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

Erratum to the Diagnostic Assessment Report

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

BMJ Technology Assessment Group

This document contains errata in respect of the EAG diagnostic assessment report (DAR) in response to factual inaccuracies raised by stakeholders. The table below lists the pages to be replaced in the original document and the nature of the change. The main changes in the DAR relate to a correction made to the economic model. References to the EAG's updated results in the table below correspond to the corrections made in the economic model.

Page No.	**Change**
vi	The EAG's results were updated. The text "…with an additional cost of £9,526 and a QALY loss of 0.06." was replaced with "…with an additional cost of £7,636 and a QALY loss of 0.10."
xiv	In Title of Figure 2, "Glycominds International" amended to "Glycominds".
xxvi	The EAG's results were updated. The text "…with an additional cost of £9,526 and a QALY loss of 0.06." was replaced with "…with an additional cost of £7,636 and a QALY loss of 0.10."
10	Two instances of "Glycominds International" amended. First citation amended to "Glycominds, LLC (hereafter referred to as Glycominds)", and second to "Glycominds".
12	"Glycominds International" amended to "Glycominds".
16	"Glycominds International" amended to "Glycominds".
26	Two instances of "Glycominds International" amended to "Glycominds".
36	"Glycominds International" amended to "Glycominds".
37	"Glycominds International" amended to "Glycominds".
128–135	The EAG's results were updated in the text and in Tables 38, 39, 40, 41 and 42. Figures 39 -41 were also updated.
142–153	The EAG's results were updated in the text, tables and figures.
157–158	The EAG's results were updated in the text.
167	"Glycominds International" amended to "Glycominds".
168	Two instances of "Glycominds International" amended to "Glycominds".

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 £7,636 and a QALY loss of 0.10.

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.

LIST OF FIGURES

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 £7,636 and a QALY loss of 0.10.

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

IBDX® (Glycominds LLC, hereafter referred to as Glycominds) 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 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;

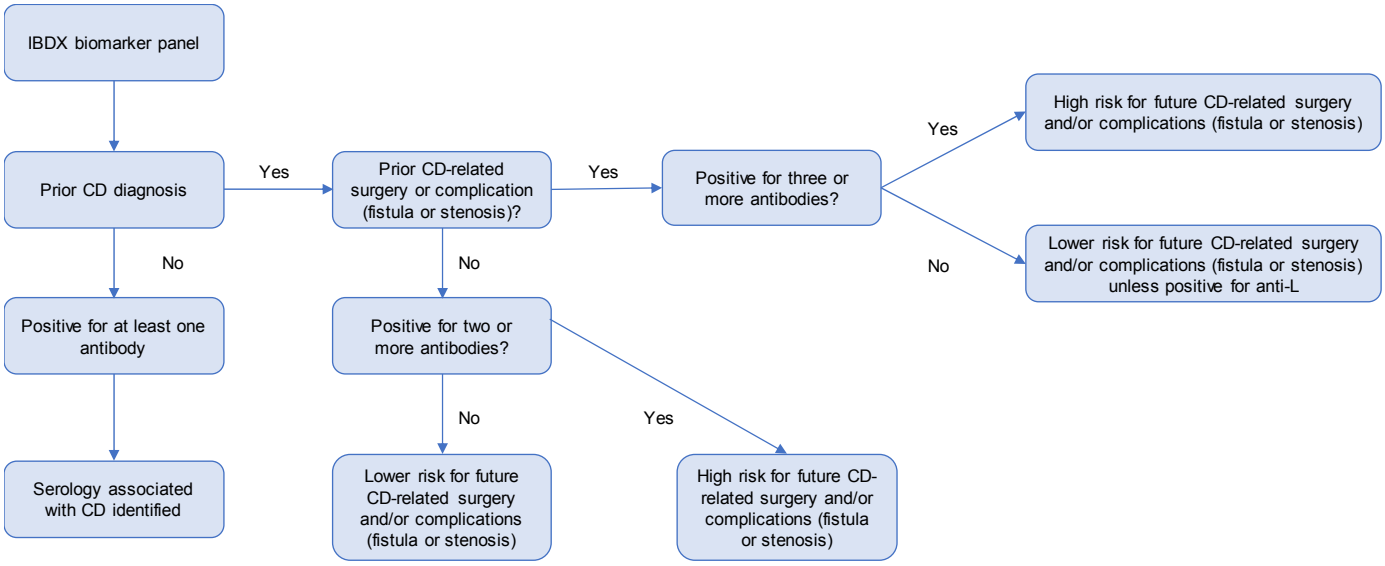
Table 1. Cut-off values for individual IBDX ELISA kits[43]

	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[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);

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) 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

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

Glycominds 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).

3.2.2.2 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.

3.3.3 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 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 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

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 £7,636and a QALY loss of 0.10.

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

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard of Care	£207,857	15.96	–	–	–
PredictSURE IBD™	£215,493	15.85	£7,636	-0.10	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.

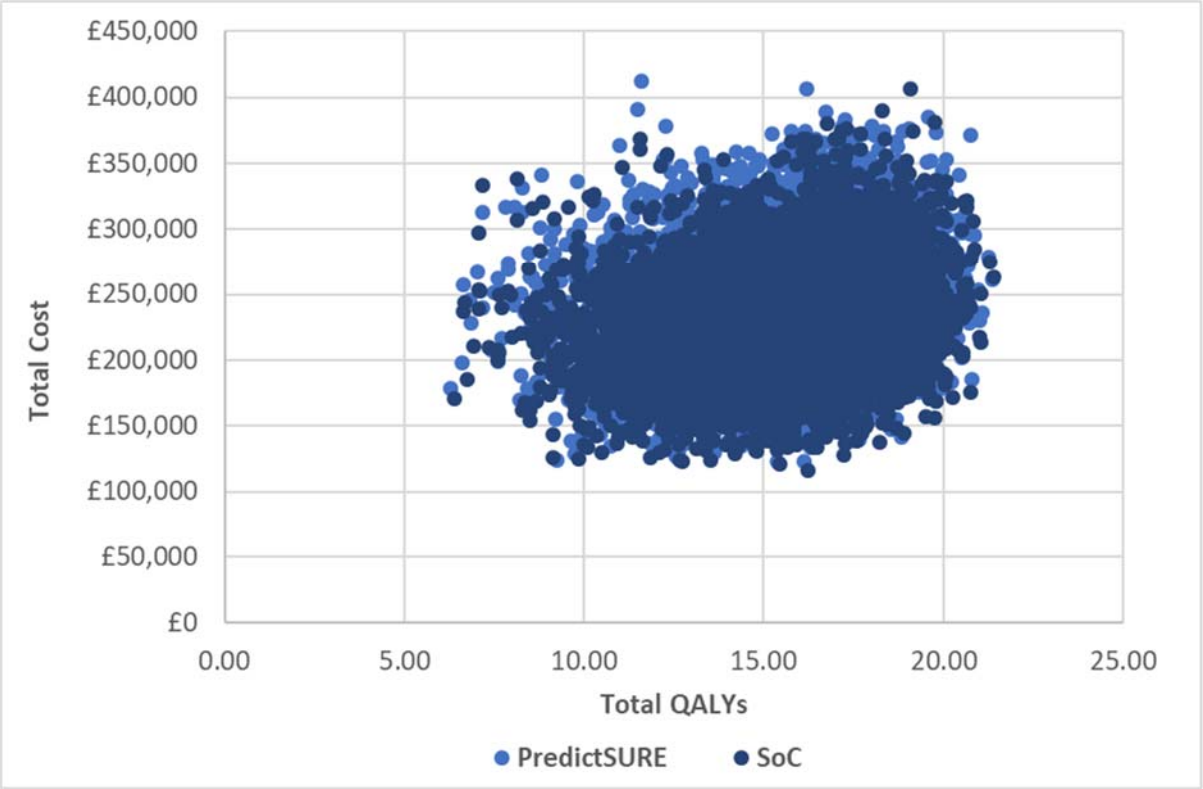
For cost estimates, gamma distributions were applied using 20% of the mean value to estimate standard errors, while for probabilities and utilities, beta distributions were applied; again, with an assumption that the standard errors are 20% of the mean estimate. A summary of the full parameterisations 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 21% against 79% for the SC arm.

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

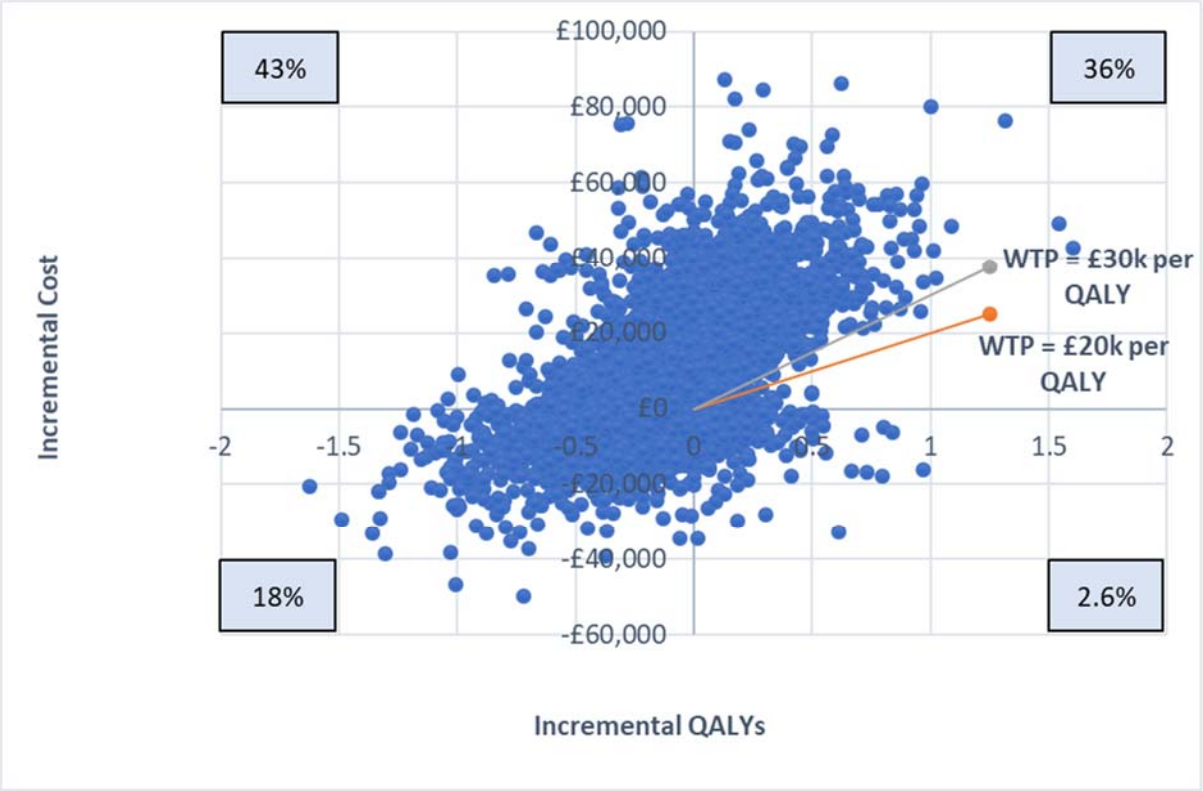
Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Standard of Care	£228,609	15.72	–	–	–
PredictSURE IBD™	£238,920	15.66	£10,312	-0.06	Dominated
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year.					

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



Abbreviations in figure: SoC, standard of car

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.

Table 40. Results of scenario analyses

Intervention	Total Costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Scenario 1: Applying IBDX cost					
Standard of Care	£207,857	15.96	–	–	–
IBDX	£214,590	15.85	£6,733	-0.10	Dominated
Scenario 2: Applying utilities from TA456					
Standard of Care	£207,857	15.68	–	–	–
PredictSURE IBD™	£215,493	15.57	£7,636	-0.11	Dominated
Scenario 3: Applying induction vectors and transition probabilities based on TA352 studies					
Standard of Care	£207,587	15.95	–	–	–
PredictSURE IBD™	£215,294	15.85	£7,707	-0.10	Dominated
Scenario 4: Applying equivalent TTS curves for top down and step up					
Standard of Care	£207,857	15.96	–	–	–
PredictSURE IBD™	£216,059	15.85	£8,202	-0.11	Dominated
Scenario 5: Removing Ara & Brazier utility adjustment					
Standard of Care	£207,857	16.03	–	–	–
PredictSURE IBD™	£215,493	15.92	£7,636	-0.11	Dominated
Scenario 6: Use the minimum induction period from the treatment class to estimate induction costs					
Standard of Care	£201,623	15.93	–	–	–
PredictSURE IBD™	£208,901	15.82	£7,278	-0.11	Dominated
Scenario 7: 100% of high-risk patients who receive SU do not respond to IM treatment					
Standard of Care	£214,678	15.85	–	–	–
PredictSURE IBD™	£215,493	15.85	£815	-0.0001	Dominated
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	£207,857	15.96	–	–	–
IBDX®	£214,590	15.85	£6,733	-0.10	Dominated
PredictSURE IBD™	£215,493	15.85	£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	£190,628	15.96	–	–	–
PredictSURE IBD™	£196,974	15.85	£6,346	-0.10	Dominated
Biologic discount: 50%					
Standard of Care	£173,399	15.96	–	–	–
PredictSURE IBD™	£178,454	15.85	£5,055	-0.10	Dominated
Biologic discount: 75%					
Standard of Care	£156,169	15.96	–	–	–
PredictSURE IBD™	£159,935	15.85	£3,765	-0.10	Dominated
Anti-TNF discount: 25%					
Standard of Care	£199,028	15.96	–	–	–
PredictSURE IBD™	£206,898	15.85	£7,870	-0.10	Dominated
Anti-TNF discount: 50%					
Standard of Care	£190,198	15.96	–	–	–
PredictSURE IBD™	£198,302	15.85	£8,104	-0.10	Dominated
Anti-TNF discount: 75%					
Standard of Care	£181,369	15.96	–	–	–
PredictSURE IBD™	£189,707	15.85	£8,338	-0.10	Dominated
Biologic and Anti-TNF discount: 25%					
Standard of Care	£181,798	15.96	–	–	–
PredictSURE IBD™	£188,378	15.85	£6,580	-0.10	Dominated
Biologic and Anti-TNF discount: 50%					
Standard of Care	£155,740	15.96	–	–	–
PredictSURE IBD™	£161,263	15.85	£5,523	-0.10	Dominated
Biologic and Anti-TNF discount: 75%					
Standard of Care	£129,682	15.96	–	–	–
PredictSURE IBD™	£134,149	15.85	£4,467	-0.10	Dominated
Abbreviations: ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year.					

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.

5.1.8 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 £67,741 per QALY gained, with PredictSURE IBD™ being more costly than SC but generating a QALY gain of 0.11. 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 £44,103 for SC vs PredictSURE IBD™, meaning that the diagnostic tool is less expensive than SC (by £4,621) but also less effective (0.10 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 £29,932 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 (-£4,621) albeit less effective (-0.10).

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 £60,056 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 79%. 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 £28,192 and £43,286, 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 18a (where PredictSURE 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 79%, then becoming dominated from below 43%. Figure 43 shows the resulting final ICERs, and the drastic variation in these at 79% non-response, when the incremental QALYs become negative.

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

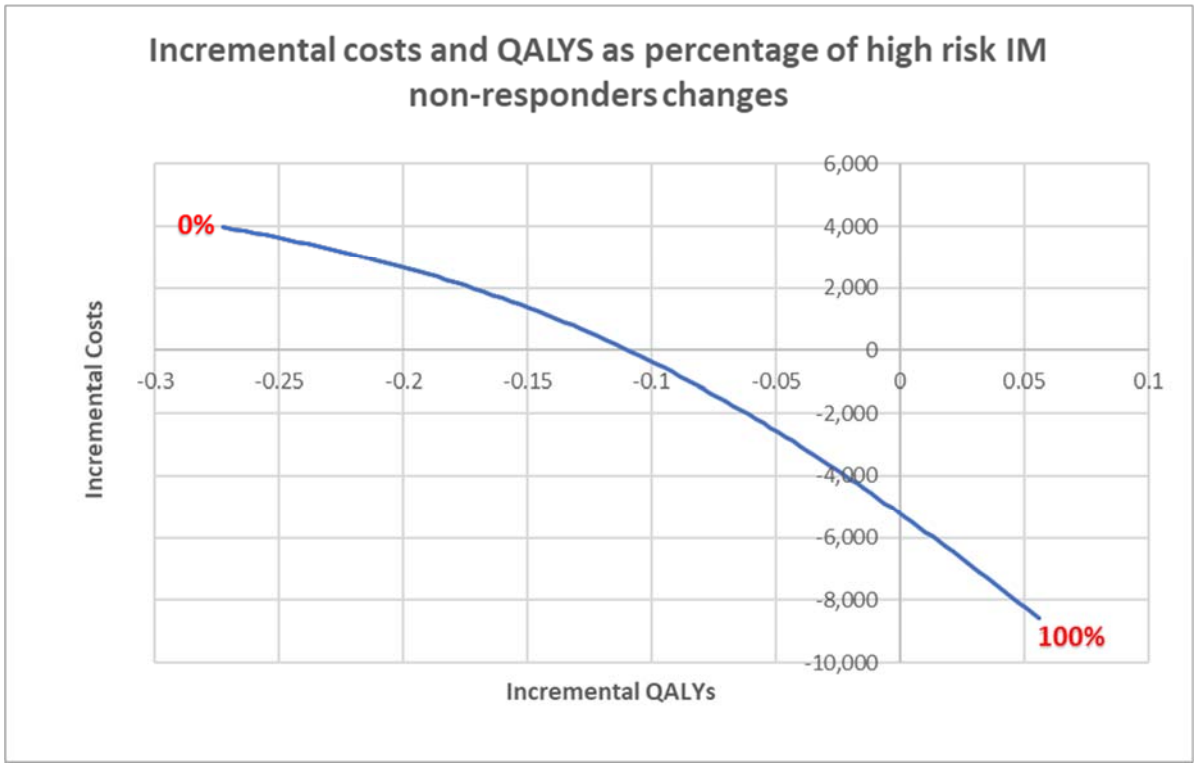
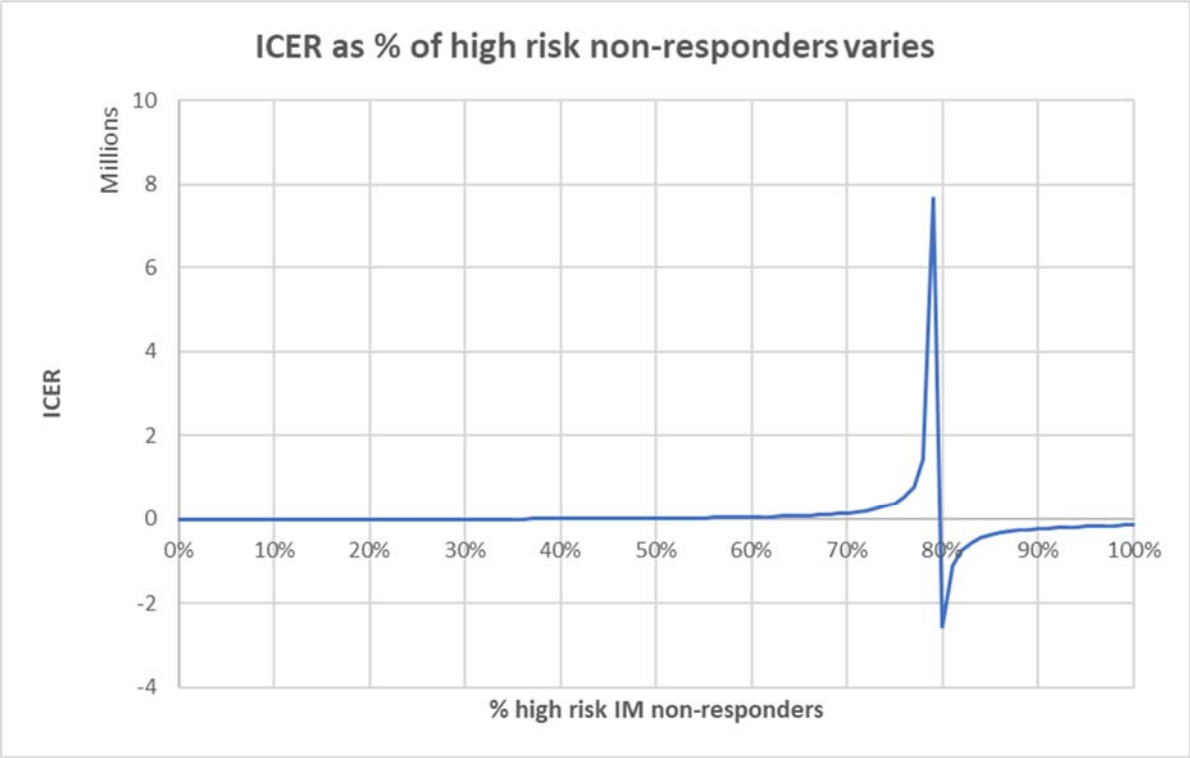


Figure 43. Resulting ICERs as the percentage of high risk IM non-responders varies



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 costs associated with biologic treatment (through assuming different rates of mucosal healing leading to remission); the ICER ranged from dominant to £47,842 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 79% (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	£207,857	15.96	–	–	–
PredictSURE IBD™	£215,516	16.07	£7,659	0.11	£67,741
Scenario 5.2.2 a i - Assuming half of the base case relative effectiveness for TD on TTE for further steps					
Standard of Care	£204,720	15.90	–	–	–
PredictSURE IBD™	£213,724	15.82	£9,004	-0.08	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	£200,403	15.82	–	–	–
PredictSURE IBD™	£210,640	15.77	£10,237	-0.05	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	£204,720	15.90	–	–	–
PredictSURE IBD™	£212,848	15.81	£8,128	-0.09	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	£200,403	15.82	–	–	–
PredictSURE IBD™	£208,949	15.74	£8,546	-0.08	Dominated
Scenario 5.2.3 a i – Assuming discontinuation of biologic treatment for 76% TD; 40% SU.					
Standard of Care	£186,932	15.96	–	–	–
PredictSURE IBD™	£182,311	15.85	-£4,621	-0.10	£44,103*
Scenario 5.2.3 a ii - Assuming discontinuation of biologic treatment for 76% TD; 76% SU.					
Standard of Care	£168,099	15.96	–	–	–
PredictSURE IBD™	£173,362	15.85	£5,263	-0.10	Dominated
Scenario 5.2.3 a iii - Assuming discontinuation of biologic treatment for 40% TD; 40% SU.					
Standard of Care	£186,932	15.96	–	–	–

PredictSURE IBD™	£193,319	15.85	£6,387	-0.10	Dominated
Scenario 5.2.3 b - Assuming discontinuation of biologic treatment for 100% TD; 100% SU.					
Standard of Care	£155,544	15.96	–	–	–
PredictSURE IBD™	£160,058	15.85	£4,514	-0.10	Dominated
Scenario 5.2.4 a – Assuming surgery as last treatment step					
Standard of Care	£209,767	16.22	–	–	–
PredictSURE IBD™	£217,480	16.13	£7,713	-0.09	Dominated
Scenario 5.2.4 b – Removing surgery from the model					
Standard of Care	£203,768	15.97	–	–	–
PredictSURE IBD™	£211,987	15.87	£8,219	-0.11	Dominated
Scenario 5.2.5 a (Scenario 5.2.1 + Scenario 5.2.3 a i)					
Standard of Care	£186,932	15.96	–	–	–
PredictSURE IBD™	£180,063	16.07	-£6,869	0.11	Dominant
Scenario 5.2.5 b (Scenario 5.2.1 + Scenario 5.2.3 a ii)					
Standard of Care	£168,099	15.96	–	–	–
PredictSURE IBD™	£171,483	16.07	£3,384	0.11	£29,932
Scenario 5.2.5 c (Scenario 5.2.1 + Scenario 5.2.3 a iii)					
Standard of Care	£186,932	15.96	–	–	–
PredictSURE IBD™	£192,341	16.07	£5,409	0.11	£47,842
Scenario 5.2.6 a (Scenario 5.2.2 a ii + Scenario 5.2.3 a i)					
Standard of Care	£180,487	15.82	–	–	–
PredictSURE IBD™	£177,932	15.77	-£2,555	-0.05	£50,936*
Scenario 5.2.6 b (Scenario 5.2.2 a ii + Scenario 5.2.3 a ii)					
Standard of Care	£162,563	15.82	–	–	–
PredictSURE IBD™	£169,411	15.77	£6,848	-0.05	Dominated
Scenario 5.2.6 c (Scenario 5.2.2 a ii + Scenario 5.2.3 a iii)					
Standard of Care	£180,487	15.82	–	–	–
PredictSURE IBD™	£188,940	15.77	£8,453	-0.05	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	£207,282	15.71	–	–	–
PredictSURE IBD™	£210,640	15.77	£3,357	0.06	£60,056

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	£186,521	15.71	–	–	–
PredictSURE IBD™	£177,932	15.77	-£8,588	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	£167,835	15.71	–	–	–
PredictSURE IBD™	£169,411	15.77	£1,576	0.06	£28,192
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	£186,521	15.71	–	–	–
PredictSURE IBD™	£188,940	15.77	£2,420	0.06	£43,286
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

5.3.9 One-way sensitivity analysis

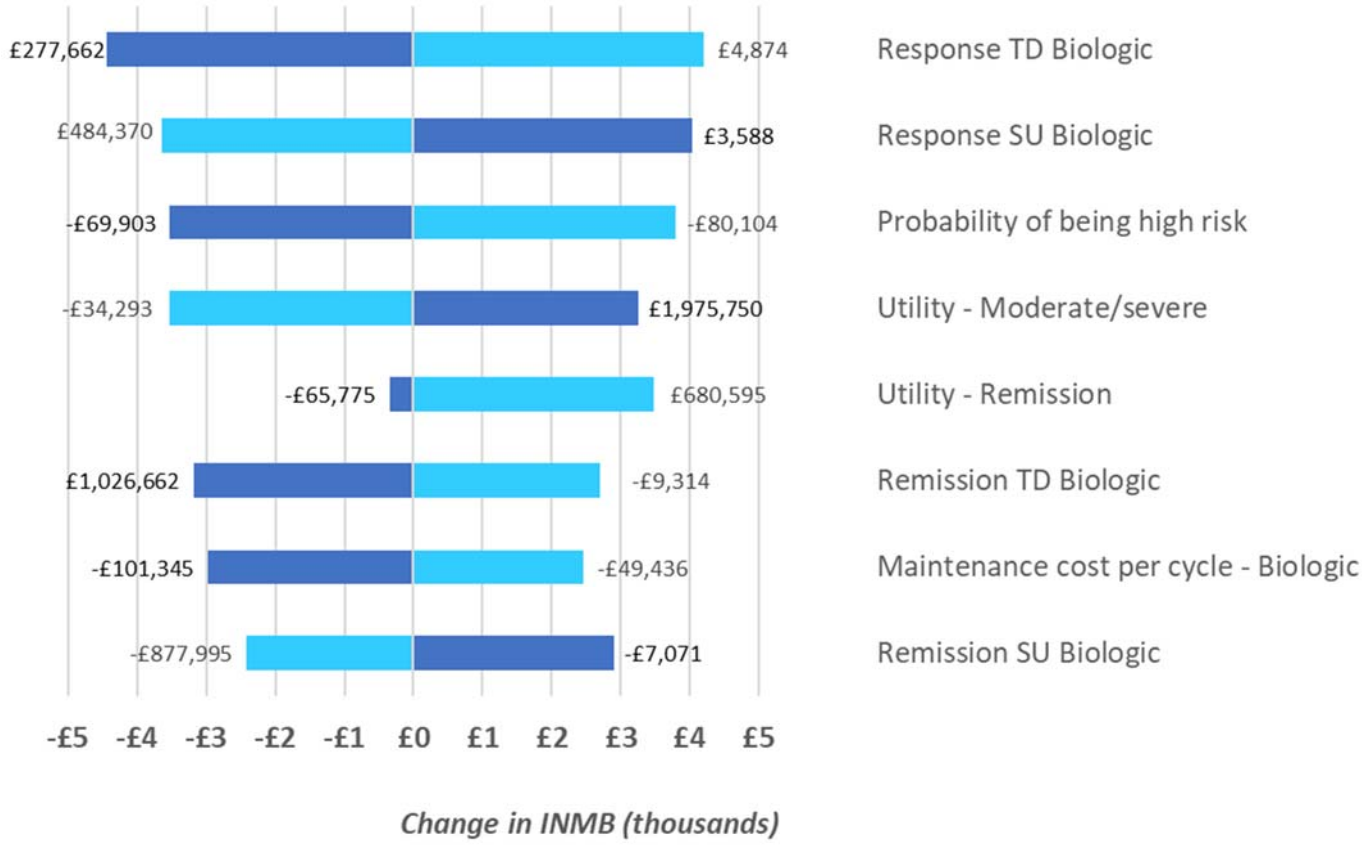
The EAG conducted a number of deterministic one-way sensitivity analyses around the model inputs as described in Table 45. Figure 46 ranks the model key drivers by their impact on the incremental net monetary benefit (INMB) of PredictSURE-IBD™ compared to standard care, based on a willingness-to-pay (WTP) threshold of £30,000 per QALY. The lower and upper bounds of each parameter input 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. The inputs with the highest impact on the model results were the response to biologic treatments in both the TD and the SU arms.

Table 45. Inputs and results of OWSAs

Model Parameter	Lower bound	Upper bound	Lower ICER	Upper ICER
Age	21.3	48.7	-£68,002	-£86,923
Crohn's disease expected body weight	43.8	100.2	-£74,787	-£70,567
Proportion of males	0.2280	0.5220	-£72,402	-£73,386
Probability of being high risk	0.3496	0.8004	-£80,104	-£69,903
Proportion on infliximab in anti-TNF biologics class	0.2432	0.5568	-£72,861	-£72,902
Proportion on vedolizumab in non-anti-TNF biologics class	0.3040	0.6960	-£70,349	-£75,413
Proportion on azathioprine for immunomodulators	0.4864	1.0000	-£73,370	-£72,641
Proportion of 6-mercaptopurine for immunomodulators	0.0608	0.1392	-£72,913	-£72,861
Proportion of anti-TNF with IM bundle	0.1824	0.4176	-£72,921	-£72,836
Proportion of Biologics with IM bundle	0.1216	0.2784	-£72,792	-£72,984
Response TD Biologic	0.1918	0.4390	£4,874	£277,662
Remission TD Biologic	0.0795	0.1821	-£9,314	£1,026,662
Response TD anti-TNF	0.1565	0.3583	-£59,548	-£110,878
Remission TD anti-TNF	0.2231	0.5108	-£55,244	-£148,135
Response SU Biologic	0.1918	0.4390	£484,370	£3,588
Remission SU Biologic	0.0795	0.1821	-£877,995	-£7,071
Response SU anti-TNF	0.1565	0.3583	-£123,227	-£40,429
Remission SU anti-TNF	0.2231	0.5108	-£160,422	-£32,784

Response SU IM	0.1380	0.3160	-£75,825	-£69,471
Remission SU IM	0.0950	0.2176	-£77,255	-£68,744
Probability of death following surgery	0.0009	0.0021	-£72,260	-£73,646
Health state cost - Remission	£10	£23	-£73,296	-£72,376
Health state cost - Mild	£16	£37	-£73,425	-£72,221
Health state cost - Moderate/severe	£74	£170	-£66,698	-£80,388
Health state cost - No response	£74	£170	-£73,516	-£72,110
Induction cost per cycle - Anti TNF	£927	£2,123	-£72,368	-£73,503
Induction cost per cycle - Biologic	£940	£2,151	-£71,130	-£75,007
Induction cost per cycle - Immunomodulator	£3	£6	-£72,923	-£72,829
Maintenance cost per cycle - Anti TNF	£326	£747	-£78,669	-£65,853
Maintenance cost per cycle - Biologic	£399	£914	-£49,436	-£101,345
Maintenance cost per cycle - Immunomodulator	£7	£17	-£73,717	-£71,866
IV administration first attendance	£121	£277	-£72,683	-£73,122
IV administration follow-up	£129	£295	-£67,441	-£79,486
Cost of surgery	£5,359	£12,268	-£75,004	-£70,303
Utility - Remission	0.50	1.00	£680,595	-£65,775
Utility - Mild	0.44	1.00	-£256,508	-£54,680
Utility - Moderate/severe	0.35	0.79	-£34,293	£1,975,750
Disutility for surgery	0.02	0.06	-£73,231	-£72,463

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



Abbreviations in figure: ICER, incremental cost effectiveness ratio; INMB, incremental net monetary benefit; OWSA, one-way sensitivity analysis; SU, step up; TD, top down.

Note: The bars in the graph represent the change in INMB and the respective ICERs are presented at both ends of the bars. Light blue bars represent the lower bound of the parameter changed while dark blue bars represent the upper bound of the parameter changed.

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 £7,636 and a QALY loss of 0.10.

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 £47,842 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 SD 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 79% (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.

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