Risankizumab for previously treated moderately to severely active Crohn's disease [ID3986]
A Single Technology Appraisal

Produced by	**Peninsula Technology Assessment Group (PenTAG)** **University of Exeter Medical School** **South Cloisters** **St Luke's Campus** **Heavitree Road** **Exeter** **EX1 2LU**
Authors	**Maxwell S. Barnish**[1] **Jae Naik**[2] **Will Sullivan**[2] **Amanda Brand**[1] **Emma Hook**[2] **Justin Matthews**[1] **Mark Edmondson-Jones**[2] **Sophie Robinson**[1] **Jonathan Digby-Bell**[3] **G.J. Melendez-Torres**[1] [1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter [2] Delta Hat Limited, Nottingham [3] Royal Devon and Exeter Hospital, Exeter
Correspondence to	Dr Maxwell S. Barnish 3.09 South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; m.s.barnish@exeter.ac.uk
Date completed	01/09/2022
Source of funding	This report was commissioned by the NIHR Systematic Reviews Programme as project number 13/56/25

Declared competing interests of the authors	None.
Acknowledgments	The authors acknowledge the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe (both PenTAG).
Rider on responsibility for document	The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.
This report should be referenced as follows	Barnish MS, Naik J, Sullivan W, Brand A, Hook E, Matthews J, Edmondson-Jones M, Robinson S, Digby-Bell J, Melendez-Torres GJ. Risankizumab for previously treated moderately to severely active Crohn's disease [ID3986] A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2022.

Author Contributions:

Maxwell S. Barnish	Project lead, lead for the ERG's appraisal of clinical evidence, drafted background and clinical sections of report, writing and editorial input.
Jae Naik	Health economic lead for the project. Led health economic review of the company's evidence submissions, implemented ERG adaptations to the company model, led writing of economic sections of this report.
Will Sullivan	Contributed to the critical appraisal of the cost effectiveness evidence and writing of the economic sections of the report.
Amanda Brand	Critical appraisal of the company submission, writing and editorial input.
Emma Hook	Contributed to the critical appraisal of the cost effectiveness evidence and writing of the economic sections of the report.
Justin Matthews	Critical appraisal of the company submission, writing and editorial input.
Mark Edmondson-Jones	Critical appraisal of the company submission, writing and editorial input.
Sophie Robinson	Critical appraisal of the literature search strategies.
Jonathan Digby-Bell	Clinical advice and review of draft report.
G.J. Melendez-Torres	Critical appraisal of the clinical and economic evidence, editorial review.

Table of Contents

List of key issues

List of tables

List of Figures

1. EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the Evidence Assessment Group (EAG) as being potentially important for decision making. It also includes the EAG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to Section 1.6 explain the key issues in more detail. Background information on the condition, technology and evidence and information on non-key issues are in the main EAG report.

All issues identified represent the EAG's view, not the opinion of NICE.

1.1. Overview of the EAG's key issues

A brief overview of the key issues identified by the EAG in their appraisal of the company submission (CS) is provided in Table 1. Further detail of the issues is provided in Sections 1.3, 1.4, 1.5, and 1.6.

Broadly speaking the key clinical issues related to the appropriateness of how the network meta-analysis was conducted. There was also a decision problem key issue related to the subgroup analysis and an 'other' key issue related to differences in the method of administration for risankizumab between the clinical trials and intended clinical practice. In terms of cost effectiveness, the EAG noted key issues with various aspects of the company's modelling approach, including the appropriateness of a model structure based on Crohn's Disease Activity Index, assumptions regarding treatment effectiveness estimates and the estimation of health state utility values.

Table 1: Summary of key issues

ID	Summary of issues	Report sections
#1	Feasibility of exploratory subgroup analysis by CD location	2.4
#2	Unexplored heterogeneity in network meta-analyses in relation to baseline risk	3.4.3
#3	Network structure in maintenance network meta-analyses should be connected	3.4.6

ID	Summary of issues	Report sections
#4	Appropriateness of the model structure	4.2.2
#5	Treatment duration and residual treatment effect assumptions	4.2.6
#6	Estimation and application of maintenance treatment effectiveness assumptions	4.2.6
#7	Health state utility value estimation	4.2.7
#8	Method of administration for risankizumab	2.3, 3.2.2.3, 4.2.4

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are outlined in Table 2.

Table 2: Key differences between the company's preferred assumptions and EAG's preferred assumptions

	Company's preferred assumption	EAG preferred assumption	Report Sections
Maximum treatment duration on biologic therapy, and residual treatment effect following biologic therapy	Assume all patients discontinue biologic therapy at 52 weeks. From this point, assume patients move to conventional care. Assume all patients experience a 52-week residual treatment effect following biologic therapy.	Highlight uncertainty around a true maximum treatment duration and residual treatment effect for biologic therapies. Assume a 20-year maximum treatment duration for biologic therapy in the base case. Assume a 26-week residual treatment effect following biologic therapy in the base case.	4.2.2 and 4.2.6.7
Network structure in maintenance NMA, and placebo CDAI-remission rates	Separate treatments into two disconnected networks, to reduce the heterogeneity in the placebo arms of maintenance studies.	Use a single maintenance network, and model placebo CDAI-remission rates using trial date as a potential candidate for explaining between-trial heterogeneity	3.4.6 and 4.2.6
Transition matrix calibration and cycle-length adjustment	Calibrate transition probabilities for each comparator, by adjusting the remission | mild cut-point in the risankizumab ordered probit model, to match 52-week remission estimates from the maintenance NMA. Estimate per-cycle (2-week) transition probabilities from implied 26-week transition probabilities, using an exponential assumption.	Calibrate transition matrices by adjusting both the remission | mild and mild | moderate-to-severe ordered probit cut-points by the same amount. Apply a transition matrix cycle length adjustment approach which does not rely on the use of the approximate exponential assumption.	4.2.6.4

	Company's preferred assumption	EAG preferred assumption	Report Sections
Health state utility values	Estimate mean CDAI-based health state utility values using OLS regression.	Estimated mean CDAI-based health state utility values using a linear mixed model, which includes a random effect to account for repeated measures.	4.2.7.1

Abbreviations: CDAI, Crohn's Disease Activity Index; EAG, External Assessment Group; OLS, ordinary least squares; NMA, network meta-analysis.

1.2. Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Affecting the expected Crohn's Disease Activity Index (CDAI) score of patients over time, and in doing so affecting the estimated distribution of patients in remission vs mild disease vs moderate-severe disease states over a lifetime perspective, with implications for patient health-related quality of life (HRQoL)
- A treatment-specific risk of adverse events, with implications for patient HRQoL

Overall, the technology is modelled to affect costs by:

- Introducing the acquisition cost of risankizumab to the treatment pathway for moderate-to-severely active CD
- Affecting the expected CDAI score of patients over time, and in doing so affecting the estimated distribution of patients in remission vs mild disease vs moderate-severe disease states over a lifetime perspective, with implications for the lifetime expected patient healthcare resource usage and associated costs
- A treatment-specific risk of adverse events, with implications for patient healthcare resource usage and associated costs

The modelling assumptions that have the greatest effect on the ICER are:

- The assumed maximum treatment duration for biologic therapies

- The estimation and application of long-term treatment effectiveness estimates; more specifically, assumptions regarding the separation of networks in the maintenance NMA and approach for calibrating and adjusting health state transition matrices
- The choice of model for estimating CDAI-based health state utility values

1.3. The decision problem: summary of the EAG's key issues

The EAG reviewed the approach of the company to addressing the NICE decision problem for this appraisal and identified the following key issues for consideration by the committee.

Key Issue 1: Feasibility of exploratory subgroup analysis by CD location

Report sections	Section 2.4
Description of issue and why the EAG has identified it as important	The final NICE scope for this appraisal includes subgroup analysis by CD location. The company excluded this from its decision problem. Clinical advice to the EAG was that CD location was likely the key prognostic factor for clinical effectiveness in this population.
What alternative approach has the EAG suggested?	Due to the clinical significance of this subgroup analysis, the EAG considered that the company should as a minimum have retained the CD location subgroup analysis in the decision problem and stated that it was unable to provide data to conduct this analysis. However, the EAG did not consider that the company's rationale for being unable to conduct subgroup analysis by CD location to be clearly justified. The EAG agreed that the numbers of participants per subgroup were fairly low but noted that this was also the case for the subgroup analysis the company presented by age and did not consider that this would preclude conducting an exploratory subgroup analysis.
What is the expected effect on the cost-effectiveness estimates?	An increase in uncertainty in clinical effectiveness, and consequently cost effectiveness, estimates based on a failure to adequately profile a key prognostic factor.
What additional evidence or analyses might help to resolve this key issue?	Provision of exploratory subgroup analysis by CD location based on the NICE scope using available data, noting the limitations of available evidence

Abbreviations: CD, Crohn's disease; EAG, Evidence Assessment Group; NICE, National Institute for Health and Care Excellence

1.4. The clinical effectiveness evidence: summary of the EAG's key issues

The EAG reviewed the clinical effectiveness and safety evidence presented in the CS and identified the following key issues for consideration by the committee.

Key Issue 2: Unexplored heterogeneity in network meta-analyses in relation to baseline risk

Report sections	Section 3.4.3
Description of issue and why the EAG has identified it as important	The company asserts that the base case network meta-analyses (NMAs) use a risk difference metric to address heterogeneity in baseline risk. However, the EAG regards that this is not an adjustment per se, and that it does not account for differences in treatment histories between trials, particularly in the group that has already experienced a biologic failure (BF). The company additionally advocates use of a fixed effects model because a random effects model produces implausibly large confidence intervals, an argument that does not unto itself have face validity in the presence of heterogeneity.
What alternative approach has the EAG suggested?	The EAG suggests that baseline risk adjustment be explored for risk difference-metric meta-analyses, and that a random effects model using an informative prior be explored.
What is the expected effect on the cost-effectiveness estimates?	The expected effect is unclear, but is likely to manifest in wider credible intervals in probabilistic sensitivity analysis (due to a random effects meta-analysis) and differences in incremental QALYs arising from baseline risk adjustment.
What additional evidence or analyses might help to resolve this key issue?	The EAG regards that an updated meta-analysis incorporating the model specification above would resolve the issue.

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; NMA, network meta-analyses; QALY, quality-adjusted life year

Key Issue 3: Network structure in maintenance network meta-analyses should be connected

Report sections	Section 3.4.6
Description of issue and why the EAG has identified it as important	The company's base case NMA for maintenance treatments separates drugs into two disconnected networks, citing rationales relating to drug mechanism of action and half-life.
What alternative approach has the EAG suggested?	The EAG has suggested using a single, joined-up network for each maintenance NMA.
What is the expected effect on the cost-effectiveness estimates?	It is difficult to disentangle the impact of this from changes to the application of these NMAs (see Key Issue 2) below, but the EAG believes this is likely to produce more stable estimates.
What additional evidence or analyses might help to resolve this key issue?	At clarification the company provided joined-up maintenance NMAs, though it retained the disconnected networks as its base case. Furthermore, related to Key Issue 2 above, estimates from these NMAs may change.

Abbreviations: EAG, Evidence Assessment Group; NMA, network meta-analyses

1.5. The cost effectiveness evidence: summary of the EAG's key issues

The EAG reviewed the economic model and cost-effectiveness evidence presented in the CS and identified the following key issues for consideration by the committee.

Key Issue 4: Appropriateness of the model structure

Report sections	Sections 4.2.2
Description of issue and why the EAG has identified it as important	The company's model structure defines health status by CDAI score (in particular, CDAI response and remission rates), yet the EAG is in receipt of expert advice that CDAI score is not used in NHS clinical practice for the management of CD, owing to its overcomplicated nature and poor correlation with endoscopy. Instead, advice to the EAG is that the Harvey Bradshaw Index and endoscopic response are used. As such, the EAG are concerned that company's model structure is not reflective of relevant patient outcomes. The company recognised this issue in their evidence submission, defending their approach in the context of limited endoscopic outcome data, which the company describe as only available from risankizumab and ustekinumab overall populations. Separately, the addition of risankizumab to the treatment options currently available would extend the plausible options available to treat each patient, yet the company assumes that after the initial therapy, patients move to conventional care, on every treatment arm. The EAG are concerned that this assumption does not reflect the treatment pathway as described by both the company and the EAG's clinical expert, which sees patients treated with every available and suitable option sequentially. Further, the modelled assumption that patients transition to conventional care after initial therapy discontinuation is at odds with the company's argument against providing a comparison to BSC, as requested in the Final Scope.
What alternative approach has the EAG suggested?	The EAG saw no alternative to the use of CDAI outcomes within the cost-effectiveness model to address the decision problem, given the data limitations described by the company. The EAG noted that it would have been possible for the company to have better captured the expected treatment pathway implications of risankizumab's proposed introduction, within a different model structure.
What is the expected effect on the cost-effectiveness estimates?	The effect of these issues upon cost-effectiveness estimates is unknown. The EAG are not able to explore the importance of these structural uncertainties within the scope of the company's cost-effectiveness model, and are not able to speculate on likely directional bias.
What additional evidence or analyses might help to resolve this key issue?	A considered, alternative approach to cost-effectiveness modelling that captures the expected pathway implications of the proposed introduction of risankizumab could serve to improve confidence in drawing cost-effectiveness conclusions in this appraisal.

Abbreviations: CD, Crohn's disease; CDAI, Crohn's Disease Activity Index; EAG, Evidence Assessment Group; NHS, National Health Service

Key Issue 5: Treatment duration and residual treatment effect assumptions

Report sections	Section 4.2.6
Description of issue and why the EAG has identified it as important	The EAG had several concerns with the company's approach to treatment discontinuation assumptions. The company's analysis assumes treatment-specific, constant rates of biologic treatment discontinuation in the maintenance phase of the model, for the first 52 weeks of maintenance therapy, then assumes all patients discontinue. From this point, patients are assumed to move to conventional care, whereby the company assume there is a further 52-week residual treatment effect in absence of biologic treatment costs. The EAG's clinical adviser found it difficult to judge whether assuming different 1-year discontinuation rates across treatments based on observed data across trials was appropriate, given differences in inclusion criteria and study design across trials. Expert advice to the EAG suggests it is implausible that all patients discontinue at 52 weeks. The EAG's clinical adviser's perspective is that if maintenance therapy is working for a patient, there is every effort and incentive to maintain treatment. The company's own TTD data from the FORTIFY study are consistent with this advice. Expert advice to the EAG suggests a residual treatment effect is plausible, with such an effect linked to the half-life of the treatment discontinued. For ustekinumab, the EAG's expert advises it can take around 24 weeks for symptoms to return.
What alternative approach has the EAG suggested?	The EAG-preferred base case assumes a 20-year maximum treatment duration, across treatments. The EAG explores different maximum treatment duration assumptions in scenario analyses, ranging from 5 to 40 years. The EAG-preferred base case assumes a 6-month residual effect across treatments, given the similar half-lives across treatments and EAG expert advice on estimated time to symptomatic return for ustekinumab.
What is the expected effect on the cost-effectiveness estimates?	Applied collectively, EAG-preferred maximum treatment duration and residual treatment effect assumptions lead to an increase in total costs and total QALYs across all biologic therapies. As such, the expected impact on cost-effectiveness results is multifaceted, and conditional on other model inputs and assumptions (such as biologic discontinuation rates and the cost of biologic maintenance treatment). In the CCF population, for risankizumab versus infliximab SC, incremental costs decrease while incremental QALYs increase, resulting in an improvement in the ICER. However, in the BF population, for risankizumab versus vedolizumab SC, incremental costs increase while incremental QALYs decrease, resulting in a higher ICER.
What additional evidence or analyses might help to resolve this key issue?	Further follow-up of FORTIFY TTD data could better inform time to treatment discontinuation assumptions in the cost-effectiveness model. Post-hoc analysis of FORTIFY patient outcomes following risankizumab discontinuation could better inform residual treatment effect assumptions in the cost-effectiveness model.

Abbreviations: EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; TTD, time-to-treatment-discontinuation; SC, subcutaneous.

Key Issue 6: Estimation and application of maintenance treatment effectiveness assumptions

Report sections	Section 4.2.6
Description of issue and why the EAG has identified it as important	Beyond issues with the maintenance phase NMA covered in Key Issue 2 and Key Issue 3 the EAG recommends that the company expand the placebo remission model to allow for plausible causes of heterogeneity, in particular a temporal association with the time at which individual clinical trials were conducted. This is consistent with an apparent improvement in remission outcomes over time, as treatments have improved. In addition, the EAG has several concerns with the company's approach to capture treatment effectiveness implications of maintenance therapy based on combining results from this NMA and observed FORTIFY data, and the implications for cost-effectiveness predictions. The company use an ordered probit model fit to FORTIFY subsample data to estimate transition probabilities. Despite company responses to EAG requests for clarity, justification for the appropriateness of the subsample data, the use of an ordered probit model, and the ordered probit model structure is weak. Conversion of implied 26-week transition matrices to model cycle-length (2-week) matrices is subject to known approximations that the company do not adequately justify. For comparator transition matrix estimation, the company calibrated the transition matrices estimated from FORTIFY data, to ensure 52-week remission rates matched the NMA-predicted 52-week remission rates, before cycle length adjustment. However, the calibration approach used is apparently arbitrary, adjusting only the balance of transitions to remission and mild at 26-weeks, and alternative approaches with different implications for long-term projections are possible. In particular, it is not considered tenable to assume that a change in the proportion of patients reaching remission does not also impact the proportion of patients moving to/remaining in moderate-to-severe disease. Separately, the company assume dose escalation affects costs but not patient outcomes, in assuming that standard dose transition probabilities apply to patients subject to biologic dose escalation. This EAG view this as an assumption that very likely biases comparative cost-effectiveness estimates in favour of risankizumab, as dose escalation applies only to comparator biologics.
What alternative approach has the EAG suggested?	The EAG prefers that placebo remission rates are modelled to include a temporal effect, and that absolute remission rates in maintenance are then based on this anchor point. The EAG also recommends an alternative approach to changing cycle length which avoids the use of the approximate exponential assumption. Additionally, the EAG prefers a calibration approach which adjusts both of the estimated ordered probit cutpoints by the same amount.
What is the expected effect on the cost-effectiveness estimates?	The isolated effect of the EAG-preferred estimation and application of maintenance treatment effectiveness assumptions is uncertain, and conditional on other preferred assumptions. The isolated impact on cost-effectiveness (when compared with the company's preferred base case in which a 52-week maximum treatment duration for biologic therapies is applied) is lower than the combined effect when implementing the EAG-preferred assumptions described in key issue 5.

Report sections	Section 4.2.6
	When also applying the EAG-preferred assumptions described in key issue 5, the effect of the EAG-preferred estimation and application of maintenance treatment effectiveness assumptions leads to higher incremental costs and lower incremental QALYs for risankizumab versus infliximab SC (in the CCF population) and versus vedolizumab SC (in the BF population). The EAG has not amended company dose escalation assumptions, and not this as a limitation of both the EAG-preferred and company base case analyses, that may bias results in favour of risankizumab.
What additional evidence or analyses might help to resolve this key issue?	Further follow up of FORTIFY patient outcomes could better inform risankizumab maintenance effectiveness assumptions in the cost effectiveness model. In lieu of these data, and for effectiveness projections for comparator treatments, a more considered and more robustly justified approach to modelling maintenance treatment effectiveness, taking into account the EAG's critique, may reduce the uncertainty around this issue. Additionally appropriate imputation methods may improve estimation of transition matrices, where CDAI data are missing, and diagnostics to assess the fit of the ordered probit model should be undertaken. The company could better inform its dose escalation assumptions, and provide further exploratory analyses, to illustrate the importance of potential bias in the company's approach, for cost-effectiveness results.

Abbreviations: EAG, Evidence Assessment Group

Key Issue 7: Health state utility value estimation

Report sections	Section 4.2.6
Description of issue and why the EAG has identified it as important	The company estimated the effect of CDAI category upon patient HRQL in ADVANCE, MOTIVATE and FORTIFY patient-reported data using ordinary least squares estimation, in order to inform cost-effectiveness model health state utility assumptions.
What alternative approach has the EAG suggested?	In the context of within-patient repeated measures, the EAG prefer to use health state utility values based on the same data but estimated using a (linear) mixed model that includes a random effect to account for repeated measures.
What is the expected effect on the cost-effectiveness estimates?	Applying EAG-preferred health state utility values leads to decrease in the total lifetime estimated QALYs across all treatment arms, as the linear mixed model predicts lower health state utility values in the remission and mild health states compared with the ordinary least squares regression used in the company's base case. The expected effect on cost-effectiveness results is variable, and depends on other assumptions regarding treatment effectiveness estimates, which determine the proportion of patents in the remission, mild and moderate-to-severe health states over time. In the CCF population, when compared with infliximab SC, applying the EAG-preferred health state utility value improves cost-effectiveness outcomes for risankizumab. However, in the BF population, when compared with

Report sections	Section 4.2.6
	vedolizumab SC, cost-effectiveness estimates are worse for risankizumab when applying EAG-preferred health state utility values.
What additional evidence or analyses might help to resolve this key issue?	The EAG feels there is no additional evidence needed to resolve this key issue, as it is a choice between alternative methods.

Abbreviations: EAG, Evidence Assessment Group, SC, subcutaneous.

1.6. Other key issues: summary of the EAG's views

The EAG identified the following additional key issues for consideration by the committee.

Key Issue 8: Method of administration for risankizumab

Report sections	2.3, 3.2.2.3, 4.2.4
Description of issue and why the EAG has identified it as important	The method of administration for risankizumab in the included clinical trials differs from the intended method of administration for clinical practice. Risankizumab was administered by intravenous clinician-administered injection in ADVANCE and MOTIVATE and by subcutaneous clinician-administered injection in FORTIFY sub-study 1. In the CS, the company stated that the intention was for risankizumab to be administered in routine practice using an on-body device. Very limited information was provided on this method of administration in the CS. In response to a clarification question by the EAG, the company stated that: "Risankizumab 600 mg intravenous (IV) induction will be administered in a hospital setting whilst risankizumab 360 mg subcutaneous (SC) maintenance will be administered through the on-body-device (OBD) either at home or in clinic. The OBD is a self-injection device which takes up to five minutes to administer from when the OBD is placed on the body at the injection site. The OBD allows for at-home treatment (where agreed with the healthcare team). The device can be placed to the abdomen or thigh and then upon pressing the button the OBD delivers a steady injection. In terms of administration the OBD should be stored in the refrigerator (at 2–8°C) and just before injecting the medication should be left to come up to room temperature. Upon activating the OBD a beeping sound will be heard, and a flashing blue status light will appear. The OBD can be secured on the injection site and the grey injection button should then be firmly pressed and released to deliver the medication. The OBD will beep, and the status light will flash green as the injection is delivered. The patient may do moderate physical activities, such as walking, reaching and bending, during the injection. The status light will change from flashing green to solid green and the device will beep once the medication has been delivered, at this stage and then the OBD can be removed by peeling the adhesive OBD off the skin. The OBD and cartridge can then be disposed by placing them into a special disposal container".

Report sections	2.3, 3.2.2.3, 4.2.4
	EAG also noted that the company stated in its clarification response that it was the [redacted] The company captures the cost implications of this administration difference, but in the model presumes no impact on clinical effectiveness parameters. The EAG considered this to be a strong assumption in the absence of evidence. It was also unclear to the EAG whether the on-body device method of administration had been considered in regulatory review for safety. The company provides no transparent (non-CIC) information on this method of administration in the CS or the clarification response. As the method of administration is a fundamental part of the delivery of the intended technology, the EAG had concerns that this could preclude effective stakeholder consultation on this appraisal and whether it could preclude NICE showing the evidential basis for its decision, given the intended method of administration does not match that used in the trials included in the submission.
What alternative approach has the EAG suggested?	The company could have considered FORTIFY sub-study 4, which according to publicly available information from clinical trial registries used an on-body injector as the method of administration, as a potential means of sourcing or adjusting clinical effectiveness parameters for the model using the intended method of administration. However, clarification would be required as to whether the on-body injector referenced in publicly available information on FORTIFY sub-study 4 is the same as the on-body device referenced in the CS.
What is the expected effect on the cost-effectiveness estimates?	There is considerable uncertainty as to whether the clinical effectiveness inputs to the cost effectiveness model remain valid given they were assessed using a different method of administration.
What additional evidence or analyses might help to resolve this key issue?	Data from FORTIFY sub-study 4 could help address this uncertainty, provided the on-body injector referenced in publicly available information on FORTIFY sub-study 4 is the same as or similar to the on-body device referenced in the CS. Clarification as to whether the on-body device method of administration was considered in the regulatory review for safety would also be useful. Some descriptive results from FORTIFY sub-study 4 were provided in the clarification response but these were not numerical in nature and they were not used to source or adjust clinical effectiveness parameters for the model using the intended method of administration. The narrative results provided were not sufficient to allow the EAG to conduct any useful critique of FORTIFY sub-study 4 results.

Abbreviations: EAG, Evidence Assessment Group

1.7. Summary of EAG's preferred assumptions and resulting ICER

As there is more than one comparator of relevance to the decision problem, the cost-effectiveness results are ideally calculated by fully incremental, probabilistic analysis. However,

for clarity and ease of calculation within the company's model, the step-by-step impact of EAG corrections to the company base and EAG preferred assumptions are summarised using deterministic changes and in pairwise analyses, in Table 3 and Table 5, for the conventional care failure (CCF) and biologic failure (BF) populations, respectively. Furthermore, the design of the company's economic model and volume of Visual Basics for Applications (VBA) code is a limiting factor for exploring probabilistic analysis. The economic model includes one 'Markov trace' (calculation) sheet for the selected comparator, and therefore must cycle through the list of included comparators using automated processes to perform incremental analysis, while also drawing recalibrated transition matrices. The above factors and number of included comparators contribute to a PSA run-time of approximately 9 hours when sampling 1,000 iterations; as such, the EAG did not consider it feasible to produce probabilistic results for each EAG preferred assumption or exploratory analysis within the EAG report timeframe. Additionally, the EAG note the company's economic model presents probabilistic results only in graphical form. In clarification question B31, the EAG requested an executable version of the cost-effectiveness model that included fully incremental probabilistic analysis (in line with the company base case presented in CS B.3.10.1); however, such model was not provided by the company. Thus, the EAG present full incremental analysis results probabilistically for the EAG preferred base case only.

In the company's and EAG's CCF population base case, adalimumab biosimilar is the 'reference' (lowest cost) treatment, and infliximab SC is the optimal comparator in the incremental analysis at a willingness-to-pay threshold of £20,000 to £30,000 per QALY gained. Thus, Table 3 presents pairwise cost-effectiveness results for risankizumab versus infliximab SC, for the CCF population. Fully incremental results for the EAG's preferred CCF population base case are presented in Table 4. For ease of reference, the EAG have excluded original forms of infliximab and adalimumab from the CCF incremental analysis table, as biosimilars are assumed by the company to provide equal QALYs at a lower cost.

Table 3: Summary of EAG's preferred assumptions and ICER (CCF population), risankizumab versus infliximab SC

Scenario	Incremental cost	Incremental QALYs	ICER (stepwise change)
Company's base case (probabilistic)	[redacted]	[redacted]	Dominated, -£81,752
Company's base case (deterministic)	[redacted]	[redacted]	Dominated, -£84,028
EAG corrected company base case	[redacted]	[redacted]	Dominated, -£102,827 (-£18,800)

Scenario	Incremental cost	Incremental QALYs	ICER (stepwise change)
+ Maximum treatment duration of 20 years for all biologic treatments	[redacted]	[redacted]	£52,499 (+£155,326)
+ Residual treatment effect of 26 weeks for all biologic treatments	[redacted]	[redacted]	£57,503 (+£5,004)
+ Single maintenance network, with an estimated maintenance placebo remission proportion that is adjusted for a temporal effect	[redacted]	[redacted]	Dominated, -£76,611 (-£134,114)
+ Transition matrices estimated by adjusting both the remission | mild and mild | moderate-to-severe cut points, and without an exponential assumption to estimate 2-week transitions	[redacted]	[redacted]	Dominated, -£75,237 (+£1,374)
+ Health state utility values estimated using a mixed linear model	[redacted]	[redacted]	Dominated, -£88,792 (-£13,555)
EAG's preferred base case (deterministic)	[redacted]	[redacted]	Dominated, -£88,792
EAG's preferred base case (probabilistic)	[redacted]	[redacted]	Dominated, -£90,018

Abbreviations: BF, biological failure; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table 4: Summary of EAG's preferred base case (CCF population), incremental analysis

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
EAG preferred deterministic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]	-	-	-	-
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£5,536	£5,536
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£56,481	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£52,086	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£1,349,539	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£4,358,832	Dominated
EAG preferred probabilistic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]	-	-	-	-
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£6,744	£6,744
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£55,111	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£48,951	Dominated

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£867,497	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	-£91,825,236	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; IV, intravenous; QALYs, quality-adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

In the company's BF population base case, all comparators are dominated by risankizumab; however, in the EAG's preferred base case, vedolizumab SC is the optimal treatment option in incremental cost-effectiveness analysis at a willingness-to pay threshold of £20,000-£30,000 per QALY gained. Table 5 therefore presents pairwise cost-effectiveness results for risankizumab versus vedolizumab SC. Fully incremental results for the EAG's preferred BF population base case are presented in Table 6.

Table 5: Summary of EAG's preferred assumptions and ICER (BF population), risankizumab versus vedolizumab SC

Scenario	Incremental cost	Incremental QALYs	ICER (stepwise change)
Company's base case (probabilistic)	[redacted]	[redacted]	Dominant, -£44,642
Company's base case (deterministic)	[redacted]	[redacted]	Dominant, -£43,738
EAG corrected company base case	[redacted]	[redacted]	Dominant, -£26,902 (+£16,836)
+ Maximum treatment duration of 20 years for all biologic treatments	[redacted]	[redacted]	£65,837 (+£92,739)
+ Residual treatment effect of 26 weeks for all biologic treatments	[redacted]	[redacted]	£66,781 (-£943)
+ Single maintenance network, with an estimated maintenance placebo remission proportion that is adjusted for a temporal effect	[redacted]	[redacted]	£55,959 (-£10,822)
+ Transition matrices estimated by adjusting both the remission | mild and mild | moderate-to-severe cut points, and without an exponential assumption to estimate 2-week transitions	[redacted]	[redacted]	£119,509 (+£63,550)
+ Health state utility values estimated using a mixed linear model	[redacted]	[redacted]	£143,088 (+£23,579)
EAG's preferred base case (deterministic)	[redacted]	[redacted]	£143,088
EAG's preferred base case (probabilistic)	[redacted]	[redacted]	£142,074

Abbreviations: BF, biological failure; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table 6: Summary of EAG's preferred base case (BF population), incremental analysis

Treatment	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
EAG preferred deterministic base case						
VDZ SC	[redacted]	[redacted]	-	-	-	-
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£2,198,195	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£252,156	Extendedly dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£143,088	£143,088
EAG preferred probabilistic base case						
VDZ SC	[redacted]	[redacted]	-	-	-	-
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£1,487,732	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£248,239	Extendedly dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£142,074	£142,074

Abbreviations: BF, biological failure; IV, intravenous; QALY, quality-adjusted life year; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

Modelling errors identified and corrected by the EAG are described throughout Section 4, and summarised in Section 6.1. For further details of the exploratory and sensitivity analyses performed by the EAG, see Section 6.2. For further details of the EAG preferred base case, see Section 6.3. For additional exploratory scenarios around the EAG preferred base case, see Section 6.4.

2. INTRODUCTION AND BACKGROUND

2.1. Introduction

In this report, the Evidence Assessment Group (EAG) provides a review of the evidence submitted by AbbVie in support of risankizumab for previously treated moderate to severe Crohn's disease.

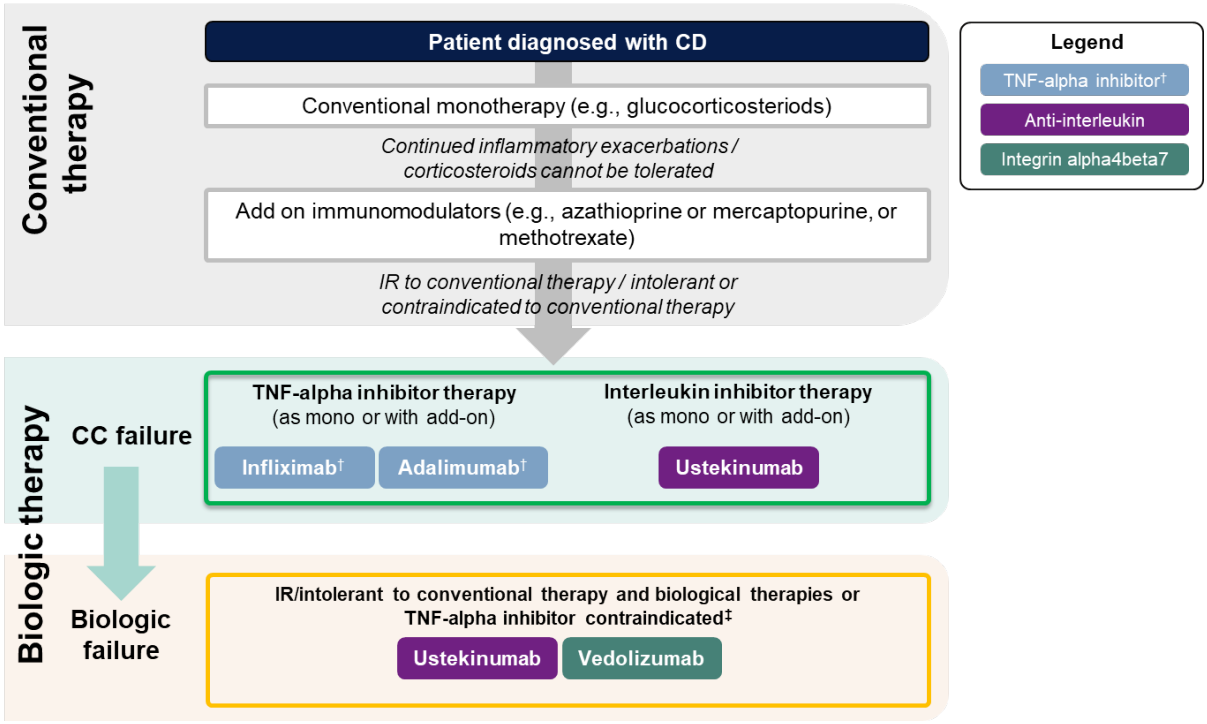
2.2. Critique of the company's description of the underlying health problem

The company's description of the underlying health problem, moderate to severe Crohn's disease (CD), is summarised in the CS Document B Section B.1.3.1. CD is a chronic relapsing systemic inflammatory bowel disease that can cause inflammation and mucosal ulceration to the entire gastrointestinal tract, but most commonly the distal small intestine. The pathogenesis of CD involves the complex interaction of immunological, microbiological, environmental and genetic factors.[1-3] Symptoms of CD can be heterogeneous, but include abdominal pain, diarrhoea, fatigue, weight loss, and blood or mucus in stools.[3, 4] Major extraintestinal manifestations for CD include ocular, renal, digestive, musculoskeletal, cardiovascular, dermatological and oral manifestations.[1] Symptoms can affect educational outcomes, work productivity, mental health and quality of life,[5-10] and result in extensive health service utilization.[6, 9-11] The prevalence of CD in the UK in 2021 was estimated to be 0.35% for males and 0.44% for females, leading to an estimated 185, 668 people aged 16 and over with CD in England.[12, 13] Around 40% of people with CD in the UK have been estimated to have moderate-to-severe disease, producing an estimated target population of 74,267 people with moderate-to-severe CD in England. The EAG considered the company's description of the underlying health problem to be adequate. Clinical advice to the EAG indicated that there was typically a dual peak in age distribution of patients encountered in routine clinical practice (late teens-early twenties and around the age of 50), that there was not considered to be an important difference in CD prevalence by gender, and that the key prognostic factors in this clinical population were age (the younger the patient is at diagnosis the less responsive the disease is likely to be), smoking status, and disease distribution (colonic disease is the most responsive to treatment and perianal disease the least responsive).

2.3. Critique of the company's overview of current service provision

The company's current care pathway is described in CS Document B Section 1.3.3. This is based on NICE Guideline NG129[14] and depicted in a flowchart. Clinical advice to the EAG was that each major centre has its own treatment pathway and that there are differences between centres, but in as much as there is a national standard of practice, the flowchart below is reasonably accurate in depicting this.

Figure 1. Treatment pathway based on CD management guidance by NICE



Risankizumab is humanised IgG1 monoclonal antibody that specifically binds with high affinity to the p19 subunit of human IL-23 cytokine blocking the binding of IL-23 to IL-23Rα without binding to IL-12.[15, 16] The recommended induction dose is 600 mg administered IV at Week 0, Week 4 and Week 8, followed by a maintenance dose of 360 mg administered SC at Week 12 and Q8W thereafter. Risankizumab was delivered IV in the risankizumab induction trials (ADVANCE and MOTIVATE) and SC in the risankizumab maintenance trial (FORTIFY) included in the CS. The EAG noted from publicly available information on clinical trials registries that an 'on-body injector' was used in FORTIFY sub-study 4, which was not included in the CS. In clinical practice, the company anticipates that risankizumab SC will be delivered using an on-body device. Clinical advice to the EAG indicated a low level of clinical familiarity with on-body injectors but identified both potential advantages and disadvantages of this approach. The EAG

considered that the description of the on-body device intended for clinical use in the CS was insufficiently detailed. The company provided further information in the clarification response QB3, as follows:

> *"Risankizumab 600mg intravenous (IV) induction will be administered in a hospital setting whilst risankizumab 360mg subcutaneous (SC) maintenance will be administered through the on-body-device (OBD) either at home or in clinic. The OBD is a self-injection device which takes up to five minutes to administer from when the OBD is placed on the body at the injection site. The OBD allows for at-home treatment (where agreed with the healthcare team). The device can be placed to the abdomen or thigh and then upon pressing the button the OBD delivers a steady injection. In terms of administration the OBD should be stored in the refrigerator (at 2–8°C) and just before injecting the medication should be left to come up to room temperature. Upon activating the OBD a beeping sound will be heard, and a flashing blue status light will appear. The OBD can be secured on the injection site and the grey injection button should then be firmly pressed and released to deliver the medication. The OBD will beep, and the status light will flash green as the injection is delivered. The patient may do moderate physical activities, such as walking, reaching and bending, during the injection. The status light will change from flashing green to solid green and the device will beep once the medication has been delivered, at this stage and then the OBD can be removed by peeling the adhesive OBD off the skin. The OBD and cartridge can then be disposed by placing them into a special disposal container".*

There are no additional tests or investigations associated with risankizumab use. Risankizumab currently holds marketing authorisation in the UK for the treatment of moderate-to-severe plaque psoriasis and active psoriatic arthritis. It has been recommended by NICE for the treatment of moderate-to-severe plaque psoriasis (TA596) and alone or in combination with methotrexate for the treatment of active psoriatic arthritis in adults who have had an inadequate response or been intolerant to one or more disease-modifying antirheumatic drugs.

2.4. Critique of company's definition of decision problem

The company statement regarding the decision problem is presented in the CS Section B.1.1, Table 1. The company position and the EAG response are provided in Table 7 below.

The EAG considered that the company’s definition of the decision problem was generally acceptable. The EAG identified one key issue related to the decision problem: feasibility of exploratory subgroup analysis by CD location.

Table 7: Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Population	People with previously treated moderately to severely active CD	As per scope	NA	NA
Intervention	RZB	As per scope	NA	NA
Comparator(s)	• TNF-alpha inhibitors (IFX and ADA) • VDZ • UST For people for whom TNF-alpha inhibitors, VDZ and UST have been ineffective, are contraindicated or are not tolerated: • BSC	• TNF-alpha inhibitors (IFX and ADA) • UST • VDZ	The scope includes BSC as a comparator for those who have failed or are contraindicated to all currently available biologics (TNF-alpha inhibitors [ADA, IFX], UST and/or VDZ). BSC is not considered an appropriate comparator; in clinical practice, if a biologic therapy has failed or is contraindicated, the individual will be offered an alternative biologic therapy.	The EAG agreed that the exclusion of BSC as a comparator was likely appropriate given BSC would not be routinely used in clinical practice, based on clinical advice provided to the EAG. The EAG agreed that the focus on comparators applicable to UK practice was appropriate.
Outcomes	• Disease activity (remission, response, relapse) • Mucosal healing • Surgery • Adverse effects of treatment • Health-related quality of life	As per scope	NA	NA

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per QALY. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and PSS perspective. The availability of any commercial arrangements in place for the intervention, comparator or subsequent treatment technologies will be taken into account.	• Cost per QALY • Lifetime horizon (suitably long to reflect differences) • NHS and PSS perspective on costs (base case) • PASs to be taken into account	N/A	The company present a non-reference case scenario analysis including societal costs
Subgroups	If evidence allows; location of CD (ileal, colonic and perianal)	• People who have had an inadequate response to conventional care (CCF) • People who have received ≥1 previous biologic and had an inadequate response (BF)	The trial design of RZB included the non-Bio-IR† and Bio-IR‡ populations, which were aligned in the model with CCF and BF populations. Separate analyses were conducted in these subpopulations as the comparators and clinical efficacy were different. Due to low	The EAG noted that the additional CCF and BF subgroup analyses in the company decision problem had not been specified in the NICE final scope for this appraisal. The company explained in the clarification response (A3) that the CCF and BF

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
			subject numbers the analysis of outcomes by CD location was deemed untenable.	subgroups were consistent with previous TAs for ustekinumab (TA456) and vedolizumab (TA352). The EAG considered this potentially justifiable but a matter for the Committee to determine as it is not in line with the NICE scope. Regarding the exclusion of subgroup analysis by CD location, the EAG noted that it is reported in the CS that this was conducted. No details are reported. Without seeing the results of this analysis, the EAG is unable to agree that no meaningful conclusions could be drawn from this subgroup analysis. Clinical advice to the EAG identified location of CD as probably the key prognostic factor. Table 12 in the CS showed 155 patients with ileocolic CS, 76 patients with colonic CD and 55 patients with ileal CD in FORTIFY across both intervention and placebo arms. While noting power may be

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
				suboptimal, the EAG considered that these numbers would likely be adequate for an exploratory subgroup analysis, noting that numbers were low in the presented subgroup analysis by patient age.
Special considerations including issues related to equity or equality	The availability and cost of biosimilars should be taken into consideration	• TNF-alpha inhibitors (ADA and IFX) are comparators which have biosimilars available	Cost of biosimilars have been taken into consideration where available i.e., for ADA and IFX.	Clinical advice to the EAG did not identify any equality concerns related to the potential introduction of risankizumab into the treatment pathway

Abbreviations: ADA, adalimumab; BF, biologic failure; BSC, best supportive care; Bio-IR, biologic inadequate response/intolerance; CD, Crohn's disease; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; NA, not applicable; NICE, National Institute for Health and Care Excellence; PAS, patient access scheme; PSS, Personal Social Services; QALY, quality-adjusted life year; RZB, risankizumab; TNcCF, tumour necrosis factor; UST, ustekinumab; VDZ, vedolizumab.; † Participants who had an inadequate response or intolerance to conventional therapy (defined as one or more of the following: aminosalicylates, oral locally acting steroids [e.g., budesonide, beclomethasone], systemic corticosteroids [prednisone or equivalent], or immunomodulators). This population may include patients who had received biologic therapy in the past but stopped therapy based on reasons other than inadequate response (IR) or intolerance (e.g., change in reimbursement coverage, well-controlled disease); ‡ Participants with documented intolerance or inadequate response (either failure to respond to induction treatment, or loss of response to maintenance therapy) to one or more biologics for CD (infliximab, adalimumab, certolizumab, natalizumab, vedolizumab, and/or ustekinumab).

3. CLINICAL EFFECTIVENESS

The sections below discuss the evidence submitted by the company in support of the clinical effectiveness of risankizumab for previously treated moderate to severe Crohn's disease.

The EAG reviewed the details provided on:

- Methods implemented to identify, screen, data extract and assess the risk of bias in relevant evidence.
- Clinical efficacy of risankizumab.
- Safety profile of risankizumab.
- Assessment of comparative effectiveness of risankizumab against relevant comparators.

A detailed description of an aspect of the CS is only provided where the EAG disagreed with the company's assessment or proposal, or where the EAG identified a particular area of concern that the EAG considered necessary to highlight for the Committee.

The following clinical effectiveness key issues were identified:

- Unexplored heterogeneity in network meta-analyses in relation to baseline risk and use of fixed effect models
- Network structure in maintenance network meta-analyses should be connected

Additionally, the EAG considered that the following key issues had relevance to the clinical effectiveness evidence:

- Feasibility of exploratory subgroup analysis by CD location (decision problem key issue)
- Method of administration for risankizumab (other key issue)

3.1. Critique of the methods of reviews

The company undertook a global systematic literature review (SLR) to identify randomised controlled trials (RCTs) providing evidence for risankizumab (summarised in Section 3.2) and other relevant comparator therapies in people with moderately to severely active Crohn's disease. The company stated that included comparators to risankizumab may not all be relevant

to the UK due to the global approach that was used. Eligible RCTs were used to inform the company's indirect treatment comparison (Sections 3.3 and 3.4). An overview of the methods used in the SLRs is provided in Table 8 below.

Table 8: Summary of EAG's critique of the methods implemented by the company to identify evidence relevant to the decision problem

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Document B, Section B.2.1, Appendix D.1.1.	The EAG considered the company searches to be well executed overall. However, the RCT filter that was used by the company is not a recognised, validated filter such as the one from the Cochrane Handbook. In clarification the company stated that they used a mixture of different filters from SIGN and NICE; but this is not how these RCT filters are designed to be used and this makes the effectiveness of the search uncertain. In clarification, the company stated that no additional searches were carried out for adverse events as these were included in the overall clinical effectiveness search results. It is possible that exclusion of cohort, case-control, cross-sectional and case series as publication types in the literature searches (due to use of an RCT filter) meant that papers reporting adverse events have been missed.
Inclusion criteria	Inclusion criteria for clinical evidence: Appendix D.1.2. Table 2 (p.15-16) Inclusion criteria for studies included in the NMA: Appendix D.1.2. Table 3 (p.17-18)	The inclusion criteria for the clinical effectiveness review are considered broadly appropriate to the decision problem. Comparators not listed in the NICE scope, i.e. brazikumab, certolizumab pegol, estrasimod, etrolizumab, filgotinib, guselkumab, mirikizumab, ozanimod and upadacitinib were listed as eligible comparators, though the EAG noted that the company undertook a 'global' SLR. The EAG noted inclusion of adults with biologic-naïve, -exposed and –refractory CD, which is aligned with the population detailed in the company's scope as detailed in Table 7; however, no specific inclusion criteria were specified to identify trials in patients with specific locations for CD as per the NICE-scoped subgroups. The EAG noted the company's position that the number of patients per disease location made subgroup analyses untenable but considered that the company may have sufficient data to enable exploratory subgroup analyses by disease location, particularly since clinical advice to the EAG indicated that disease location is an important

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
		prognostic factor. Therefore, inclusion criteria related to the location of CD for the purpose of subgroup analysis could have been useful. The additional inclusion criteria for the NMA were considered broadly reasonable by the EAG, though inclusion criteria related to the follow-up time for outcomes was considered to be potential source of heterogeneity (see Section 3.3.2.1), particularly in the induction NMAs. The EAG also noted that trials with 'treat-through' maintenance phases, as well as those re-randomising participants based on clinical remission, were excluded from the SLR. No explicit justification for this was provided, however, the EAG considered these exclusions to be appropriate given the likely impact this would have on the reduction of heterogeneity and intransitivity in the NMA (see Section 3.4.4).
Screening	Appendix D.1.2., p.16	Screening was conducted to appropriate standards to minimise selection bias, with duplicate, independent screening of identified studies and arbitration of discrepancies by a third reviewer. The EAG noted mention of the number of studies reviewed at the title and abstract screening stage as well as the full-text stage, though this staged approach was not explicitly reported.
Data extraction	Appendix D.1.2., p.16	Data extraction was conducted to appropriate standards to minimise selection bias, with extractions by a single reviewer into a pre-defined Excel-based template validated by a senior reviewer. Though data extraction was not done independently and in duplicate, the EAG noted that data validation by a second reviewer is permissible with the AMSTAR 2 critical appraisal tool.[17]
Tool for quality assessment of included study or studies	All studies included in the NMA: Appendix D.3	Quality assessments for ADVANCE, MOTIVATE and FORTIFY were conducted using the NICE clinical effectiveness quality assessment checklist for RCTs.[18] The tool was also used to assess the quality of all 13 other RCTs included in the company's NMA. The risk of bias of all 16 RCTs included in the NMA (ADVANCE, MOTIVATE and FORTIFY inclusive) was additionally assessed using the Cochrane risk of bias tool. The EAG considered these methods appropriate, though it was not clear why both methods were used, whether the outcomes of these assessments were considered together, or if the results of a

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
		specific tool were selected. Furthermore, the EAG noted that the Cochrane risk of bias assessments included domains of the updated Cochrane risk of bias 2 tool,[19] but that no outcome-level assessments were conducted. The EAG considered this to be an inappropriate application of the tool. Various errors in algorithm results for this tool were identified, e.g. Domain 4 of several trials (ACCENT 1, CHARM, CLASSIC 1, GAIN, GEMINI 2 and GEMINI 3) should not be 'Low', and Domain 2 judgments for MOTIVATE and ADVANCE are incorrectly captured.
Evidence synthesis	Document B, Section B.2.9.1, Appendix D.1.3.3.	The company conducted several NMAs to evaluate the comparative efficacy of risankizumab with other available treatments within the CCF and BF subgroups; these were further stratified by induction and maintenance phases for each subgroup. This was considered reasonable by the EAG. The results within the maintenance phase for each subgroup were further divided into one of two treatment networks: risankizumab-ustekinumab or vedolizumab-TNFi. The EAG identified this grouping of treatment networks to be an area of uncertainty, as discussed in Section 3.4. The EAG also considered that further outcomes, particularly adverse events or treatment discontinuations, could have been evaluated; however, the company did not report feasibility assessment and therefore it is not possible to determine if these outcomes were considered but found not feasible for analysis. Statistical methods were appropriate, though the EAG highlighted concerns related to the way in which the maintenance networks were structured (see Section 3.4.6), potential heterogeneity in follow-up time (see Section 3.3.2.1) and potential effect modification due to patient characteristics (see Section 3.3.2.2). Given the company's preference for fixed effects analyses, the EAG regarded that random effects analyses using informative priors should have been considered.

Abbreviations: BF, biologic failure; CCF, conventional care failure; CD, Crohn's disease; CS, Company submission; EAG, Evidence Assessment Group; NMA, network meta-analysis; RCT, randomised controlled trial; SLR, systematic literature review; TNFi, tumour necrosis factor inhibitors

3.2. Critique of trials of the technology of interest, the company's analysis and interpretation (and any standard meta-analyses of these)

3.2.1. Studies included in the clinical effectiveness review

The company presented evidence from three clinical studies: two pivotal induction studies (ADVANCE[20] and MOTIVATE[21, 22]) and one maintenance study (FORTIFY).[23, 24] These are analysed below.

3.2.2. Description and critique of the design of the studies

3.2.2.1. Design of the studies

The CS included two pivotal induction studies (ADVANCE[20, 22] and MOTIVATE[21, 22]) and one maintenance study (FORTIFY).[23, 24] The pivotal induction studies were both placebo-controlled randomised multi-centre trials conducted internationally, including UK centres. The design of the included studies is summarised in Table 9. Only sub-study one from FORTIFY[23, 24] was included in the CS.

Table 9: Clinical evidence included in the CS

Study name and acronym	Study design	Population	Intervention	Comparator	Study type
ADVANCE (NCT03105128)[20, 22]	Phase 3 multicentre, randomised induction study	People aged 16 or older with moderate-to-severe CD and inadequate response or intolerance to prior biologic therapy (Bio-IR), or with inadequate response or intolerance to conventional therapy (non-Bio-IR)	Risankizumab, 600 mg or 1200 mg IV Q4W	Placebo	RCT
MOTIVATE (NCT03104413)[21, 22]	Phase 3 multicentre randomised induction study	People aged 16 or older with moderate-to-severe CD, with a documented inadequate response or intolerance to ≥1 biologic	Risankizumab, 600 mg or 1200 mg IV Q4W	Placebo	RCT

Study name and acronym	Study design	Population	Intervention	Comparator	Study type
		therapy/therapies for CD (Bio-IR)			
FORTIFY (NCT03105102)[23, 24] Sub-study 1	Phase 3, multi-centre, partially randomised, double-blind, placebo-controlled, 52-week maintenance study with an ongoing open-label extension (OLE)	Participants who have entered and completed ADVANCE, MOTIVATE, or another AbbVie risankizumab Crohn's disease study and achieved clinical response during induction treatment with intravenous risankizumab or placebo	Randomised: participants with response to risankizumab 600 mg IV or 1200 mg IV during induction **randomised to risankizumab** 360 mg SC Q8W or 180 mg SC Q8W	Randomised: participants with response to risankizumab 600 mg IV or 1200 mg IV during induction **randomised to placebo** injection SC Q8W	RCT
			Non-randomised: participants with response to risankizumab 360 mg SC Q8W or 180 mg SC Q8W during induction continued on this dose	Non-randomised: participants with response to placebo IV during induction received placebo SC Q8W	NRS

Abbreviations: Bio-IR, biologic inadequate response/intolerance; CD, Crohn's disease; IV, intravenous; n/a, not applicable; non-Bio-IR, conventional therapy inadequate response/intolerance; NRS, non-randomised study; OBI, on-body injector; OLE, open-label extension; Q4W, every four weeks; Q8W, every eight weeks; RCT, randomised controlled trial; SC, subcutaneous

3.2.2.2. Population

In the ADVANCE study,[20, 22] eligible participants were people aged 16 or older with moderate-to-severe CD who had inadequate response or intolerance to prior biologic therapy (Bio-IR), or with inadequate response or intolerance to conventional therapy (non-Bio-IR). Detailed inclusion and exclusion criteria were provided in the CS (Appendix M.1, Table 83). Participants were randomized to receive risankizumab 600 mg IV (n=336) or placebo (n=175).

In the MOTIVATE study,[25] eligible participants were people aged 16 or older with moderate-to-severe CD, with a documented inadequate response or intolerance to ≥1 biologic therapy/therapies for CD (Bio-IR). Detailed inclusion and exclusion criteria were provided in the CS (Appendix M.1, Table 83). Participants were randomized to receive risankizumab 600 mg IV (n=191) or placebo (n=187).

In the FORTIFY sub-study 1 (SS1),[23, 24] which is included in the company submission, eligible participants were people who had entered and completed either the ADVANCE or MOTIVATE study and achieved clinical response with risankizumab or placebo. This was defined as a ≥30% decrease in average daily stool frequency and/or a ≥ 30% decrease in average daily abdominal pain score; with both not worse than at baseline for the induction study at the last visit of ADVANCE or MOTIVATE. Detailed inclusion and exclusion criteria were provided in the CS (Appendix M.1, Table 83).

In both ADVANCE[20, 22] and MOTIVATE, the proportion of patients with exposure to ustekinumab was restricted to 20%. The EAG noted that these technologies had a similar mechanism of action. The company explained (clarification response A20) that this limit was based on prior experience in the adalimumab programme with participants exposed to infliximab. It was considered that there could be reduced efficacy in participants exposed to another technology designed to inhibit the same pathway. The EAG noted that in the company's response it was stated that a rationale for this 20% limit was to ensure 'probability of success for the co-primary endpoints'. Clinical advice to the EAG was that prescription of risankizumab to a patient who had not responded to ustekinumab was unlikely, so this was not a major issue.

FORTIFY SS1 included several analysis sets: three intention-to-treat (ITT) populations and one safety population. For the former, ITT1 included both randomised and non-randomised participants who received at least one dose of the study drug; ITT1A formed the primary population for efficacy analysis and included only randomised subjects in ITT1 who had a simple endoscopic score for Crohn's disease (SES-CD) of ≥ 6 (≥ 4 for isolated ileal disease) at baseline for either induction study; and ITT1C included only non-randomised subjects in the ITT1 set. The safety population, SA1, comprised all randomised participants who received at least one dose of study risankizumab in SS1. More details of these analysis sets are provided in CS in Table 8 (p.35) and Figure 6 (p.36).

A total of 141 participants receiving the licensed dose (360 mg SC Q8W) of risankizumab and 164 participants receiving placebo were included in the ITT1A population. Power calculations indicated that a sample size of 150 participants in each group would provide power (two-sided, α=0.05) of 87%, 93% and 99% for the co-primary endpoints of CDAI clinical remission, SF/APS clinical remission, and endoscopic response, respectively, at Week 52; the assumption of remission and response rates used in these calculations are reported in the supplementary appendix of Ferrante 2022.[24]

Baseline characteristics from the three included studies were provided in the CS Table 12 and reproduced below as Table 10.

Table 10. Characteristics of participants in the studies across treatment groups (ITT1A population)

	ADVANCE		MOTIVATE		FORTIFY[¥]	
Characteristic	RZB 600 mg IV N=336	PBO IV N=175	RZB 600 mg IV N=191	PBO IV N=187	RZB 360 mg SC N=141	PBO SC[††] N=164
Age, mean years (SD)	38.3 (13.3)	37.1 (13.4)	40.2 (13.6)	39.3 (13.5)	37.0 (12.8)	38.0 (13.0)
Age category, n (%)						
16 to <18 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
18–40 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
40–65 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥65 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sex, n (%)						
Male	189 (56.3)	88 (50.3)	92 (48.2)	99 (52.9)	81 (57.4)	89 (54.3)
Female	147 (43.8)	87 (49.7)	99 (51.8)	88 (47.1)	60 (42.6)	75 (45.7)
Race						
White	258 (76.8)	134 (76.6)	176 (92.1)	162 (86.6)	111 (78.7)	126 (76.8)
Black or African American	9 (2.7)	9 (5.1)	7 (3.7)	7 (3.7)	8 (5.7)	10 (6.1)
Asian	65 (19.3)	31 (17.7)	8 (4.2)	15 (8.0)	20 (14.2)	28 (17.1)
American Indian/Alaska Native	0	0	0	0	0	0
Native Hawaiian or other Pacific Islander	0	1 (0.6)	0	2 (1.1)	0	0
Multiple	4 (1.2)	0	0	1 (0.5)	2 (1.4)	0
Ethnicity						
Non-Hispanic/Latino	325 (96.7)	165 (94.3)	175 (91.6)	168 (89.8)	134 (95.0)	157 (95.7)
Hispanic/Latino	11 (3.3)	10 (5.7)	16 (8.4)	19 (10.2)	7 (5.0)	7 (4.3)
BMI (kg/m^2), mean (SD)	24.1 (5.6)	24.3 (5.8)	25.3 (6.4)	25.1 (5.8)	23.9 (5.4)	24.8 (6.3)
CD duration (years), mean (SD)	9.0 (8.8)	8.2 (7.1)	10.9 (7.7)	12.5 (9.7)	9.3 (8.1)	9.6 (8.8)

	ADVANCE		**MOTIVATE**		**FORTIFY¥**	
Disease location						
Ileocolic	180 (53.6)	90 (51.4)	96 (50.3)	98 (52.4)	72 (51.1)	83 (50.6)
Colonic disease	76 (22.6)	39 (22.3)	38 (19.9)	45 (24.1)	32 (22.7)	44 (26.8)
Ileal	62 (18.5)	37 (21.1)	49 (25.7)	33 (17.6)	25 (17.7)	30 (18.3)
Ileal - involving upper GI tract	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Colonic disease - involving upper GI tract	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ileocolic - involving upper GI tract	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Faecal calprotectin (mg/kg), median (mean [SD])	n=141	n=284	n=150	n=146	n=114	n=140
	960 (1767.3 [2272.7])	1200 (2499.3 [4308.8])	1367 (2379.2 [3879.6])	987.5 (2648.9 [4831.2])	1543 (2182.5 [2471.7])	794.5 (1640.7 [2055.7])
Average daily SF, mean (SD)	5.8 (2.7)	6.1 (2.8)	6.2 (3.1)	6.4 (2.9) (n=186)	5.9 (2.6)	5.8 (2.7)
Average daily AP, mean (SD)	1.9 (0.6)	1.9 (0.6)	1.8 (0.5)	1.9 (0.5) (n=186)	1.8 (0.5)	1.9 (0.5)
CDAI, mean (SD)	311.2 (62.4)	319.2 (59.4)	310.7 (63.6)	319.6 (69.8) (n=186)	308.9 (61.1)	307.4 (64.9)
SES-CD, mean (SD)	14.7 (7.7)	13.8 (6.8)	14.4 (7.6)	15.0 (8.1)	14.3 (7.4)	14.0 (7.1)
Immunomodulator use, n (%)	88 (26.2)	42 (24.0)	36 (18.8)	40 (21.4)	40 (28.4)	40 (24.4)
Biologic failure, n (%)						
0	141 (42.0)	78 (44.6)	0	0	39 (27.7)	41 (25.0)
1	100 (29.8)	41 (23.4)	92 (48.2)	88 (47.1)	51 (36.2)	60 (36.6)
2	40 (11.9)	30 (17.1)	54 (28.3)	45 (24.1)	27 (19.1)	36 (22.0)
3	35 (10.4)	20 (11.4)	22 (11.5)	29 (15.5)	17 (12.1)	22 (13.4)
>1 (2-7)	95 (28.3)	56 (32.0)	99 (51.8)	99 (52.9)	51 (36.2)	63 (38.4)
TNF-alpha failure, n (%)	n=195†	n=97†			n=102†	n=123†

	ADVANCE		MOTIVATE		FORTIFY¥	
0	12 (6.2)	0	14 (7.3)	6 (3.2)	11 (10.8)	4 (3.3)
1	110 (56.4)	57 (58.8)	101 (52.9)	103 (55.1)	49 (48.0)	71 (57.7)
>1	73 (37.4)	40 (41.2)	76 (39.8)	78 (41.7)	42 (41.2)	48 (39.0)
Vedolizumab failure, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ustekinumab failure, n (%)	n=195†	n=97†			n=102†	n=123†
	43 (22.1)	19 (19.6)	36 (18.8)	40 (21.4)	17 (16.7)	15 (12.2)
CD medication‡ at baseline§, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Aminosalicylates	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Corticosteroids	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Immunosuppressants/immunomodulators	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Antibiotics	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Anti-diarrhoeal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AP, abdominal pain; Bio-IR, biologic inadequate response/intolerance; BMI, body mass index; CD, Crohn's disease; CDAI, Crohn's Disease Activity Index; IV, intravenous; PBO, placebo; RZB, risankizumab; SC, subcutaneous; SD, standard deviation; SES-CD, Simple Endoscopic Score for Crohn's Disease; SF, stool frequency; TNF, tumour necrosis factor; WHO, World Health Organization.
†Bio-IR population; ‡generic name (WHO 2018Q1); §for FORTIFY, baseline refers to baseline of the induction study; ¥Data reported for randomised subjects only from FORTIFY SS1; †† The placebo SC (withdrawal) arm consisted of subjects who achieved SF/APS clinical response to IV risankizumab induction therapy in ADVANCE or MOTIVATE and were randomised to receive placebo in FORTIFY.

The CS reports that the risankizumab Crohn's disease study programme enrolled a total of [redacted] subjects at [redacted] UK centres, with UK participants representing [redacted]%, [redacted]% and [redacted]% of the study populations in ADVANCE, MOTIVATE and FORTIFY, respectively. Clinical advice to the EAG was that findings from overseas participants were likely to generalize reasonably well to the UK clinical practice setting, although usual caveats relating to treatment pathways and population similarity should be noted. The company was unable to provide baseline characteristics or results specifically for UK participants (clarification question A17), which increases uncertainty regarding generalizability to the target UK context.

3.2.2.3. Intervention

The intervention used in all included studies was risankizumab. Dosing and method of administration differed between the pivotal induction trials (IV administration in ADVANCE and MOTIVATE) and the maintenance trial (SC administration in FORTIFY). In both ADVANCE and MOTIVATE, risankizumab was administered intravenously at a dose of 600 mg (licensed dose) or 1200 mg Q4W.

In the FORTIFY sub-study 1, which is included in the company submission, the intervention is a maintenance dose of risankizumab; administered subcutaneously to participants randomised thereto as either 360 mg Q8W (licensed dose) or 180 mg Q8W risankizumab for 52 weeks following response to induction treatment with risankizumab in the ADVANCE or MOTIVATE induction trials. Non-randomised intervention arms in FORTIFY included participants who responded to non-licensed induction doses of risankizumab in ADVANCE or MOTIVATE, i.e. 360 mg Q8W (following 12 weeks of 600 mg risankizumab IV induction plus 12 weeks of 360 mg SC risankizumab) or 180 mg Q8W risankizumab.

3.2.2.4. Comparator

The comparator in the ADVANCE and MOTIVATE studies was placebo, which was not listed as a comparator in the NICE scope. The comparator used in SS1 of FORTIFY, the sub-study included in the CS, was also placebo; comprising succinic acid (0.5 mmol/L), disodium succinate hexahydrate (3.9 mmol/L), sorbitol (275 mmol/L), polysorbate 20 (0.16 mmol/L), and water for injection (Ferrante 2022).[24] In participants randomised thereto in FORTIFY SS1 following response to induction treatment with risankizumab in ADVANCE or MOTIVATE, placebo was administered as subcutaneous injection Q8W. Non-randomised participants with response to intravenously administered placebo during ADVANCE or MOTIVATE received subcutaneous placebo Q8W during FORTIFY SS1.

Clinical advice to the EAG was that placebo is not used in routine practice. The CS states that placebo was used due to FDA and EMA requirements. As such, no directly comparative trial evidence was provided linking risankizumab to scoped comparators. Therefore, a network meta-analysis was presented by the company (see Section 3.4) to link clinical effectiveness estimates for risankizumab from included trials with clinical effectiveness estimates for scoped comparators identified from the wider literature.

3.2.2.5. Outcomes

The outcomes reported in the included studies are summarised in the CS Table 9 and reproduced below as Table 11. The EAG noted that clinical remission was measured using CDAI. Clinical advice to the EAG was that this outcome measure is of limited utility and is not used in UK clinical practice and that use of the Harvey Bradshaw Index would have been preferable. The EAG noted that company's response (clarification question A18) that CDAI is commonly used in clinical trials and that its use improved comparability of results across trials. The company explained (clarification question A16) that in the original protocol for the risankizumab Crohn's disease studies, the definition of the co-primary endpoint was patient-reported outcomes 2-item (stool frequency/abdominal pain score) (PRO2 [SF/APS]) clinical remission and endoscopic response. However, subsequent discussions with the FDA led to the creation of a US-specific protocol, which defined the co-primary endpoint as CDAI clinical remission and endoscopic response. An outside-US (OUS) protocol was created which retained the original definition of the co-primary endpoint, i.e., using PRO2 (SF/APS) to assess clinical remission. Consequently, the co-primary endpoint for the OUS protocol was clinical remission (PRO2 [SF/APS]) and endoscopic response, while the co-primary endpoint for the US protocol was clinical remission (CDAI) and endoscopic response. Both co-primary endpoints were measured at all trial sites, regardless of region. The only differences between the protocols are the outcomes used to determine clinical remission for the co-primary endpoints, the ranking of secondary endpoints and the sample size power calculation based on the revised co-primary endpoint.

The EAG noted that the definitions of CDAI remission (defined as a CDAI score of ≤ 150 points) and CDAI clinical response (defined as a reduction of 100 or more points from baseline) effectively meant that a participant could be defined as in remission without being defined as having clinical response. The EAG considered this to be a limitation of the use of this measure.

Table 11. Overview of outcomes in included trials

Trial no. (acronym)	M16-006 (ADVANCE)	M15-991 (MOTIVATE)	M16-000 Sub-Study 1 (FORTIFY)
Primary outcomes (including scoring methods and timings of assessments)	Definitions of coprimary endpoints: • CDAI clinical remission at Week 12: CDAI <150 • PRO2 (SF/APS) clinical remission at Week 12: average daily SF ≤2.8 and not worse than Baseline, and average daily AP score ≤1 and not worse than Baseline • Endoscopic response at Week 12: decrease in SES-CD >50% from Baseline (or for subjects with isolated ileal disease and a Baseline SES-CD of 4, at least a 2-point reduction from Baseline), as scored by central reviewer Assessments: • CDAI clinical remission: CDAI scores were calculated using a central laboratory Hct value from the same visit for all visits (Week 4, 8, 12/premature discontinuation, 16, 20, 24 or any unscheduled visit) except Baseline, where the most recent Screening Hct value was used‡‡ • PRO2 (SF/APS): Average daily SF, average daily AP score, and well-being were calculated from the subject diary at all visits (Baseline, Week 4, 8, 12/premature discontinuation, 16, 20, 24 or any unscheduled visit). The Screening period was a minimum of 7 days to calculate the Baseline scores. • Endoscopic response: an endoscopy was performed during screening,†† Week 12/premature discontinuation, Week 24 - The same endoscopist, where possible, performed all endoscopies		Definitions of co-primary endpoints: • CDAI clinical remission at Week 52: CDAI <150 • PRO2 (SF/APS) clinical remission at Week 52: average daily SF ≤2.8 and not worse than Baseline of the induction study, and average daily AP score ≤1 and not worse than Baseline of the induction study • Endoscopic response at Week 52: decrease in SES-CD >50% from Baseline of the induction study (or for subjects with isolated ileal disease and a SES-CD of 4 at Baseline of the induction study, at least a 2-point reduction from Baseline of the induction study), as scored by central reviewer Assessments: • The CDAI was calculated at each visit (Week 24, 52/premature discontinuation, any unscheduled visit, or rescue therapy visit). The scores calculated at the final visit in ADVANCE or MOTIVATE served as the Week 0 scores‡‡ • PRO2 (SF/APS): Average daily SF, average daily AP score, and well-being were calculated from the subject diary at each visit (Week 8, 16, 24, 32, 40, 48, 52/premature discontinuation, any unscheduled visit or rescue therapy visit). The scores calculated at the final visit in ADVANCE or MOTIVATE served as the Week 0 scores • Endoscopic response: An endoscopy was performed at Week 52/premature discontinuation

Trial no. (acronym)	M16-006 (ADVANCE)	M15-991 (MOTIVATE)	M16-000 Sub-Study 1 (FORTIFY)
	- Where possible, the Investigator or sub-Investigator was an endoscopist. All endoscopies were reviewed by a blinded central reviewer		- An endoscopy may have been performed at unscheduled visits to confirm inadequate response if hs-CRP and FCP are not elevated - The same endoscopist, where possible, performed all endoscopies - Where possible, the Investigator or sub-Investigator was an endoscopist. All endoscopies were reviewed by a blinded central reviewer
Other outcomes used in the economic model/specified in the scope	• Endoscopic remission at Week 12 • CDAI clinical response at Week 4 or Week 12 • EQ-5D-5L at Week 4 or Week 12		• Endoscopic remission at Week 52

Abbreviations: APS, abdominal pain score; CDAI, Crohn's disease activity index; FCP, faecal calprotectin; hs-CRP, high-sensitivity C-reactive protein; PRO2, patient reported outcome 2-item; SES-CD, simple endoscopic score for Crohn's disease; SF, stool frequency;

Notes: †† An endoscopy performed before the Screening visit, independently of the study, may have been used as the Screening endoscopy, with the approval of the AbbVie TA MD, if the following conditions were met: 1. Biopsy confirmation of the diagnosis was available according to section "Biopsy During Endoscopy" below, as applicable, 2. The endoscopy took place within 45 days prior to Baseline visit, 3. The endoscopy was recorded in a video format as the endoscopic eligibility will be determined by the central reviewers; ‡‡ The final CDAI for all other visits was calculated once the Hct value was received from the central lab. If the Hct was missing due to technical issues, the Hct value from the preceding visit may have been used.

3.2.2.6. Critical appraisal of the design of the studies

The company's approach to the critical appraisal of included trials was reported in the CS (Appendix D.3., Tables 26 and 27). Quality assessments for ADVANCE[20, 22]and MOTIVATE,[21, 22] the two risankizumab induction trials, as well as for FORTIFY,[23, 24] the maintenance trial for risankizumab, were conducted using the NICE clinical effectiveness quality assessment checklist for RCTs.[18] The EAG noted that the declaration of conflicts of interest was not assessed as part of the NICE guidance for quality appraisal.

The risk of bias of these trials was additionally assessed using the Cochrane risk of bias tool. The EAG considered these methods appropriate, though it was not clear why both methods were used, whether the outcomes of these assessments were considered together, or if the results of a specific tool were selected. The EAG also noted that the Cochrane risk of bias assessments included domains of the updated Cochrane risk of bias 2 tool,[19] but that no outcome-level assessments were conducted. In addition, the final question for domain 2, i.e. '2.7. If No/Probably No/No Information to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?' was entirely omitted. The EAG considered this to be an inappropriate and incomplete application of the tool that may have altered the overall risk of bias judgment of the assessed trials; however, given that this information is not used to inform any sensitivity analyses in the CS the impact is likely limited.

ADVANCE

Using the NICE guidance for quality appraisal of RCTs, the company appraised this trial as having no methodological concerns. No substantiation of these judgements were provided. The EAG considered the company's judgments related to randomisation, allocation concealment and baseline equivalence to be reasonable, given that patients were randomly assigned using interactive response technology as well as the unimportant differences between randomised groups at baseline (D'Haens 2022).[22] Furthermore, the EAG agreed with the company's judgments related to the lack of selective reporting, given the agreement between the primary publication (D'Haens 2022)[22] and the trial registry (NCT03105128), as well as the analytical approach, given that intention-to-treat analyses were conducted and conservative assumptions applied to the imputation of missing data. The EAG did not agree with the company's judgment related to blinding, since outcome assessors were not indicated as having been blinded, and considered 'No' to be a more appropriate judgment for this domain. The EAG also did not agree

with the company's judgment related to imbalances in loss to follow-up: while the overall attrition is acknowledged to be reasonably low at 7% (disregarding reasonable exclusions related to non-compliance of sites and low SES-CD participants), differential attrition was suggested by loss to follow-up of 5% and 4% in the risankizumab 600 mg and 1200 mg groups, respectively, versus loss to follow-up of 14% in the placebo group. As a result, the EAG considered 'Yes' to be a more appropriate judgment for this domain.

Using the Cochrane Risk of Bias tool, the company also appraised this trial as having no methodological concerns. As for the NICE quality appraisal, no substantiations accompanied these judgments. The EAG considered the judgments for domain 1 (Randomisation), domain 3 (Missing outcome data) and domain 5 (Selection of the reported result) to be reasonable, in line with the discussion related to the appraisal using the NICE guidance. It is not clear to the EAG, however, why the company indicated that carers and people delivering interventions were probably aware of assignment under question 2.2; particularly given the difference in the judgement for MOTIVATE, which was reported in exactly the same way. Furthermore, the answers to domain 2 (Deviations from intended interventions) represent an incorrect progression of the algorithm which could not have resulted in a domain-level judgment, unless this was overridden by the Assessor's judgment. The EAG considered that such a step should have been noted and justified. As was the case for the NICE guidance, there is no substantiating evidence to show that outcome assessors were blinded and therefore the EAG also considers that question 4.3 under domain 4 (Measurement of the outcome) was incorrectly assessed.

MOTIVATE

Using the NICE guidance for quality appraisal of RCTs, the company appraised this trial as having no methodological concerns. No substantiation of these judgements were provided. The EAG considered the company's judgments related to randomisation, allocation concealment and baseline equivalence to be reasonable, given that patients were randomly assigned using interactive response technology as well as the unimportant differences between randomised groups at baseline (D'Haens 2022).[22] Furthermore, the EAG agreed with the company's judgments related to the lack of selective reporting, given the agreement between the primary publication[22] and the trial registry (NCT03104413), as well as the analytical approach, given that intention-to-treat analyses were conducted and conservative assumptions applied to the imputation of missing data. The EAG did not agree with the company's judgment related to

blinding, since outcome assessors were not indicated as having been blinded, and considered 'No' to be a more appropriate judgment for this domain. The EAG also did not agree with the company's judgment related to imbalances in loss to follow-up: while the overall attrition is acknowledged to be reasonably low at 6% (disregarding reasonable exclusions related to non-compliance of sites and low SES-CD participants), differential attrition was suggested by loss to follow-up of 3% and 4% in the risankizumab 600 mg and 1200 mg groups, respectively, versus loss to follow-up of 14% in the placebo group. As a result, the EAG considered 'Yes' to be a more appropriate judgment for this domain.

Using the Cochrane Risk of Bias tool, the company also appraised this trial as having no methodological concerns. As for the NICE quality appraisal, no substantiations accompanied these judgments. The EAG considered the judgments for domain 1 (Randomisation), domain 3 (Missing outcome data) and domain 5 (Selection of the reported result) to be reasonable, in line with the discussion related to the appraisal using the NICE guidance. The EAG noted that the answers to domain 2 (Deviations from intended interventions) represent an incorrect progression of the algorithm which could not have resulted in a domain-level judgment, unless this was overridden by the Assessor's judgment; a step that should have been noted and justified, if this is the case. The company provided no substantiating evidence to show that outcome assessors were blinded during this trial, and therefore the EAG also considers that question 4.3 under domain 4 (Measurement of the outcome) was incorrectly assessed.

FORTIFY

Using the NICE guidance for quality appraisal of RCTs, the company appraised this trial as having no methodological concerns. No substantiation of these judgements were provided. The EAG considered the company's judgments related to randomisation, allocation concealment and baseline equivalence to be reasonable, given that patients were randomly assigned using interactive response technology as well as the unimportant differences between randomised groups at baseline[24]. The EAG agreed with the company's judgments related to blinding, given the quadruple blinding (participant, care provider, investigator and outcome assessor) reported in the trial registry (NCT03105102). Furthermore, the EAG agreed with the company's assessment regarding a lack of selective reporting, given the agreement between the primary publication[24] and the trial registry (NCT03105102), as well as the analytical approach, given that intention-to-treat analyses were conducted and conservative assumptions applied to the imputation of missing data. The EAG noted the company's judgment related to imbalances in

loss to follow-up: it considered overall attrition as fairly high at 11%, even when disregarding reasonable exclusions related to non-compliance of sites, low SES-CD participants and those with ineligible induction periods. However, no differential attrition was suggested by loss to follow-up of 8% and 12% in the risankizumab 180 mg and 360 mg groups, respectively, versus loss to follow-up of 12% in the placebo group. As a result, 'Yes' may possibly be a more appropriate judgment for this domain; however, the EAG accepted the company's judgment since no numerical cut-off value for 'high attrition' was stated in the CS.

Using the Cochrane Risk of Bias tool, the company also appraised this trial as having no methodological concerns. As for the NICE quality appraisal, no substantiations accompanied these judgments. The EAG considered the judgments for all domains to be reasonable, in line with the discussion related to the appraisal using the NICE guidance, and found no errors in the algorithm progression for this trial.

3.2.3. Description and critique of the results of the studies

3.2.3.1. Clinical effectiveness results

Disease activity (remission, response, relapse)

In ADVANCE, the co-primary endpoints of clinical remission (CDAI and PRO2 [SF/APS]) and endoscopic response were met for the risankizumab 600 mg IV arm when compared with the placebo IV arm. At week 12, a significantly greater proportion of participants in the risankizumab 600 mg IV arm achieved the co-primary endpoint of CDAI clinical remission versus the placebo IV arm (45.2% vs 24.6%, respectively; p<0.001). At week 12, a statistically significantly greater proportion of subjects in the risankizumab 600 mg IV arm achieved endoscopic response compared with the placebo IV arm (40.3% vs 12.0%, respectively; p<0.001). At week 4, significantly more participants in the riskankizumab arm achieved CDAI clinical response than those in the placebo arm (40.8% vs 25.2%, respectively; p<0.001).

In MOTIVATE, the co-primary endpoints of clinical remission (CDAI) and endoscopic response were met for the risankizumab 600 mg IV arm when compared with the placebo IV arm[21, 22]. At week 12, a significantly greater proportion of participants in the risankizumab 600 mg IV arm achieved the co-primary endpoint of CDAI clinical remission versus the placebo IV arm (42.0% vs 19.8%, respectively; p<0.001). At week 12, a statistically significantly greater proportion of participants in the risankizumab 600 mg IV arm achieved endoscopic response compared with the placebo IV arm (28.8% vs 11.2%, respectively; p<0.001). At week 4, significantly more

participants in the risankizumab arm achieved CDAI clinical response than those in the placebo arm (36.6% vs 20.9%, respectively; p=0.001).

In FORTIFY SS1, the co-primary endpoints of clinical remission (CDAI) and endoscopic response were met for the risankizumab 360 mg SC arms when compared with the SC placebo (withdrawal) arm according to the CSR[23] and primary publication[24] of this study (35.8% vs 15.9%, respectively; nominal p <0.001). The CS indicates that this result did not achieve statistical significance (NS) based on the pre-specified graphical testing procedure for the US-specific protocol, though the EAG noted the small nominal p-value. At week 52, a statistically significantly greater proportion of participants in the risankizumab 360 mg SC arm achieved CDAI clinical remission (as defined in Table 11) when compared to SC placebo (withdrawal) (52.2% vs 40.9%, respectively; p=0.005).

Among participants in FORTIFY SS1 who had CDAI clinical remission at week 0, a greater proportion in the risankizumab 360 mg SC arm achieved CDAI clinical remission at week 52 when compared to those re-randomised to SC placebo (withdrawal); however, statistical significance was not met according to the pre-defined testing procedure ([redacted] vs [redacted], respectively; nominal [redacted])

In FORTIFY SS1, a greater proportion of participants in the risankizumab 360 mg SC arm achieved SF remission at week 52 when compared to those receiving SC placebo (withdrawal); statistical significance was not met (57.0% vs 44.5%, respectively; nominal p=0.004). Similarly, a greater proportion of participants in the risankizumab 360 mg SC arm achieved AP remission at week 52 when compared to those receiving SC placebo (withdrawal), though statistical significance was also not reached (56.5% vs 46.3%, repesctively; nominal p=0.014).

A total of 29.1% of participants receiving risankizumab 360 mg SC achieved deep remission at week 52 of FORTIFY SS1 compared with 10.4% of those re-randomised to SC placebo (withdrawal) (difference of 18.8; nominal p<0.001). The CS indicates that this result did not achieve statistical significance (NS) based on the pre-specified graphical testing procedure for the US-specific protocol, though the EAG noted the small nominal p-value.

A total of 39% of participants treated with risankizumab 360 mg SC achieved endoscopic remission at week 52 of FORTIFY SS1 compared with 12.8% of those receiving SC placebo (withdrawal) (difference of [redacted]; nominal p<0.001). The CS indicates that this result did not

achieve statistical significance (NS) based on the pre-specified graphical testing procedure for the US-specific protocol, though the EAG noted the small nominal p-value.

As described in Appendix M.5.9. (p.383), participants entering FORTIFY SS1 who were treated with steroids (to a maximum of ≤20 mg/day prednisone or equivalent or ≤9 mg/day budesonide) were initiated on a mandatory steroid taper. Discontinuation of corticosteroid use in participants who were taking steroids at the baseline of ADVANCE or MOTIVATE is presented in Figure 18 of Appendix M.5.9. (p.384). The rates of steroid-free CDAI clinical remission were significantly higher with risankizumab 360 mg SC when compared to SC placebo (withdrawal) ([redacted] vs [redacted], respectively; [redacted]; Figure 19, Appendix M.5.9., p.385); as were steroid-free SF/APS remission ([redacted] vs [redacted], respectively; [redacted]; Figure 19, Appendix M.5.9., p.385), and steroid-free endoscopic remission ([redacted] vs [redacted], respectively; [redacted]; Figure 20, Appendix M.5.9., p.385).

A greater proportion of participants who received risankizumab 360 mg SC in FORTIFY SS1 achieved CDAI clinical response at week 52, when compared to those receiving SC placebo (withdrawal) (61.6% vs 48.2%, respectively; nominal p=0.002). The CS indicates that this result did not achieve statistical significance (NS) based on the pre-specified graphical testing procedure for the US-specific protocol, though the EAG noted the small nominal p-value.

At week 52, a statistically significantly greater proportion of participants in the risankizumab 360 mg SC arm achieved endoscopic response (as defined in Table 11) when compared to SC placebo (withdrawal) (46.5% vs 22.0%, respectively; p<0.001). Rates of steroid-free endoscopic response were significantly higher with risankizumab 360 mg SC when compared to SC placebo ([redacted] vs [redacted], respectively; [redacted]; Figure 20, Appendix M.5.9., p.385).

No relapse data were presented for any of the included studies.

Mucosal healing

No data for mucosal healing were presented in the CS, but were presented in data on file supplied in the company's reference pack. The proportions of participants with mucosal healing at week 12 on risankizumab 600 mg IV were [redacted] in ADVANCE and [redacted] in MOTIVATE, while on placebo IV it was [redacted] in ADVANCE and [redacted] in MOTIVATE. In the maintenance trial FORTIFY SS1, at week 52, these proportions were [redacted] on risankizumab 360 mg SC and [redacted] on SC placebo (withdrawal).

Surgery

No clinical data from the risankizumab trials are provided for surgery or colectomy in the CS or other supplied documents. The EAG noted that the CS stated that trial outcomes were reported according to the NICE scope, and its decision problem (Document B, Table 1) included surgery as an outcome. HRQoL and cost-effectiveness implications of surgery were factored into the economic model, though the EAG noted these values were based on Hospital Episode Statistics data, reported in a prior appraisal, as well as various assumptions.

Health-related quality of life

In ADVANCE, the risankizumab 600 mg IV arm was associated with statistically significant improvements in EQ-5D-5L at week 4 and week 12 compared with the placebo IV arm. For EQ-5D-5L Index Value scores, participants in the risankizumab 600 mg IV arm had a greater improvement from baseline (least squares [LS] mean) when compared with the placebo IV arm at week 4 ([redacted]) and week 12 ([redacted]). Similar results were observed for EQ-5D visual analogue scale (VAS) scores; participants in the risankizumab 600 mg IV arm had a greater improvement from baseline (LS mean) when compared with the placebo IV arm to week 4 ([redacted]) and week 12 ([redacted]).

In MOTIVATE, the risankizumab 600 mg IV arm was associated with statistically significant improvements in EQ-5D-5L from as early as week 4 and also at week 12 compared with the placebo IV arm. For the EQ-5D Index Value scores, participants in the risankizumab 600 mg IV arm had a greater improvement from baseline (LS mean) when compared with the placebo IV arm at week 4 ([redacted]) and week 12 ([redacted]).

In FORTIFY SS1, participants receiving risankizumab 360 mg SC had similar improvements in EQ-5D-5L Index Value scores from baseline of the induction study (LS mean from a mixed-effect model repeat measurement (MMRM)) to week 52 when compared to those receiving SC placebo (withdrawal). No significant differences in changes EQ-5D-5L Index Value scores from baseline to week 52 were found between these two trial arms ([redacted] vs [redacted], respectively; [redacted]). Participants receiving risankizumab 360 mg SC had a greater, but non-significant, improvement in EQ-5D-5L VAS scores from baseline of the induction study (LS mean from MMRM) to week 52 when compared to those receiving SC placebo (withdrawal). No significant

differences in changes EQ-5D-5L Index Value scores from baseline to week 52 were found between these two trial arms ([redacted] vs [redacted], respectively; [redacted]).

Participants receiving risankizumab 360 mg SC in FORTIFY SS1 had a numerically greater improvement in the Inflammatory Bowel Disease Questionnaire (IBDQ) total score from baseline of the induction study (LS mean from analysis of covariance (ANCOVA)) to week 52 when compared to those re-randomised to SC placebo (withdrawal); the difference in change from baseline between the two trial arms was not significant ([redacted] vs [redacted], respectively; [redacted]).

Similar changes were observed in the risankizumab 360 mg SC and SC placebo (withdrawal) arms of FORTIFY SS1 for the Functional Assessment of Chronic Illness Therapy (FACIT)-fatigue score with regards to change from baseline of the induction study (LS mean from ANCOVA) to week 52; the difference in change from baseline between the two trial arms was not significant ([redacted] vs [redacted], respectively; [redacted]).

Similar changes were observed in the risankizumab 360 mg SC and SC placebo (withdrawal) arms of FORTIFY SS1 for the Short Form 36-item health questionnaire (SF-36) physical component score with regards to change from baseline of the induction study (LS mean from ANCOVA) to week 52; the difference in change from baseline between the two trial arms was not significant ([redacted] vs [redacted], respectively; [redacted]).

Subgroup analyses

The company presented subgroup analysis (Appendix E) for participants who had prior TNF-alpha inhibitor failure and also for participants aged 16-17.

In ADVANCE, at week 12, a greater proportion of participants with prior TNF-alpha inhibitor failure, either participants who failed one inhibitor or those who failed >1 inhibitor, achieved CDAI clinical remission (CDAI <150) when compared with the placebo. The difference in response rate versus placebo was greater for participants who failed >1 inhibitor compared with those who failed 1 inhibitor ([redacted]% vs [redacted]%, respectively). Moreover, a greater proportion of participants with prior TNF-alpha inhibitor failure, either participants who failed one inhibitor or those who failed >1 inhibitor, achieved endoscopic response (decrease in SES-CD >50% from baseline [or for participants with isolated ileal disease and a baseline SES-CD of 4, ≥2-point reduction from baseline]) when compared with the placebo arm. The difference in response rate versus placebo was greater for participants who failed one inhibitor compared with those who failed >1 inhibitor ([redacted]% vs [redacted]%, respectively).

In MOTIVATE, at week 12, a greater proportion of participants with prior TNF-alpha inhibitor failure, either participants who failed one inhibitor or those who failed >1 inhibitor, achieved CDAI clinical remission (CDAI <150) when compared with the placebo arm. The difference in response rate versus placebo was greater for participants who failed one inhibitor compared with those who failed >1 inhibitor ([redacted]% vs [redacted]%, respectively). At week 12, a greater proportion of participants with prior TNF-alpha inhibitor failure, either participants who failed one inhibitor or those who failed >1 inhibitor, achieved endoscopic response (decrease in SES-CD >50% from baseline [or for participants with isolated ileal disease and a baseline SES-CD of 4, ≥2-point reduction from baseline]) when compared with the placebo arm. The difference in response rate versus placebo was greater for participants who failed >1 inhibitor compared with those who failed one inhibitor ([redacted]% vs [redacted]%, respectively).

In FORTIFY, at week 52, a greater proportion of participants with prior TNF-alpha inhibitor failure, either participants who failed one inhibitor or those who failed >1 inhibitor, achieved CDAI clinical remission (CDAI <150) when compared with the placebo arm. The difference in response rate versus placebo was similar for participants who failed one inhibitor compared with those who failed >1 inhibitor ([redacted]% vs [redacted]%, respectively). At week 52, a greater proportion of participants with prior TNF-alpha inhibitor failure, either participants who failed one inhibitor or those who failed >1 inhibitor, achieved endoscopic response (decrease in SES-CD >50% from baseline [or for participants with isolated ileal disease and a baseline SES-CD of 4, ≥2-point reduction from baseline]) when compared with the placebo arm. The difference in response rate versus placebo was marginally greater for participants who failed >1 inhibitor compared with those who failed one inhibitor ([redacted]% vs [redacted]%, respectively).

The company noted that there were low numbers of participants aged 16-17 in the included studies and cautions against drawing conclusions from the data. The EAG agreed that the subgroup analysis for participants aged 16-17, presented in Appendix E of the CS, does not offer robust results.

The company also presented analyses separated by Bio-IR vs Non-Bio-IR. These were presented in the main results section of the CS. However, the EAG presents these results in the subgroup analysis section, aligned with the decision problem.

In ADVANCE, CDAI clinical remission at week 12 was achieved by numerically more patients in the Non-Bio-IR group than the Bio-IR group (response rate difference vs placebo 25.8 vs 16.7, 95% CI 13.3, 38.3 vs 5.5, 27.8), although the confidence intervals overlapped. Similarly, there

was a numerically greater endoscopic response in the Non-Bio-IR group than the Bio-IR group (response rate difference vs placebo 37.7 vs 21.5, 95% CI 26.5, 48.8 vs 12.3, 30.7), although the confidence intervals overlapped. CDAI clinical response at week 12 was similar in the two groups (response rate difference vs placebo Bio-IR 24.2 vs Non-Bio-IR 21.7, 95% CI 12.4, 35.9 vs 8.2, 35.3). Endoscopic remission at week 12 was slightly higher numerically in the Non-Bio-IR group than the Bio-IR group (response rate difference vs placebo 17.9 vs 13.3, 95% CI 7.0, 28.8 vs 6.3, 20.3) and the confidence intervals overlapped.

In MOTIVATE, CDAI clinical remission at week 12 was numerically higher in those who had failed ≤1 prior biologics compared to those who failed >1 prior biologics (response rate difference vs placebo [redacted] vs [redacted], 95% CI 12.0, 38.4 vs 7.0, 31.7), although the confidence intervals overlapped. Endoscopic response at week 12 was slightly higher numerically in those who had failed ≤1 prior biologics compared to those who failed >1 prior biologics (response rate difference vs placebo [redacted] vs [redacted], 95% CI 7.6, 32.5 vs 5.5, 24.8) and the confidence intervals overlapped.

In FORTIFY sub-study 1, CDAI clinical remission at week 52 was numerically higher in the Bio-IR group than the Non-Bio-IR group (response rate difference vs placebo 12.7 vs 5.6, 95% CI -0.2, 25.6 vs -15.7, 26.9), while the confidence intervals overlapped. Endoscopic response was higher in the Non-Bio-IR group than the Bio-IR group (27.0 vs 23.4, 95% CI 6.3, 47.7 vs 11.4, 35.4).

Across studies, patients without prior biologic failure did better numerically, but the difference was not statistically significant with wide and overlapping confidence intervals indicating a lack of precision.

Adverse effects

Information on adverse events (AEs) is presented in the CS Section B.2.10. The EAG had no major concerns with the presentation of AE data.

The EAG agreed that risankizumab IV 600mg was generally well tolerated in both ADVANCE and MOTIVATE. The overall incidence of treatment-emergent adverse events (TEAEs) during the 12-week induction period was similar between the risankizumab 600 mg IV and placebo IV treatment arms (56.3% vs 56.5% in ADVANCE and 47.6% vs 66.2% in MOTIVATE). The rates of serious AEs (SAEs), severe AEs and AEs leading to discontinuation were numerically higher in the placebo IV arm than the risankizumab 600 mg IV arm. Two deaths occurred during

induction (ADVANCE), both of which were in the placebo IV arm. No deaths occurred in the risankizumab 600 mg IV arm.

The EAG agreed that risankizumab 360 mg SC as maintenance treatment for 1 year was generally well tolerated in FORTIFY sub-study 1. The incidence of TEAEs was 72.1% in the risankizumab 360 mg SC arm and 73.4% in the placebo SC (withdrawal) arm. The percentage of subjects with SAEs, severe AEs and AEs leading to discontinuation were comparable in risankizumab 360 mg SC and placebo SC (withdrawal) arms. There were no deaths reported during the maintenance study.

The EAG noted that publicly available information from clinical trial registries stated that an on-body injector was used for FORTIFY sub-study 4. This method of administration was not included in the CS, although the company stated that it intended that an on-body device would be intended to be used in clinical practice. It would be valuable to verify that this method of administration was covered in regulatory review.

3.3. Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

The company included 16 trials in its network meta-analyses (NMAs), covering a range of positions in the treatment pathway: this included nine induction, five maintenance and two induction/maintenance trials (CS Appendix D.1.3., Table 6). The CS included a summary of each (CS Appendix D.1.3.2). Most studies were multisite and international, though Watanabe (2012)[26] and Watanabe (2020)[27] were carried out in Japan only. Participants in included trials were CCF, BF or both (CS Appendix D.1.3.1.2., Table 9). The EAG noted that where trials presented findings for both CCF and BF, but did not stratify by these groups, they were reportedly excluded from analyses; however, it is not clear how many trials were excluded on this basis. A number of other exclusions are worth mentioning. One trial represented by two records was excluded on the basis of a treat-through design (i.e. without re-randomisation), and a further study was excluded on the basis of re-randomising based on remission rather than response. The EAG regarded that these exclusions were appropriate to reduce heterogeneity in the network.

Outcomes included in trials were CDAI remission and CDAI-100 response (CS Appendix D.1.3.1.2., Table 8). Doses varied as shown in CS Appendix D.1.3.1.1., Table 7; in particular, risankizumab doses are indicated as 600 mg IV for induction and 360 mg SC for maintenance.

The EAG presents two key domains for considering included trials: risk of bias of included trials, and differences across trials in design and patient population.

3.3.1. Risk of bias in included trials

The CS reported low risk of bias in all domains for included trials (CS Appendix D.3., Table 27), and largely acceptable assessments of trial quality (CS Appendix D.3., Table 26), the main limitation being several trials in which blinding was not achieved over all roles in the trial. The EAG was unable to independently replicate all assessments in the presented appraisals but noted that judgments relating to risk of bias domains followed from the presented judgments. There was no clear sign of imbalance across included treatments on risk of bias judgments.

3.3.2. Differences across trials in design and patient population

Included trials differ in a number of ways: these differences relate to design in terms of time of follow-up, and patient populations.

3.3.2.1. Time of follow-up

CS Appendix D.1.3.1.2., Table 9 details the week in which outcome data were collected for induction and maintenance. Maintenance outcome data were collected between week 44 and week 60, though networks were too sparse to comment on imbalance in time to follow-up. However, induction outcome data were collected between four and 12 weeks post-baseline. There is some evidence of imbalance in the distribution of follow-up times, with both risankizumab trials establishing post-induction outcomes at 12 weeks, while adalimumab and infliximab trials establish post-induction outcomes at four weeks. This variation is a potential source of heterogeneity, though the sparseness of networks precludes any formal meta-regression.

3.3.2.2. Patient populations

Another important way in which included trials differ is in included patient populations. The consequences of this are discussed below in Section 3.4.3 and 3.4.4. Trials varied across a range of effect modifiers. The company describes assessing included trials on the basis of these effect modifiers to establish transitivity of networks, and presented tabulated data relating to relevant effect modifiers in response to clarification question 12. The EAG regarded that there were no obvious sources of imbalance on the basis of these effect modifiers, which included

age, weight, duration of disease, CDAI score, inflammatory bowel disease questionnaire score, biomarkers and location of disease.

However, there were two important remaining sources of effect modification. The first is overall comparison group risk, which differed systematically by trial. This is important because imbalances in baseline risk across the network, which were evidence in included NMAs, creates likely effect modification. The second is that stratification by CCF and BF, while useful to distinguish between two clinically relevant subgroups, does not solve the issue of treatment history heterogeneity in the BF group. Specifically, the overall population implied by NMAs for the BF population would not strictly be at risk of every treatment of the network. This is because by definition, experiencing the failure of one biologic treatment suggests that not all subsequent treatments are appropriate treatment choices. The implication of this for NMA estimates is unclear, and the specific provenance of the BF subgroups in the analysis is unclear; that is, whether all trials contributing to the BF NMA defined the subgroup similarly.

3.4. Critique of the indirect comparison and/or multiple treatment comparison

3.4.1. General approach

Analysis of NMAs was carried out in a Bayesian framework by two binomial methods: ‘standard’ logit link and risk-difference (RD). The binary outcomes assessed were CDAI clinical remission and CDAI-100 response (defined in Document B, Table 35). The former is an absolute measurement while the latter a change in score over baseline level: alternative analysis, treating these as ordinal measurements, is therefore not appropriate. In the RD case extra steps were required and taken to ensure risk estimates are bounded between [0,1] following Warn et al. (2002). In the view of the EAG, these binomial analyses are appropriate and, depending on circumstances, recommended by TSD2.

The company provided a copy of their NMA code: logit-link analysis was carried out with *bnma* package in R; baseline modelling and RD analysis with WinBugs (driven from R). Data and control parameters were not supplied initially, resulting in the EAG requesting complete code in clarification question A7. The code provided at following clarification appeared clear and well-programmed, although on attempting to run errors were encountered with undefined variables being referenced. Furthermore, the code as configured did not replicate the results used in the model without amendment.

The CS reported difficulties with the risk-adjusted logit link analyses (section B.2.9.3.1), and in clarification question A4 the company further explained that the adjusted logit-link random-effects (RE) model failed to converge, while in the adjusted logit-link fixed-effect (FE) model the regression term was not significantly different to zero (that is, an unadjusted model was not rejected). Given the problematic adjusted logit-link analysis, the company went on to argue that the RD analysis (which the EAG believes to be unadjusted) is preferred to the unadjusted logit-link analysis (B.2.9.3.1, CQ A4). The EAG is not aware of any strong rationale for or against this point of view. This is discussed further in Section 3.4.5.

Following this reasoning, the CS only gives NMA effect estimates for the RD approach. The estimated risk differences between treatment and placebo are combined with a baseline risk to give absolute risks under treatment (see Appendix P1.1. tables). These risks are inputs to the cost-effectiveness model.

The EAG found the CS somewhat unclear about why all parts of the reported analyses were not applied, where relevant, to each of the unadjusted and adjusted logit-link and the RD analyses. The EAG understands that only in the unadjusted logit-link analyses was there an assessment of consistency, and only for the logit-link model an attempt to adjust by regression for varying baseline risk.

In the CS, network nodes were defined by treatment and dose (e.g. ADA160/80 and ADA80/40 are separate nodes). The EAG agreed with this approach to setting up nodes, which is in line with the recommendations of Dias et al. (2018). Separate networks were used for CCF and BF subgroups (also referred to as non-Bio IR and Bio IR, respectively, in the trials) and induction and maintenance phases. The company further chose to separate the maintenance phase into two networks ‘based on biologic half-life, induction duration, and study heterogeneity’ (reported in Document B, Section B.2.9.1.4): risankizumab and ustekinumab vs TNFi (adalimumab and infliximab) and vedolizumab. The EAG was not convinced by this rationale and queried this decision; this is further covered in Section 3.4.6 below.

With respect to between-trial heterogeneity, an FE framework was used in the company base case and an RE model in a scenario analysis. The company argued that the FE model was preferred given similar deviance information criterion (DIC) values with the RE models and for ease of interpretation. In clarification question A5 the company further explained that under an RE model credible intervals included values that favoured placebo over biologics, and the company concluded that the RE model therefore lacked face validity. The EAG disagrees,

however, that an RE model should have been discarded given the manifest heterogeneity in the analysis, noting that informative priors should have been considered to produce more plausible results.

The CS presents results for logit-link FE NMAs compared with an inconsistency ('Unrelated Mean Effects') model in the CS (Document B, Tables 44 and 45). The differences in residual deviance where available are small (<1) implying no evidence of inconsistency, though there are few loops in the networks to do so. The EAG further notes that the residual deviances are of similar magnitude to the number of data points, indicating an acceptable model fit.

NMAs are susceptible to bias when there is variation in effect modifiers across the network. The CS lists potential effect modifiers, and the company supplied a summary of these in clarification. Some potential effect modifiers were controlled by design (for example, all trials used outcomes on the CDAI scale). The EAG believes that potentially the most problematic are variations in previous treatments between trials and associated differences in patient populations. Section 3.4.4 contains further discussion of effect modification. The CS indicated that an adjustment was made for baseline risk, where baseline risk was a proxy for an unspecified set of effect modifiers. The EAG agreed that this step could help protect against bias, but concluded that it was not carried out in the RD NMAs (see section 3.4.5 below for further discussion).

3.4.2. NMA results

The CS presents results for the company's base case network configuration RD model with FE in Document B and with RE in Appendix P1.2. These results are discussed in the induction setting below. For the maintenance setting, the EAG preferred a different network configuration; these results were presented in clarification response. As per the CS, 'significance' denotes credible intervals not crossing zero.

3.4.2.1. Induction

CS results in Document B were provided as RDs. In the following text, the EAG used a RD threshold of [redacted] as an indication of 'substantial' magnitude (unrounded figures will be found in the tables) with the strong caveat that precision of the estimates is generally low.

Under induction, risankizumab and most comparator drugs showed statistically significant improvement over placebo with substantial point RDs. In the BF subgroup, risankizumab was substantially favoured (i.e. with point magnitude RD [redacted]) over three of four comparators, with evidence of a statistical difference for two of these (VDZ300 and UST6) but not for ADA80/40. In

the CCF subgroup, there was not statistical evidence of a difference between risankizumab and any comparators, though several point magnitudes were substantial, favouring risankizumab over VDZ300 and ADA80/40, but favouring IFX5 over risankizumab.

Table 12: Summary of treatment effect estimates (RD) with CrIs under induction on CDAI remission of risankizumab versus comparators from FE NMA

	CCF[a]	BF[b]
	RZB600	**RZB600**
PBO	[redacted]	[redacted]
ADA80/40	[redacted]	[redacted]
VDZ300	[redacted]	[redacted]
UST6	[redacted]	[redacted]
ADA160/80	[redacted]	[redacted]
IFX5	[redacted]	NA

Abbreviations: ADA, adalimumab; BF, biologic care failure; CCF, conventional care failure; CDAI, Crohn's Disease Activity Index; CrI, credible interval; FE, fixed effect; IFX, infliximab; NA, not applicable; NMA, network meta-analysis; PBO, placebo; RD, risk difference; RZB, risankizumab; UST, ustekinumab; VDZ, vedolizumab.

Sources: CS Document B, [a] Table 36; [b] Table 37

NMA results comparing risankizumab against comparators for CDAI-100 over induction under the FE model are shown in Table 13. These results were extracted from full tables in the CS. The RD estimates significantly favoured risankizumab over placebo in both CCF and BF subgroups, with substantial point magnitude in both. In BF, risankizumab was significantly favoured, with substantial point magnitudes, over all but one (ADA80/40) of its comparators. In CCF, differences from other comparators were not statistically significant, but point magnitudes were approaching substantial in one case, favouring risankizumab over VDZ300.

Results for the RD models with RE rather than FE are given in CS Appendix P.1. Confidence intervals were wider and point estimates similar, as anticipated.

Table 13: Summary of treatment effect estimates (RD) with CrIs on CDAI-100 clinical response of risankizumab versus comparators from FE NMA over induction

	BF[a]	CCF[b]
	RZB600	**RZB600**
PBO	[redacted]	[redacted]
VDZ300	[redacted]	[redacted]
UST6	[redacted]	[redacted]

	BF[a]	CCF[b]
ADA160/80	[redacted]	[redacted]
ADA80/40	[redacted]	[redacted]

Abbreviations: ADA, adalimumab; BF, biologic care failure; CCF, conventional care failure; CDAI, Crohn's Disease Activity Index; CrI, credible intervals; FE, fixed effect; NMA, network meta-analysis; PBO, placebo; RZB, risankizumab; UST, ustekinumab; VDZ, vedolizumab.

Sources: CS Document B, [a] Table 39; [b] Table 38

3.4.2.2. Maintenance

The NMA results for reconfigured maintenance networks were provided in the response to clarification question A15. The EAG preference is for a single maintenance network (issue detailed in Section 3.4.6). The single network results for CDAI remission in CCF and BF populations are reproduced here in Table 14 and Table 15, respectively. Note these results were received as absolute risks, whereas the induction results (Section 3.4.2.1) are risk differences.

Table 14 : Single maintenance NMA network results for CDAI remission: CCF population

Treatment	Median	Lower CrI	Upper CrI
ADA 40 QW	[redacted]	[redacted]	[redacted]
ADA 40 Q2W	[redacted]	[redacted]	[redacted]
IFX 5/10 Q8W	[redacted]	[redacted]	[redacted]
UST Q8W	[redacted]	[redacted]	[redacted]
VDZ IV Q8W	[redacted]	[redacted]	[redacted]
IFX5 Q8W	[redacted]	[redacted]	[redacted]
VDZ IV Q4W	[redacted]	[redacted]	[redacted]
UST Q12W	[redacted]	[redacted]	[redacted]
VDZ SC Q2W	[redacted]	[redacted]	[redacted]
RZB Q8W	[redacted]	[redacted]	[redacted]
PBO	[redacted]	[redacted]	[redacted]

Abbreviations: ADA, adalimumab; CCF, conventional care failure; CrI, credible interval; IFX, infliximab; IV, intravenous; PBO, placebo; QxW, every x weeks; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab

Source: Company clarification response, Table 24

Table 15: Single maintenance NMA network results for CDAI remission: BF population

Treatment	Median	Lower CrI	Upper CrI
ADA 40 QW	[redacted]	[redacted]	[redacted]
ADA 40 Q2W	[redacted]	[redacted]	[redacted]

Treatment	Median	Lower CrI	Upper CrI
VDZ SC Q2W	[redacted]	[redacted]	[redacted]
VDZ IV Q8W	[redacted]	[redacted]	[redacted]
VDZ IV Q4W	[redacted]	[redacted]	[redacted]
UST Q8W	[redacted]	[redacted]	[redacted]
RZB Q8W	[redacted]	[redacted]	[redacted]
UST Q12W	[redacted]	[redacted]	[redacted]
PBO	[redacted]	[redacted]	[redacted]

Abbreviations: ADA, adalimumab; BF, biologic failure; CrI, credible interval; IV, intravenous; PBO, placebo; Q2W, every 2 weeks; Q4W, every 4 weeks; Q8W, every 8 weeks; Q12W, every 12 weeks; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

Source: Company clarification response, Table 25

Remission on risankizumab is relatively low among the comparators in CCF ([redacted]) and in BF ([redacted]). Adalimumab has the highest median remission rates in both CCF and BF, and is the only treatment significantly better than placebo regardless of dose. All comparators perform better than placebo, though the difference is not always significant.

3.4.3. Baseline risk

The company modelled *'the reference treatment (placebo in all instances) …using a baseline natural history model that was constructed independently from the model of relative treatment effects'* (D1.3.3.10). The EAG agrees that the separation of modelling is as advised by TSD5.

The EAG consulted a clinician on whether or how the trial placebo arm mapped to a pathway/health state in UK clinical practice. The trial concept of placebo is active treatment withheld for the duration of the trial or withheld altogether. Clinical advice to the EAG was that withholding/delaying comparator treatments to risankizumab would be an unsatisfactory clinical practice usually only necessitated when patients are seriously ill or steroid-dependent. Also, in clinical practice, if patients did not respond overall to any of the comparator drugs they are not returned to standard care, but alternative strategies are sought (new drugs via trials, repeat TNFis or combinations of drugs). Further details on trials for comparator drugs were provided in CS Appendix D1.3.2 – this seems to indicate that patients were generally permitted concomitant medication (aminosalicylates, immunomodulators, corticosteroids, antibiotics).

Because there is no real-world 'placebo' treatment and because trial placebo arm participants generally receive conventional care medicines, the EAG believes that trial control arms, as

opposed to observational studies, are the most likely and perhaps only source of data for baseline risk, as has been used in the CS.

The posterior baseline risks estimated by the company model are summarised in Table 16. In each case, the placebo arms of every trial in the NMA contribute data for these estimates. The question arises whether a subset of these trials would give better a representation of UK clinical practice. For example, the trials by Watanabe et al (2012) [26]and (2020) [27]were carried out in Japan only.

The trial-level proportion of patients in response or remission are shown in Figure 2, based on data supplied at clarification (CQ A7). Apart from CDAI-remission at maintenance, there is considerable variation. The CS base case approach used a FE model (Appendix D D1.3.3.10), but with this level of heterogeneity the EAG prefers a RE model.

Table 16: Placebo response rates collated by EAG

Comparators in network	Treatment phase	Population	Outcome	Estimate of % attaining outcome under 'placebo'	CrI
RZB, UST, ADA, IFX, VDZ, PBO	Induction	CCF	CDAI remission	[redacted]	[redacted]
		BF		[redacted]	[redacted]
		CCF	CDAI-100 response	[redacted]	[redacted]
		BF		[redacted]	[redacted]
RZB, UST, PBO	Maintenance	CCF	CDAI remission	[redacted]	[redacted]
		BF		[redacted]	[redacted]
ADA, IFX, VDZ, PBO	Maintenance	CCF	CDAI remission	[redacted]	[redacted]
		BF		[redacted]	[redacted]

Abbreviations: ADA, adalimumab; BF: biologic failure; CCF, conventional care failure; CrI, credible interval; IFX, infliximab; PBO, placebo; RZB, risankizumab; UST, ustekinumab; VDZ, vedolizumab

Sources: Tables 113-120, Appendix P.1.1.

Arguments regarding placebo heterogeneity, in particular in the maintenance trial data, are not convincing, and the use of two different proportions reaching remission according to comparators under consideration is problematic. The maintenance trial data shown in Figure 2 have been supplemented with the trial start date (sourced from www.clinicaltrials.gov) for both the biologic failure and conventional care failure subgroups and illustrated in Figure 3 and Figure 4 respectively. Both figures suggest that the observed heterogeneity can be largely attributed to a temporal effect, with improvements in available concomitant treatment options.

Such a model is preferred for estimation of a single maintenance placebo remission proportion at a suitable timepoint, which would be the basis for absolute estimates for use in the cost-effectiveness model generated in combination with relative effects estimated from a single maintenance network.

Figure 2: Proportion responding or remitting in the placebo arm of each trial in the NMAs (EAG plot).

Abbreviations: BF, biologic failure; CCF, conventional care failure; IND, induction; MAINT, maintenance
Sources: Data supplied in company's response to clarification questions A7

Figure 3: Proportion of placebo arm patients achieving remission following biologic failure (EAG plot)

Abbreviations: BF: Biologic failure
Sources: Data supplied in company's response to clarification questions A7; www.clinicaltrials.gov

Figure 4: Proportion of placebo arm patients achieving remission following conventional care failure (EAG plot)

Abbreviations: CCF: Conventional care failure
Sources: Data supplied in company's response to clarification questions A7; www.clinicaltrials.gov

3.4.4. Effect modifiers

Differences in effect modifiers across trials can lead to bias in NMA estimates. The CS identified potential effect modifiers in Appendix D (pp.56-57).

Some potential effect modifiers were addressed with varying success by the design of the NMAs: the outcomes scale was homogeneous (CDAI used throughout); CCF vs BF subgroups were analysed separately; and follow-up periods were at some variance: between 4 and 12 weeks for induction studies and, more consistently, between 44 and 60 weeks for maintenance studies (App D Table 9). The separate analysis of CCF and BF patients in the CS has the effect of creating a crude subgrouping by prior treatment. Nevertheless, CCF may have included patients who had biologic treatment in the past and stopped for reasons other than inadequate response/failure (Appendix D1.3.3.3). BF patients will have received one or more biologics

previously and had inadequate response/treatment failure, but the line of treatment and their composition may be heterogeneous.

Maintenance trial study design was a potential effect modifier, but in all included maintenance trials, patients were randomised at maintenance (and perhaps also at induction). Another design potentially applied in maintenance trials is a ‘treat-through’ design in which patients are randomised prior to induction only: this was excluded by the company’s SLR (Appendix D Table 5, records 226-227), resulting in exclusion of 2 records, both reporting on the SEAVUE trial comparing adalimumab and ustekinumab. In most of the maintenance trials, the participants were responders to induction treatment. In the CHARM study all patients, regardless of response, were randomized to one of three maintenance treatment groups (adalimumab 40 mg Q2W, adalimumab 40 mg QW, or placebo) for an additional 52 weeks following a 4-week induction phase; however, randomisation was stratified by response status at week 4. The EAG also noted the exclusion of the CLASSIC II trial, where participants were re-randomised based on clinical remission instead of response. Overall this aspect of design heterogeneity was judged by the EAG to be well-controlled by the company’s SLR and exclusions were considered to reduce heterogeneity and intransitivity in the NMA.

Trial-level values of potential effect modifiers were supplied in response to CQ A12 (Tables 9-14). The EAG observed consistency in a number of the variables, including weight (usually averaging around 70 kg, though a study in Japan (Watanabe et al 2012) [26] was clearly different averaging around 55 kg), and age (averaging 30 to 40 years). Clinical advice to the EAG suggested that age, duration of disease, C-reactive protein (CRP) and gastrointestinal areas involved are effect modifiers for response to treatment. Of these, most showed considerable variability: duration of disease ranged 4.4 to 12.7 years; CRP levels from 7.8 to 30 mg/L; ileal involvement from 9 to 75%; colonic involvement from 14 to 68% and ileo-colonic involvement from 9 to 70%. Only age was considered to be homogeneous enough to have limited implications for effect modification.

Prior treatment was by necessity recorded only crudely in the summary tables, but appeared to be rather variable. There is also variability in the exclusion criteria of the NMA trials (see D1.3.2): for example, CLASSIC excludes patients if TNFi previously received, while GEMINI2 excludes patients previously on vedolizumab, natalizumab, efalizumab or rituximab. In the summary tables a division is clearly seen between trials with no history or no failures of TNFi

recorded as zero or perhaps NR, perhaps due to exclusion criteria (e.g. ACCENT 1), and others with counts over zero.

An example of variation in prior treatment history across a network is seen when comparing RZB with placebo with ustekinumab (Document B, Figure 8). Patients previously receiving IL-12 or IL-23 antagonists in the IM-UNITI trials were excluded, while in contrast up to 20% patients entering induction via the ADVANCE and MOTIVATE trials and prior to the maintenance period of the FORTIFY trial might receive ustekinumab, an IL-23/IL-12 inhibitor.

3.4.5. Adjustment for baseline risk

The analysis of the CS aims to adjust for risk in the placebo arm (D1.3.3.6), this acting as a proxy for the combined influence of known and unknown effect modifiers. The EAG agrees with this aim, because the baseline risk is a logical proxy and because it is known to be heterogeneous (see Figure 2, Section 3.4.3).

The logit-link NMA adjusted for baseline-risk using 'standard' code supplied by Dias et al/TSDs (coded in R with package '*bnma*'). The baseline-risk adjusted model for logit-link contains a coefficient to represent a linear change in risk with trial-level difference from overall average (treatment x covariate interaction, with baseline risk as the covariate). This regression term is coded for using the *bnma* package, and also referred to in Appendix D1.3.3.6, equation 6, with respect to the logit-link analysis. The CS indicated the use of a 'common baseline' model as a response to the sparsity of the network (CQ A6) – the EAG accepts this reasoning.

No analogous regression modelling of baseline risk appears to have been used in the risk difference model. The EAG believes the RD model *does not* adjust for baseline risk, because there is no regression term of a form similar to beta * (x – mx) (see Dias 2018- [28] p243 ff.) in the company RD code. The CS argues (B.2.9.3.1) *'absolute probabilities of treatment response were subtracted across interventions in RD models, minimising potential impacts of overly low or high placebo efficacy. This may help minimise bias when there are imbalances in the number of studies with low placebo response rates across pairwise contrasts in the network'.* The EAG agrees that RD model *does account* for differences in baseline risk in the usual way because there is an uninformative prior on baseline risk, but it *does not adjust* for baseline risk via meta-regression.

The CS indicates that *"The NMAs used in this submission utilised the RD method, which was used in this instance as it is recommended where baseline risk-adjusted models are deemed*

inappropriate due to lack of convergence or face validity" referring to TSD2. The EAG was not able to locate this recommendation in TSD2, though it is a logical response to the difficulties. Additionally, CS section B.2.9.3.1 says *"TA521 concluded that baseline-risk adjusted models and risk difference NMAs should yield less biased estimates of effect than the unadjusted NMA analyses on the relative scale"*. The EAG identified the following passage in TA521 that makes the following argument*: "We also presented an alternative approach to adjust for cross-trial differences using risk differences, as opposed to relative effects. Rather than divide by low placebo response rates, which inflate relative effects, differences in absolute probabilities across treatments are subtracted (i.e., treated as risk differences). This may help minimize bias when there are imbalances in the number of studies with low placebo response rates across pairwise contrasts in the network*." (response to clarification 7(f), Committee Papers pp196-197). The EAG did not find this argument wholly persuasive, since the logit-link transforms to a linear scale in which treatments effects are also additive. The EAG accepts the company point that in the comparison of RD versus logit-link results on an OR scale presented in response to clarification question A10, there are some 'extreme' variances for logit-link estimates on the OR scale, though the EAG also notes that the RD variances can also be very large (e.g. IFX IV vs PBO [redacted], clarification question response, Table 2).

The EAG concludes that both logit-link and RD models take account of varying baseline risk in the standard way, with the inclusion of modelling terms for the control arm risk and accompanying uninformative priors. However, while *adjusting* for baseline-risk as a proxy for various effect modifiers would be desirable, this adjustment has *not* been included in the company's base case RD model. Baseline-risk adjustment *was* carried out using the logit-link model (D.1.3.3.6) but analysis was problematic (B2.9.3.1 and CQ A4 and A10) and not reported in the CS.

3.4.6. Separation of maintenance network

The CS separated the maintenance evidence into two networks risankizumab/ustekinumab vs adalimumab/infliximab/vedolizumab) *'based on biologic half-life, induction duration, and study heterogeneity'* (B2.9.1.4). The EAG disagrees with this approach in general because network formation is recommended on the basis of comparator connections (Dias et al 2018, section 1.6.1), not drug characteristics. On top of this, the EAG noted in clarification question A15 the similarity in the half-life and induction period between vedolizumab and ustekinumab, and that they are seen as similar therapy options.

In response to clarification question A15 the company argue that vedolizumab has a different biological mechanism to IL inhibitors, ustekinumab and risankizumab, and is therefore not appropriate to include in that network. The EAG finds the argument inconsistent because the alternative network is made up of the TNFis, infliximab and adalimumab, which also have different biological mechanism.

In clarification question A15 and CS B2.9.3.2 the company argues that their chosen separation of the networks, grouping vedolizumab with adalimumab and infliximab, mitigates placebo arm heterogeneity. However, the EAG notes from the data plotted in Figure 2 that the placebo arm remission rate for the VISIBLE2 trial (VDZ vs PBO) is actually rather high, and closer to placebo rates in FORTIFY (RZB vs PBO) and IM-UNITI (UST vs PBO). The EAG considered this lack of placebo arm dissimilarity as further evidence that the networks should be combined.

A final argument made by the company in favour of its base case separated networks is that "*the single maintenance NMA network does not stand up to basic face validity as the outputs suggest that in some cases placebo is more effective than ustekinumab, vedolizumab and risankizumab; this observation goes against the results presented in the Phase 3 clinical trials of the respective biologic therapies"*. The EAG noted that the company's preferred NMA base case, comprising two disconnected networks, also had cases where active treatments were not significantly better than placebo. It considered that this, by the company's reasoning, also lacks face validity when compared to individual trial results. As such, the EAG considered these findings for both the disconnected and combined network to be methodologically driven, and not an issue of face validity.

The EAG requested (clarification question A15) further analyses (1) grouping vedolizumab in the network with risankizumab and ustekinumab instead of the TNFis, and (2) grouping all treatments (TNFis, vedolizumab, ustekinumab and risankizumab) together in a single network. The single-network results are outlined in 3.4.2.2 and forms the EAG's preferred base case.

3.5. Additional work on clinical effectiveness undertaken by the EAG

None.

3.6. Conclusions of the clinical effectiveness section

The EAG considered that the company's SLR had generally been conducted adequately, although certain limitations were noted, particularly with regard to the assessment of risk of bias. The company's decision problem generally aligned with the NICE scope, but the EAG noted in

particular that subgroup analysis by CD location had been excluded from the company decision problem. The EAG did not consider that this exclusion was suitably justified. The EAG also noted that no results were presented for surgical outcomes. The EAG considered that generally the company's SLR and included trials were adequately described, although certain information was not described in sufficient detail.

Three clinical trials were included in the CS. There were two phase 3 multicentre, randomised placebo-controlled induction trials (ADVANCE and MOTIVATE) plus one Phase 3, multi-centre, partially randomised, double-blind, placebo-controlled, 52-week maintenance study with an ongoing open-label extension (FORTIFY). Only sub-study one from FORTIFY was included in the CS. In ADVANCE and MOTIVATE, riskankizumab was administered intravenously 600 mg or 1200 mg Q4W by a clinician. In FORTIFY sub-study one, risankizumab was administered subcutaneously 360 mg Q8W or 180 mg Q8W by a clinician. The EAG noted that the proposed method of administration for clinical practice using an on-body device differed from the method of administration used in the included trials. However, publicly available information from clinical trials registries stated that an on-body injector was used in FORTIFY sub-study four, which was not used in the CS. No studies in the CS directly compared riskankizumab with any scoped comparators. The EAG was satisfied that based on the included trials in the CS there was evidence for a benefit for risankizumab against placebo for remission, response, mucosal healing and health-related quality of life.

NMA methods were broadly appropriate, though the EAG regarded that RE models would have been more suitable, and highlighted challenges with the body of evidence (prior history of treatments, baseline risk) that challenge interpretation of analysis. The EAG considered the use of a single, connected maintenance network to be preferable to the split networks provided as the company base case NMA, as it was unconvinced by the company's clinical rationale for splitting the network. In induction meta-analyses, risankizumab was not significantly better than any other active comparator for remission in the CCF population, though risankizumab was numerically superior to most. In the BF population, risankizumab was numerically superior to all comparators and significantly better than several of them. In the maintenance meta-analyses of the connected network, risankizumab was numerically superior only to placebo in the CCF population and only placebo and ustekinumab Q12W in the BF population for remission; it was not significantly better than any comparator in either of these populations.

The EAG considered the methods used to assess the quality of the three risankizumab trials (ADVANCE, MOTIVATE and FORTIFY), as well as trials included in NMA, to be an appropriate selection of methodological approach. However, the application of the Cochrane risk of bias tool was considered to have limitations (see Section 3.2.2.6). The EAG noted that these may have altered the overall risk of bias of assessed trials, though the impact was judged to be minimal since no approaches assessing the robustness of effectiveness results (e.g. sensitivity analyses) were informed by methodological quality of the included trials. In terms of specific trial-level judgments the EAG was mostly in agreement with the company's appraisal, though it disagreed with the assessment of risk of bias related to blinding for ADVANCE and MOTIVATE and considered that attrition may also have been of concern for these induction trials; it also flagged uncertainty around the judgment of attrition bias in FORTIFY. The EAG did not assess quality assessments of other trials included in the NMA independently (see Section 3.3.1).

The following clinical effectiveness key issues were identified:

- Unexplored heterogeneity in network meta-analyses in relation to baseline risk
- Network structure in maintenance network meta-analyses should be connected

Additionally, the EAG considered that the following key issues also had relevance to the clinical effectiveness evidence:

- Feasibility of exploratory subgroup analysis by CD location (decision problem key issue)
- Method of administration for risankizumab (other key issue)

4. COST-EFFECTIVENESS

4.1. EAG comment on company's review of cost-effectiveness evidence

Appendices G, H and I of the CS detail systematic searches of the literature used to identify cost effectiveness, health-related quality of life, healthcare resource use and costs evidence, critique is provided in Table 17, Table 18, and Table 19. Searches and eligibility criteria were appropriate and therefore it is unlikely that relevant studies were missed.

Table 17. Summary of EAG's critique of the methods implemented by the company to identify cost-effectiveness evidence

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix G, Table 40.	The company literature search appears to be well conducted. The cost effectiveness filter that was used does not appear to be a tested filter;[29] this makes the effectiveness of the search uncertain and it is possible that some relevant papers may have been missed.
Inclusion criteria	Appendix G, Table 41	The inclusion criteria were broad and therefore likely to have captured the available evidence. The company included a total of 69 studies of which seven analyses were relevant to the UK.[30-36] A summary was provide in Table 55 of the CS (Document B). None of the identified cost-effectiveness analyses evaluated Risankizumab. The company made reference to two previous NICE technology appraisals – TA352 and TA456.
Screening	Appendix G, Section G.1.2	Titles and abstracts were screened by two independent reviewers and disagreements were resolved by consensus or by a third reviewer. Full texts were also screened by the two reviewers and disagreements resolved in the same way.
Data extraction	Appendix G, Section G.1.2	Data extraction was completed by one reviewer with a senior reviewer checking the extraction and disagreements resolved through discussion.
QA of included studies	Appendix G, Section G.2	The methodological quality of included full text publications was assessed using the Drummond checklist for cost-effectiveness studies.[37]

Abbreviations: CS, Company Submission; EAG, Evidence Assessment Group; HRQoL, health-related quality of life; QA, quality assessment

Table 18. Summary of EAG's critique of the methods implemented by the company to identify health related quality of life

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix H, Table 46.	The company literature search appears to be well conducted and a good range of sources were searched.
Inclusion criteria	Appendix H, Table 47	The inclusion criteria were broad and therefore likely to have captured the available evidence. A total of 142 studies (reported in 204 publications) were included. The majority of studies were conducted in US, Canada, EU-5, Australia and Japan, and were selected for data extraction. The remaining studies were not extracted as the geography was not relevant.
Screening	Appendix H, Section H.1.2	Titles and abstracts were screened by two independent reviewers and disagreements were resolved by consensus or by a third reviewer. Full texts were also screened by the two reviewers and disagreements resolved in the same way.
Data extraction	Appendix H, Section H.1.2	Data extraction was completed by one reviewer with a senior reviewer checking the extraction and disagreements resolved through discussion.
QA of included studies	Appendix H, Section H.2	The methodological quality assessment for utility studies was performed using the NICE checklist, while the quality assessment for HRQoL studies was performed using the Efficace checklist.[38]

Abbreviations: CS, Company Submission; EAG, Evidence Assessment Group; HRQoL, health-related quality of life; QA, quality assessment

Table 19. Summary of EAG's critique of the methods implemented by the company to identify healthcare resource use and costs

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix I.1	The company literature search appears to be well conducted and a good range of sources were searched. The same literature search strategy was used for the cost effectiveness searches, see Table 17
Inclusion criteria	Criteria reported in Appendix G, Table 41 – healthcare resource use and cost outcomes were collected in the review for economic evaluations	The inclusion criteria were broad and therefore likely to have captured the available evidence. A total of 91 studies (91 records), were included that reported cost and healthcare resource use (HCRU) outcomes relevant to the UK were identified. A total of 14 records were found to be relevant to the UK.[30, 39-51]

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Screening	Referred to Appendix G, Section G.1.2	Titles and abstracts were screened by two independent reviewers and disagreements were resolved by consensus or by a third reviewer. Full texts were also screened by the two reviewers and disagreements resolved in the same way.
Data extraction	Referred to Appendix G, Section G.1.2	Data extraction was completed by one reviewer with a senior reviewer checking the extraction and disagreements resolved through discussion.
QA of included studies	Referred to Appendix G, Section G.2	The methodological quality of included full text publications was assessed using the Drummond checklist for cost-effectiveness studies.[37]

Abbreviations: CS, Company Submission; EAG, Evidence Assessment Group; HRQoL, health-related quality of life; QA, quality assessment

4.2. Summary and critique of company's submitted economic evaluation by the EAG

4.2.1. NICE reference case checklist

Table 20: NICE reference case checklist

Attribute	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	✓ Not explicitly stated in the company submission
Perspective on costs	NHS and PSS	✓ The company presented a non-reference case scenario analysis including societal costs
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	✓ No comments
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	✓ A lifetime horizon is suitable for decision making, owing to plausibly lifetime implications of the intervention upon patient health outcomes and NHS and PSS costs
Synthesis of evidence on health effects	Based on systematic review	✓ No comments
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	✓ No comments

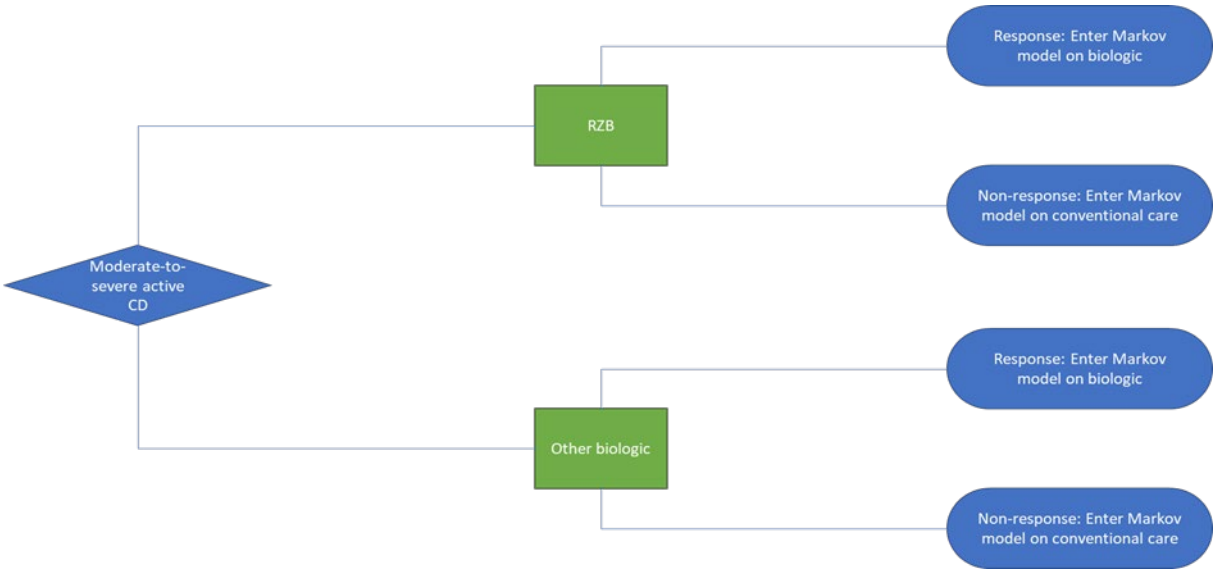
Attribute	Reference case	EAG comment on company's submission
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	✓ No comments
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	✓ No comments
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	✓ No comments
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	✓ No comments
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	✓ No comments

Key: CD, Crohn's disease; EQ-5D, EuroQol 5 dimension; HRQoL, health-related quality of life; NHS, National Health Service; PSS, Pseronal Social Services; QALY: quality-adjusted life year; TA: technology appraisal
Note(s):
Source(s):

4.2.2. Model structure

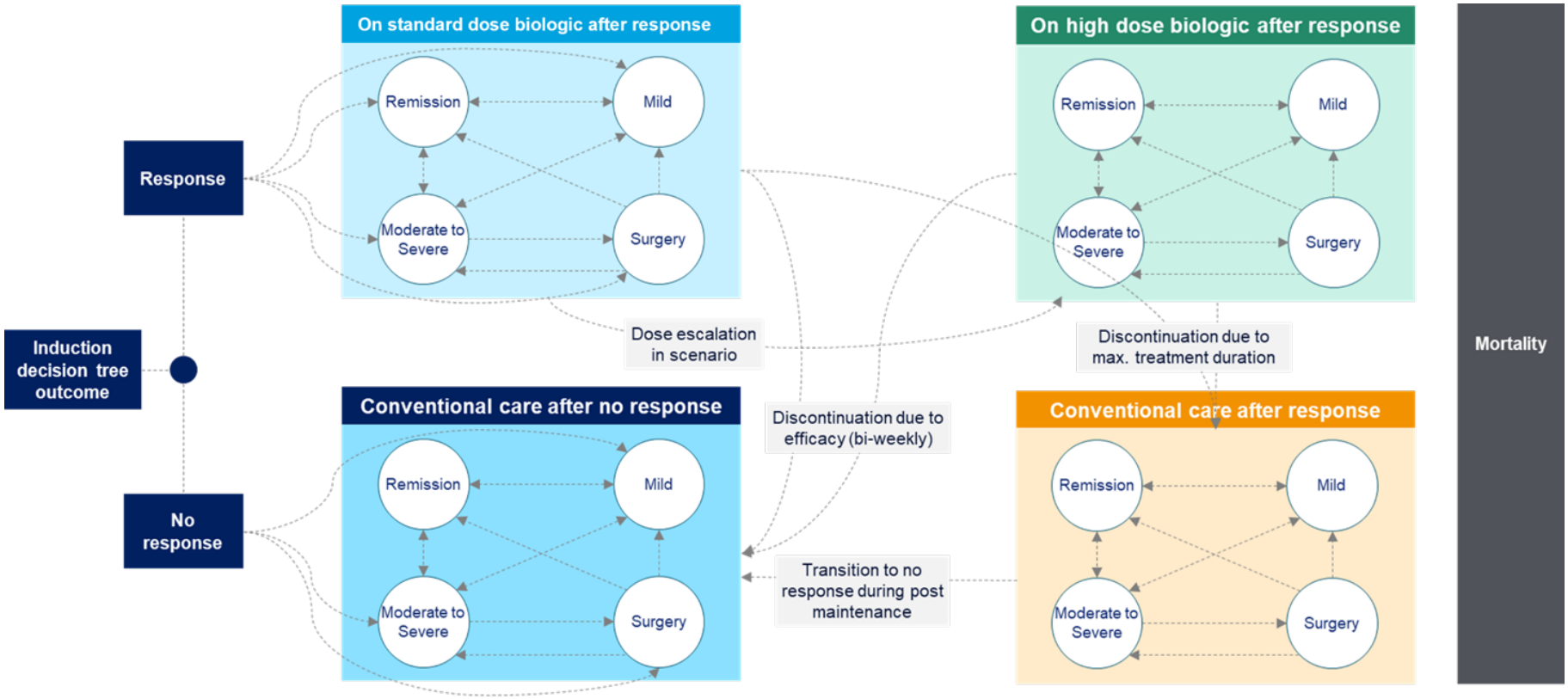
The company's *de novo* economic analysis comprises a cohort-level model developed in Microsoft Excel®, consisting of two distinct phases: i) a decision tree reflecting a short-term induction treatment phase (Figure 5), and ii) a Markov model (as described by the company) representing long-term maintenance treatment and post-maintenance phases.

Figure 5: Company's decision tree model structure diagram (CS Figure 12)



Key: CD, Crohn's disease; CS, company submission; RZB, risankizumab.
Note (CS Figure 12): Squares represent decision nodes, circles are chance nodes, and triangles are termini of the decision tree. The baseline of the induction trials is aligned with the model baseline, which occurs at the first square (decision node) on the left in the figure above.

Figure 6: Company's Markov model structure diagram (CS Figure 13)



Key: CS, company submission.
Note (CS Figure 13): Patients may remain in the health state in which they began a cycle. Surgery includes one surgical (2 weeks) and three post-surgical tunnel (6 weeks) states, such that a surgical episode lasts 8 weeks. Patients may transition to death at any time. Dose escalation in the base case only affects patient biologic costs; patients do not transition to the high-dose matrix as they have failed standard-dose treatment and therefore escalate to achieve standard-dose efficacy.

Patients with moderately to severely active CD enter model via the decision tree, where they receive treatment with risankizumab or comparator biologic therapy (described in Section 4.2.4). The length of the decision-tree differs by treatment arm, depending on the duration of induction treatment and response assessment for each biologic therapy. Efficacy outcomes are assessed at the end of the induction period; at the end of the decision tree, patients enter the Markov model either as responders on biologic treatment or as non-responders on conventional care.

Response at the end of the induction phase is defined in the company's economic analysis as a ≥100-point drop in CDAI score from baseline to end of induction (CR-100). The proportion assumed to achieve induction response is based on selected results from the company's NMA (described in Section 3.4, and discussed further in Section 4.2.6). The EAG notes that the abbreviations "CDAI-100" and "CR-100" appear to be used interchangeably throughout the CS; the EAG understand both CDAI-100 and CR-100 to refer to response determined by a ≥100-point drop in CDAI score from baseline to end of induction.

The company note that different definitions of response were used across trials in the network but justify their use of the CDAI-100 criterium as "a similar approach" (CS, B.3.2.2.1) was taken in the two most recent NICE appraisals in moderately to severely active CD, TA456 (ustekinumab) and TA352 (vedolizumab). The company present a scenario analysis (company scenario #7) in which a ≥70-point drop in CDAI score from baseline (CDAI-70) is used to define response, though no rationale or explanation of the relative merits of CDAI-70 versus CDAI-100 are provided. Importantly, as noted in 3.2.2.5, the EAG's clinical adviser has highlighted that CDAI score is not used in NHS clinical practice for the management of CD, owing to its overcomplicated nature and poor correlation with endoscopy. Instead, the Harvey Bradshaw Index and endoscopic response are used. The company justifies the use of CDAI-100 as a key outcome in their analysis based on its common use as an outcome across CD trials. The company acknowledges that "an NMA performed using endoscopic outcomes would potentially be more relevant to UK clinical practice" (CS, B.3.7.4), but explain their approach in the context of limited endoscopic data, which the company state was only available for risankizumab and ustekinumab overall populations.

At the end of the induction phase, all patients move to the long-term Markov model, which is characterized by CDAI-based health states and the need for surgery. The model structure adopted by the company is based on that presented by Bodger et al. (2009) [30], variants of which were used in TA456 and TA352. The long-term model health states are defined as follows:

- Remission (CDAI <150)
- Mild CD (150 ≤ CDAI <220)
- Moderate to severe CD (220 ≤ CDAI <600)
- Surgery (comprising one surgery model cycle and three post-surgery model cycles)
- Death

The company selected a 2-week cycle length for the long-term model in their analysis. The choice of cycle length was not justified in the company submission; however, the EAG consider a 2-week cycle length short enough to adequately capture the available data.

Each model cycle, patients can remain in their current health state, transition to another CDAI-based health state or experience death (which is an absorbing state). The company's economic model assumes that only patients in the 'moderate to severe CD' health state can experience surgery. Patients who experience surgery remain in the 'surgery' health state for one model cycle (2 weeks), and post-surgery tunnel states for three model cycles (6 weeks), after which patients return to a CDAI-based health state.

The company's economic analysis assumes that the mortality rate of patients with CD is equivalent to that of the age- and sex-matched general population (based on Office for National Statistics [ONS] 2018-20 national life table data for the UK). Consequently, the company's analysis assumes that mortality is not dependent on CDAI score, nor affected by treatment. Clinical advice to the EAG suggests that CD is not generally considered a life-shortening disease, and that it is reasonable to assume that patients with CD have equivalent survival to the general population. However, published evidence identified by the EAG is indicative of a heightened mortality risk for CD patients versus the general population, including risk related to higher rates of colorectal-cancer, pulmonary disease, and nonalcoholic liver disease.[52]

The company describe the long-term model as consisting of "four Markov model matrices that estimated the long-term course of CD including maintenance therapy and post-maintenance phases using clinical trial data" (CS, B.3.2.2). The four sets of transition matrices informing the long-term model are summarised in Table 21; the approach for modelling treatment effectiveness is more fully described and critiqued in Section 4.2.6.

Table 21: Summary of maintenance and post-maintenance transition matrices

Transition matrix	Application	Data informing transitions
Standard-dose biologic after response	Health state occupancy is determined using this transition matrix for patients who experience a CR-100 response at the end of induction and receive standard-dose biologic therapy in the maintenance phase. In the company base case, this matrix is also used for patients who start the maintenance phase on high-dose biologic therapy, as the target remission rate used to inform the calibrated transition matrix is based on weighted standard-dose and high-dose data.	Maintenance NMAs (standard dose and high dose), ordered probit models and 'calibration' (discussed further in Section 4.2.6)
High-dose biologic after response	Health state occupancy is not determined using this transition matrix in the company's base case analysis, as it is assumed the efficacy of those who dose escalate is equivalent to those who receive standard-dose maintenance therapy. This approach assumes that patients who dose escalate have lost response to standard-dose biologic treatment, and therefore the benefit from the increased dose is to match standard-dose efficacy. As such, dose escalation increases comparator costs without changing effectiveness estimates	Maintenance NMA (high dose), ordered probit models and 'calibration' (discussed further in Section 4.2.6)
Conventional care after response	Health state occupancy is determined using this transition matrix when responders discontinue biologic therapy at the point of a maximum treatment duration. The maximum treatment duration for biologics used in the company's analysis (52 weeks) is discussed further Section 4.2.6. This transition matrix assumes a residual treatment effect for patients who discontinue biologic therapy. The residual treatment effect period for biologics used in the company's analysis (52 weeks) is discussed further Section 4.2.6	Maintenance NMA (conventional care after response), ordered probit models, and 'calibration' (discussed further in Section 4.2.6) Ordered probit based on re-randomized placebo SC [withdrawal] arm in FORTIFY (n = 164); patients who received risankizumab IV for induction, had a response at the end of the initial 12-week induction period, and were subsequently randomized to the placebo SC arm in maintenance
Conventional care after no response	Health state occupancy is determined using this transition matrix for: i) non-responders who subsequently receive conventional care in the maintenance phase, and ii) patients for whom the residual treatment effect period has ended	True placebo group from FORTIFY. Namely, IV placebo responders at end of the initial 12-week induction period in ADVANCE and MOTIVATE, who were assigned to receive maintenance placebo SC in FORTIFY. The true placebo group consisted of n = 24 patients

Abbreviations: IV, intravenous; n, number; NMA, network meta-analysis; SC, subcutaneous.

In the company's model structure diagram (CS Figure 13), and as described in Table 21 above, it is assumed that dose escalation in the base case only affects patient biologic costs; patients do not transition to the high-dose matrix as they have failed standard-dose treatment and therefore escalate to achieve standard-dose efficacy. It is also stated in CS Figure 13 that a consistent assumption is applied to patients who initiate maintenance with high dose ustekinumab, as the higher dose is administered where a patient is expected to not respond adequately to the standard dose. However, the EAG interpret from the company's cost-effectiveness model that the 'standard dose' transition matrix for ustekinumab is calibrated using weighted standard dose and high dose NMA data. The EAG prefer the assumption whereby the 'standard dose' transition matrix calibration target is weighted the proportion of patients starting on standard and high dose therapy (in line with the EAG's interpretation of the company's model). Nevertheless, the EAG are concerned that the company's assumption of dose escalation affecting costs but not patient outcomes biases comparative cost-effectiveness estimates in favour of risankizumab, as dose escalation applies only to comparator biologics.

The EAG are concerned with the choice of data used to inform the conventional care after no response transitions. As described in Table 21, health state occupancy for i) non-responders who subsequently receive conventional care in the maintenance phase, and ii) patients for whom the residual treatment effect period has ended is informed using data from the 'true placebo' group from FORTIFY. Firstly, the EAG has concerns with the relatively small sample size of the true placebo group (n = 24), which is used to estimate transitions over a lifetime horizon. Secondly, the EAG has concerns as to whether placebo responders from the pivotal trial are representative of patients in practice who are non-responders or have discontinued biologic therapy. This is particularly important when applying the company's maximum treatment duration and residual treatment effect assumptions (described in further detail in Section 4.2.6), whereby all patients experience 'conventional care after no response' transitions from a maximum of 2 years (despite conventional care not being reflective of the treatment pathway as described by both the company and the EAG's clinical expert). Nonetheless, in the absence of alternative data, the EAG use the company's conventional care after no response transitions in the EAG preferred base case.

4.2.3. Population

The company's economic analysis considers a population in line with the anticipated license for risankizumab; that is, [redacted]

[redacted] The company notes that the patient population considered within the economic analysis is also aligned with the eligibility criteria for the pivotal risankizumab CD induction (ADVANCE and MOTIVATE) and maintenance (FORTIFY) trials.

The final scope issued by NICE specified that the subgroups by location of CD (ileal, colonic and perianal) may be considered, subject to data availability. However, the company did not present subgroups by location of CD, stating that the analysis was untenable due to low subject numbers. Clinical advice to the EAG indicated that location of CD is a key prognostic factor in CD.

The company instead presented the following subgroups in their economic analyses:

- Conventional care failure (CCF) population
- Biological failure (BF) population

As described in Section 2.3, clinical advice to the EAG indicated that the flowchart of current treatment practices presented by the company (Figure 1) was broadly reflective of a national standard of practice (while acknowledging potential differences between centres at the local level). Thus, the EAG considers the two populations presented in the company's economic analysis (CCF and BF) appropriate for addressing the decision problem outlined in the final scope issued by NICE.

Clinical data informing the CCF subgroup in the economic model is sourced from the ADVANCE and FORTIFY studies, while clinical data informing the BF population is taken from the ADVANCE, MOTIVATE and FORTIFY studies.

The company report that ADVANCE included both patients with inadequate response/intolerance to prior biologic therapy (described as the 'Bio-IR' population) and patients with inadequate response/intolerance to conventional therapy (described as the 'non-Bio-IR' population) for CD, whereas MOTIVATE was solely in a Bio-IR population.

The company state that the non-Bio-IR population is analogous to the CCF population; however, the EAG notes the non-Bio-IR population includes patients "who had received biologic therapy in the past but stopped therapy based on reasons other than inadequate response". The company report that [redacted] of patients in the non-Bio-IR population had not received a prior biologic therapy, implying that up to [redacted] of patients informing the CCF population had received prior biologic therapy.

The company describe the Bio-IR population, which includes patients “with documented intolerance or inadequate response (either failure to respond to induction treatment, or loss of response to maintenance therapy) to one or more biologics for CD”, as analogous to the BF population.

As ADVANCE, MOTIVATE and FORTIFY were international multicentre studies, it is unclear whether the trial populations can be considered generalizable to patients with moderately-to-severely active CD in NHS England practice. This is particularly in the context of prior treatments and concomitant conventional care received in the clinical trials compared with NHS England practice (discussed further in Section 4.2.8).

4.2.4. Interventions and comparators

The intervention considered in the company’s economic analysis is risankizumab 600 mg administered intravenously as induction therapy in Weeks 0, 4, and 8, followed by a maintenance period of risankizumab 360 mg administered subcutaneously Q8W, up to a maximum treatment duration of 52 weeks.

As described in Section 2.3, in clinical practice, the company anticipates that risankizumab SC will be delivered using an on-body device. Clinical advice to the EAG indicated a low level of clinical familiarity with on-body injectors but identified both potential advantages and disadvantages of this approach. As the company capture the cost implication of this administration difference, but assume no impact on clinical effectiveness parameters, the EAG have significant concerns with regards to the clinical effectiveness estimates informing the risankizumab arm of the economic model. More specifically, it is uncertain whether it is reasonable to assume there are no effectiveness implications from the different administration methods between the trials informing the analysis and expected clinical practice.

The comparators considered in the company’s economic analysis are dependent on the subgroup evaluated. In the CCF population, risankizumab is compared with infliximab, adalimumab and ustekinumab. In the BF population, risankizumab is compared with ustekinumab and vedolizumab. Dosing information for the intervention and comparators (including induction dose, induction duration, response assessment, maintenance dose, and escalated maintenance dose) are presented in Table 22 (adapted from Table 58 of the CS).

The final scope issued by NICE indicated that the availability and cost of biosimilars should be taken into consideration; and as such, the company compares risankizumab with infliximab and

adalimumab biosimilars in the CCF population. The company's economic analysis also considers both IV and SC forms of infliximab, adalimumab and vedolizumab. Furthermore, the company considers two alternative adalimumab induction dosing regimens (referred to in the CS as ADA 160/80 and ADA 80/40). In the company's economic analysis, treatments with biosimilars, IV and SC formulations and alternative induction doses are treated as standalone comparators (as summarized in Table 22).

Table 22: Intervention and comparator dosing information (adapted from CS Table 58)

Treatment	Induction				Maintenance	
	Induction dosing		Induction duration (weeks)	Response assessed (weeks)	Maintenance dosing	Maintenance dose escalation
RZB	600 mg IV at weeks 0, 4 and 8		12	12	360 mg SC Q8W from week 12	N/A
UST	Weight based IV dosing at week 0	<55 kg: 260 mg	8	6 and 8†	90 mg SC Q12W from week 8	90 mg SC Q8W
		>55 kg and <85 kg: 390 mg				
		>85kg: 520 mg				
VDZ IV	300 mg IV at weeks 0, 2 and 6		10	6 and 10‡	300 mg IV Q8W from week 14	300 mg IV Q4W
VDZ SC	300 mg IV at weeks 0, 2 and 6		10	6 and 10‡	108 mg SC Q2W from week 14	N/A
ADA 160/80 biosimilar	160 mg SC at week 0; 80 mg SC at week 2		4	4	40 mg SC Q2W from week 4	40 mg SC QW
ADA 160/80	160 mg SC at week 0; 80 mg SC at week 2		4	4	40 mg SC Q2W from week 4	40 mg SC QW
ADA 80/40	80 mg SC at week 0; 40mg SC at week 2		4	4	40 mg SC Q2W from week 4	40 mg SC QW
IFX IV	5 mg/kg IV at weeks 0 and 2		6	2	5 mg/kg IV Q8W from week 14	10 mg/kg IV Q8W
IFX IV biosimilar	5 mg/kg IV at weeks 0 and 2		6	2	5 mg/kg IV Q8W from week 14	10 mg/kg IV Q8W
IFX SC§	5 mg/kg IV at weeks 0 and 2		6	2	120 mg SC Q2W from Week 6	N/A

Key: ADA, adalimumab; INF, infliximab; IV, intravenous; N/A, not applicable; QxW, every x weeks; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.
Note (CS Table 58): † Manufacturer indicates response assessed at weeks 6 and 8, in the model week 8 is used; ‡ Manufacturer indicates response assessed at weeks 6 and 10, in the model week 10 is used. The biologic labels allow for continued biologic therapy to patients after induction therapy, even for non-responders, for a specified period of time. § For infliximab subcutaneous, only a biosimilar formulation is available, but is referred to as IFX SC throughout the CS.

In the final scope issued by NICE, BSC was specified as a relevant comparator, for people in whom TNF-alpha inhibitors, vedolizumab and ustekinumab have been ineffective, are contraindicated or are not tolerated. However, the company's economic analysis does not include a comparison of risankizumab with BSC. The company argue that BSC is not considered an appropriate comparator as, in clinical practice, if a biologic therapy has failed or are contraindicated, patients would be offered an alternative biologic therapy. The EAG notes that the anticipated license for risankizumab includes "[redacted]". Considering both the anticipated risankizumab license and final scope, the EAG requested the company provide further rationale for excluding BSC as a comparator from the economic analysis (clarification question B4). At the clarification questions stage, the company did not provide further justification but reiterated that BSC is not deemed an appropriate comparator based on clinical feedback and patients who are intolerant or unsuitable for biologic therapy would be considered for a different class of biologic in practice.

Although this is the case, the EAG notes that, while all other comparators in the company's submitted cost-effectiveness model may be included or excluded by the user, conventional care is a mandatory comparator in both the CCF and BF populations.

Nevertheless, clinical advice to the EAG suggested that, in practice, BSC is unlikely to be a relevant comparator to risankizumab for patients with CD for whom TNF-alpha inhibitors, vedolizumab and ustekinumab have failed, are contraindicated or not tolerated. Clinical advice to the EAG suggested that treating clinicians would instead explore every available and suitable biologic option sequentially.

The EAG is satisfied to an extent with the exclusion of BSC as a comparator for patients with moderately to severely active CD in NHS England practice, but notes an issue in the scope of this CD evidence submission and those that have come before (TA456 and TA352).[53, 54] The addition of risankizumab to the treatment options currently available would extend the plausible options available to treat each patient. For example, in the BF population, the EAG understands it would be plausible for a patient to sequentially receive risankizumab, ustekinumab and vedolizumab. In this instance, the availability of risankizumab would increase NHS/PSS treatment acquisition and administration costs while hopefully increasing the HRQL of the affected patient. Yet, the company's submission does not address this decision problem; instead, it assumes that after the initial therapy, patients move to conventional care, on every

treatment arm. In light of the company's argument that BSC is not a relevant comparator as patients would be offered an alternative biologic therapy, this simplistic approach to modelling the treatment pathway appears even more problematic. In the company's analysis, patients are not offered an alternative biologic therapy.

In the company's updated cost-effectiveness model submitted at the clarification question stage (6a. ID3986_Risankizumab CD_NICE_CEM v0.2 040822 v1.2 [ACIC]), when the BF population is selected on the 'Model Setup' worksheet, it is suggested via checkboxes that ustekinumab, vedolizumab IV, and vedolizumab SC are included as comparators. However, vedolizumab IV and vedolizumab SC are excluded from the incremental analysis (see worksheet 'List', range 'list_regimen_active_inc_all'). In EAG correction #1, summarised alongside other EAG corrections in 6.1, vedolizumab IV and SC are included as comparators in the model's incremental analysis.

4.2.5. Perspective, time horizon and discounting

In line with the NICE reference case, the perspective of the company's base case economic analysis is that of the NHS and PSS on costs (as reported Section B.3.2 of the CS), and direct health effects for patients (the perspective on outcomes is not explicitly stated in the CS).

The company present a non-reference case scenario analysis (company scenario #6), which is described in the CS as including "societal (indirect) costs". The company's justification for including indirect costs in a non-reference case scenario is to assess the burden of CD onto society; however, none of the inputs or methods for estimating indirect costs are described in the company submission. In Section B.3.5.4 of the CS, it is stated that "no additional miscellaneous costs are considered in the cost-effectiveness model".

A time horizon of 60 years is used in the company's economic analysis, which the company describe as a lifetime horizon based on a mean age at baseline of 38.83 and 38.22 years in the CCF and BF populations, respectively. Therefore, the company's analysis tracks the cohort of patients to a maximum average age of 98.83 and 98.22 years in the CCF and BF populations, respectively. The EAG consider a lifetime horizon appropriate for decision making, due to the chronic nature of CD and plausibly lifetime implications of treatment. The company assumes no excess mortality due to CD compared with the age- and sex-matched general population; and consequently, the company's economic model estimates that >97.7% and >97.1% of patients will have entered the death state in the CCF and BF arms after 60 years, respectively. In

scenario analysis, the company explore several alternative time horizons between 1 and 10 years in (company scenario #1a-d).

In Section B.3.2.2.2 of the CS, it is stated that a half-cycle correction is applied in the cost-effectiveness model to "account for the fact that events and transitions could occur at any point during the cycle". Typically, in a discrete-time, cohort-level model, a half cycle correction is applied by averaging rows of the 'Markov trace' (i.e., for each health state, the average of the proportion of patients at time T and time T+1 is taken, consequently assuming transitions occur at the mid-point of a cycle, instead of the beginning or end).

In the company's cost-effectiveness model, rather than adjusting the proportion of patients in each health state to half-cycle correct, the company include an additional row of the Markov trace beyond the time horizon in the model calculations and half the 'number of years per cycle' in the first and final row of the long-term model. As the model cycle length is 2 weeks, the 'number of years per cycle' (which is combined with health state occupancy to determine LYs and QALYs) in the long-term model is 0.04 years (to 2 decimal places, calculated as 2/52 weeks). However, when the half-cycle correction switch in the company's model is set to 'yes', the years per cycle in the first and last long-term model cycle are equal to 0.02 years (to 2 decimal places, calculated as (2/52) * 0.5).

The EAG believe the company's half-cycle correction application (i.e., capturing an extra cycle and assuming the first and last cycle of the Markov trace is equivalent to a 1-week duration) is inaccurate when considering time-preference discounted results.

Furthermore, the company apply a half-cycle correction to drug acquisition and administration costs, despite the 1-year dosing schedules for biologic therapies being known and outlined in the "Calc - Dosing" sheet of the company's economic model. The EAG do not consider a half-cycle correction appropriate for costs or outcomes known to occur at the start of a model cycle. In the company's analysis, biologic acquisition and administration costs are marginally underestimated using in the base case.

In Section 6.1 of this report, the modification of the half-cycle correction application is referred to as EAG correction #2.

4.2.6. Treatment effectiveness and extrapolation

The clinical parameters and data sources informing treatment effectiveness estimates in the company's cost-effectiveness model are summarized in Table 23, and described in further detail throughout this section of the report.

Table 23: Summary of treatment effectiveness parameters

Parameter	Source	Assumptions
CDAI response and remission rates at the end of induction	Induction NMA	Observed data (CDAI-100 response definition in the base case)
Percentage of responders and non-responders with CDAI moderate-to-severe CD at the end of induction	Risankizumab CD trials (ADVANCE and MOTIVATE)	Observed data for risankizumab, and assumed equivalence for comparator biologics
CDAI remission rates at the end of maintenance	Maintenance NMA	Observed data, with assumptions regarding the formation of the network
Transition probabilities in the maintenance phase	Ordered probit models and calibration	Derived using the distribution of patients across health states at the end of the induction phase and the end of the maintenance phase (52 weeks), estimated using ordered probit models with calibration of the remission | mild cut-point parameter
Proportion of patients starting standard-dose maintenance therapy and dose escalation	Clinical expert opinion	Assumed dose escalation only increases comparator biologic costs, without increasing efficacy
Biologic discontinuation rates	Risankizumab and comparator CD trials	Observed data and assumed constant discontinuation rate up to assumed maximum treatment duration
Maximum treatment duration and residual treatment effect	Assumption	Assumed maximum treatment duration of 52 weeks for all biologic therapies. Assumed residual treatment effect of 52 weeks following discontinuation of biologic therapy
Surgery	NHS Hospital Episode Statistics	Assumed only patients with moderate-to-severe CD experience surgery. Equivalent rates of surgery assumed across treatments. Assumed constant rate of experiencing surgery
Mortality	Life tables	Assumed the same as the age- and sex-matched general population

Abbreviations: CD, Crohn's disease; CDAI, Crohn's Disease Activity Index; NMA, network meta-analysis.

4.2.6.1. CDAI response and remission rates (induction NMA)

As described in Section 4.2.2, patients enter the model with moderately-to-severely active CD and efficacy outcomes are assessed at the end of the induction treatment period. The distribution of patients across health states at the end of the induction decision tree is estimated using the following parameters, as set out in Table 24.

- CDAI-remission rate (α)
- CDAI-response rate (β)
- Proportion of responders with moderate-to-severe CD (γ)
- Proportion of non-responders with moderate-to-severe CD (δ)

Table 24: Distribution of patients across health states at the end of induction

Responders			Non-responders		
Remission	**Mild CD**	**Moderate-to-severe CD**	**Remission**	**Mild CD**	**Moderate-to-severe CD**
α	β - α - (β * γ)	β * γ	0	(1 - β) – ([1 - β] * δ)	(1 - β) * δ

Abbreviations: α, CDAI-remission rate; β, CDAI-response rate; γ, proportion of responders with moderate to severe CD; δ, proportion of non-responders with moderate-to-severe CD; CD, Crohn's disease.

4.2.6.2. CDAI-remission (α) and CDAI-response (β) rates

The CDAI-remission and -response rates at the end of the induction period, which are derived from the induction NMA, are presented in Table 25.

As noted in Section 3.4.1, NMA results are provided using a risk difference method, rather than the more usual logit scale. While the rationale for this is not entirely clear this is not, in itself, expected to have a notable impact on the cost-effectiveness impact. The conversion of relative treatment effects to absolute levels of CDAI-response and CDAI-remission are likely to be more susceptible to modelling assumptions, although these will affect all treatments similarly.

As described in Section 4.2.2, given that clinical advice to the EAG indicated that CDAI-scores are not used in clinical practice, and the absence of commentary or explanation from the company on the relative merits of CDAI-70 versus CDAI-100 as a measure of response, the EAG feel unable to comment on relative suitability of CR-100 versus CR-70 response data.

Table 25: CDAI-remission and CDAI-response rates from the induction NMA

Treatment	Remission (CDAI <150)	CDAI-response (CDAI-100, company base case)	CDAI-response (CDAI-70, company scenario analysis)
CCF population			
RZB	[redacted]	[redacted]	[redacted]
UST	[redacted]	[redacted]	[redacted]
ADA 160/80 & biosimilar	[redacted]	[redacted]	[redacted]
ADA 80/40	[redacted]	[redacted]	[redacted]
INF IV & biosimilar	[redacted]	[redacted]	[redacted]
IFX SC	[redacted]	[redacted]	[redacted]
BF population			
RZB	[redacted]	[redacted]	[redacted]
UST	[redacted]	[redacted]	[redacted]
VDZ IV	[redacted]	[redacted]	[redacted]
VDZ SC	[redacted]	[redacted]	[redacted]

Abbreviations: ADA, adalimumab; BF, biological failure; CCF, conventional care failure; CDAI, Crohn's Disease Activity Index; IFX, infliximab; IV, intravenous; NMA, network meta-analysis; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

4.2.6.3. Proportion of responders (γ) and non-responders (δ) with moderate-to-severe CD

In the company's analysis, patients who are not in remission at the end of the induction phase are distributed between the mild CD and moderate-to-severe CD health states.

The company use a post-hoc analysis of ADVANCE and MOTIVATE risankizumab trial data to estimate the proportion of responders who remain in the moderate-to-severe CD state in the CCF population (8.4%) and BF population (7.8%), and similarly the proportion of non-responders who remain in the moderate-to-severe state in the CCF population (71.8%) and BF population (73.5%). In the absence of equivalent reported data from the relevant comparator studies, these proportions are assumed to also apply to all comparators in the company's model.

In NICE TA456, a similar approach was taken, using the proportion of moderate-to-severe responders from the IM-UNITI study. These parameters are commercial-in-confidence and not publicly available, although the Evidence Review Group in TA456 did note that the percentage of moderate-to-severe responders was reported in the NICE appraisal of vedolizumab (TA352).

The proportion of moderate-to-severe responders on vedolizumab, as reported in TA352, is 17.8% and 24.3% in the CCF and TNF-failure populations, respectively.

Nevertheless, in the absence of available data for all relevant comparators for both moderate-to-severe responders and non-responders, the EAG consider the company's approach, assuming the proportions from the risankizumab trials are applicable to all biologics, to be reasonable.

4.2.6.4. CDAI remission rates (maintenance NMA)

Within the maintenance phase, the company splits the evidence network into two separate sets of treatments/doses. A part of the rationale given for this is heterogeneity which it is suggested is seen in the wide range of placebo remission rates. Arguably, this heterogeneity could be modelled, at least in part, which may negate the purported need to split the network, and yield more relevant absolute estimates of CDAI-remission. It would also remove the presence of two different CDAI-remission parameters for conventional care after response, according to the estimates arising from these separate networks. As noted in Section 3.6, the EAG prefer the use of a single maintenance network; and as is noted in Section 3.4.3, modelling placebo CDAI-remission rates by trial date appears a suitable candidate to explain between-trial heterogeneity, and may be justifiable in terms of improvements in available concomitant treatments over time. Modelling placebo CDAI-remission in this way would uplift all treatments CDAI-remission by a similar amount using the risk difference approach.

4.2.6.5. Maintenance phase transition matrix estimation

The company use an ordered probit model to estimate state transition probabilities based on data from the FORTIFY trial for three separate subgroups: those randomised to risankizumab 360 mg SC ('biologic', n = 141); those who were randomised to risankizumab 360 mg SC for induction and placebo SC for maintenance ('placebo withdrawal', n = 164); and those randomised to placebo SC for both induction and maintenance ('true placebo', n = 24). The ordered probit model the company specified has main effects for lagged health state (factor with levels: remission, mild, and moderate-severe), and a linear term for the number of days since the previous (lagged) health state assessment. The ordered probit model then estimates cutpoints for the linear predictor to indicate the boundary between remission and mild, and between mild and moderate-severe health states. The company do not provide detailed justification for choosing an ordered probit, rather than an ordered logit model, although it is likely that the differences would be minimal, and quite likely trivial. However, this is not demonstrated.

The company justify the ordered probit model selection on the basis of its simplicity, rather than any formal model selection. A linear term for the number of days since the previous health state assessment may be reasonable if this variable shows little variability, as might be expected in the absence of missing data with health states recorded at 0, 24 and 52 weeks. However, in cases with only observations at 0 and 52 weeks, this linear term may be inappropriate. Appropriate imputation of missing observations (e.g., multiple imputation, potentially involving CDAI scores) may mitigate this problem. Furthermore, the use of a lagged health state term is potentially more problematic, as it makes certain assumptions, including regarding the absence of any interaction with the other terms (days since previous assessment, and the two cutpoints) relating to the lagged health state. An alternative that could have been investigated would be to fit three separate models according to the previous health state.

In addition to the limitations noted above, it should also be noted that the 'true placebo' ordered probit model is estimated on a particularly small sample (n = 24) and so the estimates may be unreliable, as is suggested by the associated standard errors and the fact that none of the estimated parameters approach statistical significance.

The results of the ordered probit model are used by the company to estimate (uncalibrated) 26-week transition matrices for each of the three subject subgroups. For the 'biologic' group, the linear predictors generated for a 182-day period are -0.00669 x 182 = -1.21758 (from remission); 1.07098 – 0.00669 x 182 = -0.14660 (from mild); 1.68745 – 0.00669 x 182 = 0.46987 (from moderate-severe). With regard to the remission | mild and mild | moderate-severe cutpoints of -0.33324 and 0.47878, respectively, by reference to the standard normal cumulative distribution function (the 'probit' link) we obtain an estimated transition matrix as presented immediately below, where the rows correspond to originator health states (remission, mild, moderate-severe from top to bottom) and columns correspond to destination health states (remission, mild, moderate-severe from left to right)

$$\begin{pmatrix} 0.81174 & 0.14335 & 0.04491 \\ 0.42597 & 0.30817 & 0.26586 \\ 0.21096 & 0.29260 & 0.49645 \end{pmatrix}$$

According to company responses to clarification questions, each 182-day transition matrix is then converted to a 14-day transition matrix using an exponential assumption. For example, the 182-day probability of transition from a remission health state to a mild health state is calculated as $1 - (1 - 0.14335)^{14/182} = 0.01183$. Other transitions are calculated similarly with the

probability of remaining in each state being calculated such that each row sums to one. Applying this method to the above 182-day transition matrix we obtain an estimated 14-day transition matrix of

$$\begin{pmatrix} 0.98464 & 0.01183 & 0.00353 \\ 0.04180 & 0.93471 & 0.02349 \\ 0.01806 & 0.02628 & 0.95566 \end{pmatrix}$$

It is known that this method of changing cycle durations introduces error (Chhatwal et al, 2016);[55] for example, it fails to account for subjects passing through one health state to reach another. The company acknowledges that alternative solutions are possible, for example eigen-decomposition. The exponential assumption method was stated to be used in the interests of convenience. However, in this case if we multiply up the 14-day transition matrix back to 182-days we obtain the following

$$\begin{pmatrix} 0.85082 & 0.10248 & 0.04670 \\ 0.36691 & 0.46527 & 0.16782 \\ 0.22284 & 0.19017 & 0.58699 \end{pmatrix}$$

Quite large discrepancies have been introduced. For example, the probability of transitioning from a mild to a moderate-severe health state over a 182-day period is reduced from 0.26586 to 0.16782 as a result of this approximation.

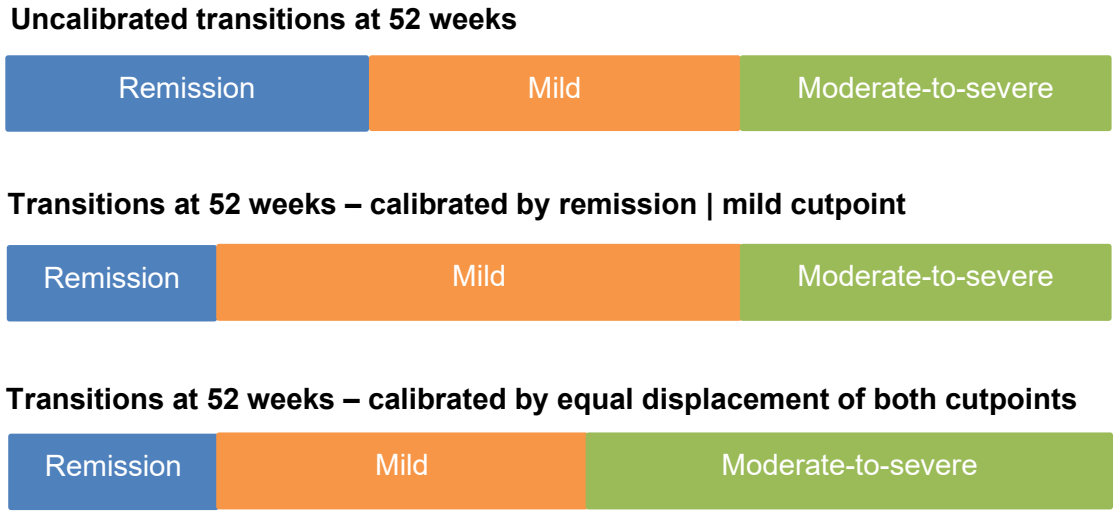
The EAG note that the subsequent calibration process (described below) will adjust the proportion of patients in a remission health state to hit a target level at 52 weeks. However, there is no rationale that this calibration process will adequately correct for the above source of error, and there are no grounds to assume that the individual transition probabilities will not retain significant levels of error, especially for transitions to mild or moderate-to-severe disease, where such adjustment is not made or for different durations of follow up other than 52 weeks.

In the absence of patient-level data for comparators, the company have used a calibration process to adjust the risankizumab transition probabilities for each comparator treatment in order that the proportion of patients in remission at 52 weeks matches the estimates obtained from the maintenance NMA. This calibration process adjusts the remission | mild cutpoint estimated from the biologic ordered probit model and then applies the exponential assumption cycle length change method to obtain 14-day transition probabilities.

Similar calibrations are performed on the 'placebo withdrawal' and 'true placebo' subgroups. The EAG note that the split maintenance network results in different target remission levels according to the placebo estimates resulting from these two sub-networks.

The adjustment of the remission | mild cutpoint is apparently arbitrary, and is illustrated below in Figure 7. Alternative parameter estimates in the ordered probit model could be adjusted to achieve the same 52-week remission proportion calibration. The company justify their approach on the basis of simplicity, as only one value needs to be adjusted, and consequently this is computationally convenient. The EAG note that this adjustment only directly rebalances the 182-day transitions to the remission and mild health states. There is no rationale for why this might adequately reflect the differences in state transitions between different comparators. Indeed, it seems implausible that only the balance between remission and mild health states would be rebalanced.

Figure 7: Methods for calibration of ordered probit estimates



In response to clarification questions, the company provided a comparison with three alternative calibration methods. The first made an equal adjustment to both cutpoints, the second adjusted the probability of remaining in remission only (rather than transitioning to mild or moderate-severe), and the third method rescaled the transitions to/remaining in remission and then scaled the probabilities of transition to/remaining in other health states accordingly for each row. No rationale for preferring any of these approaches is provided. As the ordered probit model

assumes a latent variable with cutpoints, and the health states are defined based on CDAI with thresholds, the first scenario (equal adjustment to both cutpoints) may be more justifiable, albeit not without further simplifying assumptions regarding the relationship between CDAI and the latent variable. The EAG consider that the first of these methods, which is also illustrated in Figure 7 provides a more plausible adjustment than the company's base case.

4.2.6.6. Standard-dose maintenance therapy and dose escalation

The company report that ustekinumab, vedolizumab, infliximab and adalimumab have both standard- and high-dose maintenance regimens. The company note that clinical expert opinion suggested only a small proportion of patients start on high-dose maintenance therapy, with the exception of ustekinumab.

Therefore, based on clinical expert opinion, the company assume that 92.5% of ustekinumab patients begin the maintenance phase on high-dose therapy. For all other biologics, the company assume that all patients start on standard-dose maintenance therapy.

In the company's economic model, the NMA-derived 52-week remission rate, which is used to estimate the calibrated transition matrices, is weighted by the proportion of patients who start on standard-dose and high-dose therapy. As such, for ustekinumab, the transition matrix described as 'response - standard dose maintenance' in the company's economic model incorporates both the standard- and high-dose maintenance NMA.

The company's economic analysis also considers dose escalation throughout the maintenance period, which is applicable to all biologics other than risankizumab. Dose escalation rates are based on clinical expert opinion for infliximab, adalimumab, ustekinumab and vedolizumab. For ustekinumab, the annual probability of dose escalation is the equivalent to the probability of starting on high dose ustekinumab (92.5%). Based on the information reported in CS, it is unclear to the EAG whether clinical advice to the company indicated that both 92.5% of patients start on high-dose maintenance ustekinumab *and* the annual probability of dose escalation is 92.5%, or whether the company assume equivalence to inform these parameters. The EAG is concerned that the company's approach of assuming 92.5% of patients start on high-dose ustekinumab and assuming an annual ustekinumab dose escalation rate of 92.5%, may overestimate the proportion of patients receiving high-dose ustekinumab.

In the company's base case, it is assumed that the treatment effectiveness estimates for those patients who dose escalate are equivalent to those who receive standard-dose maintenance

therapy. This approach assumes that patients who dose escalate have lost response to standard-dose biologic treatment, and therefore the only benefit from the increased dose is to match standard-dose efficacy. As such, dose escalation increases comparator costs without changing effectiveness estimates. As described in Key Issue 6, the EAG view this as an assumption that very likely biases comparative cost-effectiveness estimates in favour of risankizumab, as dose escalation applies only to comparator biologics.

4.2.6.7. Biologic discontinuation rates, maximum treatment duration and residual treatment effect

The company's analysis assumes treatment-specific, constant rates of biologic treatment discontinuation in the maintenance phase of the model, for the first 52 weeks of maintenance therapy, based on available trial data across treatments. The discontinuation probability assumptions applied in the model and their sources are summarised in Table 75 of the CS, and range from 4.3% in the first year (risankizumab) to 41.3% in the first year (vedolizumab IV or SC). Importantly, the company assumes a maximum biologic maintenance treatment duration of 52 weeks. From this point, patients are assumed to move to conventional care, where as noted in 4.2.2, the company assume there is a further 52-week residual treatment effect.

The EAG have several concerns with the company's approach to treatment discontinuation assumptions. First, the EAG's clinical adviser found it difficult to judge whether assuming different 1-year discontinuation rates across treatments based on observed data across trials was appropriate, given differences in inclusion criteria and study design across trials.

Second, and importantly, clinical advice to the EAG indicated that in practice, discontinuation rates are low, discontinuation becomes less likely as treatment duration increases, and that an assumption that all patients discontinue after 52 weeks of maintenance therapy is false. The EAG's clinical adviser's perspective is that if maintenance therapy is working for a patient, there is every effort and incentive to maintain treatment.

Figure 8 shows FORTIFY time to treatment discontinuation (TTD) data, provided by the company in response to an EAG request. From these data and the expert clinical advice received by the EAG, it is clear to the EAG that assuming a 52-week maximum maintenance treatment duration is inappropriate.

Figure 8: Time to treatment discontinuation due to lack of efficacy (FORTIFY ITT1A population, clarification question B8)

Key: ITT, intent-to-treat; IV, intravenous; SC, subcutaneous; SES-CD, Simple Endoscopic Score for Crohn's disease.+ censored observations.
ITT1A population includes randomised subjects in the ITT population who received risankizumab IV for only one period of 12 weeks in ADVANCE or MOTIVATE, and ≥1 dose of the study drug in FORTIFY substudy 1 and had eligible SES-CD of ≥6 (≥4 for isolated ileal disease) at baseline of the induction study.
Note: Subjects who discontinued the study due to lack of efficacy are considered as events.

The company argue that universal discontinuation from biologic therapy at 52 weeks "reflects clinical practice and NICE guidance, which states that patients should be re-assessed at 12 months to determine whether continuing with biologic treatment is appropriate" (CS, B.3.2.2). At the clarification question stage, the company went on to cite two primary reasons to assume maximum treatment duration of 52 weeks. Firstly, the 52-week timepoint reflects the available of trial data and modelling outcomes beyond 1 year would require assumptions regarding clinical effectiveness. Secondly, a consistent approach was used in recent NICE appraisals in CD (TA456 and TA352). The EAG do not consider the need to extrapolate beyond the trial period as sufficient justification for assuming a universal maximum treatment duration across biologic therapies: the company's base case analysis adopts a lifetime horizon - by design, outcomes are extrapolated beyond the trial period. Further, the EAG do not feel precedent is a rationale to use assumptions that lack clinical plausibility in this appraisal.

The 52-week discontinuation period in the company's preferred base case grossly underestimates time on biological treatment in clinical practice, as supported by clinical opinion and evidenced in Figure 8. Consequently, both the costs and efficacy of biologic treatments are misrepresented in the company's analysis. In Section 6.2, the EAG explore several alternative treatment discontinuation scenarios, whereby the maximum duration of biologic treatment is increased to align more closely with clinical practice. Furthermore, the EAG explores assuming equivalent rates of discontinuation for biologic therapies, to remove potential confounding issues due to differences in study design.

The company provide little rationale for the assumed 52-week residual treatment effect post-discontinuation, with clinical advice to the EAG estimating a 6-month time to symptomatic return for ustekinumab. Given the similar half-lives across treatments, the EAG considers a 52-week period likely overestimates the residual treatment effect post-discontinuation, with the modelled patients residing in the conventional care after response matrix for longer than is reflective of clinical practice. In Section 6.2, the EAG explore reducing the residual treatment effect to 26-weeks, to align with clinical opinion and a scenario presented by the company (company scenario #2).

4.2.6.8. Surgery

The company report the following inputs used in TA456, by using a Hospital Episode Statistics (HES)-sourced annual surgery rate estimate of 7%, converted to a 2-week cycle probability of 0.28% using an exponential formula (CS, B.3.3.4.3). In TA456, the 2011-14 HES dataset informed surgery risk assumptions,[54] while the HES data cited in the CS for this appraisal is from 2019-20 (CS, B.3.3.4.3 and B.4). It is unclear to the EAG whether the annual rate of surgery was equivalent in the 2011-14 and 2019-20 Hospital Episode Statistics data sets, or whether the value used by the company was lifted from TA456 materials, or identified by the company in the 2019-20 HES dataset.

As described in Section 4.2.2 and in line with TA456, the company assume a risk or surgery only applies to patients in the moderate-to-severe CDAI-based health state. Patients who experience surgery are routed through post-surgery tunnel states for three model cycles before being re-assigned to a CDAI-based health state in the company's model. Post-surgery transition matrices, which were sourced from TA456 and TA352, used data from Bodger (2009).[30]

4.2.6.9. Mortality

As described in Section 4.2.2, the company assume that there is not a heightened risk of death for patients with CD, compared with the general population. As such, age- and sex-matched general population mortality rates are applied each model cycle, regardless of population, health state or treatment. Literature identified by the EAG indicates a heightened risk of mortality for CD patients, though clinical opinion to the EAG advised CD is often not a life-shortening disease. The EAG consider equivalent mortality to the general population to be a reasonable assumption, though explore the relaxation of this assumption in Section 6.2.7 to align with the literature identified.

While reviewing the company model, the EAG identified an error in the calculation of general population mortality risk. The proportion of males and females alive at each year of age is incorrectly calculated in the company model, resulting in slight errors in the general population mortality risk. Though unlikely to have a large impact on the results, EAG correction #3, summarised alongside other EAG corrections in 6.1, corrects the proportion of males and females alive at each year of age.

4.2.7. Health-related quality of life

4.2.7.1. CDAI-based health state utility values

Patient-reported health-related quality of life (HRQoL) data were collected in the risankizumab CD induction and maintenance studies, including data collected using the EQ-5D-5L descriptive system. In line with the NICE reference case, health state utility values informing the company's economic analysis were calculated by mapping EQ-5D-5L data onto the EQ-5D-3L value set, using the algorithm developed by Hernández Alava et al. (2020).[56] The company assume HRQoL in the economic analysis is determined by CDAI-based health state, and not directly determined by patient population (CCF or BF), treatment arm (biologic therapy) or treatment status (on- or off-biologic therapy).

As the number of EQ-5D-5L observations from the pivotal risankizumab trials by CDAI-based health state was not reported in the CS, it was challenging for the EAG to assess the validity of the predicted health state utility values for informing the economic analysis. However, in response to clarification question B24b, the company report that [redacted], [redacted] and [redacted] EQ-5D-5L observations were recorded by patients in risankizumab trials samples assumed to represent the remission, mild CD and moderate-to-severe CD health states, respectively.

In the CS, it is reported that average health state utility values were estimated using ordinary least squares (OLS) regression; however, no rationale was provided for this approach. In clarification question B24a, the EAG requested the company provide justification for OLS estimation of utility values, particularly in the context of within-patient repeated measures. In clarification question B24c, the EAG specifically requested the company provide utility values estimated using a linear mixed model, including a random effect to account for repeated measures.

In response to clarification question B24, the company report that OLS is “simple, straightforward and commonly used (for estimated health state utilities)” and subsequently state it is believed that “allowing for correlated errors at the patient level would yield similar coefficient estimates and utility predictions”. However, the company provided health state utility values estimated using a linear mixed model as requested. Table 26 compares CDAI-based health state utility values estimated using OLS (company base case) and a linear mixed model (clarification question B24c).

Table 26: Estimated health state utility values (OLS versus linear mixed model)

Health state	OLS (CS, Table 83), mean (95% CI)	Linear mixed model (clarification question B24c, Table 39), mean (95% CI)
Remission	[redacted]	[redacted]
Mild CD	[redacted]	[redacted]
Moderate-to-severe CD	[redacted]	[redacted]

Abbreviations: CD, Crohn’s disease; CI, confidence interval; CS, company submission; OLS, ordinary least squares.

The company conducted an SLR to identify studies reporting HRQoL data for patients with moderate-to-severe CD, and although results of the included studies are presented in CS Appendix K, neither interpretation of these results nor assessment of suitability for inclusion in the economic model is provided in the CS. As such, the EAG requested further information on the relevance of included studies to this appraisal in clarification question B25. The company’s response cited previous NICE appraisals in CD and Bodger et al. (2009) [30] as the most relevant HRQoL sources (beyond the risankizumab pivotal trials), based on their alignment with the company’s modelled health states, previous use in NICE appraisals and relevance to a UK population.

In the CS, two scenario analyses are presented using alternative sources of health state utility values from the literature. In Section B.3.4.2 of the CS, the company state that the methods for these alternative scenarios are described in Section B.3.11.3; however, very little information is provided. The alternative sources are described as i) mapped ustekinumab Inflammatory Bowel Disease Questionnaire scores (company scenario #4a), and ii) utility values from Bodger (2009)[30] (company scenario #4b). For each of these scenarios, the utility values themselves were not reported in the CS.

The EAG broadly agrees with the company's approach of using patient reported HRQoL data collected in the risankizumab studies to inform health state utility values in the base case, with utility values from the previous NICE appraisals and the literature tested in scenario analysis. However, the EAG considers the linear mixed model a more robust (and therefore more appropriate) approach than OLS for estimating health state utility values, given the ability to account for differences between observations at the patient level.

The EAG notes a range of non-base-case utility sources are available in the company's submitted economic model, beyond the two described above. In clarification question B26, the EAG requested additional information on the process undertaken for selecting the two alternative sources for scenario analyses. The company note that the two sources were used in previous Health Technology Assessment submissions, but do not provide a descriptive comparison of the relative merits or appropriateness of the two sources compared with the additional sources identified by the EAG in the company's model (summarized in Table 27).

Table 27: Utility values from the literature (company model, worksheet: 'Library - HU')

Label: description (company model)	Health state utility value			Scenario in CS?
	Remission	Mild CD	Moderate-to-severe CD	
IBDQ: Data from IM-UNITI mapped using Buxton et al. (2007)[57]; company scenario #4a	0.800	0.680	0.550	Yes
SF36: Buxton et al. (2007)[57]	0.540	0.480	0.420	No
CDAI: Buxton et al. (2007)[57]	0.820	0.700	0.540	No
GEMINI: NICE TA352 (Table 7.4.3.1)[53]	0.820	0.730	0.570	No
Bodger et al. (2009)[30]; company scenario #4b	0.832	0.700	0.550	Yes

Abbreviations: CD Crohn's disease; CDAI, Crohn's Disease Activity Index; CS, company submission; IBDQ, Inflammatory Bowel Disease Questionnaire; TA, technology appraisal; SF36, Short Form 36.

4.2.7.2. Surgery health state utility

The company's economic analysis assumes the health-related quality of life of a patient with CD who experiences surgery is equivalent to that of a patient with CDAI moderate-to-severe CD for one model cycle (2 weeks), and subsequently equivalent to that of a patient in CDAI remission for three model cycles (6 weeks).

The company's description and justification of this approach is somewhat contradictory and unclear. In B.3.2.2.2 of the CS, it is stated that patients experience "surgery-related disutilities and costs". Conversely, in B.3.4.4 of the CS, it is reported that "surgical complications did not incur health utility decrements in the model but only affected costs". The company's rationale for excluding surgery-related utility decrements was that, as surgery is modelled as a health state, the utility value would include the expected utility loss from complications (CS, B.3.4.4). However, the utility value for the surgery health state is assumed equal to the mild-to-severe CD utility, not derived from surgery-specific data. In B.3.9.2 of the CS, the company report the rationale for assuming surgical complications do not incur health losses as a lack of data (rather than assuming health state utility implicitly capturing the health-related quality of surgery patients).

Overall, the EAG infer that the company's approach very likely underestimates the HRQoL implications of surgery, and explore alternative assumptions in Section 6.2.

4.2.7.3. Adverse event disutility values

The company's economic analysis captures utility decrements associated with experiencing treatment-related AEs. Differential AE rates are assumed across biologic treatment arms, based on observed data. The EAG has concerns with this approach as the observational data collected across studies may be affected by confounding, through differences in eligibility criteria and study design. Assuming differential AE rates across arms introduces a treatment effect into the model, and with no direct evidence to support this assumption, the EAG have concerns around the validity of the company's approach. In Section 6.2, the EAG explore the impact of assuming equivalent AE rates across biologic treatment arms.

The company use an exponential formula to convert 52-week AE probabilities from the relevant clinical trials to 2-week probabilities, in order to apply AE decrements each model cycle. Utility decrements are applied to the proportion of patients experiencing AEs in the standard dose and

high dose remission, mild CD and moderate-to-severe CD health states each model cycle. As such, the EAG interprets that the impact of experiencing any adverse event on a patient's HRQL is assumed to last one full model cycle (2 weeks). In clarification question B27, the company confirm that there is no clinical justification for the 2-week AE duration, beyond the assumption that AEs would be resolved quickly.

The EAG consider sourcing AE-specific durations from the literature a more accurate approach to applying disutility values; however, anticipate that the impact on the cost-effectiveness results is likely to be minimal. Nevertheless, the EAG trials alternative AE durations in Section 6.2 to explore the effect on the results.

4.2.7.4. Age-related utility decrement

Although not explicitly described in the CS, the company's economic analysis incorporates an age-related utility adjustment, to account for an expected natural decline in health-related quality of life over time, based on general population data. In Section B.3.4.1.1 of the CS, age-adjusted utility coefficients of -0.000173 (age) and -0.000034 (age^2) are reported. These age and age^2 coefficients are referenced as "NICE TA456, EAG Report Table 63. Data from Ara and Brazier 2010". However, the EAG were unable to identify the reported values in the primary source (Ara and Brazier 2010)[58]; and as such, requested that the company provide further detail in clarification question B28. At the clarification question stage, the company submitted an updated cost-effectiveness model which included corrected coefficients (age: -0.0002587, age^2: 0.0000332), as cited in Ara (2010).[58]

However, beyond this, the EAG identified additional errors with the company's age-adjustment approach. Firstly, the company report that the "average age of utility research" is 40 years (CS, B.3.4.1.1, Table 81), and consequently assume an age-adjustment multiplier >1 for model cycles in which the age is below 40 years. The EAG are unable to identify the reported average age of utility research in the primary source. Secondly, the company included regression coefficients for age and age^2 in their model but did not include the 'constant' (0.950857) or 'male' (0.021213) coefficients reported in Ara and Brazier 2010[58]. In EAG correction #4 (Section 6.1), the EAG update the company's utility age-adjustment approach by calculating the general population utility at baseline age in the model, and the general population utility each subsequent cycle, using the full regression equation reported in Ara and Brazier 2010.[58]

4.2.8. Resources and costs

The company report that an SLR of cost and resource use data identified 14 studies relating to the management of CD that were relevant to the UK. However, none of the identified studies were used to inform cost and resource use data or assumptions in the company's economic analysis. The company's justification for disregarding the output of the SLR was that, compared with the sources described throughout Section B.3.5 of the CS, none of the systematically identified studies had more recent data available. The sources and data informing the company's cost inputs are critiqued throughout this section of the report.

The company consider the following cost categories in their economic analysis:

- Drug acquisition costs
- Administration costs
- Concomitant medication costs
- Resource use costs
- Adverse event, surgery and surgical complication costs

4.2.8.1. Drug acquisition, administration and concomitant medication costs

Risankizumab unit costs (including a simple PAS discount) are provided by the company, while unit costs for comparator biologics are sourced from the British National Formulary (BNF). Drug acquisition unit costs are presented in Section B.5.1 (Table 85) of the CS.

The EAG notes that, per the BNF website, risankizumab is currently available as a 150 mg/ml pre-filled pen/syringe. In response to clarification question B29, the company confirmed that risankizumab will be available in 600 mg vials for induction, and as a 360 mg solution for maintenance therapy. The company model submitted at clarification included an incorrect price of [redacted] (a difference of [redacted] for risankizumab. EAG correction #5, as described in Section 6.1, aligns the risankizumab price to that reported in the CS.

Drug acquisition costs are calculated in line with the dosing schedules reported in Table 22 of this report. The only treatments subject to weight-based dosing schedules are ustekinumab (induction only) and infliximab (induction and maintenance). For ustekinumab, weight distributions based on the usteknimuab induction dosing schedule were calculated from a post-

hoc analysis of MOTIVATE and ADVANCE data (risankizumab induction trials) and used to calculate the average required induction dose. For infliximab, wastage (with regards to weight-based dosing) was considered by rounding to the nearest number of whole vials required, based on the average weight from the risankizumab CD trials. The EAG assumes the company considers wastage for infliximab only due to the weight-based dosing schedule throughout both induction and maintenance. As a fixed dose is administered for ustekinumab in the maintenance phase, no wastage is assumed. The EAG consider the company's approach to wastage and weight-based dosing to be acceptable. Average induction and 52-week maintenance costs are summarized in CS Table 86.

In the company's analysis, administration costs for treatments administered subcutaneously include an initial training cost on first administration (based on one one-hour of nurse time) and no subsequent costs. However, IV treatments are assumed to incur per administration cost based on the NHS Payment by Results tariff 2020/21 (item code FD02H). Risankizumab will be administered using an OBD, as defined in Section 2.3. As the method of administration differs from that in the clinical trials, the EAG have concerns that the efficacy and discontinuation rate of risankizumab may not be consistent with the observed data. However, in the absence of alternative data, the EAG preferred base case accepts the company's assumption of no efficacy and discontinuation rate implications from a different administration method.

Concomitant medication costs were sourced from the Drugs and pharmaceutical electronic market information tool (eMIT) if possible, else from the BNF. An average concomitant medication cost per 2-week cycle was calculated (£13.76) using per-day doses for individual treatments (sourced from TA352)[53] and usage estimates (sourced from TA456).[54] The company assume 61% of patients on biologic also receive conventional care, based on data from FORTIFY.

The company's economic model calculates treatment costs in the maintenance phase using a per-cycle approach, which the EAG considers inaccurate. In cases where the dosing schedule is known (i.e., X vials administered every Y weeks) and can be aligned with the model cycle length, it is unnecessary to estimate a per-cycle cost. The EAG understands that the company's approach may underestimate biologic acquisition and administration costs, as splitting costs which are known to be applied up-front across several cycles will overestimate treatment discontinuation and time-preference discount factors. The EAG's approach to modelling

treatment acquisition and administration costs, in line with the dosing schedules outlined the company's model, is described as EAG correction #6 in Section 6.1.

4.2.8.2. Health care resource use costs

In Section B.3.5.2 of the CS, there is very little information presented that describes the company's approach to modelling health care resource use costs. The company simply state that health-state costs were taken from TA456 and that costs were inflated to a 2020/21 cost year and adjusted to a 2-week cycle. The company do not report any of the following information:

- Resource use items by CDAI-based health state (i.e., itemised list of healthcare resource use requirements for patients with CD)
- Resource use proportions by CDAI-based health state (i.e., the proportion of patients assumed to experience each healthcare resource use item)
- Resource use frequencies by CDAI-based health state (the frequency at which patients with CD are assumed to require each healthcare resource use item)
- Original aggregate healthcare resource use costs from the reference source (TA456).[54]

The company note that in TA456, health care resource usage was gathered from a modified Delphi panel, in which 12 clinicians estimated resource use for each model health state. In TA456,[54] information was collected via telephone interviews and a face-to-face meeting to determine frequency of usage for all items.

Based on the information reported in the CS, the EAG are unable to verify the suitability of the aggregate health care resource use cost estimates. The EAG consider a more robust approach would be to model resource use costs using a 'bottom-up' approach (i.e., combine individual resource use estimates with the latest available unit costs), rather than uplifting aggregate health state costs. As such, in clarification question B30, the EAG requested an itemised list of resources and frequencies assumed for each health state. Furthermore, the EAG asked the company to confirm whether any clinical input was sought to validate the resources and frequencies sourced from TA456 for current practice in 2022.

In response, the company noted that the individual cost components (reported in TA456 Appendix 13)[54] were not publicly available, and that "UK clinicians were invited to review model

inputs used in the CS, but no clinicians provided comments on them". Overall, the EAG feel the company's approach to costing resource use is somewhat lacking, with the response received at clarification showing little understanding of the resources, frequencies and costs used to inform the cost-effectiveness model. The TA456 health care resources and frequencies have not been validated by either the company or clinical experts, nor has the EAG been able to assess the appropriateness of the resources included. In the absence of itemized resource use, the EAG find the company's approach acceptable, though note the limitations of the inability to perform validation.

4.2.8.3. Adverse event, surgery and surgical complication costs

Unit costs associated with surgery, managing surgical complications, and managing AEs from the company's analysis are presented in Table 28.

The cost of surgery is applied each model cycle to the proportion of patients in the surgery health state. Surgical complication and AE costs are applied each model cycle based on the estimated 2-week probabilities (CS Doc B, Table 79 and Table 80).

Table 28: Surgery, surgical complication and adverse event costs

Item	Cost	Reference
Surgery	£9,947	NICE TA456, EAG Report Table 68.[54] Values inflated to 2020/21
Surgical complications		
Wound infection	£986	NHS reference costs 2019/20 (WH07G)[59]
Prolonged ileus / bowel obstruction	£839	NHS reference costs 2019/20 (FD10M)[59]
Intra-abdominal abscess	£986	NHS reference costs 2019/20 (WH07G)[59]
Anastomotic leak	£986	NHS reference costs 2019/20 (WH07G)[59]
Adverse events		
Serious infections	£1,531	NHS reference costs 2019/20 (WJ06J)[59]
Tuberculosis	£1,894	NHS reference costs 2019/20 (DZ14J)[59]
Lymphoma	£842	NHS reference costs 2019/20 (SA31F)[59]
Hypersensitivity	£412	NHS reference costs 2019/20 (WH05Z)[59]
Skin reactions	£986	NHS reference costs 2019/20 (WH07G)[59]

4.2.8.4. Miscellaneous unit costs and resource use

As described in Section 0, the company present a non-reference case scenario analysis (company scenario #6), which is described in the CS as including “societal (indirect) costs”. However, none of the inputs or methods for estimating such costs are described in the company submission. In Section B.3.5.4 of the CS, it is stated that “no additional miscellaneous costs are considered in the cost-effectiveness model”.

5. COST-EFFECTIVENESS RESULTS

5.1. Company's cost-effectiveness results

In this section of the report, the company's cost-effectiveness results are presented for the CCF and BF populations. Clinical advice to the EAG indicated that, in practice, risankizumab would likely be used in the BF population, unless there was a strong contraindication to anti-TNF therapy. However, results for both populations are presented for completeness.

5.1.1.1. Base case results

The results reported by the company for the CCF and BF populations are shown in Table 29 and Table 30, respectively. Where biosimilar products are available, the product with the lowest cost is presented. The company report that probabilistic results are presented in the base case (CS B.3.10), based on the updated NICE methods guide. Deterministic base case results, calculated using the company's submitted economic model, are also presented in Table 29 and Table 30.

As the company's cost-effectiveness analysis compares more than two technologies, the company conduct a fully incremental analysis to identify the most cost-effective treatment option. The company's process for conducting incremental analysis is described as follows:

- Treatments are ordered from least to most expensive
- Check for strong dominance. Treatments are dominated if they are both costlier and less effective than another treatment included in the analysis.
- Check for extended dominance. Treatments are extendedly dominated if an alternative treatment can provide more QALYs for a lower cost per QALY. This is because decision makers prefer a more effective treatment with a lower ICER

When using the risankizumab PAS price, the deterministic and probabilistic results for patients with CD in a CCF population indicate risankizumab is dominated (i.e., less effective, more costly) when compared with adalimumab (80/40, 160/80, 160/80 biosimilar), infliximab (SC, IV, IV biosimilar) and ustekinumab. In the BF population, the results indicate that risankizumab is dominant (i.e., more effective and less costly) when compared with ustekinumab, vedolizumab SC, and vedolizumab IV.

The cost-effectiveness analysis reported in the CS uses list prices for all comparator treatments; however, the company notes that ustekinumab and vedolizumab each have confidential PASs.

Table 29: Company base case results (CCF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Company deterministic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	-	-
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£4,387	Dominated
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£31,259	£31,259
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£55,406	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£283,020	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£195,929	Dominated
Company probabilistic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	-	-
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£26,314	£26,314
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£53,236	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£155,894	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£208,134	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; IFX, infliximab; IV, intravenous; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

Table 30: Company base case results (BF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Company deterministic base case						
RZB	[redacted]	[redacted]	[redacted]	[redacted]	-	-
UST	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
VDZ SC	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Company probabilistic base case						
RZB	[redacted]	[redacted]	[redacted]	[redacted]	-	-
UST	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
VDZ SC	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Abbreviations: BF, biological failure; IV, intravenous; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

5.2. Company's sensitivity analyses

5.2.1. One-way sensitivity analysis

The company note that the parameters varied in their 'deterministic sensitivity analysis (DSA)' included baseline patient characteristics, efficacy and safety parameters, health-state utility values and costs (direct medial costs, AE costs, indirect costs). It is unclear to the EAG why the company varies indirect costs in the DSA, as the economic analysis is conducted from an NHS and PSS perspective on costs in line with the NICE reference case, and such changes have no effect upon results.

The company state that 'efficacy outputs' were varied in DSA using the upper and lower 95% confidence/credible intervals (CI/Cr) where possible, but that other inputs were sampled at ± 20% of their mean. The company do not provide rationale for varying any inputs by ± 20%, as opposed to within an estimated 95% CI. The EAG consider a more suitable approach would be to sample the lower and upper bounds from the 95% CI of an assigned probability distribution for each parameter, using the mean and standard error where available (or an assumed standard error where neccesary). The range of values tested when using ± 20% may be smaller than would typically be expected, though the EAG do not anticipate the company's approach to have a great impact on the DSA results.

The company provide a summary of model parameters and corresponding "DSA (low; high)" values in Table 99 of the CS. The company present tornado diagrams summarizing the top 20 most influential parameters on pairwise incremental net monetary benefit (INMB) for the CCF and BF populations based on a willingess-to-pay threshold of £30,000. Tornado diagrams are presented for risankizumab versus ustekinumab, adalimumab 160/80 biosimilar, infliximab IV

biosimilar and infliximab SC, respectively, for the CCF population, and versus ustekinumab, vedolizumab IV and vedolizumab SC, respectively, for the BF population.

The company report that the most influential parameters, in both the CCF and BF populations, are the efficacy inputs dervied from the NMAs (specifically, risankizumab probability of response and remission).

The company also report that body weight is a key driver of incremental NMB in the comparison of risankizumab and ustekinumab. However, the EAG are concerned with the company's approach to varying body weight. Ustekinumab induction dosing is based on weight-bands (i.e. <55 kg; >55 kg and ≤ 85 kg; >85 kg), and in the base case, the proportion of patients in each band is based on risankizumab trial data. However, for DSA, the company report lower bound values assuming 100% of patients are < 55 kg and upper bound values assuming 100% of patients are > 85 kg. The EAG do not believe that such extreme value testing is truly reflective of parameter uncertianty, and the company's approach is likely to overestimate the influence of weight distribuitions on cost-effectiveness results.

5.2.2. Probabilistic sensitivity analysis

The company undertook probabilistic sensitivity analysis (PSA) to explore parametric uncertainty by assigning various distributions to input parameters and running the model for 1,000 simulations. In the CS, no justification is provided for the chosen number of PSA iterations, nor are PSA convergence diagrams for costs, QALYs or incremental NMB provided in the company's cost-effectiveness model. The EAG believe the company should have performed an assessment of the stability of probabilistic outcomes, to determine whether 1,000 iterations are suitable for decision making.

In B.3.11.1 of the CS, it is reported that the parameters varied in PSA were baseline patient characteristics, health utilities, efficacy rates, and costs. To inform the PSA, the company assign a probability distribution to all included parameters (reported in B.3.9.1 of the CS), except for induction and maintenance treatment efficacy, for which Convergence Diagnostic and Output Analysis (CODA) samples are used to capture uncertainty in the NMA output. The EAG consider the company's approach, drawing CODA samples with replacement, appropriate for capturing uncertainty around NMA outputs in the PSA.

In addition to reporting tabulated, probabilistic results in the base case, the company present cost-effectiveness acceptability curves (CS B.3.11.1). The company report risankizumab (PAS

price) is associated with [redacted] and [redacted] probabilities of being the most cost-effective treatment option at a willingness-to-pay threshold of £30,000 per QALY gained in the CCF and BF populations, respectively. The EAG infer from the company's submitted cost-effectiveness model that risankizumab (PAS price) was the most cost-effective treatment option in [redacted] and [redacted] of simulations in the CCF and BF populations, respectively, at a willingness-to-pay threshold of £20,000 per QALY gained.

5.2.3. Scenario analyses

The company provide a series of deterministic scenario analyses to assess structural and methodological uncertainty in the cost-effectiveness analysis. In CS B.3.11.3, the company describes seven scenario analyses settings, which include: model time horizon, residual treatment effect, NMA, utility values, dose escalated regimens (start of maintenance), indirect costs and CDAI score.

The company presented the results of the scenario analyses in Section B.3.11.3.1 of the CS, and note that, in the CCF population, the TNF-alpha inhibitors remain cost-effective versus risankizumab. In the BF population, the company notes that risankizumab (PAS price) remains either dominant or is the cost-effective treatment option in all scenarios tested.

5.3. Model validation and face validity check

In CS Section B.3.14, the company describe internal validity checks, with regards to verification of the cost-effectiveness model. However, the company do not provide evidence of external validation, with regards to a comparison of modelled outcomes and trial-observed outcomes over time.

To justify approaches and assumptions throughout the CS, advice from a clinical expert advisory board meeting is cited by the company. The EAG notes the report for this meeting is citation 80 in Document B of the CS; however, the report itself is not provided. In clarification question B2, the EAG requested the company provide this meeting report (as commercial-in-confidence material). The company response indicated that the report could not be provided in full, as elements of the report include proprietary and confidential information that is not relevant for the purposes of the appraisal. The company noted that, where referenced in the CS, relevant excerpts of the advisory board report are disclosed within the Document B reference pack. In CS B.2.3.4, it is noted that eight experts (six clinicians and two health economic experts) were approached to join in a virtual advisory board meeting, all of whom participated. The company

report that “the criteria for selecting suitable experts were expertise and experience of treating CD in the UK (clinician) and specialised technical expertise in economic evaluation and health technology assessment (health economic expert)” (CS B.2.3.4).

The company report that their model was prepared according to several best practice guidelines, and is aligned with NICE guidance (CS, B.3.14.1). Furthermore, the company note that the results of the cost-effectiveness model were verified through an independent review of the model for coding errors, inconsistencies and the plausibility of model inputs (CS, B.3.14.1).

6. EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

The EAG identified a number of limitations within the company's base case and has explored the impact of alternative assumptions which the EAG believes are plausible. The EAG note that addressing all of the identified issues with the company's approach was not possible within the scope of the EAG's review. Specifically, the EAG has not explored key issues around uncertainty around the company's chosen model structure (Key Issue 4), nor around the company's approach to dose escalation (Key Issue 6). The EAG noted, with concern, that this will likely bias results in favour of risankizumab.

This section is organised as follows: Section 6.1 details the impact of errors identified in the EAG's validation of the executable model. Section 6.2 details a series of exploratory analyses investigating the robustness of the cost-effectiveness results to specific assumptions and additional uncertainties identified by the EAG. These analyses were conducted within the company corrected base-case analysis. The scenario analyses presented in Section 6.2 focus on exploring the following issues and uncertainties:

- Company's choice of maximum treatment duration for biologics
- Company's choice of residual treatment effect duration following biologics
- Company's approach to treatment discontinuation rates
- Company's choice of network structure in the maintenance NMA
- Company's decision to calibrate transition matrices by adjusting the remission | mild cut-point estimated from the biologic ordered probit model
- Company's approach to adjust transition matrices for a model 2-week cycle length using an exponential assumption
- Company's background mortality assumptions
- Company's approach to capturing AEs costs and consequences
- Company's approach to estimating health state utility values using OLS regression
- Company's assumption regarding patient HRQL in the surgery state

In Section 6.3, the EAG base-case is presented based on a combination of the exploratory analyses presented in Section 6.2. In Section 6.4, additional EAG scenarios are presented around the EAG preferred base case.

6.1. EAG corrections and adjustments to the company's base case model

The company implemented an amendment to general population utility parameters in an updated version of the cost-effectiveness model submitted alongside EAG clarification question responses; however, the EAG made further corrections to the company's utility age-adjustment approach (see EAG correction #4). Beyond this, a small number of additional errors were identified by the EAG in the company's cost-effectiveness model submitted at clarification question stage. The EAG have made corrections for these errors, which are described as EAG correction #1 to #6 throughout Section 4, and summarized below.

- EAG correction #1, as described in Section 4.2.4, includes vedolizumab IV and vedolizumab SC as comparators in the incremental analysis for the BF population
- EAG correction #2, as described in Section 0, corrects the half-cycle correction application
- EAG correction #3, as described in Section 4.2.6.9, corrects the approach to estimating general population mortality
- EAG correction #4, as described in Section 4.2.7.4, corrects the utility age-adjustment application
- EAG correction #5, as described in Section 4.2.8.1, aligns the risankizumab pack price with the cost reported in the CS
- EAG correction #6, as described in Section 4.2.8.1, applies biologic treatment acquisition and administration costs per the reported dosing schedules, without estimating an average per 2-week model cycle cost

EAG-corrected company base case results are presented for the CCF and BF populations in Table 31 and Table 32, respectively. In the CCF population, risankizumab remains dominated (more costly and less effective) when the EAG's corrections are applied. In the BF population, risankizumab remains a dominant (less costly and more effective) treatment option when the EAG's corrections to the company's base case are implemented.

The design of the company's economic model and volume of VBA code is a limiting factor for exploring probabilistic analysis. The economic model includes one 'Markov trace' (calculation) sheet for the selected comparator, and therefore must cycle through the list of included comparators using automated processes to perform incremental analysis, while also drawing recalibrated transition matrices. The above factors and number of included comparators contribute to a PSA run-time of approximately 9 hours when sampling 1,000 iterations; as such, the EAG did not consider it feasible to produce probabilistic results for each EAG preferred assumption or exploratory analysis within the EAG report timeframe. Additionally, the EAG note the company's economic model presents probabilistic results only in graphical form. In clarification question B31, the EAG requested an executable version of the cost-effectiveness model that included fully incremental probabilistic analysis (in line with the company base case); however, such model was not provided by the company. As such, the EAG present incremental analysis results deterministically (except for the EAG preferred base case in Section 6.3, where incremental analysis results are presented deterministically and probabilistically).

Table 31: EAG-corrected company base case results (CCF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
EAG-corrected company deterministic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	-	-
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£4,229	Dominated
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£32,556	£32,556
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£57,977	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£329,812	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£211,356	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; IV, intravenous; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

Table 32: EAG-corrected company base case results (BF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
EAG-corrected company deterministic base case						
RZB	[redacted]	[redacted]	[redacted]	[redacted]	-	-

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
VDZ SC	[redacted]	[redacted]	[redacted]	[redacted]	-£26,902	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	-£51,865	Dominated
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£34,655	Dominated

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; IV, intravenous; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

6.2. Exploratory and sensitivity analyses undertaken by the EAG

This section explains and interprets results from the additional analyses conducted by the EAG in turn. Pairwise, deterministic results from each individual exploratory analysis are presented in Table 33 and Table 34 for risankizumab compared to the optimal comparator (as described in Section 1.7) in the CCF (infliximab SC) and BF (vedolizumab SC) populations, respectively.

The volume of VBA code implemented in the company's model is a limiting factor for exploring additional EAG scenarios. When applying alternative assumptions in the model, the VBA code used to calculate the cost-effectiveness results often overwrites changes to settings in favour of the company's base case assumptions. As such, adapting the model to explore alternative scenarios can be a lengthy process, requiring careful checking to ensure the results correspond to the desired settings.

6.2.1. Increased maximum treatment duration

As discussed in Section 4.2.6.7, the EAG considered the maximum treatment duration of 52 weeks to be inappropriate based on the patient-level data observed in the FORTIFY clinical study (Figure 8) and clinical advice provided to the EAG. Clinical opinion indicated that patients would continue to receive treatment while remaining in remission or exhibiting controlled disease, with a high proportion of patients expected to remain on treatment for several years following treatment initiation. Given the lifetime horizon modelled, the EAG considered a maximum treatment duration of 20 years a more realistic estimate of duration, with alternative durations ranging between 5 and 40 years explored in sensitivity analysis.

In the CCF population, increasing the maximum treatment from 1 to 20 years for all biologic therapies results in lower incremental costs and higher incremental QALYs for risankizumab versus infliximab SC. As such, risankizumab moves from the north-west quadrant (dominated,

more costly and less effective) to the north-east quadrant (more costly and more effective) on the incremental cost-effectiveness plane versus infliximab SC, with an ICER of £52,449.

In the BF population, increasing the maximum treatment duration resulting in higher incremental costs and QALYs for risankizumab compared with vedolizumab SC. Therefore, risankizumab moves from the south-east quadrant (less costly and more effective) to the north-east quadrant of the cost effectiveness plane versus vedolizumab SC, with an ICER of £65,837.

6.2.2. Residual treatment effect

As discussed in Section 4.2.6.7, following discontinuation of biologic therapy, the company assumes a residual treatment effect lasting 52 weeks. Clinical advice to the EAG estimated a 6-month time to symptomatic return for ustekinumab. Given the similar half-lives across treatments, the EAG anticipates that a 52-week period likely overestimates the residual treatment effect post-discontinuation, with 26-weeks a more realistic time point in clinical practice. To align with clinical opinion, the EAG reduced the residual treatment effect duration from 52 to 26 weeks (consistent with company scenario #2).

In the CCF population, risankizumab remains dominated by infliximab SC (more costly and less effective) when assuming a 26-week residual treatment effect duration.

In the BF population, risankizumab remains dominant over vedolizumab SC (less costly and more effective); however, incremental costs and QALYs are lower (relative to the EAG-corrected company base case) when assuming a 26-week residual treatment effect for biologics.

6.2.3. Treatment discontinuation

The EAG considered differences in treatment discontinuation rates between biologic treatments could be an artifact of confounding between study designs, rather than a true difference. The EAG’s clinical adviser found it difficult to judge whether assuming different 1-year discontinuation rates across treatments based on observed data across trials was appropriate, given differences in inclusion criteria and study design across trials. Consequently, the EAG explored the impact of applying risankizumab discontinuation rates to all biologic treatments considered in the analysis.

Consistent with the EAG-corrected company base case, when assuming equivalent biologic discontinuation rates across treatments, risankizumab was dominated by infliximab SC. In the BF population, risankizumab remained dominant over vedolizumab SC.

6.2.4. Single maintenance network

As discussed in Section 3.4.6, the EAG considered a single network more appropriate than a split network for estimating efficacy in the maintenance phase. The EAG disagreed with the company's approach to splitting the evidence into two networks, and found the rationale to support the approach inconsistent. Aligned with the basis that network formulation should be based on comparator connections, the EAG implements a single maintenance network in the analysis, using data requested at clarification.

In a version of the cost-effectiveness model submitted by the company at clarification stage, a scenario was presented using a single NMA network in the maintenance phase (described as 'Scenario 2: Single NMA network (all biologics)' in the 'Results – Deterministic (Pair)' worksheet of the model. The EAG note that, while the company updated NMA inputs on the 'Model NMA inputs' worksheet for the standard dose NMAs in this scenario, high dose NMA inputs were unchanged. As transition matrices for ustekinumab are estimated using weighted standard-dose and high-dose NMA inputs, the EAG include a scenario described in the cost-effectiveness model as 'Scenario 3: Corrected single NMA network (all biologics)' in which the high-dose single network inputs are also updated, using data provided by the company in response to clarification question A15. Within the timeframe of the EAG's review of the cost-effectiveness analysis and implementation of the additional and exploratory analysis, the EAG were unable to reflect the parametric uncertainty around the single maintenance network inputs.

When using the EAGs corrected single NMA network scenario, risankizumab remains dominated (more costly and less effective) by infliximab SC in the CCF population. When compared with corrected company base case, the single network results in higher incremental costs ([redacted] versus [redacted] and a larger QALY decrement for risankizumab ([redacted] versus [redacted]).

In the equivalent scenario in the BF population, risankizumab remains dominant (less costly and more effective) when compared with vedolizumab SC. However, compared with the corrected company base case, risankizumab cost savings are lower ([redacted]versus [redacted]) and QALY gains are lower ([redacted] versus [redacted]).

6.2.5. Single maintenance network, adjusted for a temporal effect

Beyond the use of a single maintenance network, the EAG notes heterogeneity is a key limitation of the maintenance phase NMA, as discussed in Section 4.2.6. Generally over time, remission outcomes have improved as treatments themselves have improved. As such, an EAG analysis models the placebo remission rate to include a temporal association with the time at which clinical trials were conducted and bases the absolute remission rates in maintenance on this anchor point (described as 'Scenario 4: Temporal trend single NMA network (all biologics)' on the 'Model NMA Inputs' sheet of the cost-effectiveness model). Within the timeframe of the EAG's review of the cost-effectiveness analysis and implementation of the additional and exploratory analysis, the EAG were unable to reflect the parametric uncertainty around the single maintenance network (adjusted for a temporal effect) inputs.

When applying a single network with temporal effect to the maintenance phase, risankizumab remains dominated (more costly and less effective) by infliximab SC in the CCF population. When compared with corrected company base case, the single network results in higher incremental costs ([redacted] versus [redacted] and a larger incremental QALY decrement for risankizumab ([redacted] versus [redacted]).

In the equivalent scenario in the BF population, risankizumab remains dominant (less costly and more effective) when compared with vedolizumab SC, with marginally lower risankizumab cost savings ([redacted] versus [redacted]) and QALY gains ([redacted] versus [redacted]) company with the corrected base case.

6.2.6. Maintenance phase transition matrix estimation

As discussed in Section 4.2.6.5, the company convert 182-day transition matrices to 14-day transition matrices using an exponential assumption. However, as demonstrated by the EAG, this approach is limited as discrepancies are introduced through the methods inability to account for patients passing through health states to reach others. As such, the EAG proposes an alternative approach to changing cycle length, as suggested in Chhatwal et al., (2016)[55], to avoid the use of an approximate exponential assumption. The EAG's alternative approach estimates the 14-day transition probabilities which, when multiplied repeatedly for 13 cycles, more closely approximate the 182-day transition matrix. This approach minimizes the sum of differences between the observed 182-day transition probabilities and that implied by the 14-day transition probabilities.

In the cost-effectiveness model, the EAG estimates transition matrices without the use of exponential assumption for the single network scenario described in Section 6.2.4 and the single maintenance network, adjusted for a temporal effect scenario described in Section 6.2.5.

In the CCF population, risankizumab remains dominated by infliximab SC in i) the single maintenance network, with non-exponential transition matrix estimation and ii) the single maintenance network (adjusted for a temporal effect), and non-exponential transition matrix estimation.

In the BF population risankizumab remains dominant over vedolizumab SC in i) the single maintenance network, with non-exponential transition matrix estimation and ii) the single maintenance network (adjusted for a temporal effect), with non-exponential transition matrix estimation.

As discussed in Section 4.2.6, the company adjust the ordered probit remission | mild cut point for each treatment to calibrate transition probabilities, in order that the proportion of patients in remission at 52-weeks matches the estimates obtained from the maintenance NMA. As a result, changes to the proportion of patients in the remission and mild health states are allowed in each cycle transition, though no impact is assumed upon the proportion in the moderate-to-severe group. The EAG found this approach unrealistic and preferred instead to adjust both the remission | mild and mild | moderate-to-severe cut points by the same amount.

Without an exponential assumption to adjust cycle length, the EAG estimate transition probabilities by adjusting both cut points for the single network scenario described in Section 6.2.4 and the single maintenance network, adjusted for a temporal effect scenario described in Section 6.2.5. Within the timeframe of the EAG’s review of the cost-effectiveness analysis and implementation of the additional and exploratory analysis, the EAG were unable to reflect the parametric uncertainty around the EAG-derived transition matrices.

In the CCF population, risankizumab remains dominated by infliximab SC in i) the single maintenance network, with non-exponential transition matrix estimation and adjustment of both ordered probit cut points and ii) the single maintenance network (adjusted for a temporal effect), with non-exponential transition matrix estimation and adjustment of both ordered probit cut points.

In the BF population, when exploring a single maintenance network, with non-exponential transition matrix estimation and adjustment of both ordered probit cut points, risankizumab is

associated with higher incremental costs and lower incremental QALYs compared with vedolizumab SC. Furthermore, on the incremental cost-effectiveness plane versus vedolizumab SC, risankizumab moves from the south-east quadrant (dominant, less costly and more effective) to the north-east quadrant (more costly and more effective), with an ICER of £63,812.

However, in the BF population when assuming a single maintenance network (adjusted for a temporal effect), with non-exponential transition matrix estimation and adjustment of both ordered probit cut points, risankizumab remains dominant versus vedolizumab SC.

6.2.7. Increased mortality for CD

Advice to the EAG concurred with the company assumption that CD is not a life-shortening disease. While the EAG consider it reasonable to assume patients with CD have equivalent survival to the general population, applying an SMR to increase CD mortality was explored in sensitivity analysis. Based on published evidence identified by the EAG, it is possible that patients with CD are at a heightened mortality risk versus the general population thus, the EAG considered the exploration necessary. Bewtra et al. (2013) [52] report all-cause mortality SMRs varying from 0.71 to 3.20 for CD, with a summary SMR of 1.38.

In the CCF population, applying SMRs of 1.38 and 3.20 to general population mortality results in a change in ICERs of -£33 and -£192, respectively (as such, risankizumab remains dominated by infliximab SC).

Similarly, in the BF population, applying SMRs of 1.38 and 3.20 to general population mortality results in increased ICERs by £11 and £63, respectively (as such, risankizumab remains dominant over vedolizumab SC).

6.2.8. Equivalent AEs across biologic treatments

As discussed in Section 4.2.7.3., differential AEs are assumed across biologic treatments in the company analysis. The EAG considered this a limitation given the differential AEs are based on observed data across studies which could be affected by confounding. This limitation was further affirmed by clinical advice to the EAG, indicating that comparing observed data naively may be inappropriate due to differences in study design. To align with clinical opinion, using the risankizumab observed AEs, the EAG explored the impact of assuming equivalent AEs between biologic treatments.

Incorporating equivalent AEs between biologics results in consistent results with the corrected company base case, with risankizumab remaining dominated (more costly and less effective) by infliximab SC in the CCF population and risankizumab remaining dominant (less costly and more effective) over infliximab SC in the BF population.

6.2.9. AE duration

The company indicated at clarification that the 2-week AE duration assumed in the model was arbitrarily chosen (as discussed in Section 4.2.7.3). In absence of AE-specific durations sourced from the literature, the EAG reduced the AE duration to 1 week, and increased the AE duration to 4-weeks and 8-weeks in various scenario analyses to investigate the impact on the cost-effectiveness results. The EAG implemented the AE duration exploratory analysis in the company's cost-effectiveness model by adjusting the per-cycle weight attributed to AEs.

In the CCF population, when comparing risankizumab with infliximab SC, changing the AE duration did not have a large impact on cost-effectiveness results, with risankizumab remaining dominated (more costly and less effective). When compared with the corrected company base case, increasing the assumed AE duration to 8 weeks marginally reduces incremental costs for risankizumab versus infliximab SC ([redacted] versus [redacted]), while also marginally reducing the incremental QALY decrement ([redacted] versus [redacted]).

In the BF population, risankizumab remains dominant (less costly and more effective) over vedolizumab SC when exploring alternative AE durations; increasing the assumed AE duration marginally increases cost savings and QALY gains for risankizumab versus vedolizumab SC.

6.2.10. Utility estimation for CDAI-based health states

As discussed in Section 4.2.7.1, within-patient repeated EQ-5D-5L observations are not adjusted for in the company base case health state utility estimates. At clarification question stage, the company provided rationale for the use of an OLS regression to estimate utility values, but also presented utility values with a linear mixed model including a random effect to account for repeated measures. Although the estimated values are reasonably similar between methods (Table 26), the EAG considers the linear mixed model a more robust (and therefore more appropriate) approach, given the ability to account for differences between observations at the patient level.

In the CCF population, when applying the linear mixed model estimated utility values in the analysis, the predicted incremental lifetime QALY loss associated with risankizumab compared

with infliximab SC marginally decreases relative to the EAG-corrected company base case; however, risankizumab remains dominated by infliximab SC (more costly and less effective).

In the BF population, when applying the linear mixed model estimated utility values in the analysis, the predicted incremental lifetime QALY gain associated with risankizumab versus vedolizumab SC marginal decreases compared with the corrected company base case ([redacted] versus [redacted]). However, risankizumab remains dominant (less costly and more effective) when compared with vedolizumab SC.

6.2.11. Utility assumptions for the surgery health state

As discussed in Section 4.2.7.2, the company's analysis assumes the surgery health state utility value is equivalent to the CDAI moderate-to-severe utility for one model cycle (2 weeks), and subsequently equivalent to that of CDAI remission for three model cycles (6 weeks). Furthermore, the company's rationale for excluding surgery-related utility decrements was that, as surgery is modelled as a health state, the utility value would include the expected utility loss from complications (CS, B.3.4.4).

Overall, the EAG infer that the company's approach very likely underestimates the HRQoL implications of surgery, and explore cost-effectiveness results when the health state utility value for surgery is assumed to be 80% and 90% of the health state utility value for CDAI moderate-to-severe CD.

In the CCF population, incremental QALYs remain generally consistent with the corrected company base, and when applying a surgery utility multiplier of 80% and 90%, the ICER for risankizumab versus infliximab SC changes by +£155 and +£310, respectively (with risankizumab remaining dominated).

Similarly, in the BF population, incremental QALYs remain generally consistent with the corrected company base, and when applying a surgery utility multiplier of 80% and 90%, the ICER for risankizumab versus vedolizumab SC changes by +£23 and +£46, respectively (with risankizumab remaining dominant).

6.2.12. Impact on the ICER of additional clinical and economic analyses undertaken by the EAG

The EAG made the changes described in Sections 6.2.1 to 6.2.11 individually. The effect of each change upon the EAG-corrected company base case for the optimal comparator in each

population (infliximab SC and vedolizumab SC, CCF and BF respectively – as discussed in Section 1.7) are provided in Table 33 and 34.

In the CCF population, risankizumab remains dominated by infliximab SC in most exploratory analyses performed. The greatest difference in incremental costs and QALYs is observed when exploring assumptions regarding the use of a single maintenance network (adjusted for a temporal effect), with non-exponential transition matrix estimation and adjustment of both ordered probit cut points. This scenario is associated with incremental costs for risankizumab versus infliximab SC of [redacted] (compared with [redacted] in the corrected base case), and incremental QALYs of [redacted] (compared with [redacted] in the company base case). Furthermore, compared with the corrected base case, increasing the maximum treatment duration for biologic therapies to 20 years had a large impact on incremental QALYs for risankizumab versus infliximab SC ([redacted]). Relatively small differences in results are observed when changing assumptions around background mortality rates, adverse event rates and durations and the utility value in the surgery health state.

In the BF population, the cost-effectiveness results appear most sensitive to assumptions regarding the maximum treatment duration. Assuming a 20-year maximum treatment duration for biologic therapies results in incremental costs for risankizumab versus vedolizumab SC of [redacted] (compared with [redacted] in the corrected base case) and incremental QALYs of [redacted] (compared with [redacted] in the company base case), with a resulting ICER of £65,837.

Table 33: EAG's exploratory analyses – CCF population (risankizumab versus infliximab SC)

Scenario	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
EAG corrected company base-case	[redacted]	[redacted]	Dominated, -£102,827	N/A
Maximum treatment duration, 5 years	[redacted]	[redacted]	Dominated, -£70,999	+£31,828
Maximum treatment duration, 10 years	[redacted]	[redacted]	£109,669	+£212,496
Maximum treatment duration, 20 years	[redacted]	[redacted]	£52,499	+£155,326
Maximum treatment duration, 40 years	[redacted]	[redacted]	£61,486	+£164,313
Residual treatment effect, 26 weeks	[redacted]	[redacted]	Dominated, -£100,343	+£2,484

Scenario	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
Treatment discontinuation rate equivalent to risankizumab for all biologics	[redacted]	[redacted]	Dominated, -£97,765	+£5,062
Single maintenance network	[redacted]	[redacted]	Dominated, -£81,619	+£21,208
Single maintenance network, with non-exponential transition matrix estimation	[redacted]	[redacted]	Dominated, -£81,870	+£20,957
Single maintenance network, with non-exponential transition matrix estimation and adjustment of both ordered probit cut points	[redacted]	[redacted]	Dominated, -£83,597	+£19,231
Single maintenance network (adjusted for a temporal effect)	[redacted]	[redacted]	Dominated, -£78,107	+£24,720
Single maintenance network (adjusted for a temporal effect), and non-exponential transition matrix estimation	[redacted]	[redacted]	Dominated, -£78,005	+£24,822
Single maintenance network (adjusted for a temporal effect), with non-exponential transition matrix estimation and adjustment of both ordered probit cut points	[redacted]	[redacted]	Dominated, -£76,763	+£26,064
SMR for CD compared with the general population = 1.38	[redacted]	[redacted]	Dominated, -£102,860	-£33
SMR for CD compared with the general population = 3.20	[redacted]	[redacted]	Dominated, -£103,019	-£192
AEs equivalent to risankizumab	[redacted]	[redacted]	Dominated, -£100,355	£2,472
AE duration, 1 week	[redacted]	[redacted]	Dominated, -£98,592	£4,235
AE duration, 4 weeks	[redacted]	[redacted]	Dominated, -£113,414	-£10,587
AE duration, 8 weeks	[redacted]	[redacted]	Dominated, -£149,448	-£46,621
Health state utility values, risankizumab trials, EQ-5D, linear mixed model	[redacted]	[redacted]	Dominated, -£123,458	-£20,630
Surgery versus moderate-to-severe, health state utility multiplier = 0.9	[redacted]	[redacted]	Dominated, -£102,672	+£155
Surgery versus moderate-to-severe, health state utility multiplier = 0.8	[redacted]	[redacted]	Dominated, -£102,517	+£310

Abbreviations: AE, Adverse event; CCF, conventional care failure; CD, Crohn's Disease; EAG, Evidence Assessment Group; EQ-5D, European Quality of Life Five Dimension; ICER, incremental cost-effectiveness ratio; LMM, linear mixed model; N/A, not applicable; QALY, quality adjusted life year; SC, subcutaneous; SMR, standardized mortality ratio.

Table 34: EAG's exploratory analyses – BF population (risankizumab versus vedolizumab SC)

Scenario	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
EAG corrected company base-case	[redacted]	[redacted]	Dominant, -£26,902	N/A
Maximum treatment duration, 5 years	[redacted]	[redacted]	£32,798	+£59,699
Maximum treatment duration, 10 years	[redacted]	[redacted]	£53,111	+£80,013
Maximum treatment duration, 20 years	[redacted]	[redacted]	£65,837	+£92,739
Maximum treatment duration, 40 years	[redacted]	[redacted]	£71,529	+£98,431
Residual treatment effect, 26 weeks	[redacted]	[redacted]	Dominant, -£19,550	+£7,352
Treatment discontinuation rate equivalent to risankizumab for all biologics	[redacted]	[redacted]	Dominant, -£32,609	-£5,707
Single maintenance network	[redacted]	[redacted]	Dominant, -£12,547	+£14,355
Single maintenance network, with non-exponential transition matrix estimation	[redacted]	[redacted]	Dominant, -£9,709	+£17,193
Single maintenance network, with non-exponential transition matrix estimation and adjustment of both ordered probit cut points	[redacted]	[redacted]	£63,812	+£90,714
Single maintenance network (adjusted for a temporal effect)	[redacted]	[redacted]	Dominant, -£22,368	+£4,534
Single maintenance network (adjusted for a temporal effect), and non-exponential transition matrix estimation	[redacted]	[redacted]	Dominant, -£19,745	+£7,157
Single maintenance network (adjusted for a temporal effect), with non-exponential transition matrix estimation and adjustment of both ordered probit cut points	[redacted]	[redacted]	Dominant, -£3,869	+£23,033
SMR for CD compared with the general population = 1.38	[redacted]	[redacted]	Dominant, -£26,891	+£11
SMR for CD compared with the general population = 3.20	[redacted]	[redacted]	Dominant, -£26,839	+£63
AEs equivalent to risankizumab	[redacted]	[redacted]	Dominant, -£29,641	-£2,739
AE duration, 1 week	[redacted]	[redacted]	Dominant, -£27,615	-£713
AE duration, 4 weeks	[redacted]	[redacted]	Dominant, -£25,635	£1,267

Scenario	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
AE duration, 8 weeks	[redacted]	[redacted]	Dominant, -£23,596	+£3,306
Health state utility values, risankizumab trials, EQ-5D, linear mixed model	[redacted]	[redacted]	Dominant, -£31,061	-£4,159
Surgery versus moderate-to-severe, health state utility multiplier = 0.9	[redacted]	[redacted]	Dominant, -£26,879	+£23
Surgery versus moderate-to-severe, health state utility multiplier = 0.8	[redacted]	[redacted]	Dominant, -£26,856	+£46

Abbreviations: AE, Adverse event; BF, biologic failure; CD, Crohn's Disease; EAG, Evidence Assessment Group; EQ-5D, European Quality of Life Five Dimension; ICER, incremental cost-effectiveness ratio; LMM, linear mixed model; N/A, not applicable; QALY, quality adjusted life year; SC, subcutaneous; SMR, standardized mortality ratio.

6.3. EAG's preferred assumptions

The EAG's preferred adaptations to the EAG-corrected company base case draw on several of the exploratory analyses described and presented in Section 6.2. Table 35 and Table 36 demonstrate the deterministic, pairwise, step-by-step impact of the EAG-preferred assumptions, from the EAG-corrected company base case to the EAG preferred base case, against the optimal in the CCF (infliximab SC) and BF (vedolizumab SC) populations (as described in Section 1.7), respectively.

The EAG note that neither the company's base case nor the EAG's preferred base case address issues with the company's chosen model structure (Key Issue 4) and approach to dose escalation (Key Issue 6).

Table 37 and Table 38 summarise incremental deterministic and probabilistic results for the EAG base case in the CCF and BF populations, respectively.

Table 35: EAG's preferred model assumptions – CCF population (risankizumab versus infliximab SC)

Preferred assumption	Section in EAG report	Cumulative ICER, £/QALY (stepwise change)
Company's base case (probabilistic)	Section 5.1.1.1	Dominated, -£81,752
Company's base case (deterministic)	Section 5.1.1.1	Dominated, -£84,028
EAG corrected company base case	Section 6.1	Dominated, -£102,827 (-£18,800)

Preferred assumption	Section in EAG report	Cumulative ICER, £/QALY (stepwise change)
+ Maximum treatment duration of 20 years for all biologic treatments	Section 4.2.6.7 and 6.2.3	£52,499 (+£155,326)
+ Residual treatment effect of 26 weeks for all biologic treatments	Section 4.2.6.7 and 6.2.3	£57,503 (+£5,004)
+ Single maintenance network, with an estimated maintenance placebo remission proportion that is adjusted for a temporal effect	Section 4.2.6 and 6.2.1	Dominated, -£76,611 (-£134,114)
+ Transition matrices estimated by adjusting both the remission | mild and mild | moderate-to-severe cut points, and without an exponential assumption to estimate 2-week transitions	Section 4.2.6 and 6.2.6	Dominated, -£75,237 (+£1,374)
+ Health state utility values estimated using a mixed linear model	Section 4.2.7.1 and 6.2.10	Dominated, -£88,792 (-£13,555)
EAG's preferred base case (deterministic)	Section 6.2 and 6.3	Dominated, -£88,792
EAG's preferred base case (probabilistic)	Section 6.2 and 6.3	Dominated, -£90,018

Abbreviations: CCF, conventional care failure; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SC, subcutaneous.

Table 36: EAG's preferred model assumptions – BF population (risankizumab versus vedolizumab SC)

Preferred assumption	Section in EAG report	Cumulative ICER, £/QALY (stepwise change)
Company's base case (probabilistic)	Section 5.1.1.1	Dominant, -£44,642
Company's base case (deterministic)	Section 5.1.1.1	Dominant, -£43,738
EAG corrected company base case	Section 6.1	Dominant, -£26,902 (+£16,836)
+ Maximum treatment duration of 20 years for all biologic treatments	Section 4.2.6.7 and 6.2.3	£65,837 (+£92,739)
+ Residual treatment effect of 26 weeks for all biologic treatments	Section 4.2.6.7 and 6.2.3	£66,781 (-£943)
+ Single maintenance network, with an estimated maintenance placebo remission proportion that is adjusted for a temporal effect	Section 4.2.6 and 6.2.1	£55,959 (-£10,822)
+ Transition matrices estimated by adjusting both the remission | mild and mild | moderate-to-severe cut points, and without an exponential assumption to estimate 2-week transitions	Section 4.2.6 and 6.2.6	£119,509 (+£63,550)
+ Health state utility values estimated using a mixed linear model	Section 4.2.7.1 and 6.2.10	£143,088 (+£23,579)

Preferred assumption	Section in EAG report	Cumulative ICER, £/QALY (stepwise change)
EAG's preferred base case (deterministic)	Section 6.2 and 6.3	£143,088
EAG's preferred base case (probabilistic)	Section 6.2 and 6.3	£142,074

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SC, subcutaneous.

Table 37: EAG incremental base case results – CCF population

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
EAG preferred deterministic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]	-	-	-	-
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£5,536	£5,536
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£56,481	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£52,086	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£1,349,539	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£4,358,832	Dominated
EAG preferred probabilistic base case						
ADA 160/80 biosimilar	[redacted]	[redacted]				
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£6,744	£6,744
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£55,111	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£48,951	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£867,497	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	-£91,825,236	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; IV, intravenous; QALYs, quality-adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

Table 38: EAG incremental base case results – BF population

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
EAG preferred deterministic base case						
VDZ SC	[redacted]	[redacted]	-	-	-	-
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£2,198,195	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£252,156	Extendedly dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£143,088	£143,088
EAG preferred probabilistic base case						
VDZ SC	[redacted]	[redacted]	-	-	-	-
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£1,487,732	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£248,239	Extendedly dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£142,074	£142,074

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; IV, intravenous; QALYs, quality-adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

6.4. EAG scenarios around the EAG preferred base case

In Section 6.1 of this report, EAG corrections to the company's executable cost-effectiveness model are described. In Section 6.2, several exploratory analyses around the EAG-corrected company base case are individually presented using pairwise cost-effectiveness analysis. In Section 6.3, both the step-by-step effect of EAG preferred changes on pairwise cost-effectiveness results, and fully incremental EAG preferred base case results are reported.

Here, in Section 6.4, additional EAG scenario analyses applied to the EAG-preferred base case are presented for the CCF and BF populations, using fully incremental, deterministic analysis.

6.4.1. Maximum treatment duration assumption

As outlined in Key Issue 5, the EAG has significant concerns with the company's treatment discontinuation assumptions; in particular, assuming all patients discontinue biologic therapy at 52 weeks. As such, the EAG's preferred base case assumes a 20-year maximum treatment duration rate for all biologic therapies.

However, Table 39 (CCF population) and Table 40 (BF population) present full incremental analysis results from a scenario around the EAG's preferred base case, in which the EAG's preferred 20-year maximum treatment duration assumption is relaxed to both 1 and 5 years.

In the CCF population, consistent with the EAG preferred base case, risankizumab is dominated by TNF-alpha inhibitors when the maximum treatment duration is assumed to be 1 or 5 years. In the BF population, when reducing the maximum treatment duration from 20 years to 5 years the ICER for risankizumab versus vedolizumab SC falls from £143,088 (EAG base case) to £103,081. Moreover, lowering the maximum treatment duration to 1 year leads to a further reduction in the ICER for risankizumab versus vedolizumab SC (£568).

Table 39: Maximum treatment duration scenarios around EAG preferred base case (CCF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Maximum treatment duration of 1 year						
ADA 160/80 biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	-	-
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£3,825	£3,825
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£19,503	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£23,242	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	-£358,121	Dominated
UST6	[redacted]	[redacted]	[redacted]	[redacted]	-£868,516	Dominated
Maximum treatment duration of 5 years						
ADA 160/80 biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	-	-
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£49,495	Dominated
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£8,824	£8,824
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£42,057	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	-£560,218	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	-£6,699,647	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; IV, intravenous; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

Table 40: Maximum treatment duration scenarios around EAG preferred base case (BF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Maximum treatment duration of 1 year						
VDZ SC	[redacted]	[redacted]	[redacted]	[redacted]	-	-
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£568	£568
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£1,332,202	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£65,355	Dominated
Maximum treatment duration of 5 years						
VDZ SC	[redacted]	[redacted]	-	-	-	-
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£2,018,541	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£215,997	Extendedly dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£103,081	£103,081

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; IV, intravenous; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

6.4.2. Estimation and application of maintenance treatment effectiveness assumptions

As outlined in Key Issue 6, the EAG has significant concerns with the company's maintenance treatment effectiveness estimates and assumptions. The EAG recommends the use of a single network for the maintenance NMA and a placebo remission model allowing for plausible causes of heterogeneity (in particular, a temporal association with the time at which individual clinical trials were conducted). Furthermore, the EAG prefers transition matrices that are calibrated by adjusting both the remission | mild and mild | moderate-to-severe ordered probit cut points, and a transition matrix cycle length adjustment approach that does not rely on an exponential assumption.

However, a scenario analysis is presented around the EAG-preferred base case in Table 41 (CCF) and Table 42 (BF), which relaxes the EAG's preferred assumptions around the single maintenance network, placebo remission temporal adjustment, transition matrix calibration method and transition matrix cycle length adjustment approach.

In the CCF population, when the majority of EAG preferred assumptions are combined with the company's base case NMA and transition matrix calibration and adjustment approach, risankizumab is associated with an ICER of £62,821 versus infliximab SC. In the equivalent BF population scenario, risankizumab is associated with an ICER of £79,559 versus vedolizumab SC.

Table 41: NMA and transition matrix calibration scenario around EAG preferred base case (CCF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
ADA 80/40	[redacted]	[redacted]	-	-	-	-
ADA 160/80 biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£46,941	Extendedly dominated
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£34,456	£34,456
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£41,283	£62,821
UST	[redacted]	[redacted]	[redacted]	[redacted]	£72,392	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£230,961	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; IV, intravenous; NMA, network meta-analysis; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

Table 42: NMA and transition matrix calibration scenario around EAG preferred base case (BF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Maximum treatment duration of 5 years						
VDZ SC	[redacted]	[redacted]				
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£3,190,924	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£208,011	Extendedly dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£79,559	£79,559

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; IV, intravenous; NMA, network meta-analysis; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

6.4.3. Health state utility estimation

As described in Key Issue 7, the company use OLS regression to estimate CDAI-based health state utility values from patient-reported risankizumab trial data in their base case. In the context of within-patient repeated measures, the EAG prefer to use health state utility values based on the same data but estimated using a (linear) mixed model that includes a random effect to account for repeated measures.

However, a scenario is presented in Table 43 (CCF) and Table 44 (BF) around the EAG-preferred base case, which combines the majority of EAG preferred assumptions with the company preferred OLS-estimated health state utility values. In this scenario, consistent with the EAG preferred base case, risankizumab is dominated in the CCF population incremental analysis. In the equivalent scenario in the BF population, the risankizumab ICER versus vedolizumab SC is £119,509 (compared with £143,088 in the EAG preferred base case).

Table 43: Health state utility estimation scenario (OLS estimation) around EAG preferred base case (CCF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
ADA 160/80 biosimilar	[redacted]	[redacted]	-	-	-	-
IFX SC	[redacted]	[redacted]	[redacted]	[redacted]	£4,947	£4,947
ADA 80/40	[redacted]	[redacted]	[redacted]	[redacted]	-£45,377	Dominated
IFX IV biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	£46,553	Dominated
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£5,472,524	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£13,855,370	Dominated

Abbreviations: ADA, adalimumab; CCF, conventional care failure; EAG, Evidence Assessment Group; IFX, infliximab; IV, intravenous; OLS, ordinary least squares; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab.

Table 44: Health state utility estimation scenario (OLS estimation) around EAG preferred base case (BF population)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
Maximum treatment duration of 5 years						
VDZ SC	[redacted]	[redacted]	-	-	-	-

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained	
					Versus baseline	Incremental analysis
VDZ IV	[redacted]	[redacted]	[redacted]	[redacted]	-£1,876,962	Dominated
UST	[redacted]	[redacted]	[redacted]	[redacted]	£193,271	Extended dominance
RZB	[redacted]	[redacted]	[redacted]	[redacted]	£119,509	£119,509

Abbreviations: BF, biologic failure; EAG, Evidence Assessment Group; IV, intravenous; OLS, ordinary least squares; QALYs, quality adjusted life years; RZB, risankizumab; SC, subcutaneous; UST, ustekinumab; VDZ, vedolizumab.

6.5. Conclusions of the cost-effectiveness section

The company's cost-effectiveness analysis estimates that risankizumab is dominated by (more costly and less effective than) relevant comparator treatment options in NHS England and, as such, is not a cost-effective treatment option for patients with moderately-to-severely active CD in a CCF population. However, the company's cost-effectiveness analysis estimates that risankizumab is dominant (generates more QALYs at a lower cost), when compared with relevant NHS England treatment options, for patients with moderately-to-severely active CD in a BF population.

The EAG was not satisfied that the cost-effectiveness evidence submitted by the company fully addressed the decision problem at hand. Although the company's cost-effectiveness analysis provides an estimate of the lifetime cost and QALY implications of introducing risankizumab to NHS England practice for moderately-to-severely active CD (from an NHS and PSS cost perspective and a direct health effect perspective for patients), the EAG has significant concerns with the cost-effectiveness evidence that neither the company's base case nor the EAG's preferred base case can address. Primarily, as outlined in Key Issue 4, the EAG are concerned that company's CDAI-based model structure is not reflective of relevant patient outcomes. Furthermore, adding risankizumab to the list of currently available treatment options currently available would extend the plausible biologic options available to treat each patient, yet the company assumes that after the initial therapy (up to 52 weeks in the company's base case), all patients move to conventional care. The EAG are concerned that this assumption does not reflect the treatment pathway as described by both the company and the EAG's clinical expert.

The EAG was not satisfied that the company's cost-effectiveness results provide an unbiased estimate of the likely cost-effectiveness of moderately-to-severely active CD. The company's

cost-effectiveness analysis is largely driven by assumptions regarding the estimation and application of treatment effectiveness, and the assumed maximum treatment duration for biologic therapies.

The EAG was unable to provide alternative solutions for all identified issues; namely, the chosen model structure as described above, the company's dose escalation assumptions which the EAG worry bias in favour of risankizumab (Key Issue 6) and the method of administration for risankizumab (Key Issue 8). However, the EAG was able to carry out several exploratory analyses (as described throughout Section 6.2); some of which were preferred adaptations which were used to form the EAG preferred base case (as described in Section 6.3). The EAG's preferred analysis increases the maximum treatment duration for all biologic therapies to 20 years, reduces the residual treatment effect following biologic therapy to 26 weeks, uses a single maintenance network that is combined with an estimated maintenance placebo remission proportion that is adjusted for a temporal effect, estimates transition matrices by adjusting both the remission | mild and mild | moderate-to-severe cut points in the ordered probit model, adjusts transition matrices for a 2-week cycle length without using an exponential assumption, and estimates health state utility values estimated using a mixed linear model.

In line with the company's cost-effectiveness estimate, in the CCF population, risankizumab is dominated by (more costly and less effective than) relevant NHS England treatment options in the EAG's preferred base case. In the BF population, vedolizumab IV is dominated by vedolizumab SC, and ustenkinumab extendedly dominated risankizumab. The ICER for risankizumab versus vedolizumab SC (the optimal comparator in the BF population incremental analysis), in the EAG's preferred base case, falls above the typical NICE willingness-to-pay threshold of £20,000 to £30,000 per QALY gained.

References

1. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet. 2012;380(9853):1590-605.

2. Petagna L, Antonelli A, Ganini C, Bellato V, Campanelli M, Divizia A, et al. Pathophysiology of Crohn's disease inflammation and recurrence. Biol Direct. 2020;15(1):23.

3. NHS Inform. Crohn’s Disease – Illness & Conditions 2020 [Available from: https://www.nhsinform.scot/illnesses-and-conditions/stomach-liver-and-gastrointestinal-tract/crohns-disease (last accessed: August 2022.)

4. Perler BK, Ungaro R, Baird G, Mallette M, Bright R, Shah S, et al. Presenting symptoms in inflammatory bowel disease: descriptive analysis of a community-based inception cohort. BMC Gastroenterol. 2019;19(1):47.

5. Crohn's, Colitis UK. Mental health and wellbeing. 2021. Available from: https://www.crohnsandcolitis.org.uk/about-crohns-and-colitis/publications/mental-wellbeing (Accessed June 2022).

6. Floyd DN, Langham S, Séverac 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(2):299-312.

7. Gay M, Barrison I, Bean K, Bevan S, Carroll R, Clark S. Crohn’s, Colitis and Employment-from career aspirations to reality. Hertfordshire: Crohn’s and Colitis UK. 2011.

8. Irving P, Barrett K, Nijher M, de Lusignan S. Prevalence of depression and anxiety in people with inflammatory bowel disease and associated healthcare use: population-based cohort study. Evid Based Ment Health. 2021;24(3):102-9.

9. Lönnfors S, Vermeire S, Greco M, Hommes D, Bell C, Avedano L. IBD and health-related quality of life -- discovering the true impact. J Crohns Colitis. 2014;8(10):1281-6.

10. Norton BA, Thomas R, Lomax KG, Dudley-Brown S. Patient perspectives on the impact of Crohn's disease: results from group interviews. Patient Prefer Adherence. 2012;6:509-20.

11. NHS Digital. Hospital Episode Statistics for England. Admitted Patient Care statistics, 2019-20 2022 [Available from: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2019-20 (last accessed: August 2022.)

12. King D, Reulen RC, Thomas T, Chandan JS, Thayakaran R, Subramanian A, et al. Changing patterns in the epidemiology and outcomes of inflammatory bowel disease in the United Kingdom: 2000-2018. Aliment Pharmacol Ther. 2020;51(10):922-34.

13. Office for National Statistics. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland 2021 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland (last accessed: August 2022.)

14. National Institute for Health and Care Excellence. Crohn's disease: management (NG129) Manchester: NICE; 2019 [Available from: https://www.nice.org.uk/guidance/ng129/resources/crohns-disease-management-pdf-66141667282885 (last accessed: August 2022.)

15. Kashani A, Schwartz DA. The Expanding Role of Anti-IL-12 and/or Anti-IL-23 Antibodies in the Treatment of Inflammatory Bowel Disease. Gastroenterol Hepatol (N Y). 2019;15(5):255-65.

16. Ma C, Panaccione R, Khanna R, Feagan BG, Jairath V. IL12/23 or selective IL23 inhibition for the management of moderate-to-severe Crohn's disease? Best Pract Res Clin Gastroenterol. 2019;38-39:101604.

17. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Bmj. 2017;358:j4008.

18. National Institute for Health and Care Excellence. Single technology appraisal and highly specialised technologies evaluation: User guide for company evidence submission template (PMG24) 2022 [Available from: https://www.nice.org.uk/process/pmg24/chapter/clinical-effectiveness#qualityassessment-of-the-relevant-clinical-effectiveness-evidence.)

19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Bmj. 2019;366:l4898.

20. AbbVie. A Multicenter, Randomized, Double-Blind, Placebo Controlled Induction Study of the Efficacy and Safety of Risankizumab in Subjects with Moderately to Severely Active Crohn's Disease (ADVANCE [M16-006]). Clinical Study Report. Risankizumab. 2021.

21. AbbVie. A Multicenter, Randomized, Double-Blind, Placebo Controlled Induction Study to Assess the Efficacy and Safety of Risankizumab in Subjects with Moderately to Severely Active Crohn's Disease Who Failed Prior Biologic Treatment (MOTIVATE [M15-991]). Clinical Study Report. Risankizumab. 2021.

22. D'Haens G, Panaccione R, Baert F, Bossuyt P, Colombel J-F, Danese S, et al. Risankizumab as induction therapy for Crohn's disease: results from the phase 3 ADVANCE and MOTIVATE induction trials. The Lancet. 2022;399(10340):2015-30.

23. AbbVie. Multicenter, Randomized, Double-Blind, Placebo-Controlled 52-Week Maintenance and an Open-Label Extension Study of the Efficacy and Safety of Risankizumab in Subjects with Crohn's Disease (FORTIFY [M16-000]). Clinical Study Report. Risankizumab. 2021.

24. Ferrante M, Panaccione R, Baert F, Bossuyt P, Colombel JF, Danese S, et al. Risankizumab as maintenance therapy for moderately to severely active Crohn's disease:

results from the multicentre, randomised, double-blind, placebo-controlled, withdrawal phase 3 FORTIFY maintenance trial. Lancet. 2022;399(10340):2031-46.

25. AbbVie. Crohn’s disease HTA advisory board report. Data on file. 2021. (MAP ID: 2021-GB-MNP-00091).

26. 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(2):160-73.

27. Watanabe K, Motoya S, Ogata H, Kanai T, Matsui T, Suzuki Y, et al. Effects of vedolizumab in Japanese patients with Crohn's disease: a prospective, multicenter, randomized, placebo-controlled Phase 3 trial with exploratory analyses. J Gastroenterol. 2020;55(3):291-306.

28. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Network meta-analysis for decision-making: John Wiley & Sons; 2018.

29. Glanville J, Kaunelis D, Mensinkai S. How well do search filters perform in identifying economic evaluations in MEDLINE and EMBASE. International Journal of Technology Assessment in Health Care. 2009;25(4):522-9.

30. 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. Aliment Pharmacol Ther. 2009;30(3):265-74.

31. Catt H, Bodger K, Kirkham JJ, Hughes DA. Value Assessment and Quantitative Benefit-Risk Modelling of Biosimilar Infliximab for Crohn's Disease. Pharmacoeconomics. 2019;37(12):1509-23.

32. Lindsay J, Punekar YS, Morris J, Chung-Faye G. Health-economic analysis: cost-effectiveness of scheduled maintenance treatment with infliximab for Crohn's disease--modelling outcomes in active luminal and fistulizing disease in adults. Aliment Pharmacol Ther. 2008;28(1):76-87.

33. Loftus EV, Jr., Johnson SJ, Yu AP, Wu EQ, Chao J, Mulani PM. Cost-effectiveness of adalimumab for the maintenance of remission in patients with Crohn's disease. Eur J Gastroenterol Hepatol. 2009;21(11):1302-9.

34. Robson RE, Vanoli A, Tolley K. Cost-effectiveness of vedolizumab compared with conventional therapy in patients with Crohn's disease in the United Kingdom (UK). Value Health. 2018;21(Supplement 3):S455-S6.

35. 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. J Crohns Colitis. 2013;7(2):167-74.

36. Wilson M, Lucas A. Cost-effectiveness of vedolizumab versus ustekinumab in United Kingdom Crohn's disease patients who previously failed TNF inhibitor therapy. Value Health. 2018;21(Supplement 3):S145.

37. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ. 1996;313(7052):275-83.

38. Efficace F, Bottomley A, Osoba D, Gotay C, Flechtner H, D'Haese S, et al. Beyond the development of health-related quality-of-life (HRQOL) measures: a checklist for evaluating HRQOL outcomes in cancer clinical trials--does HRQOL evaluation in prostate cancer research inform clinical decision making? J Clin Oncol. 2003;21(18):3502-11.

39. Bassi A, Dodd S, Williamson P, Bodger K. Cost of illness of inflammatory bowel disease in the UK: a single centre retrospective study. Gut. 2004;53(10):1471-8.

40. Buchanan J, Wordsworth S, Ahmad T, Perrin A, Vermeire S, Sans M, et al. Managing the long term care of inflammatory bowel disease patients: The cost to European health care providers. J Crohns Colitis. 2011;5(4):301-16.

41. Burisch J, Vardi H, Pedersen N, Brinar M, Cukovic-Cavka S, Kaimakliotis I, et al. Costs and resource utilization for diagnosis and treatment during the initial year in a European inflammatory bowel disease inception cohort: an ECCO-EpiCom Study. Inflamm Bowel Dis. 2015;21(1):121-31.

42. Ding Z, Muser E, Hoops T, Izanec JL, Ma T, Gao L-L, et al. S723 Improvements in Work Productivity Loss and Associated Cost Reductions with Ustekinumab and Adalimumab in Biologic-Naïve Patients with Moderately-to-Severely Active Crohn's Disease: Results From the SEAVUE Study. Official journal of the American College of Gastroenterology| ACG. 2021;116:S330.

43. Ghosh N, Premchand P. A UK cost of care model for inflammatory bowel disease. Frontline Gastroenterol. 2015;6(3):169-74.

44. Jin R, Nduka C, Courmier D, Knight H, Meadows R, Piercy J, et al. REAL-WORLD UTILIZATION OF ADALIMUMAB BIOSIMILAR (ABP 501) IN PATIENTS WITH INFLAMMATORY BOWEL DISEASE IN EUROPE. Gastroenterology. 2022;162(3):S109-S10.

45. Lim S, Gros Alcade B, Lehmann A, Lindsay J, Gaya D, Caulfield L, et al. P554 Vedolizumab Associated quality of Life and inflammatory bowel disease control Upon Entering a subcutaneous switching program (VALUE) – Results from a multi-centre UK study. Journal of Crohn's and Colitis. 2022;16(Supplement_1):i499-i500.

46. Louis E, Löfberg R, Reinisch W, Camez A, Yang M, Pollack PF, et al. Adalimumab improves patient-reported outcomes and reduces indirect costs in patients with moderate to severe Crohn's disease: Results from the CARE trial ☆. Journal of Crohn's and Colitis. 2013;7(1):34-43.

47. Moore AC, Russell D, Dart R, Dias S, Ghuman S, Petri B, et al. P411 Stable pharmacokinetics and biochemical markers of inflammatory bowel disease activity: Results from a real-world adalimumab biosimilar switching program. Journal of Crohn's and Colitis. 2022;16(Supplement_1):i401-i2.

48. Obando C, Ding Z, Gao L, Pires A, Muser E, Gasink C. Reduced Healthcare Costs of Crohn's Disease-Related Hospitalizations and Surgeries in Patients Initiated with Ustekinumab

in the IM-UNITI LTE Study. 2018. Presented at Academy of Managed Care Pharmacy Nexus 2018 (abstract #K4). Available at: https://www.jmcp.org/doi/pdf/10.18553/jmcp.2018.24.10-a.s1 (Accessed June 2022).

49. Peyrin-Biroulet L, Vermeire S, D'Haens GR, Panés J, Dignass A, Magro F, et al. DOP90 Efficacy of the treat-to-target approach in modifying disease course with ustekinumab in patients with moderate-to-severe Crohn's Disease: Results from the STARDUST trial. Journal of Crohn's and Colitis. 2022;16(Supplement_1):i132-i4.

50. Ponsioen CY, de Groof EJ, Eshuis EJ, Gardenbroek TJ, Bossuyt PMM, Hart A, et al. Laparoscopic ileocaecal resection versus infliximab for terminal ileitis in Crohn's disease: a randomised controlled, open-label, multicentre trial. Lancet Gastroenterol Hepatol. 2017;2(11):785-92.

51. Teeple A, Muser E, Naessens D, Lueza B, Kamal Z, Colonier AF, et al. S0464 Cost Reduction Associated With Improvement in Work Productivity Loss From Ustekinumab Treatment by Disease Severity States in Crohn's Disease Patients: Results From the IM-UNITI Clinical Trial. Official journal of the American College of Gastroenterology| ACG. 2020;115:S233.

52. Bewtra M, Kaiser LM, TenHave T, Lewis JD. Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. Inflamm Bowel Dis. 2013;19(3):599-613.

53. National Institute for Health and Care Excellence. Vedolizumab for treating moderately to severely active Crohn's disease after prior therapy [TA352]. Manchester: NICE; 2014 [Available from: https://www.nice.org.uk/guidance/ta352/documents/crohns-disease-moderate-to-severe-vedolizumab-committee-papers2 (last accessed: August 2022.)

54. National Institute for Health and Care Excellence. Ustekinumab for treating moderately to severely active Crohn's disease after prior therapy STA (TA456). Manchester: NICE; 2017 [Available from: https://www.nice.org.uk/guidance/ta456/documents/committee-papers (Accessed June 2022) (last accessed: August 2022.)

55. Chhatwal J, Jayasuriya S, Elbasha EH. Changing Cycle Lengths in State-Transition Models: Challenges and Solutions. Med Decis Making. 2016;36(8):952-64.

56. Hernández Alava M, Wailoo A, Pudney S, Gray L, Manca A. Mapping clinical outcomes to generic preference-based outcome measures: development and comparison of methods. 2020;24:34.

57. Buxton MJ, Lacey LA, Feagan BG, Niecko T, Miller DW, Townsend RJ. Mapping from disease-specific measures to utility: an analysis of the relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. Value Health. 2007;10(3):214-20.

58. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health. 2010;13(5):509-18.

59. National Health Service. National Schedule of NHS Costs 2019/20 2022 [Available from: https://www.england.nhs.uk/publication/2019-20-national-cost-collection-data-publication/ (last accessed: August 2022.)