

Population Adults with either severe active CD or fistulising CD resistant to conventional treatment. **Intervention** Infliximab infusions as single, episodic and maintenance **Comparator** Placebo or other treatment for CD.	**Perspective** NHS **Cost year** NR ***Company's model in chronic Active CD*** **Discounting** Costs and benefits were discounted at 6% and 1.5% respectively. **DAM** A Markov Model with seven states was developed using the software package Decision Maker. **Health States** 1. Drug dependent severe 2. Drug refractory disease state 3. Drug Responsive. 4. Medical Remission 5. Mild disease	***Company's model in chronic Active CD*** Response rates (not reported) from the clinical trials, combined with transitional probabilities for seven different disease states extracted from Silverstein's 24-year follow-up of a population-based 'inception cohort' of 174 patients with CD in Olmstead County, USA. Silverstein provided data on the progress of patients from remission through mild and more severe disease states. To make the results relevant to the UK, UK life tables were applied to the USA data. Efficacy data were based on two published trials (names not mentioned). ***Fistulising CD model*** Efficacy data obtained from Present et al 1999. Variations related to providing re-treatment doses to those whose fistulae re- open after 14 weeks, with various assumptions on success and closure rates (varied between 70-90%).	The only available relevant UK data were from an unpublished study of a small group of 38 UK patients for whom an average cost over 12 months was estimated. This average cost was distributed across the seven health states using relative cost data from the Olmstead County study.	Both models mapped disease-specific scores on to utility scores, relying mainly on published and partly on in-house data. ***Company's model in chronic Active CD*** **Sources and Elicitation methods** Gregor et al 1997 collected CDAI, IBDQ and utility data from a sample of 180 CD patients in a single Canadian tertiary centre during 1995–96. All three methods of elicitation in Gregor et al. 1997 were considered by the company in SA: SG, TTO and VAS. The authors noted the differences between the utility scales and suggested that SG should be used owing	***Treatment of chronic active CD*** **ICER:** Single dose treatment £6,700 Episodic retreatment £10,400 Maintenance treatment £84,400 **Results of SA** **Differences from the company model** The ICER for episodic dosage fell from £72,000 to just under £50,000 with strong assumptions on either duration of benefit or a higher utility gain per patient. *ICER for scenario: single dose; episodic* Scenario 1 (all doses): 135,333; 72,261 Duration 120 days for 80: –; 48,174 Utility 0.20 for 0.13: –;

	6. Surgical Remission 7. Surgery. **Cycle length** 2 months **Time Horizon** lifetime (40 years) Two major assumptions were made: (a) that patients who achieved remission or mild health states due to infliximab then moved through seven health states to death as though they had been naturally in remission; (b) that the patient utility gains were aggregated over their lifetimes or about 40 years. No data were available on duration of clinical			to a higher percentage of patients responding to it compared with TTO. **Utility values** State 1 Drug-dependent severe-0.86 disease State 2 Drug-refractory disease state- 0.74 State 3 Drug responsive -0.77 State 4 Medical remission- 0.88 State 5 Mild disease 0.86 State 6 Surgical remission- 0.88 State 7 Surgery -0.60 ***Fistulising CD*** Used a combination of two disease-specific scores (CDAI and	46,969 50% surgery averted: –; 60,636 Scenario 2 (5 mg/kg): 93,244; 62,016 Duration 120 days for 80; –; 41,344 Utility 0.20 for 0.13: –; 40,310 50% surgery averted: –; 50,090 The company model considered patients over their lifetimes. The authors approach limited the timeframe to one in which three re-treatments could occur, which could be 1 or more years. The company model had seven health states (including surgery, surgery remission and drug responsive and drug-dependent severe disease), each with different utilities. Patients who achieved remission or mild health states due to infliximab were assumed to spend time in each of these disease states

	response (i.e. mild disease) but, for modelling purposes, this was assumed to be 80 days, equal to that for those patients achieving remission. ***Fistulising CD*** **Discounting** NA **DAM** No DAM. Translated efficacy data from the pivotal trial for fistulising Crohn's disease into time spent with closed fistulae in the first 12 months after treatment, attaching costs and benefits **Cycle length** NA **Time Horizon** 12 months			PDAI), using a formula based on unpublished work to derive utility values	accumulating QALYs. In contrast, the authors estimates assumed that patients reverted back to their original drug-refractory states. The non-fistulising model was highly sensitive to rate of 'flare' for episodic treatment. The flare rate chosen was 10%, which seemed reasonable based on clinical opinion. If more frequent flare was seen, then costs increased substantially: the ICER was £55,000 with a 50% likelihood of flare. At 1 year the cost/QALY was high at between £35,000 (for a single treatment) and £59,000 (retreatment for those relapsing from either remission or mild disease states). Over 5 years the values reduced to £16,000 (single dose treatment) and £32,000 (episodic treatment.

					Fistulising CD ICERs ranged from £102,000 to £123,000 for Initial treatment and from £82,000 to £96,000 for retreatment with the most favourable re-treatment assumptions on closure rates. The results were relatively insensitive to the costs offset (due to surgery averted), even when 100% offset was assumed.
Dretzke 2011 **NICE TA187 Infliximab and adalimumab for the treatment of Crohn's disease**					
Dretzke provided a critique of the submission on infliximab by Schering-Plough and a critique of the submission on adalimumab by Abbott, but only the independent analysis by Dretzke is extracted. The three models in the Schering-Plough submission were: a version considering cost effectiveness of IMT compared with ICD and SC without infliximab among patients with severe active CD (CDAI scores 220–400) in England and Wales (MODEL A); a second version comparing IMT against SC in fistulising CD (MODEL B); and a third model considering paediatric CD patients (MODEL C). The two models in the Abbott submission were: one comparing the cost-effectiveness of adalimumab as a maintenance therapy against SC and one comparing the cost-effectiveness of adalimumab and infliximab as maintenance therapies.					
Population Patients with moderate-to-severe CD that is refractory to	**Perspective** NHS **Discounting** N/A	In the absence of time-dependent transition probabilities for the natural history of the cohort with moderate-to-severe CD at onset, a natural history cohort that reflects the average transition rate over	Infliximab administration costs were taken from a previous HTA report.	**Sources and elicitation methods** TTO values obtained	**Costs (£); QALYs; ICER for interventions vs SC; ICERs**

standard treatment. **Interventions and comparators:** 1. Induction therapy (IND) with adalimumab or infliximab 2. Maintenance (MNT) therapy with adalimumab or infliximab 3. Standard care (SC): conventional treatment without TNF-α inhibitors including no treatment, placebo, dietary intervention, drug treatment with aminosalicylates, methotrexate, corticosteroids (prednisolone, budesonide and	**Cost year** 2006 **DAM** Markov model (moderate and severe disease considered separately) **Health States** The models for induction and maintenance treatments were conceptually similar: each involved one set of states for anti-TNF treatment and another for SC treatments (where necessary). For IND there were three treatment-related states – remission and surgery-related (IND remission, IND surgery, IND post-surgery remission). There were also two relapse states, one for each 4-week	time was used. The methods to derive those transition probabilities and the resulting values are reported in detail in the paper. **Standard care transition probabilities**: From Silverstein's 1999 transition matrix transition probabilities have been provided between the following states: remission, mild, drug-responsive, drug-dependent, drug-refractory, surgery, post-surgery, death. Various steps were then made to produce a four-step transition matrix: **Step 1**- Death was removed. This was because death from all states were equally likely. Chance of death was removed from each state and divided by six. These values were then added to the six non-mild, non-death states. **Step 2-** The mild state was then removed using a three-step process **Step 3**- The states remission, drug responsive, and drug dependent were combined to a single remission state. **Step 4**- A matrix was them produced in four states (remission, relapse, surgery and post-surgery remission for 2 monthly cycles. **Step 5**- This was then modified to produce a 4-week transition matrix.	No administration costs were included for adalimumab. Drug acquisition cost source NR. Direct NHS costs were modelled as the sum of anti-TNF costs and type-specific health-state costs; no costs related to PSS were identified as part of the modelling process. Where possible, these health-state costs were taken from the NHSRC. Surgery costs were modelled as the cost of inpatient IBD interventions, while moderate and severe relapse costs were modelled as the cost of IBD outpatient major and intermediate interventions. Post-surgery remission costs were based on outpatient surgical gastrointestinal follow-up. Relapse costs were	from Shoor 2006. It was assumed that the average utility for individuals in the major surgery state would be equivalent to EQ-5D state 22222 with utility weight of 0.516. **Utility values per cycle** Calculated by taking the annual utility value and dividing by the number of cycles (13) run by the model: All remission states 0.073 Relapse states (moderate disease) 0.068 Relapse states (severe disease) 0.056 Surgery states 0.039	**Severe disease** SC 13,415; 0.8119; –; Dominated Infliximab IND 12,051; 0.8943; Dominates; Baseline Infliximab MNT 19,143; 0.8957; £68,315 per QALY; £5.03M per QALY SC 13,421; 0.8118; –; Dominated Adalimumab IND 7053; 0.8942; Dominates; Baseline Adalimumab MNT 14,047; 0.8956; £7749 per QALY; £4.98M per QALY **Moderate disease** SC 6615; 0.8926; –; Baseline Infliximab IND 9573; 0.9240; £94,321 per QALY; £94,321 per QALY Infliximab MNT 16,751; 0.9245; £317,991 per QALY; £13.9M per QALY

hydrocortisone), azathioprine, metronidazole or surgical intervention.	period in which response was assessed (IND relapse and IND relapse 2). Those failing to respond to treatment after 8 weeks of continued relapse (i.e. passing through both IND relapse and then IND Relapse 2) transited to the SC states (SC remission, SC relapse, SC surgery, SC post-surgery remission). In order to correctly assign anti-TNF costs, this occurred through a temporary transition state which was identical to SC relapse in its transition probabilities and utility and which only differed in its costs. Maintenance treatment used the same structure as induction. Both the induction and	**Induction transition probabilities**: The main difference between the transition matrices for standard care and induction was the probability of transiting from relapse to remission. **Maintenance transition probabilities**: The incremental benefit of maintenance treatment over induction treatment was expected to be a reduction in the probability of relapse from remission states. The treatments were assumed to have an equivalent effect on the probability of remaining in remission or relapse for both moderate and severe CD.	based on a gastrointestinal admission to hospital. Remission costs were modelled using Bassi et al. 2004 and the Bassi et al. cost for quiescent CD was used as the cost for the remission state.		SC 6615; 0.8922; –; Dominated Adalimumab IND 4583; 0.9231; Dominates; Baseline Adalimumab MNT 11,657; 0.9236; £160,079 per QALY; £13.9M per QALY **SA** The SC transition matrix was modified to allow the following: incorporation of different relapse rates from remission states; consideration of the effect of 'implausible' transitions in the SC matrix; and, provision of alternative transitions from the surgical states. Other analyses included extending the time horizon of analysis up to 20 years and modifying the effectiveness estimates used in the base case.

	maintenance model started with a cohort of patients suffering from an SC-refractory relapse. **Cycle length** 4 weeks **Time horizon** 1 year (13 cycles)				
Hodgson 2018 **NICE TA456: Ustekinumab for moderately to severely active Crohn's disease after previous treatment**					
Population Two sub-populations (conventional care failure and anti-TNF alfa failure patients) of adults with moderate to severe CD. **Interventions and comparators** In the anti-TNF failure population, ustekinumab was compared with	**Perspective** NHS **Discounting** Costs and benefits discounted at 3.5% per year. **Cost year** 2014/15 **DAM** The model structure adopted by the company was based on a previous economic model developed by Bodger et al.2009 The de novo analysis	Clinical effectiveness inputs used in the induction phase of the model were based on UNITI-1, UNITI-2 and the NMA, which included data from induction trials for all relevant comparators. Clinical effectiveness data used in the maintenance phase of the model was based on the IMUNITI trial and a treatment sequence analysis which included data from the induction and maintenance trials for all comparators of interest. No evaluation of the relative efficacies of the biologics beyond 1 year was possible due to lack of data. The company used a calibration technique to estimate the transition probabilities of patients in the maintenance phase of the model. An Excel solver function was then used to estimate transition	Drug acquisition costs were sourced from the BNF and MIMS. Drug administration costs were also included for treatments delivered by intravenous drips (infliximab, vedolizumab and ustekinumab induction treatment). No drug administration costs were included for those delivered by subcutaneous injection (adalimumab and	**Sources and elicitation methods** EQ-5D utility values mapped from IBD data collected during the UNITI-1 and UNITI-2 pivotal trials. Decrements due to AEs were based on the sources used in TA352 **HSUVs (taken from the CS)**	**ERG's results** In the ERG's base case analysis, which assumed a maximum duration of 1 year, the ICER for ustekinumab compared with conventional care was £109,279 in the conventional care failure subpopulation and £110,967 in the anti-TNF alfa failure subpopulation. Exploring the impact of alternative assumptions regarding maximum duration of therapy, the ICER for ustekinumab compared with

conventional care and vedolizumab. In the conventional care failure population, ustekinumab was compared with conventional care and adalimumab.	presented by the company consists of two parts: a decision tree to represent the short-term induction phase, and a long term maintenance phase represented by the Markov model. **Health states in the Markov model:** 1. moderate to severe (CDAI >220), 2. mild (CDAI 150–220) 3. remission (CDAI <150) 4. surgery 5. death Patients who had achieved a 100-point improvement in CDAI during anti-TNF induction were classed as responders and continued to receive maintenance therapy for a maximum of 1	probabilities that fitted the clinical data available. The methods and values are reported in detail in the company's submission. *The ERG had concerns with the company's long-term effectiveness of treatment including the assumption that non-responders remain in the moderate to severe health state for the entire maintenance period and the reliability of the calculation of transition probabilities.*	infliximab maintenance treatment). The health state costs associated with CD included in the model were estimated based on an elicitation exercise involving 12 clinicians. This included costs associated with outpatient visits, radiological tests, endoscopies and hospitalisations. *The ERG was concerned that the health state costs were too high.* AE costs were included for five AEs: serious infection (defined as septicaemia, pneumonia, urinary tract infections, respiratory infections and bronchitis), tuberculosis, hypersensitivity, injection site reactions	Remission 0.82 Mild 0.70 Moderate-severe 0.55 Surgery 0.55	conventional care ranged between £131,811 and £160,165 in the conventional care failure subpopulation and between £111,122 and £116,268 in the anti-TNF alfa failure subpopulation. In all scenarios and in both populations, ustekinumab yielded similar or greater QALY gains than other biologics at lower total costs.

	year. After this point, biologic therapy was assumed to stop, with all patients going on to receive conventional care for the remaining duration of the model. Patients who failed to respond in the induction period were assumed to move directly to conventional care. Patients who initiated on conventional care were assumed to continue to receive conventional care regardless of their induction phase response. At any point during the maintenance phase of the model, patients in the moderate to severe health state can receive surgery. Patients in the moderate to severe health state were		and lymphoma. The costs for all AEs, except for lymphoma, were sourced from NHSRC and based on the values used in TA352.		

	assumed to be at constant risk of surgery in the model and could receive multiple surgeries throughout their lifetime. Surgery therefore did not impact on the likelihood of future surgeries or affect future prognosis. *The ERG had several concerns with the company's model structure:* 1. *the model failed to capture the progressive and chronic nature of CD and did not account for the relapsing–remitting nature of the condition.* 2. *the model structure fails to distinguish between different types of surgery and*				

	does not recognise that patients who receive surgery are likely to have a quite different prognosis and treatment pathway to patients receiving drug therapy. The model also fails to recognise the significant impact of surgery on HRQoL and chance of future surgery. *3. the model makes a number of structural assumptions that are inconsistent with UK clinical practice. These included the assumptions that all non-*				

	responders have moderate to severe disease; that responders with moderate to severe CD and non-responders have equal costs and HRQoL; that patients cannot re-initiate biologic treatment upon future relapse; and the use of a drop of more than 100 points in the CDAI score to define response to induction treatment. **Cycle length** 2 weeks **Time horizon** 1 year *The ERG was concerned that the duration of treatment*				

	was too short as it is typical for patients to continue on therapy for much longer than a year				
Rafia 2016 **NICE TA352: Vedolizumab for treating moderately to severely active Crohn's disease after prior therapy**					
Population patients with moderate to severe active CD in 3 subpopulations: 1. the mixed ITT population, which comprised patients who had previously received anti-TNF-a therapy and those who were anti-TNF-a naive; 2. patients who were anti-TNF-a naive only 3. patients who had previously received anti-TNF-a therapy	**Perspective** NHS **Discounting** Costs and benefits discounted at 3.5% per year **Cost year** 2012 **DAM** The company's model structure was based on the structure published by Bodger et al. 2009 A decision tree was used to evaluate outcomes at the end of the initial induction period. A Markov structure was used to evaluate subsequent	Transition probabilities in the maintenance phase were calibrated using the Solver function in Excel so that: • the proportion of patients in remission at the end of the maintenance treatment (approximately at 1 year) predicted by the model matched the 'expected' proportion of patients in remission at the end of the maintenance phase • the proportion of patients with mild disease at the end of the maintenance phase predicted by the model matched the 'expected' percentage of responders in the induction phase with a drop of C70 points in the CDAI score and not in remission at the end of the maintenance phase. *The ERG observed that the calibration approach was complex and may have been unnecessary, as patient-level data from GEMINI II were available and could have been used to estimate the transition probabilities in the maintenance phase in patients treated with conventional non-biologic therapy and*	Management costs (healthcare resource use associated with inpatient treatment, outpatient visits, investigations and medications) for the different health states were taken from Bodgers and Hughes 2009. In response to the ACD, resource use was estimated through a survey conducted among clinical experts instead of Bodger and Hughes 2009. *The ERG was concerned that the use of costs estimated from the clinician survey conducted by the*	For the CDAI health states" the model uses the observed EQ-5D scores from the GEMINI II and GEMINI III studies **HSUVs (taken from the CS)** Remission 0.820 Mild disease 0.730 Moderate-severe disease 0.570 Surgery 0.570 (assumed equal to the moderate-severe health state because patients in GEMINI II AND GEMINI III were not followed for	Changes made by the company following the ACD (namely increasing the time horizon from 10 years to a lifetime; and updating the health state costs, using resource use estimated through a survey conducted among clinical experts) reduced the company's base-case deterministic (probabilistic) ICER from £98,452 to £21,620 (£27,428) in the anti-TNF alfa failure population SA following the ACD NR

only. *The ERG was concerned that the populations in the trials were not reflective of UK practice (patients who had previously received anti- TNF alfa agents and those who were anti-TNF alfa naive should be* *two distinct, defined patient groups). In response to the ACD, the company focused on patients for whom anti-TNF-a therapy had* *failed* **Intervention** Vedolizumab **Comparator** Conventional nonbiologic therapies (a	outcomes. The model was composed of a total of 12 mutually exclusive and exhaustive health states according to the treatment received, severity of condition and surgery. **Health States in the Markov model**: 1. Remission (CDAI ≤150) 2. Mild (CDAI 150-220) 3. Moderate - severe (CDAI 220-600) 4. Discontinue (the reasons for discontinuation include lack of response and AEs. Discontinuation due to AEs is applicable only to responders receiving	*vedolizumab. The ERG identified a number of limitations with the calibration approach used by the company—* *notably, that the target datapoints used in the* *fitting process seemed inconsistent with the datapoint the model was fitted to and that the derivation of the transition probabilities was dependent on structural assumptions and input parameters. Transition probabilities were* *assumed to be constant and applied to the remainder of the model, which was uncertain, given the lack of evidence after 1 year.*	*company may also have been inaccurate.* **Other resource use sources (taken from the CS)** Treatment acquisition costs including the estimated doses and unit costs for conventional non-biological therapy were taken from the BNF *The ERG was concerned that the treatment regimens for biologics were not costed accurately* The patient access scheme was applied to the cost of vedolizumab. Administration costs of £308 per administration in the maintenance phase and £616 in the induction phase were included for	surgery) Utility decrements for AEs were identified through a targeted review of the published literature	

combination of 5-amino salicylic acids [5-ASAs], immunomodulators and corticosteroids)	biologic treatments, because no responders on biologics switch to conventional therapy and continue receiving such until the end of the model's time horizon) 5. Surgery 6. Death *The ERG was concerned that the model structure was not appropriate:* *1. ignored relapsing (exacerbation) and remitting (some patients may improve spontaneously).* *2. Surgery modelled as a single health state representing a*		vedolizumab. Surgery related complication costs were estimated by applying NHSRC to resource use as reported by clinical expert opinion. AEs costs taken from NHSRC (all patients with an AE were assumed to be hospitalised)		

	mix of procedures. *3. Assumptions and adjustments that were not adequately justified by the evidence.* *4. Key structural assumptions were debatable. These included the assumption that non-responders had moderate-to-severe disease; the lack of distinction between responders and non-responders with moderate-to-severe CD; the assumption of the same induction phase duration for all therapy; the relevance to clinical practice of a drop of C70*				

	points in the CDAI score to identify patients going on to receive maintenance treatment; the end of scheduled maintenance at approximately 1 year; and a potentially optimistic assumption following discontinuation during biologic therapy and omission of discontinuation due to lack of efficacy **Cycle length** The induction period was assumed to be 6 weeks for all biologic and non-biologic therapy. Markov cycles were 8 weeks				

	Time horizon 10 years in the in the company's first submission and lifetime in response to the ACD				
Mayberry 2013 **NICE NG129: Crohn's disease: management**					
MODEL 1: **Population** people with an acute exacerbation of CD, with a CDAI score above 150. **Intervention and comparators** 9 treatment strategies: 1. SUL, CS, AZA+CS, BIO 2. SUL, CS, MTX+CS, BIO	**Perspective** NHS **Discounting** Costs and benefits discounted at 3.5% per year **Cost year** NR **DAM** A decision tree was constructed for each treatment strategy. Each treatment strategy included 4 lines of treatment. For the first 3 lines, individual could either enter remission (without AEs) or	The aim of the NMA was to calculate treatment-specific probabilities for withdrawal and remission conditional on non withdrawal. It is assumed that people who withdraw cannot go into remission, and similarly people counted as 'a remission' have not withdrawn due to adverse events, in other words, the two events are mutually exclusive. Treatment effects for the model had to be accounted for such that the number of withdrawals and remissions could not exceed the number of people in the trial. This negative correlation in outcomes is taken account of by carrying out a conditional logistic regression NMA. Baseline log odds of withdrawal and remission conditional on non-withdrawal were calculated using a logistic regression conducted on the placebo arms in the trials and then adjusted by the treatment specific log odds ratios calculated by the NMA. To reflect the populations explored in the clinical review, two separate analyses were conducted; one for people on first-line monotherapy treatment, and one	Drug costs used in the model came from the BNF, GDG and prescription cost analysis data. Other tests (including liver function tests and DEXA scan) came from NHSRC. NICE TB guideline was used as the source of costs for chest x ray and test for latent tuberculosis.	**Sources and elicitation methods** Utility weights derived by Stark et al 2010 were used due to the comparative lack of limitations and the directness of the population. In particular, the Stark data was favoured due to: use of EQ-5D to elicit utility weights directly from patients, UK EQ-5D tariff and use of CDAI thresholds that mirrored those used in most of the papers in the clinical	**Total QALYS** 1. CS, MTX+CS, BIO, 0.461 2. SUL, CS, MTX+CS, BIO ,0.442 3. BUD, CS, AZA+CS, BIO,0.455 4. SUL, CS, AZA+CS, BIO, 0.443 5. CS, AZA+CS, BIO,0.463 6. BUD, CS, MTX+CS, BIO, 0.454 7. MES, CS, AZA+CS, BIO, 0.450 8. MES, CS, MTX+CS, BIO, 0.448 9. CS, BIO, 0.457 **Total costs** No treatment £4,185

3. MES, CS, AZA+CS, BIO 4. MES, CS, MTX+CS, BIO 5. CS, AZA+CS, BIO 6 .CS, MTX+CS, BIO 7.BUD, CS, AZA+CS, BIO 8. BUD, CS, MTX+CS, BIO 9. CS, BIO ICD, Infliximab clinical discretion; BIO, biologic; TX, methotrexate; BUD budesonide; TX, methotrexate; AZA, azathioprine; SUL, sulfasalazine; MES, mesalazine; CS, corticosteroid.	move onto the next treatment line if they withdrew due to an AE or did not respond to treatment. Remission was defined as not withdrawing due to an AE and a CDAI score of < 150. All people who do not enter remission by the end of the time horizon are assumed to undergo surgery. **Time Horizon** 30 weeks	for people on second-line treatment in combination with a glucocorticosteroid, having failed first-line glucocorticosteroid monotherapy. Among first-line treatments, sulfasalazine was associated with the highest probability of withdrawal- 35%- but with the 95% confidence interval ranging from 5% to 80%. Glucocorticosteroid treatment was associated with the highest probability of remission conditional on non-withdrawal- 66%.out of the two second-line treatments, methotrexate + a glucocorticosteroid was associated with the highest probability of withdrawal- 11%. Azathioprine + a glucocorticosteroid was associated with a higher probability of remission conditional on no withdrawal- 66% **Input Probability of withdrawal due to adverse events; probability of achieving remission conditional on no withdrawal** Glucocorticosteroid 13%; 66% Sulfasalazine 31%; 44% Mesalazine 7%; 41% Budesonide 5%; 55% Azathioprine + glucocorticosteroid 6%; 66% Methotrexate + glucocorticosteroid 11%; 61%	Weighted average costs of drug preparations was used in the model. Weights were calculated using prescription cost analysis data in order to calculate drug costs which reflect how they are prescribed in clinical practice. With the cost of surgery weights were calculated using HES data, selected OPCS codes and assuming that 10% of all operations would be associated with a major comorbidity or complication. An average cost per operation was calculated by multiplying these weights by the costs attached to selected HRG codes and adding in pre-operative and	review. **Mean utility values** Disease remission 0.89 Active disease 0.61	MES £4,168 BUD £4,335 Glucocorticosteroid £4,908 AZA/Mercaptopurine £3,021 Olsalazine £5,525 **ICER** Olsalazine and glucocorticosteroid treatment were dominated by no treatment. BUD, MES and AZA were associated with ICERs of £40,000, £25,000 and £21,000 respectively compared with no treatment. **SA** Utility weights, disease remission 0.848, active disease 0.634 Consultations, Consultant gastroenterologist: 0-100 visits per 100 people every 2 weeks, Nurse specialist: 0-100 visits per 100 people

		Biologic 11%; 62%	post-operative consultation for each operation.		every 2 weeks, Specialist registrar: 0-100 visits per 100 people every 2 weeks Efficacy of Biologics, 33% TPMT cost, £26 **Results of PSA** The probability that each treatment was cost effective in terms of incremental net benefits at a WTP threshold of £20,000 per QALY: Placebo 0% MES 1% BUD 0% Glucocorticosteroid 1% AZA 98% Olsalazine 0%

MODEL 2 **Population** People with active CD in medically-induced remission. **Intervention and Comparators** 1) No treatment 2) Azathioprine 3) Mesalazine 4) Olsalazine 5) Budesonide 6) Glucocorticosteroids	**Perspective** NHS **Discounting** Costs and benefits discounted at 3.5% per year **Cost year** NR **DAM** A Markov model was constructed where the QALY gain was driven by the amount of time people spent in remission and active disease. **Health States** 1) Remission-Maintenance Treatment. 2) Remission- No maintenance. 3) Active disease 1ST line induction (Cs) 4) Active disease-Surgery 7) Active disease 2nd line induction	Two separate analyses were conducted for the clinical review, a non-conservative analysis where only the relapse outcome was analysed, and conservative analysis where "relapse and withdrawals" was analysed. Treatment effects in the economic model were parameterised so as to account for these two different methods. For the non-conservative analysis in the economic model withdrawals and relapses were treated separately so that people who withdrew from treatment were still assumed to be in remission. For the conservative analysis in the economic model, people who withdrew were assumed to be in relapse. The GDG advised that people in relapse should be treated with the induction sequence that was found to be most cost-effective in the induction of remission model (a glucocorticosteroid, followed by azathioprine + a glucocorticosteroid then a biologic). The GDG were uncertain as to what the induction sequence should be for people who relapse while on azathioprine maintenance treatment. People who relapse on azathioprine treatment are likely to have a glucocorticosteroid or biologic induction therapy added to their azathioprine regimen, and therefore initiation of azathioprine induction therapy in these people is not relevant as they are already taking azathioprine. One plausible alternative was to	Drug costs in the model came from the BNF and NHSRC. The PSSRU were the source of the costs for specialist nurse appointments and For GP consultations. NHS Reference costs were used for the following (phone call to IBD nurse, DEXA scan, microbiology test for chickenpox, full blood count, renal and liver function tests) NICE TB guidelines were used for (chest X ray, interferon gamma test for latent TB)	**Sources and elicitation methods** Utility weights derived by Stark et al 2010 were used due to the comparative lack of limitations and the directness of the population. In particular, the Stark data was favoured due to: use of EQ-5D to elicit utility weights directly from patients, UK EQ-5D tariff and use of CDAI thresholds that mirrored those used in most of the papers in the clinical review. **Mean utility values** Disease remission 0.89 Active disease 0.61	**Total QALYS:** No treatment 1.67 MES 1.68 BUD 1.68 Glucocorticosteroid 1.65 Azathioprine/mercaptopurine 1.71 Olsalazine 1.64 **Total costs** No treatment £2,198 Mesalazine £2,628 Budesonide £2,675 Glucocorticosteroid £2,627 Azathioprine /mercaptopurine £3,021 Olsalazine £3,131 **ICER** (versus no treatment) No treatment Comparator Mesalazine £25,133 Budesonide £40,392 Glucocorticosteroid Dominated 8) Azathioprine

	(Azathioprine 5) +Cs) 6) Active disease 3rd line induction (biologic) 7) Remission (biologic) 8) Remission (No maintenance) **Cycle length** 2 months **Time horizon** 2 years	assume a three-line induction sequence for azathioprine (a glucocorticosteroid – a biologic – surgery) but a four-line sequence (a glucocorticosteroid – azathioprine + a glucocorticosteroid- a biologic – surgery) for the other maintenance strategies but this three-line sequence is less cost effective and this may potentially bias the assessment.			/mercaptopurine £20,319 Olsalazine Dominated **SA** • Time horizon increased from two years to 10 years • QALY discount rate decreased from 3.5% to 1.5% • Average number of drug-specific tests over two years used from Year 2 to average • Average estimated resource use over three years was calculated and used in the model instead of the estimated year two resource use • Since data were not available to inform a model based on varying levels of patient severity, it was decided to explore the effects of a utility decrement for each stage of failed induction therapy from 0% to 10%.

					• A higher baseline risk for relapse and relapse + withdrawal was explored in the non-conservative and conservative analyses respectively, non-conservative: relapse=39% conservative: relapse and withdrawal =52%, 90% • A lower baseline risk for relapse and relapse withdrawal was explored in the non-conservative and conservative analyses respectively, Non-conservative: Relapse=39%, conservative: relapse+ withdrawal= 52%, 10% *Cost-effectiveness rankings changed using:* • 10-year time horizon (Mesalazine ranked first, no treatment ranked second, budesonide ranked third) • yearly baseline risk of

					relapse = 90% (Mesalazine first azathioprine second, budesonide third) • yearly baseline risk of relapse = 10% (no treatment first, glucocorticosteroid second, azathioprine third) **PSA** Cost effectiveness ranks at £20,000 per QALY and their associated 95% confidence intervals were calculated by Monte-Carlo simulation. Probability of being most cost-effective: Placebo 22% Mesalazine 7% Budesonide 8% Glucocorticosteroid 23% Azathioprine 41% Olsalazine 0% Olsalazine and glucocorticosteroids were dominated by no treatment. BUD was associated with an

					ICER of £15,000 per QALY gained compared with no treatment and both MES and BUD were cost effective versus no treatment.

Abbreviations: AE, adverse event; ACD, Appraisal Committee Decision; CD, Crohn's disease; CDAI, Crohn's disease activity index; CS, company submission; DAM, decision analytic model; DRG, decision resource group; ERG, Evidence Review Group; HSUV, health state utility value; IBD, inflammatory bowel disease; ICER, incremental cost-effectiveness ratio; IND, induction; MNT, maintenance; NHSRC, NHS Reference Costs; QALY, quality-adjusted life year; SA, sensitivity analysis; SC, standard care; SG, standard gamble; TTO, time trade off; PSA, probabilistic sensitivity analysis;

9.5.3 Health-related quality of life (HRQoL) evidence

Study	Elicitation method	Valuation method	Population	Health states and utility values
Benedini et al 2012	Patients assessed their own HRQoL using the EQ-5D-3L	EQ-5D scores were converted into utilities using a UK source (Dolan et al. 1995)	162 patients with CD in the active phase and a CDAI score >150 were recruited from 25 Italian centres. CD was diagnosed within at least 6 months of study entry. 50% male, mean age 43 years (range 21 to 73)	Proportion male; mean utility (SD) Enrolment: (n=162), 50%; 0.558 (0.310) 6 months: (n=154), 50.6%; 0.682 (0.254) 12 months: (n=154), 50.6%; 0.728 (0.251) 18 months: (n=155), 51%; 0.739 (0.262)
Casellas et al 2005 Impairment of Health-related Quality of Life…	Patients assessed their own HRQoL using the IBDQ and EQ-5D-3L	EQ-5D scores were converted into utilities using representative sample of 12,245 members of the Spanish	628 CD patients from 9 different hospitals in different geographical areas in Spain. Mean disease duration 60 months (range 22 to 161), 274 male, 354 female, mean age 34 years (range 26 to 44)	Mean utility (range) Remission: (n=360) 0.8 (0.7 to 1) Mild: (n=151) 0.72 (0.5 to 0.8) Moderate-Severe: (n=115) 0.6 (0.5 to 0.7)

		general population (Badía et al. 1999)		
Casellas et al 2000	Patients assessed their own HRQoL using the IBDQ, PGWBI and EQ-5D-3L.	EQ-5D scores were converted into utilities using a Spanish source (Rue and Badia 1996)	119 Consecutive patients with CD were seen at the Digestive Diseases Department in Spain Health controls N=63, 32 years (27 to 36), 69% female Operated, in remission N=29, 33 years (32 to 40), 68% female Nonoperated, in remission N=48, 29.5 years (28 to 35), 66% female Active N=42, 29 years (29 to 38), 67% female	Median utility (95% CI) Operated CD patients: 0.87 (0.79 to 0.92) Remission CD patients: 0.86 (0.78 to 0.87) Active CD patients: 0.67 (0.62 to 0.72)
Casellas et al 2007	Patients assessed their own HRQoL using the IBDQ-36 and EQ-5D-3L.	EQ-5D scores were converted into utilities using a sample of the Spanish population (Badia et al. 1999).	Outpatients with CD who had been infliximab and azathioprine induced and maintained clinical remission for at least six months. Medians and (percentiles 25-75) Age 40 years (31 to 49), male/female 26/23, disease duration 96 months (40 to 161) Forty-nine patients with CD in stable clinical remission were included at baseline. At 12 months, 42 patients remained in remission, at 24 months 32 patients, at 36 months 13, and in the last visit at 48 months 6 patients remained in clinical remission.	Median utility for patients in remission (range) Inclusion: 1 (0.8 to 1) 12 months: 1 (1 to 1) 24 months: 1 (1 to 1) 36 months: 1 (0.9 to 1) 48 months: 1 (0.8 to 1)
Casellas et al 2005 Relevance of the Phenotypic Characteristics…	Patients assessed their own HRQoL using the IBDQ-36, PGWBI, and EQ-5D-3L	NR	142 CD outpatients and 56 CD inpatients treated at a Crohn Colitis Care Unit in Spain Median age 30 years (23 to 38), male/female 76/122, disease duration 36 months (5 to 82)	Median utility values (percentiles 25-75): <40 years old 0.72 (0.57 to 100), >= 40 years old 0.72 (0.57 to 0.87) disease location: terminal ileum 0.72 (0.57 to 0.85), colon 0.72 (0.57 to 1.00), ileocolon

			Age: <40 years n=177, ≥40 years n=21 Disease location: L1 (terminal ileum) n= 53, L2 (colon) n=62, L3 (ileocolon) n=72, L4 (upper gastrointestinal) n=11 Disease behaviour: B1 (inflammatory) n=99, B2 (stricturing) n=32, B3 (penetrating) n=67	0.72 (0.57 to 1.00), upper GI 0.71(0.54 to 0.87) disease behaviour: inflammatory 0.72(0.57 to 1.00), stricturing 0.85 (0.57 to 1.00), penetrating 0.72 (0.57 to 0.85)
Holko et al 2016	Patients assessed their own HRQoL using the EQ-5D-3L	EQ-5D scores were converted into utilities using Polish tariffs (Golicki et al. 2010)	Undertaken in Poland. Remission (N = 93); Active disease (N = 105); All patients with CD (N = 200) Age, mean (SD): 30.60 (9.94); 32.86 (10.75); 31.80 (10.41) Male, n (%): 40 (43.0); 44 (41.9;) 84 (42.2) Age at diagnosis, mean (SD): 24.14 (10.26) 24.73 (10.88) 24.46 (10.57) 47% of all patients in remission	Mean utility (SD) All: 0.839 (0.171) Remission: 0.908 (0.084) Active: 0.777 (0.203) Mild: 0.859 (0.094) Moderate: 0.754 (0.203) Severe: 0.462 (0.353)
Huaman et al 2010	Patients assessed their own HRQoL using the IBDQD-36 and EQ-5D-3L.	EQ-5D scores were converted into utilities using representative a sample of 12,245 members of the Spanish general population (Badía et al. 1999)	114 patients with CD at the Unitat Atencio Crohn Colitis Hospital in Spain. 61% in remission and 39% in relapse Mean (SD) age 32.21 years (12.13), male/female 47/67, disease duration 37.41 months (48.94) Location, n (%): L1=terminal ileum 20 (17.55), L2 = colon 33 (28.95), L3 =ileocolon 59 (51.75), L4 = upper GI2 (1.75)	Mean utility in CD 0.76, SD 0.18 Mean utility ≥ 0.9 in 33 patients and <0.9 in 81 patients
Mozzi A et al 2016	Patients assessed their own HRQoL using the EQ-5D-3L	Utilities calculated by applying Italian, US and UK general population based preference weights. The TTO	The study is based on a survey conducted in Italy from 2012 to 2013. 552 patients with moderate to severe active CD enrolled in the SOLE survey referred to gastroenterological centres	Mean utility obtained from UK tariffs (SD) HBI 8-11 (moderate burden of disease): (n=389), 0.63 (0.28)

		technique was used for the preference of the scoring function in adults randomly sampled from the general population. Algorithms were developed by Scalone et al. 2013 (Italy) Shaw et al 2005 (US) and Badia et al 2001 (UK)	Mean (SD) age 41.18 years (13.77), male/female 251/249, years from CD onset 2.19 (4.44)	HBI 12-16 (moderate/severe burden of disease): (n=84), 0.45 (0.37) HBI >16 (severe burden of disease): (n=27), 0.22 (0.40) Total sample: (n=500), 0.57(0.32)
Rencz et al 2019	Patients assessed their own HRQoL using the EQ-5D-3L and EQ-5D-5L	To obtain 3L values and 5L values, the UK tariffs by Dolan et al 1997 and the English tariffs by Devlin et al 2018 were used	Outpatients with CD from three academic gastroenterology departments and an inflammatory bowel disease centre in three large cities in Hungary. Mean (SD) age 34.7 years (10.5), 54.9% male, mean disease duration 10.5 years (6.3)	5L scores: mean (SD); median (IQR) Symptomatic remission (CDAI<150), Mild (CDAI 150–219), Moderate-to-severe (CDAI 220≤) Symptomatic remission n=154, 0.88 (0.11); 0.92(0.81-1) Mild n=32, 0.86 (0.14); 0.9 (0.79-0.99) Moderate to Severe n=18, 0.79(0.19); 0.8(0.73-0.94) 3L scores: mean (SD); median (IQR) Symptomatic remission n=154, 0.82(0.15); 0.8 (0.69-1) Mild n=31 0.77 (0.21); 0.8 (0.69-1) Moderate to Severe n=18, 0.7 (0.23); 0.71(0.65-0.86)

Saro et al 2017	Patients assessed their own HRQoL using the EQ-5D and the IBDQ.	NR	126 patients with CD were recruited from 33 centers located in 11 geographical regions in Spain. Included patients who are naïve to any biological treatment for whom adalimumab treatment was prescribed as part of regular clinical practice Female 49.2%, mean age was 39.1 ± 13.8 years, diagnosed with CD over a median of seven years Based on the Montreal classification by location of the disease, 35.7% of CD was in L1; 10.3%, in L2; 47.6%, in L3; 2.4%, in L4; 3.2%, in L1 + L4; and 0.8%, in L2 + L4	Mean utility (SD) baseline 0.680 (0.288) 1 month 0.785 (0.2) 3 months 0.800 (0.208) 6 months 0.813 (0.203) 9 months 0.805 (0.238) 12 months 0.815 (0.214)
Stark et al 2010	Patients assessed their own HRQoL using the EQ-5D-3L. The questionnaire was completed by the patients and then 4 weeks later	A TTO tariff was calculated for Germany by Greiner et al in 339 persons. German tariffs valued health states better than UK tariffs especially for poorer health states. EQ-5D UK Index tariffs obtained from Dolan et al 1997.	A random sample of 270 CD patients were drawn from the DCCV (German IBD association) according to the last digit of their membership number. Mean (SD) age 41 years (11), 37.4% male, age at diagnosis 27 years (10), disease duration 14 years (8), remission 57.1%, active 42.9%	Mean utility (SD) All patients at baseline n=269 0.77 (0.24) Remission n=128 0.89 (0.13)
Abbreviations: CD, Crohn's disease; HBI, Harvey Bradshaw Index; IBDQ, Inflammatory Bowel Disease Questionnaire; HRQoL, health-related quality of life; NR, not reported; PGWBI, Psychological General Well-Being Index SD, standard deviation; TTO, time trade-off				

9.6 Appendix 6. Quality assessment

9.6.1 Prognostic accuracy and clinical impact

9.6.1.1 IBDX

Reviewer and study information			
Reviewer name	Sam Barton		
Study ID (Author name, year)	Harrell 2010		
Study details (journal, year, volume, page range)	*Gastroenterology*, 2010, 138, (5), S529		
Type of report (full paper/only abstract/conference abstract)	Conference abstract		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes.	Unclear
	Description of the source population or population of interest	No, insufficient detail in abstract.	
	Description of the baseline study sample	No, insufficient detail in abstract.	
	Adequate description of the sampling frame and recruitment	No, insufficient detail in abstract.	
	Adequate description of the period and place of recruitment	No, insufficient detail in abstract.	
	Adequate description of inclusion and exclusion criteria	No, insufficient detail in abstract.	

Study attrition	Adequate response rate for study participants	Yes.	Unclear
	Description of attempts to collect information on people who dropped out	Unclear.	
	Reasons for loss to follow-up are provided	Unclear.	
	Adequate description of those lost to follow-up	Unclear.	
	There are no important differences between people who completed the study and those who did not	Unclear.	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes.	Unclear
	Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record)	Yes.	
	Continuous variables are reported or appropriate cut points are used	Unclear.	
	The method and setting of measurement of prognostic factor is the same for all those in the study	Unclear.	
	Adequate proportion of the study sample has	Unclear.	

	complete data for the prognostic factor		
	Appropriate methods of imputation are used for missing prognostic factor data	Unclear.	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes.	Unclear
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Unclear.	
	The method and setting of outcome measurement are the same for all those in the study	Unclear.	
Study confounding	Most important confounders are measured	Confounders are not discussed in the abstract. Assessment of efficacy of tool is based on event rate of complication of disease (intestinal fistula and/or stricture) and need for surgery.	Unclear
	Clear definitions of the important confounders measured are provided		
	Measurement of all important confounders is adequately valid and reliable		
	The method and setting of confounding measurement are the same for all those in the study		
	Appropriate methods are used if imputation is		

	used for missing confounder data		
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)		
	Important potential confounders are accounted for in the analysis		
Statistical analysis and reporting	Sufficient presentation of data to assess the adequacy of the analytic strategy	No	Unclear
	Strategy for model building is appropriate and is based on a conceptual framework or model	Unclear	
	The selected statistical model is adequate for the design of the study	Unclear	
	There is no selective reporting of results	Unclear	
Abbreviation: IBD, inflammatory bowel disease.			

Reviewer and study information	
Reviewer name	Sam Barton
Study ID (Author name, year)	Paul 2015
Study details (journal, year, volume, page range)	*J Crohns Colitis*, 2015, 9, (6), 445–451
Type of report (full paper/only	Full paper

abstract/conference abstract)			
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes Study enrolled those with ulcerative colitis and those with Crohn's disease. Those enrolled had a diagnosis of disease for more than one year.	Low
	Description of the source population or population of interest	Yes	
	Description of the baseline study sample	Yes	
	Adequate description of the sampling frame and recruitment	Yes	
	Adequate description of the period and place of recruitment	Yes	
	Adequate description of inclusion and exclusion criteria	Yes	
Study attrition	Adequate response rate for study participants	Yes Relevant samples collected from all those enrolled.	Low
	Description of attempts to collect information on people who dropped out	Not applicable.	
	Reasons for loss to follow-up are provided	Not applicable.	
	Adequate description of those lost to follow-up	Not applicable.	
	There are no important differences between people who completed	Not applicable.	

	the study and those who did not		
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes Changes in serum levels of individual antibodies.	Low
	Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record)	Yes Authors followed the manufacturer's instructions to generate unit of measurement.	
	Continuous variables are reported or appropriate cut points are used	Yes.	
	The method and setting of measurement of prognostic factor is the same for all those in the study	Yes.	
	Adequate proportion of the study sample has complete data for the prognostic factor	Yes.	
	Appropriate methods of imputation are used for missing prognostic factor data	Not applicable.	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes for the outcome of interest to the systematic review reported here. The authors of the review appreciate that it will be difficult to determine the true clinical impact of the tool.	Low
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital	Yes Analyses of clinical data and serological assessments were carried out in a masked manner without knowledge of patient's diagnosis and medical history.	

	record or record linkage)		
	The method and setting of outcome measurement are the same for all those in the study	Yes	
Study confounding	Most important confounders are measured	Confounders are not discussed in the full publication	Unclear
	Clear definitions of the important confounders measured are provided		
	Measurement of all important confounders is adequately valid and reliable		
	The method and setting of confounding measurement are the same for all those in the study		
	Appropriate methods are used if imputation is used for missing confounder data		
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)		
	Important potential confounders are accounted for in the analysis		
	Sufficient presentation of data to assess the	Yes	Low

Statistical analysis and reporting	adequacy of the analytic strategy		
	Strategy for model building is appropriate and is based on a conceptual framework or model	No model built.	
	The selected statistical model is adequate for the design of the study	Yes	
	There is no selective reporting of results	Yes	
Abbreviation: IBD, inflammatory bowel disease.			

Reviewer and study information			
Reviewer name	Sam Barton		
Study ID (Author name, year)	Rieder 2010b		
Study details (journal, year, volume, page range)	*Inflamm Bowel Dis*, 2010, 16, 263–274. Related paper: Rieder 2011a *PLoS ONE*, 2011, 6, (5), e18172 (presents data on subgroup of people reported in Rieder 2010b as a longitudinal analysis). Related paper: Rieder 2010d *Gastroenterology*, 2010, 138, (5), S522 (conference abstract for Rieder 2011a).		
Type of report (full paper/only abstract/conference abstract)	Full paper.		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes All people analysed have a diagnosis of CD. However, there is a mixed population in terms of those with a new diagnosis and with an established diagnosis, as well as presence of complicated disease at baseline versus no complications. Data are not	Moderate

		reported separately for the various subgroups.	
	Description of the source population or population of interest	Yes	
	Description of the baseline study sample	Yes	
	Adequate description of the sampling frame and recruitment	Yes	
	Adequate description of the period and place of recruitment	Yes	
	Adequate description of inclusion and exclusion criteria	Yes	
Study attrition	Adequate response rate for study participants	Yes Samples were available for all those enrolled with CD.	Low
	Description of attempts to collect information on people who dropped out	Not applicable.	
	Reasons for loss to follow-up are provided	Not applicable.	
	Adequate description of those lost to follow-up	Not applicable.	
	There are no important differences between people who completed the study and those who did not	Not applicable.	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes	Low
	Method of prognostic factor measurement is adequately valid and	Yes	

	reliable (i.e., direct ascertainment; secure record, hospital record)		
	Continuous variables are reported or appropriate cut points are used	Yes	
	The method and setting of measurement of prognostic factor is the same for all those in the study	Yes	
	Adequate proportion of the study sample has complete data for the prognostic factor	Yes	
	Appropriate methods of imputation are used for missing prognostic factor data	Not applicable	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes	Low
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Yes Analyses based on stored blood samples. The authors sent the samples to Glycominds for analysis, which was carried out in a masked manner.	
	The method and setting of outcome measurement are the same for all those in the study	Yes	
Study confounding	Most important confounders are measured	Important baseline characteristics are adjusted for in statistical analyses. Factors adjusted for were age, gender,	Moderate

		BMI, disease activity and duration, age at diagnosis, and disease location as potential confounders. The authors comment that it is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels. Assessment of efficacy of tool is based on event rate of complication of disease or need for surgery.	
	Clear definitions of the important confounders measured are provided	Yes	
	Measurement of all important confounders is adequately valid and reliable	Yes	
	The method and setting of confounding measurement are the same for all those in the study	Yes	
	Appropriate methods are used if imputation is used for missing confounder data	Not applicable	
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)	No	
	Important potential confounders are accounted for in the analysis	Yes – see comment above	
	Sufficient presentation of data to assess the	Yes	Low

Statistical analysis and reporting	adequacy of the analytic strategy		
	Strategy for model building is appropriate and is based on a conceptual framework or model	No model built.	
	The selected statistical model is adequate for the design of the study	Yes	
	There is no selective reporting of results	Yes	
Abbreviation: IBD, inflammatory bowel disease.			

Reviewer and study information			
Reviewer name	Sam Barton		
Study ID (Author name, year)	Rieder 2010c		
Study details (journal, year, volume, page range)	*Inflamm Bowel Dis*, 2010, 16, (8), 1367–1375.		
Type of report (full paper/only abstract/conference abstract)	Full paper.		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes All people analysed have a diagnosis of CD. However, there is a mixed population in terms of those with no complications of disease at baseline, and no prior surgery versus those with complications and/or prior surgery. People had to have at least 3 years of follow-up to be eligible. Subgroup data	Low

		are reported for subgroups of potential interest.	
	Description of the source population or population of interest	Yes	
	Description of the baseline study sample	Yes	
	Adequate description of the sampling frame and recruitment	Yes	
	Adequate description of the period and place of recruitment	Yes	
	Adequate description of inclusion and exclusion criteria	Yes	
Study attrition	Adequate response rate for study participants	Yes Samples were available for all those enrolled with CD.	Low
	Description of attempts to collect information on people who dropped out	Not applicable.	
	Reasons for loss to follow-up are provided	Not applicable.	
	Adequate description of those lost to follow-up	Not applicable.	
	There are no important differences between people who completed the study and those who did not	Not applicable.	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes	Low
	Method of prognostic factor measurement is adequately valid and	Yes	

	reliable (i.e., direct ascertainment; secure record, hospital record)		
	Continuous variables are reported or appropriate cut points are used	Yes	
	The method and setting of measurement of prognostic factor is the same for all those in the study	Yes	
	Adequate proportion of the study sample has complete data for the prognostic factor	Yes	
	Appropriate methods of imputation are used for missing prognostic factor data	Not applicable	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes	Low
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Yes Analyses based on stored blood samples. The authors sent the samples to Glycominds for analysis, which was carried out in a masked manner.	
	The method and setting of outcome measurement are the same for all those in the study	Yes	
Study confounding	Most important confounders are measured	The authors comment that it is unclear as to what extent antibody levels change over time in individual people	Moderate

		and what factors influence changes in levels. Assessment of efficacy of tool is based on event rate of complication of disease or need for surgery. The authors proposed age, gender, BMI, disease activity and duration, age at diagnosis, and disease location as potential confounders.	
	Clear definitions of the important confounders measured are provided	Yes	
	Measurement of all important confounders is adequately valid and reliable	Yes	
	The method and setting of confounding measurement are the same for all those in the study	Yes	
	Appropriate methods are used if imputation is used for missing confounder data	Not applicable	
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)	No	
	Important potential confounders are accounted for in the analysis	Yes Authors carried out a regression analysis to account for the potential confounders.	
Statistical analysis and reporting	Sufficient presentation of data to assess the adequacy of the analytic strategy	Yes	Low

	Strategy for model building is appropriate and is based on a conceptual framework or model	No model built.	
	The selected statistical model is adequate for the design of the study	Yes	
	There is no selective reporting of results	Yes	
Abbreviation: IBD, inflammatory bowel disease.			

Reviewer and study information			
Reviewer name	Sam Barton		
Study ID (Author name, year)	Rieder 2012		
Study details (journal, year, volume, page range)	*Inflamm Bowel Dis*, 2012, 18, (7), 1221–1231. Related paper: Rieder 2011b J Crohns Colitis, 2011, 5, (1), S48 (conference abstract).		
Type of report (full paper/only abstract/conference abstract)	Full paper.		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes All children analysed have a diagnosis of CD. However, there is a mixed population in terms of those with a new diagnosis and with an established diagnosis, as well as presence of complicated disease at baseline versus no complications. Data are not reported separately for the various subgroups.	Moderate

	Description of the source population or population of interest	Yes	
	Description of the baseline study sample	Yes	
	Adequate description of the sampling frame and recruitment	Yes	
	Adequate description of the period and place of recruitment	Yes	
	Adequate description of inclusion and exclusion criteria	Yes	
Study attrition	Adequate response rate for study participants	Yes All people were eligible for analysis and samples were available for all children.	Low
	Description of attempts to collect information on people who dropped out	Not applicable	
	Reasons for loss to follow-up are provided	Not applicable	
	Adequate description of those lost to follow-up	Not applicable	
	There are no important differences between people who completed the study and those who did not	Not applicable	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes	Low
	Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record)	Yes	

	Continuous variables are reported or appropriate cut points are used	Yes	
	The method and setting of measurement of prognostic factor is the same for all those in the study	Yes	
	Adequate proportion of the study sample has complete data for the prognostic factor	Yes	
	Appropriate methods of imputation are used for missing prognostic factor data	Not applicable	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes	Low
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Yes Analyses based on stored blood samples. Samples were analysed as per the manufacturer's instructions in a blinded manner without knowledge of diagnosis or other clinical information.	
	The method and setting of outcome measurement are the same for all those in the study	Yes	
Study confounding	Most important confounders are measured	Important baseline characteristics are adjusted for in statistical analyses. Factors adjusted for were age, gender, BMI, disease activity and ileum involvement. as potential confounders. However, it is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels. Assessment of efficacy of tool is based on event rate of complication of disease or need for surgery.	Moderate

	Clear definitions of the important confounders measured are provided	Yes	
	Measurement of all important confounders is adequately valid and reliable	Yes	
	The method and setting of confounding measurement are the same for all those in the study	Yes	
	Appropriate methods are used if imputation is used for missing confounder data	Not applicable	
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)	No	
	Important potential confounders are accounted for in the analysis	Yes – see comment above	
Statistical analysis and reporting	Sufficient presentation of data to assess the adequacy of the analytic strategy	Yes	Low
	Strategy for model building is appropriate and is based on a conceptual framework or model	No model built.	
	The selected statistical model is adequate for the design of the study	Yes	
	There is no selective reporting of results	Yes	
Abbreviations: IBD, inflammatory bowel disease; IBDX, Crohn's disease Prognosis Test.			

Reviewer and study information	
Reviewer name	Sam Barton

Study ID (Author name, year)	Seow 2009		
Study details (journal, year, volume, page range)	*Am J Gastro*, 2009, 104, (6), 1426–1434.		
Type of report (full paper/only abstract/conference abstract)	Full paper.		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes However, the study enrolled a mixed population of adults and children, those with a new diagnosis and an established diagnosis of CD, and varying degrees of existing complicated disease. Data for different subgroups are not reported separately.	Moderate
	Description of the source population or population of interest	Yes	
	Description of the baseline study sample	Yes	
	Adequate description of the sampling frame and recruitment	Yes	
	Adequate description of the period and place of recruitment	Yes	
	Adequate description of inclusion and exclusion criteria	Yes	
Study attrition	Adequate response rate for study participants	Yes Number of people with a positive test for one or more antibody totals 378 out of 517 people for whom samples were available (73.1%). It is unclear from	Low

		the details in the full publication whether the remaining 139 people did not test positive for an antibody, or whether their samples were not analysed.	
	Description of attempts to collect information on people who dropped out	Not applicable	
	Reasons for loss to follow-up are provided	Not applicable	
	Adequate description of those lost to follow-up	Not applicable	
	There are no important differences between people who completed the study and those who did not	Not applicable	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes	Low
	Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record)	Yes	
	Continuous variables are reported or appropriate cut points are used	Yes	
	The method and setting of measurement of prognostic factor is the same for all those in the study	Yes	
	Adequate proportion of the study sample has complete data for the prognostic factor	Yes	
	Appropriate methods of imputation are used for missing prognostic factor data	Not applicable	

Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes	Low
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Yes Frozen serum samples were forwarded to Glycominds Limited for analysis in a masked manner.	
	The method and setting of outcome measurement are the same for all those in the study	Yes	
Study confounding	Most important confounders are measured	No Confounders are not discussed in the publication. It is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels.	Moderate
	Clear definitions of the important confounders measured are provided		
	Measurement of all important confounders is adequately valid and reliable		
	The method and setting of confounding measurement are the same for all those in the study		
	Appropriate methods are used if imputation is used for missing confounder data		
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)		
	Important potential confounders are accounted for in the analysis		

Statistical analysis and reporting	Sufficient presentation of data to assess the adequacy of the analytic strategy	Yes	Low
	Strategy for model building is appropriate and is based on a conceptual framework or model	No model built.	
	The selected statistical model is adequate for the design of the study	Yes	
	There is no selective reporting of results	Yes	
Abbreviation: IBD, inflammatory bowel disease; IBDX, Crohn's disease Prognosis Test.			

Reviewer and study information			
Reviewer name	Sam Barton		
Study ID (Author name, year)	Wolfel 2017		
Study details (journal, year, volume, page range)	*Gastroenterology*, 2017, 152, (5), S605		
Type of report (full paper/only abstract/conference abstract)	Conference abstract.		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	Yes. All people have a diagnosis of CD and have undergone one CD-related surgery.	Unclear
	Description of the source population or population of interest	No, insufficient detail provided in the abstract.	
	Description of the baseline study sample	No, only limited detail on age and gender provided in the abstract.	

	Adequate description of the sampling frame and recruitment	No, insufficient detail provided in the abstract.	
	Adequate description of the period and place of recruitment	No, insufficient detail provided in the abstract.	
	Adequate description of inclusion and exclusion criteria	No, insufficient detail provided in the abstract.	
Study attrition	Adequate response rate for study participants	Unclear.	Unclear
	Description of attempts to collect information on people who dropped out	Unclear.	
	Reasons for loss to follow-up are provided	Unclear.	
	Adequate description of those lost to follow-up	Unclear.	
	There are no important differences between people who completed the study and those who did not	Unclear.	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes.	Unclear
	Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record)	Yes.	
	Continuous variables are reported or appropriate cut points are used	Unclear.	

	The method and setting of measurement of prognostic factor is the same for all those in the study	Unclear.	
	Adequate proportion of the study sample has complete data for the prognostic factor	Unclear.	
	Appropriate methods of imputation are used for missing prognostic factor data	Unclear.	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes.	Unclear
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Unclear.	
	The method and setting of outcome measurement are the same for all those in the study	Unclear.	
Study confounding	Most important confounders are measured	The authors state that “Multivariable Cox regression analysis was performed to assess the associations between markers and recurrence of surgery adjusting for potential confounders”. Potential confounders are not listed.	Unclear
	Clear definitions of the important confounders measured are provided		
	Measurement of all important confounders is adequately valid and reliable		

	The method and setting of confounding measurement are the same for all those in the study		
	Appropriate methods are used if imputation is used for missing confounder data		
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)		
	Important potential confounders are accounted for in the analysis		
Statistical analysis and reporting	Sufficient presentation of data to assess the adequacy of the analytic strategy	No	Unclear
	Strategy for model building is appropriate and is based on a conceptual framework or model	Unclear	
	The selected statistical model is adequate for the design of the study	Unclear	
	There is no selective reporting of results	Unclear	
Abbreviation: IBD, inflammatory bowel disease.			

9.6.1.2 PredictSURE-IBD

Reviewer and study information

Reviewer name	Sam Barton		
Study ID (Author name, year)	Biasci 2019		
Study details (journal, year, volume, page range)	*Gut*, 2019, (68), 1386–1395		
Type of report (full paper/only abstract/conference abstract)	Full paper		
Domain	**Aspects of trial for consideration in assessment of bias**	**Comment in support of assessment of bias**	**Rating of risk of bias**
Study participation	Adequate participation in the study by eligible persons	People must have active disease for the tool to detect the desired sequence. Validation cohort predominantly comprises those with newly diagnosed CD and disease is active.	Low
	Description of the source population or population of interest	Yes.	
	Description of the baseline study sample	Not supplied for validation cohort in the full publication. Baseline characteristics provided by authors on request.	
	Adequate description of the sampling frame and recruitment	Yes.	
	Adequate description of the period and place of recruitment	Yes.	
	Adequate description of inclusion and exclusion criteria	Unclear reporting of inclusion criteria for validation cohort in full publication. Authors helpfully confirmed that inclusion criteria for validation cohort were the same as those for the training cohort.	

Study attrition	Adequate response rate for study participants	Systematic literature search identified three conference abstracts that referred to a validation cohort comprising 85 people rather than the 66 reported in the full publication.[146-148] During the DAP process, the company clarified that the cohort comprising 85 people refers to the validation work at an earlier stage of research and additional samples were added before publication of the full text. The EAG considers that it is unclear whether there are two cohorts or people have been lost to follow-up.	Unclear
	Description of attempts to collect information on people who dropped out	Not reported.	
	Reasons for loss to follow-up are provided	Not reported.	
	Adequate description of those lost to follow-up	Not reported.	
	There are no important differences between people who completed the study and those who did not	Unclear.	
Prognostic factor measurement	A clear definition or description of the prognostic factor is provided	Yes	Low
	Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record)	Yes Gene expression analyses. However, people must have active disease.	
	Continuous variables are reported or	Not applicable	

	appropriate cut points are used		
	The method and setting of measurement of prognostic factor is the same for all those in the study	Yes	
	Adequate proportion of the study sample has complete data for the prognostic factor	Yes	
	Appropriate methods of imputation are used for missing prognostic factor data	Unclear	
Outcome measurement	A clear definition of the outcome of interest is provided (including time of death)	Yes for the clinical outcome of time to event. The authors of the systematic review appreciate that it will be difficult to determine the true clinical impact of the tool.	Low
	Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage)	Yes Treating clinicians were masked to the biomarker results, and to gene expression analyses.	
	The method and setting of outcome measurement are the same for all those in the study	Yes	
Study confounding	Most important confounders are measured	Assessment of efficacy of tool is based on time to an event involving treatment escalation based on clinical judgement. Confounders are not discussed in the full publication.	Unclear
	Clear definitions of the important confounders measured are provided		

	Measurement of all important confounders is adequately valid and reliable		
	The method and setting of confounding measurement are the same for all those in the study		
	Appropriate methods are used if imputation is used for missing confounder data		
	Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching)		
	Important potential confounders are accounted for in the analysis		
Statistical analysis and reporting	Sufficient presentation of data to assess the adequacy of the analytic strategy	Yes	Low
	Strategy for model building is appropriate and is based on a conceptual framework or model	In the development of the whole-blood sample test (IBDHi versus IBDLo), IBD1/IBD2 status was not included as a covariate in the batch normalisation of whole blood samples to reduce any downward bias in estimating the generalisation error during leave-one-out cross-validation. The impact of this is unclear.	
	The selected statistical model is adequate for the design of the study	Yes. A statistical (machine) learning method was applied to the whole blood	

		transcriptomic data to identify genes that could be used to calculate the probability of an individual belonging to the IBD1/IBD2 subgroups.	
	There is no selective reporting of results	Unclear. The full publication presents data for both the training and validation cohorts and the reporting of the data is considered to be unclear in some aspects.	
Abbreviations: IBD, inflammatory bowel disease.			

9.6.2 Economic evaluations

Drummond checklist for economic evaluations

Criteria	Study				
	Clark 2003	**Dretzke 2011 (TA187)**	**Hodgson 2018 NICE TA456**	**Rafia 2016 NICE TA352**	**NICE NG129**
1. Was a well-defined question posed in answerable form?	Yes	Yes	Yes	Yes	Yes
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes (infliximab versus standard care)	Yes (adalimumab or infliximab versus standard care)	Yes (ustekinumab versus standard care)	Yes (vedolizumab versus standard care)	Yes induction of remission compared nine treatment strategies, and six treatments were compared for maintenance of remission)
1.3. Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context?	Yes (NHS perspective)	Yes (NHS perspective)	Yes (NHS perspective)	Yes (NHS perspective)	Yes (NHS and personal social services perspective)

2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)?	Yes	Yes	Yes	Yes	Yes
2.1. Were there any important alternatives omitted?	No	No	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	No	No	No	No	No
3. Was the effectiveness of the programme or services established?	Yes	Yes	Yes	Yes	Yes
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	Yes (ACCENT trial and Targan trial)	Yes (effectiveness for infliximab and adalimumab treatment were derived from ACCENT I and CHARM, respectively)	Yes (induction treatment assessed in UNITI-1, UNITI-2, CERTIFI and maintenance treatment assessed in IM-UNITI)	Yes (GEMINI II, GEMINI III)	Yes (various RCTs, treatment effectiveness of interventions for induction of and maintenance of remission derived from network meta-analyses)
3.2. Was effectiveness established through an overview of clinical studies?	Yes	Yes	Yes	Yes	Yes

3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Yes (lack of observational data on the history of patients treated with infliximab has led to the reliance on data from one study, which involves two major assumptions: (i) QALY gains are reduced for people who revert to the more severe states, and (ii) the time patients spend in the various health states can be aggregated over their lifetimes, which, given the average age used, implies gains spread over about 40 years, which is a considerable extrapolation of the	Yes (different assumptions in key studies made: Arsenau *et al*., Clark *et al*. and in the adalimumab model).	Yes (structural assumptions in the model inconsistent with UK clinical practice)	Yes (structural assumptions in the model influence outcomes)	No

	benefits of infliximab).				
4. Were all the important and relevant costs and consequences for each alternative identified?	Yes	Yes	Yes	Yes	Yes
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of patients and third-party payers. Other viewpoints may also be relevant depending upon the particular analysis.)	Yes	Yes	Unclear	Yes	Yes
4.3. Were the capital costs, as well as operating costs, included?	Unclear	Yes	Yes	Unclear	Yes
5. Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life years)?	Yes	Yes	Yes	Yes	Yes

5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	No	Unclear	No	No	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	Unclear	No	Unclear	No	No
6. Were the cost and consequences valued credibly?	Yes	Yes	Yes	Yes	Yes
6.1. Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers' views and health professionals' judgements)	Yes	Yes	Yes	Yes	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Unclear	Unclear	Unclear	No	Unclear
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	Unclear	Unclear	Yes	Unclear	Unclear
6.4. Was the valuation of consequences appropriate for the question posed (i.e. has the	Yes	Yes	Yes	Yes	Yes

appropriate type or types of analysis – cost-effectiveness, cost-benefit, cost-utility – been selected)?					
7. Were costs and consequences adjusted for differential timing?	Unclear	Unclear- UK study from 2004 calculated the costs of CD. Unknown if adjusted.	Unclear	Unclear	Unclear
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	Yes (6 % for costs and 1.5% benefits respectively)	No	Yes (3.5% for costs and benefits)	Yes (3.5% for costs and benefits)	Yes (3.5% for costs and benefits)
7.2. Was there any justification given for the discount rate used?	No	No discounting required for 1-year time horizon.	Yes (in accordance with NICE reference case)	Yes (in accordance with NICE guidance)	Yes (in accordance with NICE guidance)
8. Was an incremental analysis of costs and consequences of alternatives performed?	Yes	Yes	Yes	Yes	Yes
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes
9. Was allowance made for uncertainty in the estimates of costs and consequences?	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)

9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	NA	NA	NA	NA	NA
9.2. If a sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	Yes	Yes	Yes	Yes	Yes
9.3. Were the study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences)?	Yes (chronic active model was highly sensitive to rate of 'flare' for episodic treatment. The flare rate chosen was 10%, which seemed reasonable based on clinical opinion. If more frequent flare was seen, then costs increased substantially)	Yes	Yes (model was sensitive to the duration of treatment and the analytic time horizon)	Yes	Yes

10. Did the presentation and discussion of study results include all issues of concern to users?	Yes	Yes	Yes	Yes	Yes
10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion?	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	Yes	Yes	Yes	No	No

10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	No	Yes	Yes	Yes	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	No	Yes	Yes	Yes	Yes
10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	No	No	No	No	No

Criteria	Study					
	Marchetti 2013	**Freeman 2016**	**Saito 2013**	**Bodger 2009**	**Loftus 2009**	**Lindsay 2008**
1. Was a well-defined question posed in answerable form?	Yes	Yes	Yes	Yes	Yes	Yes
1.1. Did the study examine both costs and effects of the service(s) or programme(s)?	Yes	Yes	Yes	Yes	Yes	Yes
1.2. Did the study involve a comparison of alternatives?	Yes (top-down versus step up)	Yes (monitoring of serum anti-TNF-alpha antibody levels versus no testing/standard care)	Yes (infliximab monotherapy versus infliximab plus azathioprine)	Yes (infliximab and adalimumab for versus standard care)	Yes (adalimumab versus non biologic therapies in maintenance of CD)	Yes (infliximab versus standard care)
1.3. Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context?	Yes (Italian Healthcare System)	Yes (NHS perspective)	Yes (NHS perspective)	Yes (NHS perspective)	Yes (NHS perspective and from the perspective of the social decision maker)	Yes (NHS perspective)
2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)?	Yes	Yes	Yes	Yes	Yes	Yes

2.1. Were there any important alternatives omitted?	No	No	No	No	No	No
2.2. Was (should) a do-nothing alternative be considered?	No	No	No	No	No	No
3. Was the effectiveness of the programme or services established?	Yes	Yes	Yes	Yes	Yes	Yes
3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	Yes (trial by D'Haen's *et al*.)	Yes (ACCENT I, ACCENT II)	Yes (ACCENT I, SONIC, Lemann trial)	Yes (CHARM and ACCENT I)	Yes (CHARM and CLASSIC 1)	Yes (ACCENT I, ACCENT II, Targan trial, Present trial)
3.2. Was effectiveness established through an overview of clinical studies?	Yes	Yes	Yes	Yes	Yes	Yes

3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	Unclear	Yes	Yes	Yes (surgical rates based on observational data)	Yes	Yes
4. Were all the important and relevant costs and consequences for each alternative identified?	Yes	Yes	Yes	Yes	Yes	Yes
4.1. Was the range wide enough for the research question at hand?	Yes	Yes	Yes	Yes	Yes	Yes
4.2. Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of patients and third-party payers. Other viewpoints	Yes	Yes	Yes	Yes	Yes	Yes

may also be relevant depending upon the particular analysis.)						
4.3. Were the capital costs, as well as operating costs, included?	Yes	Yes	Yes	Yes	Yes	Yes
5. Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life years)?	Yes	Unclear	Yes	Yes	Yes	Yes
5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	Unclear	Unclear	Unclear	Unclear	No	No
5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	No	No	Unclear	No	No	Yes (true placebo effect could not be estimated from the ACCENT trials)

6. Were the cost and consequences valued credibly?	Yes	Yes	Yes	Yes	Yes	Yes
6.1. Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers' views and health professionals' judgements)	Yes	Yes	Yes	Yes	Yes	Yes
6.2. Were market values employed for changes involving resources gained or depleted?	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?	Unclear	Unclear	Unclear	Unclear	No	Yes
6.4. Was the valuation of consequences appropriate for the question posed (i.e. has the appropriate type or types of analysis – cost-effectiveness,	Yes	Yes	Yes	Yes	Yes	Yes

cost-benefit, cost-utility – been selected)?						
7. Were costs and consequences adjusted for differential timing?	Unclear	Yes	Yes	Yes	Yes	Yes
7.1. Were costs and consequences that occur in the future 'discounted' to their present values?	Yes (3.5% for costs and benefits)	Yes (3.5% for costs and benefits)	NA	Yes (3.5% for costs and benefits)	Yes (3.5% for costs and benefits)	Yes (3.5% for costs and benefits)
7.2. Was there any justification given for the discount rate used?	Yes (international guidelines)	Yes (NICE reference case)	NA	Yes (NICE reference case)	No	Yes (NICE reference case)
8. Was an incremental analysis of costs and consequences of alternatives performed?	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)
8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Yes	Yes	Yes	Yes	Yes	Yes
9. Was allowance made for uncertainty in the estimates of costs and consequences?	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)	Yes (sensitivity analysis)

9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed?	NA	NA	NA	NA	NA	NA
9.2. If a sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)?	Yes	Yes	Yes	Yes	Yes	No
9.3. Were the study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences)?	Yes	Yes (sensitive to a 10 percent increase in the utility value for patients who regain response in both reflex and concurrent testing)	Yes (analyses showed that the quality of life utility associated with nonresponding active disease was the most influential parameter on the cost-effectiveness of the therapies)	Yes	Yes	Yes (in OWSA, because of the weight-based dosing of infliximab, patient weight had the most impact on the ICER)
10. Did the presentation and discussion of study results include all issues of concern to users?	Yes	Yes	Yes	Yes	Yes	Yes

10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion?	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)	Yes (cost per QALY gained)
10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?	No	Yes	Yes	Yes	Yes	Yes
10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups?	Yes	Yes	Yes	Yes	Yes	Yes
10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?	No	Yes	Yes-	Yes	No	Yes

10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	No	No	No	No	No	No

9.7 Appendix 7. Search strategies and list of excluded studies for literature review to inform estimates of clinical effectiveness of treatments

9.7.1 Search strategies

Database searched: Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily and Versions(R) **Database searched from inception through to 14 June 2019**		
#	**Terms**	**Hits**
1	Crohn Disease/	37169
2	Crohn*.mp	53162
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	42992
4	Inflammatory bowel diseases/	20151
5	IBD.mp.	22462
6	Inflammatory bowel disease*.mp.	48138
7	or/1-6	84595
8	(top-down or top down or step-up or step up).ti,ab.	15774
9	7 and 8	191

Database searched: EMBASE **Database searched from inception through to 14 June 2019**		
#	**Terms**	**Hits**
1	Exp Crohn Disease/	83531
2	Crohn*.mp	94568
3	((Crohn$ adj2 (disease or syndrome)) or regional enteritis).tw.	68633
4	Exp Inflammatory bowel disease/	134801
5	IBD.mp.	46227
6	Inflammatory bowel disease*.mp.	79562
7	or/1-6	168160
8	(top-down or top down or step-up or step up).ti,ab.	18369
9	7 and 8	472

Database searched: Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR) **Database searched from inception through to 14 June 2019**		
#	**Terms**	**Hits**
1	Crohn:ti,ab,kw	4482
2	MeSH: [Inflammatory bowel diseases] explode all trees	2889
3	IBD:ti,ab,kw	1738
4	"Inflammatory bowel disease":ti,ab,kw	2650
5	#1 or #2 or #3 or #4	7295
6	"top-down" or "top down" or "step-up" or "step up":ti,ab,kw	1194
7	#5 or #6	43

9.7.2 List of excluded studies with reason for exclusion

Study	Reason for exclusion
Chau 2015[155]	Focuses on treatment with biological therapy rather than SU versus TD.
Colombel 2018[156]	Focuses on treatment with biological therapy rather than SU versus TD.
Fan 2014[157]	RCT included in chosen SR.
Hirschmann 2017[158]	Not SR.
Hommes 2006[159]	Not SR.
Hutfless 2014[160]	Book chapter.
Katz 2007[161]	Not SR.
Kuznar 2013[162]	Not SR.
Lee 2017[146]	Not SR.
Meier 2009[163]	Not SR.
Parkes 2018[164]	Not SR.
Peyrin-Biroulet 2018[165]	Not SR.
Sucong 2013[166]	Not SR.
Xiao 2012[167]	Not SR.
Abbreviations: RCT, randomised controlled trial; SR, systematic review; SU, step up; TD, top down.	

9.8 Appendix 8. Effectiveness of induction and maintenance therapies

Reasons for exclusion of studies identified from TA352[118] (vedolizumab) and TA456[119] (ustekinumab) from the EAG's analyses are presented in Table 46. The EAG notes that the key differences across analyses carried out by the EAG and those presented in TA352 and TA456 are exclusion by the EAG

of the study carried out by Targan and colleagues (single dose of infliximab 5 mg administered) and inclusion of subgroup data from the anti-TNF naïve subgroup of the study reported by Watanabe and colleagues (Table 46). In addition, the EAG notes that studies of ustekinumab were not included in TA352 whereas they were included in both TA456 and the EAG analyses.

Table 46. Inclusion and exclusion decisions for studies identified from TA352 and TA456.

Study Name	Intervention	Induction EAG analysis	Notes	Maintenance EAG analysis	Notes
Studies from TA352[118]					
ACCENT I[126]	Infliximab	N/A	–	Included	Data available on use of infliximab in anti-TNF naïve patients at the start of induction therapy
CHARM[168]	Adalimumab	N/A	–	Excluded	47.7% patients in the study had a history of anti-TNF use before the induction study. Subgroup data were not available for maintenance treatment of those who were anti-TNF-naïve at induction
CLASSIC-I[120]	Adalimumab	Included	Data extracted for anti-TNF naïve subgroup for adalimumab 160/80 mg	N/A	–
CLASSIC-II[169]	Adalimumab	N/A	–	Excluded	All patients required to be in remission at start of maintenance treatment rather than have achieved a set level of response to induction therapy: other studies specify a cut off for response
EXTEND[170]	Adalimumab	Excluded	46.9% of patients had prior anti-TNF exposure and	Excluded	Maintenance adalimumab arm includes patients with non-response from

			subgroup data were not available for the anti-TNF naïve patients. It was noted that prior exposure did not include patients with primary non-response		induction (CDAI did not decrease by 70 or more). In addition, 46.9% of people had received prior anti-TNF, but it is acknowledged that they were not classed as "primary nonresponse"
GAIN[171]	Adalimumab	Excluded	Prior failure or intolerance to infliximab was required therefore the patients were not anti-TNF naïve	N/A	–
Watanabe 2012[121]	Adalimumab	Included	Data extracted for adalimumab 160/80 mg from the anti-TNF naïve subgroup	Excluded	52% of patients in the study had a history of anti-TNF use before entering the induction study. Data were not available for maintenance therapy in the subgroup of anti-TNF-naïve patients.
Targan 1997[172]	Infliximab	Excluded	Single dose of infliximab which is not standard protocol or in keeping with other drugs in the analysis; typically, more than one dose would be expected to be given for induction therapy	N/A	–
GEMINI II[123]	Vedolizumab	Included	Data on vedolizumab	Included	Data on vedolizumab.
GEMINI III[125]	Vedolizumab	Included	Data on vedolizumab	N/A	–
Additional studies from TA456[119]					
CERTIFI[124]	Ustekinumab	Included	Data were available for ustekinumab from the prior anti-TNF failure subgroup.	N/A	The study had some maintenance end points but they were assessed at 22 weeks and not 52

			Note: data from the 6 mg/kg arm have been used as this dose was deemed to be the most similar to the licensed dose		weeks as in other studies and were therefore excluded from analyses of maintenance
UNITI-1[119, 122]	Ustekinumab	Included	Data extracted on ustekinumab for the subgroup of those failing prior anti-TNF. Note: data from the 6 mg/kg arm have been used as this dose was deemed to be the most similar to the licensed dose	N/A	–
UNITI-2[119, 122]	Ustekinumab	Excluded	Less than 40% of patients had a history of prior anti-TNF treatment and the study inclusion criteria restricted the patients who had previously received one or more TNF antagonists to those who had not had unacceptable side effects and had not met the criteria for primary or secondary nonresponse to treatment.	N/A	–
IM-UNITI[122]	Ustekinumab	N/A	–	Included	Data extracted on ustekinumab for the subgroup of those failing prior anti-TNF
Abbreviations: CDAI, Crohn's Disease Activity Index; N/A, not applicable; TNF, tumour necrosis factor.					