Ozanimod for treating moderately to severely active ulcerative colitis [ID3841]

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	**Amanda Brand**[1] **Brian O'Toole**[1] **Madhusubramanian Muthukumar**[1] **Justin Matthews**[1] **Fraizer Kiff**[1] **Naomi Shaw**[1] **Edward CF Wilson**[1] **Jonathan Digby-Bell**[2] **Louise Crathorne**[1] **G.J. Melendez-Torres**[1] [1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter [2] Consultant Gastroenterologist, Royal Devon & Exeter Hospital
Correspondence to	Amanda Brand 3.09 South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; a.brand2@exeter.ac.uk
Date completed	22/03/2022
Source of funding	This report was commissioned by the NIHR Systematic Reviews Programme as project number 13/54/43.

Declared competing interests of the authors

Dr Digby-Bell received payment of delegate fee from AbbVie (a stakeholder in technology appraisal ID3841) for conference attendance at the British Society of Gastroenterology annual meeting 2021.

Acknowledgments

The authors acknowledge the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe (both PenTAG) and clinical expert advice provided by Dr James Goodhand (consultant gastroenterologist, Royal Devon & Exeter Hospital). Dr Goodhand received grant funding over the last 18 months from Biogen and Celltrion (companies associated with the evaluation) for work completed assessing outcomes from COVID-19 infection and vaccination in patients treated with infliximab and vedolizumab.

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

Brand, O'Toole, Muthukumar, Matthews, Kiff, Shaw, Wilson, Digby-Bell, Crathorne, Melendez-Torres. Ozanimod for treating moderately to severely active ulcerative colitis [ID3841] A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2022.

Author Contributions:	
Amanda Brand	Project lead, lead for the ERG's critical appraisal of the clinical evidence of the company submission, writing and editorial input
Brian O'Toole	Lead for the ERG's appraisal of the economic evidence, drafted economic sections of the report, writing and editorial input
Madhusubramanian Muthukumar	Critical appraisal of the economic evidence, checked and re-analysed the economic model, carried out ERG base case analyses and further scenario analyses, and drafted economic sections of the report
Justin Matthews	Critical appraisal of the clinical evidence, conducted additional clinical work on the NMA for the submission, and drafted sections of the report
Fraizer Kiff	Critical appraisal of the clinical evidence and drafted sections of the report
Naomi Shaw	Critical appraisal of the literature search strategies, conducted additional literature searching, and editorial review
Edward CF Wilson	Independent appraisal of the ERG model for errors and editorial review
Jonathan Digby-Bell	Clinical advice and review of draft report
Louise Crathorne	Critical appraisal of the company submission, writing and editorial input, and co-supervised the final report
G.J. Melendez-Torres	Critical appraisal of the company submission, writing and editorial input, and co-supervised the final report. Guarantor of the report

Abbreviations

ADA	adalimumab
AE	adverse event
AG	Assessment Group
ALT	alanine aminotransferase
AMSTAR	A MeaSurement Tool to Assess systematic Reviews
AST	aspartate aminotransferase
BID	*L. bis in die*/twice a day
β-hCG	beta-human chorionic gonadotropin
BMI	body mass index
BSC	best supportive care
CD	Crohn's disease
CEAC	cost-effectiveness acceptability curve
CI	confidence interval
CODA	Convergence Diagnosis and Output Analysis
CRD	Centre for Reviews and Dissemination
CrI	credible interval
CRP	C-reactive protein
cPAS	confidential patient access scheme
CS	company submission
CSR	clinical study report
CV	cardiovascular
CvT	conventional therapy
ECG	electrocardiogram
EMA	European Medicines Agency
EQ-5D(-5L)	European Quality of Life Five Dimension (Five Level)
ERG	Evidence Review Group
EU	European Union
FDA	Food and Drug Administration
FE	fixed effect
GOL	golimumab
HES	Hospital Episode Statistics
HRQoL	health-related quality of life
HTA	health technology assessment

ICER	incremental cost-effectiveness ratio
ICTRP	International Clinical Trials Registry Platform
IFX	infliximab
ITT	intention-to-treat
IV	intravenous
JAGS	Just another Gibbs sampler
JAK	Janus kinase
LYG	life years gained
MCID	minimally clinically important difference
MCMC	Markov chain Monte Carlo
MCS	mental composite summary score
MVH	Measurement in Valuation of Health
N/A	not applicable
NHB	net health benefit
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	network meta-analysis
NR	not reported
OLE	open-label extension
OR	odds ratio
OWSA	one-way sensitivity analysis
OZA	ozanimod
PBO	placebo
PCS	physical component summary
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
Q2W	*L. quaque* 2W/once every two weeks
QA	quality assessment
QALY	quality-adjusted life year
QD	*L. quaque die*/once a day
QoL	quality of life
RCT	randomised controlled trial
RBS	rectal bleeding sub-score
RE	random effects
SAE	serious adverse event

SC	subcutaneous
SD	standard deviation
SF-36	36-item Short Form Health Survey
SFS	stool frequency sub-score
SLR	systematic literature review
SmPC	Summary of Product Characteristics
SW	southwest
TA	technology appraisal
TDM	therapeutic drug monitoring
TEAE	treatment-emergent adverse event
TNFi	tumour necrosis factor inhibitor
TOF	tofacitinib
UC	ulcerative colitis
UCSS	ulcerative colitis symptom score
UST	ustekinumab
VAS	visual analogue scale
VEDO	vedolizumab
vs.	versus
WHO	World Health Organization
WTP	willingness-to-pay

1. EXECUTIVE SUMMARY

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

- Section 1.1 provides an overview of the key issues and the differences in the assumptions of the company and the ERG in economic analysis.
- 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 1.5 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 ERG report.
- Sections 1.6 and 1.7 provide an overview of the ERG's preferred base case and sensitivity analyses undertaken by the ERG.

All issues identified represent the ERG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1. Overview of the ERG's key issues

A brief overview of the key issues identified by the ERG 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 and 1.5.

Broadly speaking, the key clinical issues related to the exclusion of tofacitinib as a comparator as well as uncertainty surrounding the assumptions made in the analytical approach comparing ozanimod and its comparator treatments. This omission of tofacitinib and uncertainties around the network meta-analysis (NMA) have implications for the cost-effectiveness of ozanimod as well as for the positioning of ozanimod in a highly individualised treatment pathway. Furthermore, the ERG was of the opinion that random effects (RE) models should have been used to estimate clinical effectiveness in the NMAs, and that the use of fixed effect (FE) models may have resulted in inaccurate inputs of clinical effectiveness into the economic model. This, in turn, may have biased the results of the ICERs and increased the overall uncertainty of the cost-effectiveness evidence in the context of the decision problem. In terms of cost-effectiveness

issues, the ERG considered the exclusion of tofacitinib to have a high impact on cost-effectiveness and that its inclusion might result in considerably different fully incremental ICERs. The ERG also noted several concerns pertaining to the company's estimation of modelled transition probabilities and response rates for best supportive care (BSC) and uncertainty around the company's handling of subsequent treatments. Furthermore, the ERG did not consider the probabilistic sensitivity analysis (PSA) provided by the company to be helpful in decision-making, due to the exclusion of tofacitinib as a comparator and uncertainties around the NMA.

Table 1: Summary of key issues

ID	Summary of issues	Report sections
Key Issue 1	Tofacitinib was excluded as a comparator in TNFi-naïve and -experienced subgroups	Sections 1.3, 2.2.1, 2.3 and 6.1.1
Key Issue 2	Baseline risks for placebo anchors in the NMAs taken from the same trials those used for relative risk	Section 1.4, 3.3.2.4, 3.4.2.4 and 3.5.3.1
Key Issue 3	A RE model may be more appropriate for use in the maintenance phase NMAs	Section 1.4, 3.4.2.2 and 3.5.3.2
Key Issue 4	Modelled efficacy estimates for BSC in the post-active treatment phase	See Section 4.2.6, 4.2.6.3 and 6.3
Key Issue 5	There is uncertainty surrounding the handling of subsequent treatments/treatment sequencing within the model	See Sections 1.5 and 4.2.2.3
Key Issue 6	The PSA provided by the company was not considered helpful for decision making	See Sections 1.5 and 5.2.2

Abbreviations: NMA, network meta-analysis; PSA, probabilistic sensitivity analysis; RE, random effects; TNFi, tumour necrosis factor inhibitor

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

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

Company's preferred assumption	ERG preferred assumption	Report Sections
To exclude tofacitinib as a comparator	The ERG's preference was to include tofacitinib as a relevant comparator in both the TNFi-naïve and TNFi experienced subgroups. However, it was not possible to include tofacitinib into the company's model.	4.2.4 and 6.1.1
Use of FE model in both TNFi-naïve and TNFi-experienced subgroups for the maintenance phase NMAs as well as FE model in the TNFi-experienced subgroup for the induction phase NMA	The ERG preferred the use of RE models for maintenance phase NMAs. The ERG was unable to produce RE models with sufficient convergence (without using an informative prior distribution) and were therefore unable to use a RE model as part of its preferred base case.	1.4, 3.4.2.2, 3.5.3.2 and 4.2.6.4
Baseline risks for placebo anchors included in the NMAs calculated from the identical set of trials used to calculate the relative treatment effect	The ERG preferred to use the placebo arm values from individual trials included in the NMA that were more generalisable to the UK context. This has been considered as part of the ERG preferred base case.	1.4, 3.3.2.4, 3.4.2.4, 3.5.3.1 and 6.3
Remission transition probabilities in the BSC arm were estimated via loss of overall response (including remission). Furthermore, for BSC, loss of response and loss of response (No remission) were based on pooled population estimates.	The ERG preferred revised post-active treatment transition probabilities for BSC which include an alternative means of estimating remission probabilities for BSC based on 'loss of remission' (directly from the sustained remission estimates) and different BSC response rates for the TNF-naïve and TNF-experienced populations. These changes were incorporated into the ERG's preferred base case.	4.2.6.3 and 6.3

Abbreviations: BSC, best supportive care; ERG, Evidence Review Group; FE, fixed effect; NMA, network meta-analysis; RE, random effects; TNFi, tumour necrosis factor inhibitor

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:

- Moving patients to the remission and response (no remission) health states at the end of the induction phase and by sustaining remission and response (no remission) for a

proportion of patients at the end of the maintenance phase. [redacted].

- Assuming loss of response to treatment whilst in remission and response (no remission) i.e. probability of loss of response during the maintenance phase is considered for all treatments.
- Considering discontinuation due to adverse events during the maintenance phase. Patients discontinuing treatment received best supportive care, comprising components of conventional therapy, and entered the 'Active' ulcerative colitis (UC) health state accruing costs and QALYs associated with this health state.

Overall, the technology is modelled to affect costs by:

- Having lower administration costs compared to IV comparators (see Section 4.2.8.4) and by having lower drug acquisition costs compared some comparator treatments. However, the ERG noted that when cPAS discounts were included, the cost effectiveness of ozanimod compared to comparator treatments changed considerably (see cPAS Appendix).
- Monitoring requirements were assumed to be similar for all treatments, however ozanimod was assumed to require an electrocardiogram (ECG) during induction.

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

- Alternative utility values for modelled health states, variation in dose escalation assumptions, % of patients receiving subcutaneous (SC) vedolizumab and the inclusion of extended induction (as evident by the company's scenario analyses).

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

The ERG 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: Tofacitinib was excluded as a comparator in TNFi-naïve and -experienced subgroups

Report sections	Sections 1.3, 2.2.1, 2.3 and 6.1.1
Description of issue and why the ERG has identified it as important	The company excluded tofacitinib as a comparator to ozanimod in the TNFi-naïve subgroup, indicating that this treatment is not used at this line of treatment. Clinical advice to the ERG indicated that tofacitinib is used at this line of treatment in the UK clinical practice due to its rapid action and oral administration and confirmed that its use is increasing. Tofacitinib was also excluded as a comparator in the TNFi-experienced subgroup, with the company citing safety concerns. The company further indicated that the exclusion of tofacitinib was accepted by the committee for TA633, though the ERG noted that tofacitinib was included in the clinical effectiveness results and economics to provide the full picture of cost-effectiveness for ustekinumab. Clinical opinion to the ERG acknowledged these concerns but advised that these predominantly impact clinical practice in the US. Safety concerns regarding tofacitinib impact UK clinical practice far less, with concerns managed at the individual patient level. Furthermore, the ERG considered the treatment landscape to have evolved since TA633, with clinical experts advising that the use of tofacitinib is increasing. As a result, the ERG considered the exclusion of tofacitinib as a comparator in either subgroup to be an outstanding area of uncertainty.
What alternative approach has the ERG suggested?	During clarification, the ERG requested that tofacitinib be included as a relevant comparator in the model, for both subgroups, using the treatment efficacy estimates already derived from NMAs. However, this analysis was not provided to the ERG. For completeness, the ERG conducted a naïve cost comparison vs. tofacitinib, which assumed clinical equivalency between treatments in terms of efficacy. See Section 2.2.1, 2.3 and 6.1.1 for further discussion. Tofacitinib was found to be cost saving compared to ozanimod over lifetime of treatment, when considering the PAS price for tofacitinib.
What is the expected effect on the cost-effectiveness estimates?	The company did not provide the ERG with an additional analysis including tofacitinib as a relevant comparator. As such the base case results provided by the company should be interpreted with caution. The ERG expects the inclusion of tofacitinib to have a high impact on cost-effectiveness results. This may result in substantially different fully incremental ICERs.
What additional evidence or analyses might help to resolve this key issue?	The inclusion of clinical efficacy estimates for tofacitinib within the economic analysis would have sufficiently resolved this issue. The ERG was unable to amend the company's model to include tofacitinib, due to the lack of flexibility and time constraints.

Abbreviations: cPAS, confidential patient access scheme; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; NMA, network meta-analysis; TA, technology appraisal; TNFi, tumour necrosis factor inhibitor; vs., versus

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

The ERG 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: Baseline risks for placebo anchors in the NMAs taken from the same trials those used for relative risk

Report sections	Section 1.4, 3.3.2.4, 3.4.2.4 and 3.5.3.1
Description of issue and why the ERG has identified it as important	The baseline risks for placebo anchors included in the NMAs (in this context, the probability of being in non-response and non-remission, under placebo) were calculated from the identical set of trials used to calculate the relative treatment effect. The ERG noted that this was contrary to NICE guidance (Dias et al 2013),[1] which recommends separate modelling and sources of information for relative treatment and baseline effects. Several trials included in the NMA did not match well to the decision problem due to diverse settings, demographic as well as clinical features of participants, and concomitant medication use as described in Sections 3.3.2.4 and 3.4.2.4.
What alternative approach has the ERG suggested?	The baseline risk in the placebo anchors could be estimated more accurately through the use of a baseline risk in the placebo arm(s) of a study or studies that match the decision problem more closely. Though not fully in line with NICE guidance, given the timeframe of the appraisal, the ERG used the placebo arm values from individual trials included in the NMA that were more generalisable to the UK context to generate estimates of clinical effectiveness for its base case. This approach is described in Section 3.4.2.4.
What is the expected effect on the cost-effectiveness estimates?	The ERG's base case NMA, utilising placebo baseline risks from a more UK-appropriate trial, resulted in lower response rates for placebo, as well as for ozanimod and most of the active treatment comparators, as discussed in Section 3.5.3.1. The impact of the revised baseline placebo risk on cost-effectiveness in conjunction with the alternative transition probabilities for BSC in the post-active treatment phase has been discussed in Table 3 and Table 4.
What additional evidence or analyses might help to resolve this key issue?	Conducting the NMAs using baseline risk in the placebo arm, derived from a study that is highly generalisable to the UK context and was identified through a proper, protocol-driven systematic review would result in more generalisable estimations of treatment efficacy. The ERG could not conduct a comprehensive systematic review for highly generalisable evidence to inform its base case approach, as prescribed by NICE guidance, due to the timeframe of the appraisal, but used values from broadly representative single trials included in the NMA, as described in Section 3.4.2.4 and 3.5.3.1. As such, the ERG would like to highlight that there is residual uncertainty in its approach and recommends using placebo baseline risk values from a systematically identified trial that is highly specific to the decision problem to inform the NMA.

Abbreviations: BSC, best supportive care; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; NICE, National Institute for Health and Care Excellence; NMA, network meta-analysis

Key Issue 3: A random effects model may be more appropriate for use in the maintenance phase NMAs

Report sections	Section 1.4, 3.4.2.2 and 3.5.3.2
Description of issue and why the ERG has identified it as important	Clinical efficacy parameters (clinical response and remission) for all treatments were derived from the induction and maintenance NMAs conducted by the company. In the base case analysis the company opted to use a FE model in both TNFi-naïve and TNFi-experienced subgroups for the maintenance phase NMAs, as well as for the TNFi-experienced subgroup during induction. The ERG acknowledged the company's rationale for using the FE model for these subgroups and phases (namely that the fit was reasonable, RE models did not converge or had highly uncertain posterior SD). However, due to the high degree of heterogeneity amongst the studies included in the NMA, the ERG considered FE models to be inappropriate.
What alternative approach has the ERG suggested?	The ERG suggests RE models to estimate clinical response and remission for both the TNFi-naïve and TNFi-experienced subgroups in the maintenance phase, as well as for the TNFi-experienced subgroup in the induction phase to address the heterogeneity in the evidence base, as discussed in Section 3.4.2.2. The ERG attempted to re-run the company NMAs using RE models with alternative baseline placebo risk; these also failed to converge (see Section 3.5.3.2). To address non-convergence, the ERG suggests the use of appropriate informative prior distributions from literature.
What is the expected effect on the cost-effectiveness estimates?	Using a RE model is likely to result in different clinical efficacy estimates for all treatments. This is anticipated to have an impact on the cost effectiveness results, however the directional impact of this analysis could not be determined in the timeframe of the appraisal.
What additional evidence or analyses might help to resolve this key issue?	The ERG noted that overall, there is uncertainty surrounding the clinical data used in the economic model, due to heterogeneity amongst studies and the lack of direct trial data. The ERG recommends that, because of heterogeneity, the NMA be run using RE models with informative prior distributions in the event of non-convergence.

Abbreviations: ERG, Evidence Review Group; FE, fixed effect; ICER, incremental cost-effectiveness ratio; NMA, network meta-analysis; RE, random effects; SD, standard deviation; TNFi, tumour necrosis factor inhibitor

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

The ERG 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: Modelled efficacy estimates for BSC in the post-active treatment phase

Report sections	Section 4.2.6, 4.2.6.3 and 6.3
Description of issue and why the ERG has identified it as important	The ERG noted several concerns pertaining to the company's estimation of modelled transition probabilities and response rates for BSC (in the post active treatment phase), probability of remission, and the use the same pooled estimate for the BSC remission and response rates (for both TNF-naïve and TNFi-experienced subgroups). These concerns are discussed further in 4.2.6.3.
What alternative approach has the ERG suggested?	The ERG used an alternative approach to estimate remission state transition probabilities for BSC in the post active treatment phase i.e. these were calculated directly from the sustained remission estimates via 'loss of remission'. This approach has been incorporated into the ERG base case (see Section 6.3 for results). As noted in 4.2.6.3, the concerns surrounding the estimation of probability of remission, and the use of same pooled estimate for the BSC remission and response rates (for both TNF-naïve and TNF-experienced subgroups) were addressed as a result of using the alternative baseline placebo risk estimates which are different for TNF-naïve and TNF-experienced subgroups in the ERG's base case.
What is the expected effect on the cost-effectiveness estimates?	Due to the use of alternative placebo risk estimates derived by including only trials which are relevant to decision making, the overall response decreases across all treatments. For a complete description of the impact of these changes see Section 6.3.
What additional evidence or analyses might help to resolve this key issue?	The ERG base case analysis partly addressed this issue, however the uncertainty around the true remission estimates in the post active treatment phase remained.

Abbreviations: BSC, best supportive care; ERG, Evidence Review Group; QALY, quality-adjusted life years; TNFi, tumour necrosis factor inhibitor; UC, ulcerative colitis

Key Issue 5: There is uncertainty surrounding the handling of subsequent treatments/treatment sequencing within the model

Report sections	Section 1.5 and 4.2.2.3
Description of issue and why the ERG has identified it as important	The company did not consider subsequent treatment use/treatment sequencing in the base case. The ERG noted that the company provided some scenario analyses for the TNFi-naïve subgroup which assumed that patients who do not respond to their initial TNFi treatment can go on to received either vedolizumab or ustekinumab. Results were not overly sensitive to this analysis, however the ERG considered the scenario analysis to be somewhat limited (See Section 4.2.2.3).
What alternative approach has the ERG suggested?	The ERG considered undertaking additional scenario analyses using various treatment sequencing strategies, including within

Report sections	Section 1.5 and 4.2.2.3
	class switching and step up/ step down approaches. However the model was not flexible enough to allow for this. As such there is some uncertainty surrounding the impact of treatment sequencing on the base case ICER. See Section 4.2.2.3 for further discussion.
What is the expected effect on the cost-effectiveness estimates?	It is anticipated that the inclusion of alternative treatment sequence options will have a moderate impact on total treatment costs, and a minor impact on total QALYs (See Section 4.2.2.3). Given the small differences in costs between the modelled treatments, ICERs may vary once treatment sequencing is considered.
What additional evidence or analyses might help to resolve this key issue?	Updating the economic model to allow for the consideration of various treatment sequencing options would help to further explore uncertainty.

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life years; TNFi, tumour necrosis factor inhibitor

Key Issue 6: The PSA provided by the company was not considered helpful for decision making

Report sections	Section 1.5 and 5.2.2
Description of issue and why the ERG has identified it as important	Due to concerns relating to the NMA and the omission of tofacitinib as a relevant comparator in the economic model, the ERG considered that the PSA provided by the company is of limited use for decision-making and should be interpreted with caution.
What alternative approach has the ERG suggested?	Ideally the probabilistic analysis could have been done with the baseline risks of placebo arms from only the trials relevant to the decision problem and addressing the heterogeneity in the placebo arms adequately as this impacts the correlation of parameters. Further, tofacitinib could have also been included in the analysis presenting the true cost-effectiveness of relevant treatment alternatives both in the fully incremental analysis as well as with respect to the cost-effectiveness acceptability curve (CEAC).
What is the expected effect on the cost-effectiveness estimates?	The expected effect is that the probabilistic analysis would account for the correlation of treatment effects adequately and produce a CEAC including the relevant comparators presenting a true picture of cost-effectiveness of ozanimod closer to the reality.
What additional evidence or analyses might help to resolve this key issue?	Updating the economic model and re-running the PSA, addressing the concerns surrounding the NMA especially for the baseline risk estimates along with including tofacitinib as a relevant comparator would render the uncertainty analysis more suitable for decision making.

Abbreviations: CEAC, cost-effectiveness acceptability curve; ERG, Evidence Review Group; NMA, network meta-analysis; PSA, probabilistic sensitivity analysis

1.6. Summary of ERG's preferred assumptions and resulting ICER

The results below present the incremental and cumulative impact of the ERG's preferences. The ERG's preference would have been to include tofacitinib as a comparator within the economic analysis. However, due to the lack of model flexibility, it was not possible for the ERG to include tofacitinib in the economic model. As an exploratory analysis, the ERG has conducted a cost comparison versus tofacitinib (see Table 60 and Table 61 for results).

As part of the ERG preferred base case, the ERG used the following assumptions:

- Revised remission and response probability estimates for the treatments and BSC derived from the ERG run of the NMA using the alternative placebo baseline risks (as per Section 3.4.2.4)
- Revised post-active treatment transition probabilities for BSC which include an alternative means of estimating remission probabilities for BSC based on 'loss of remission' (directly from the sustained remission estimates) and different BSC response rates for the TNF-naïve and TNF-experienced populations, as opposed to an overall pooled estimate in the company's base case.

Table 3: Summary of ERG's preferred assumptions and ICER (TNFi-naïve subgroup)

Scenario	Total costs	Total QALYs	Incremental cost	Incremental QALYs	ICER: Ozanimod vs. comparators (£/QALY)
Company's base case					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£28,686
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£167,024*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£71,023*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£52,736*
ERG's preferred base case assumptions (applied incrementally over company's base case)					
Re-estimation of baseline placebo risks					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,479
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£169,098*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£82,608*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£56,298*
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,794
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£169,791*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£82,863*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£56,640*
Cumulative impact of ERG preferences (deterministic)					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,794
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£169,791*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£82,863*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£56,640*
Cumulative impact of ERG preferences (probabilistic)					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,842
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£1578721*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£87,452*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£68,470*

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; vs., versus

Note: * ICER in SW quadrant

Table 4: Summary of ERG's preferred assumptions and ICER (TNFi-experienced subgroup)

Scenario	Total costs	Total QALYs	Incremental cost	Incremental QALYs	ICER: Ozanimod vs. comparators (£/QALY)
Company's base case					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£33,725)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£199,551*
ERG's preferred base case (applied incrementally over company's base case)					
Re-estimation of baseline placebo risks					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£71,524)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£427,683*
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£70,807)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£436,080*
Cumulative impact of ERG preferences (deterministic)					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£70,807)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£436,080*
Cumulative impact of ERG preferences (probabilistic)					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£56,635)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£12,926)

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; vs., versus

Note: * ICER in SW quadrant

Including several labelling issues, the ERG noted a discrepancy between the CS Document B and the model in the distribution used for utilities in the PSA, as discussed in Section 5.2.2; however, it did not have any material impact on the results. Further, during clarification (clarification question B14) the ERG indicated that a fully incremental analysis with the associated CE frontier was missing, following which it was added to the model. Otherwise, no serious errors were found in the company's model that impacted the results.

1.7. Summary of exploratory and sensitivity analyses undertaken by the ERG

A summary of exploratory and sensitivity analyses undertaken by the ERG is provided in Table 5 (TNFi-naïve subgroup) and Table 6 (TNFi-experienced subgroup).

Table 5: ERG scenario analysis (TNFi-naïve subgroup)

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
Company base-case					
ozanimod	5.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£28,686	
infliximab		[redacted]	[redacted]	£167,024*	
golimumab		[redacted]	[redacted]	£71,023*	
vedolizumab		[redacted]	[redacted]	£52,736*	
Cost comparison with tofacitinib					
Incremental cost associated with ozanimod	6.1	[redacted]	-	-	-
Spontaneous remission (0.75% per model cycle)					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£29,830	4%
infliximab		[redacted]	[redacted]	£169,731*	2%
golimumab		[redacted]	[redacted]	£72,123*	2%
vedolizumab		[redacted]	[redacted]	£53,983*	2%
Ozanimod AE discontinuation rate in maintenance phase (5% that of induction)					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£29,790	4%
infliximab		[redacted]	[redacted]	£137,368*	-18%
golimumab		[redacted]	[redacted]	£65,285*	-8%

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
vedolizumab		[redacted]	[redacted]	£51,677*	-2%
Ozanimod AE rate in the maintenance phase (20% increase)					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£28,750	0%
infliximab		[redacted]	[redacted]	£166,869*	
golimumab		[redacted]	[redacted]	£70,961*	
vedolizumab		[redacted]	[redacted]	£52,720*	
% patients receiving SC vedolizumab (80% after year 1)					
ozanimod	6.1	-	-	-	-
adalimumab		Not applicable			
infliximab					
golimumab					
vedolizumab		[redacted]	[redacted]	£44,204*	-16%
Treatment regimen costs applied per treatment cycle					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£33,815	18%
infliximab		[redacted]	[redacted]	£188,210*	13%
golimumab		[redacted]	[redacted]	£71,528*	1%
vedolizumab		[redacted]	[redacted]	£53,501*	1%
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	6.3	-	-	-	-
adalimumab		[redacted]	[redacted]	£28,797	0%
infliximab		[redacted]	[redacted]	£167,294*	0%
golimumab		[redacted]	[redacted]	£71,133*	0%
vedolizumab		[redacted]	[redacted]	£52,859*	0%

Abbreviations: AE, adverse events; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SC, subcutaneous

Table 6: ERG scenario analysis (TNFi-experienced subgroup)

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
Company base-case					

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
ozanimod	5.1	-	-	-	-
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,725)	
vedolizumab		[redacted]	[redacted]	£199,551*	
Cost comparison with tofacitinib					
Incremental cost associated with ozanimod	6.1	[redacted]	Not applicable		
Spontaneous remission (0.75% per model cycle)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£34,594)	3%
vedolizumab		[redacted]	[redacted]	£198,146*	-1%
Ozanimod AE discontinuation rate in maintenance phase (5% that of induction)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£41,096)	22%
vedolizumab		[redacted]	[redacted]	£160,695*	-19%
Ozanimod AE rate in the maintenance phase (20% increase)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,689)	0%
vedolizumab		[redacted]	[redacted]	£199,367*	
% patients receiving SC vedolizumab (80% after year 1)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,725)	0%
vedolizumab		[redacted]	[redacted]	£161,152*	-19%
Treatment regimen costs applied per treatment cycle					

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£47,464)	41%
vedolizumab		[redacted]	[redacted]	£208,721*	5%
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	6.3	-	-	-	-
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,354)	-1%
vedolizumab		[redacted]	[redacted]	£200,192*	0%

Abbreviations: AE, adverse events; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SC, subcutaneous

2. INTRODUCTION AND BACKGROUND

2.1. Introduction

In this report, the Evidence Review Group (ERG) provides a review of the evidence submitted by Celgene, a Bristol Myers Squibb company, in support of ozanimod for treating moderately to severely active UC. The company provided an overview of the disease and burden of ulcerative colitis in the target population in Sections B.1.3.1 to B.1.3.3 in the CS.

UC is an inflammatory condition which affects the gastrointestinal tract.[2] The exact aetiology of the condition remains unknown, but the most popular hypothesis for its cause is centred around a complex interplay between genetic susceptibility, gastrointestinal microbiota, mucosal or generalised immune responses and environmental factors.[3] Ultimately, these factors cause chronic inflammation, which involves the degradation of the cells lining the lumen of the large intestine. As these cells are damaged, ulcers, which are the main cause of the symptoms associated with UC, form.

The symptoms of UC vary between people, depending largely on the extent and severity of their disease. The most common are characterised by symptoms related to an inflamed rectum and include bloody diarrhoea, abdominal pain, urgency and tenesmus.[4] However, the symptoms of UC are not limited to the GI tract as there can be extra-intestinal manifestations causing issues in the joints, eyes, bone, skin and liver, as well as anaemia and fatigue.[5-8] In addition, the condition is characterised by periods of remission interspersed by active disease relapses[4], with 50% of patients have at least one relapse per year.[9]

The degree of symptoms experienced by patients is largely dictated by the extent of their disease, although up to 25% will require hospitalisation at some point during the disease course.[10] The least severe category is ulcerative proctitis, where only the region closer to the rectum is affected. In cases of proctosigmoiditis the rectum and sigmoid colon are affected, whereas in left-sided colitis the rectum as well as sigmoid and descending colon are affected. The most extensive category is pancolitis, where the entire colon is inflamed.[11] This appraisal is focused on those with moderately to severely active UC. Mild-moderate UC is defined as less than six bowel movements per day with few systemic symptoms. Severe UC is more than six bowel movements per day with one or more of the following: a body temperature exceeding 37.8°C; pulse of more than 90 beats per minute; haemoglobin <10.5 g/dL or an erythrocyte sedimentation rate >30 mm/hour.[12]

Short-term symptoms are not the only concern associated with UC; there are also a number of potential longer-term complications. Some of these include bowel cancer, haemorrhage, perforation, strictures, abscesses, anorectal disease, primary sclerosing cholangitis, osteoporosis and toxic megacolon.[13,14] One of the main aims of treatment is to achieve remission in order to avoid the development of these longer-term conditions.

UC usually develops between the ages of 15 and 25 years, though there is a second peak between 55 and 65 years.[15] Historically, incidence has been highest in more economically developed countries, however more recently it has been increasing in developing countries while decreasing in western countries. In the UK specifically, UC affects 1 in 420 people, 52% of which have moderate to severe disease.[16] However, UK incidence rates are falling by 1.6% per year; this decrease is largely seen in the second peak while incidence in those under the age of 17 continues to increase.[2] In the UK, rates of UC are highest in the Northeast, East and Midlands.[2] UC is most common in black people and Caucasian people of European descent while it is less frequently seen in those from Asian communities. There is an equal split in prevalence between men and women,[2] although women with UC are at greater risk of relapse.

UC is initially diagnosed according to a patient's symptoms, in addition to a physical examination for anaemia, which can be confirmed with a blood test, and tenderness in the stomach. A stool sample can also be used to allow clinicians to rule out infections of the stomach or bowel, which can be mistaken for UC. Where UC is suspected, patients are referred for either an X-ray or CT scan to further rule out any serious complications. The diagnosis can then be confirmed with a sigmoidoscopy to determine the level and extent of the inflammation in the bowel, this may also involve a biopsy. If it is suspected that a greater portion of the large intestine is affected, a further colonoscopy may be carried out which can also involve a biopsy.[17]

Once a UC diagnosis is confirmed, the current treatment pathway is highly individualised. In the UK, patients will most often initially receive conventional therapies (CvTs) including corticosteroids, aminosalicylates and immunosuppressants. If CvTs are failing to manage a patient's condition, they will often progress to their first biologic treatment. In those in which they are suitable, tumour necrosis factor inhibitors (TNFi), also known as anti-TNF, treatment will be the first biologic. If patients become intolerant, or fail to respond to, their first biologic treatment, they will usually be treated with a second. The remaining treatments are biologics, vedolizumab and ustekinumab, and a small molecule drug, tofacitinib.[18] Which is used depends on many factors including comorbidities, rate of action and patient preference. If patients fail this

subsequent line of biologic treatment, they may require surgery to remove part of their large intestine, often leaving them in need of a stoma bag.

2.2. Background

2.2.1. Current treatment for ulcerative colitis

The company provides an overview of current treatment options for UC in Section B.1.3.4 of the CS.

The description of the current treatment pathway presented in the CS is broadly aligned with feedback from the ERG's clinical experts and a guidance algorithm of the NHS England (NSHE).[19] However, the CS lacks nuance in certain areas of the pathway and the use of tofacitinib within the NHS is underrepresented. Clinical advice to the ERG indicated that treatment of UC is highly individualised with factors such as comorbidities, contraindications and patient preference all relevant in establishing the optimal treatment for each patient. The CS largely describes the most common pathway and though it acknowledges the individualised nature of advanced treatment, it fails to account for even the most common patient-specific variations.

The CS suggested that CvTs are typically used as the first line treatment in UC patients. While this is the case in the majority of outpatients with UC, with most patients receiving CvT in the first line and moving on to their first biological treatment following relapse or contraindication, there are a small minority who receive a biological treatment in the first line. Clinical advice to the ERG indicated that, within this minority of patients, either vedolizumab or tofacitinib are most commonly used. The company did not consider that patients for whom CvT had failed would progress to treatment with another CvT non-concurrent with active treatment. As CvT was included as a comparator in TA633,[20] and TA547,[21] the ERG sought clinical advice which confirmed that, in UK clinical practice, CvT is not a relevant comparator in participants for whom CvTs have failed. The ERG therefore agreed with the company's exclusion of CvT as a comparator.

The company also indicated in the CS that the majority of patients will receive a TNFi as the first biological treatment. Clinical advice to the ERG concurred with this, stating that TNFis are unsuitable in 10-20 per cent of patients. The CS suggested that, in the minority of patients who do not receive a TNFi as the first biologic, vedolizumab will be prescribed. It does not, however, consider the use of tofacitinib despite its rapid mode of action and the convenience of

administration. Following relapse, the fast action of tofacitinib can reduce the need for steroid treatment in the interim and, being an oral treatment, it is often preferred by patients. Clinical advice to the ERG confirmed that the use of tofacitinib as a first treatment following CvT failure is increasing for the reasons described here.

Furthermore, the CS does not consider the use of a second TNFi following failure of a first TNFi to be routine practice. Clinical opinion to the ERG advised that standard practice is more nuanced than this simplified pathway. The CS accurately identified that TNFi therapeutic drug monitoring (TDM) is used to rationalise decision-making in this regard, with monitoring allowing clinicians to determine the reason for failure on a TNFi. Therefore, if discontinuation is due to immunogenicity, patients are eligible to receive a second TNFi. Clinical advice to the ERG indicated that TDM is much more commonplace in the UK than the US and, as a result, treatment with a second TNFi is more prevalent in the UK, though still fairly uncommon. Furthermore, the ERG noted the inclusion of adalimumab as a comparator in the subgroup of patients who had experienced failure of a biologic treatment in the appraisal of ustekinumab (TA633),[20] indicating that this TNFi is used in the second line. Clinical advice to the ERG further confirmed that, while uncommon, patients receiving a second TNFi would receive adalimumab following treatment failure with infliximab.

As stated previously, the CS does not consider tofacitinib a comparator in either the TNFi-naïve or -experienced patients. As a result of this exclusion, only vedolizumab and ustekinumab are considered relevant comparators following TNFi failure in the CS. The clinicians consulted by the ERG felt that the exclusion of tofacitinib from both subgroups did not reflect clinical practice. The CS states that tofacitinib is not routinely used in TNFi-experienced patients due to its safety profile, citing concerns raised in TA633.[20] The ERG noted, however, that tofacitinib was included as a comparator in TA633 to enable a full picture of cost-effectiveness for ustekinumab; though it further noted that the committee for this appraisal agreed with its subsequent exclusion as a comparator. Clinical experts consulted by the ERG did acknowledge the complex safety profile associated with tofacitinib, but also indicated that the treatment can be very beneficial for some patients. The CS refers to safety warnings regarding tofacitinib from the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA). Monitoring of safety warnings regarding tofacitinib by the Medicines and Healthcare products Regulatory Agency (MHRA) were also referenced by the company in its response to ERG clarification question B.9. Clinical advice to the ERG mentioned the more conservative approach to the safety of tofacitinib in the US and, notably, that use of tofacitinib is increasing in the UK,

driven largely by patients' preference for an oral treatment and its fast-acting nature, and estimated the use of tofacitinib to be approximately 5% in the first line and 25% in the second line in the Royal Devon and Exeter NHS Foundation Trust. Based on these factors, the ERG considered the treatment landscape for people with moderately to severely active UC to be changing, and the exclusion of tofacitinib as a comparator in either subgroup to be an area of particular uncertainty. Finally, real-world evidence from a recent multicentre UK cohort study[22] reported that adverse events requiring curtailment of the treatment were uncommon in the studied population, with no occurrence of thromboembolic events; the authors concluded that tofacitinib was well-tolerated. The ERG therefore requested during clarification that tofacitinib be added to the comparative cost-effectiveness evidence through its inclusion in the model; the company maintained its position regarding safety and opted not to include tofacitinib (company clarification response, question B9). Clinical advice to the ERG did not agree with the company's argument that tofacitinib would not be used in patients over the age of 65 years, those who are past or current smokers, or those who have cardiovascular or malignancy factors, instead indicating that tofacitinib may be offered to such patients following patient-involved decision-making.

Though golimumab is excluded from the decision problem table presented in the CS (Document B, Table 1, p.12), the ERG noted the inclusion of golimumab as a comparator for both TNFi-naïve and TNFi-experienced patients in the NMA and the company model. Clinical advice to the ERG indicates that, while golimumab is currently used in practice, its use is extremely limited. In most cases where it is currently used, it is predominantly in patients with comorbidities for which golimumab is appropriate. The ERG, therefore, agreed with the company that the use of golimumab is limited in UK practice, though it considered its inclusion in the NMA and model for the sake of completeness to be appropriate.

Finally, at the time of the appraisal, filgotinib (GID-TA10600) was under appraisal by NICE as a treatment for patients with moderately to severely active UC who have had an inadequate response, loss of response or were intolerant to a previous biologic agent or conventional therapy. It was not clear where in the treatment pathway filgotinib would be positioned if recommended.

2.2.2. The technology

The CS provided an overview of the mechanism and dosage of ozanimod (Zeposia®) in Section B.1.2; the company also presented the proposed positioning of the treatment in clinical practice in Section B.1.3.4.1 of the CS.

Ozanimod is a sphingosine 1-phosphate (S1P) receptor modulator that is hypothesised to sequester lymphocytes in lymph nodes by binding with high affinity to G protein-coupled S1P receptors 1 and 5 ($S1P_1$ and $S1P_5$).[23,24] Through this receptor binding, it is thought to prevent lymphocyte trafficking via the periphery to, inter alia, the intestine; thereby inhibiting inflammation of the area.[25] As a result of binding to $S1P_1$, also found in cardiac muscle and smooth arterial muscle tissue, ozanimod may have safety considerations related to the heart, in particular bradyarrythmias, as well as blood pressure increases.[25] Its affinity for $S1P_1$ also increases the risk of macular oedema, though this mechanism is more poorly understood. Further safety considerations include increased susceptibility to infections, related to the sequestration of lymphocytes; transient increases in liver enzymes; reduction in in forced expiratory volume; possible foetal harm and, in rare cases, posterior reversible encephalopathy syndrome (PRES).[25] The latter, however, was reported in a patient treated with ozanimod for multiple sclerosis.[26]

The company indicated in the CS that ozanimod is an orally administered medication taken at a dose of 1 mg daily, following an up-titration regimen of 0.25 mg on Days 1 to 4, 0.5 mg on Days 5 to 7, and a 1 mg maintenance dose thereafter. This is in line with the Summary of Product Characteristics (SmPC) for the treatment.

In the CS, the company proposed that ozanimod may be used to treat people with moderately or severely active UC, whether they had prior exposure to TNFis or not. As the line of treatment for the target population is highly individualised, the appropriate positioning of this treatment is dependent on the clinician's perspective on its efficacy and safety relative to comparators, as well as the patient's preference. In this regard, the company indicated that ozanimod satisfies an unmet need through its novel mechanism of action, tolerable safety profile and oral route of administration. Given the individualised nature of the treatment landscape for this condition, the ERG did not consider there to be a fixed position for ozanimod in the treatment pathway.

2.3. Critique of company's definition of decision problem

The company statement regarding the decision problem is presented in Section B.1.1 of the CS. The company position and the ERG response is provided in Table 7 below.

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	ERG comment
Population	Adults with moderately to severely active UC who are intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic therapy (a tumour necrosis factor-alpha inhibitor TNFi, ustekinumab or vedolizumab), a JAK inhibitor (tofacitinib), or CvT (oral corticosteroids and/or immunomodulators)	Adults with moderately to severely active UC who have had an inadequate response, lost response, or were intolerant to either CvT or a biologic agent This comprises two mutually exclusive sub-populations: •TNFi-naïve: patients who have not previously received a TNFi •TNFi-experienced: patients who have previously received a TNFi and experienced treatment failure due to intolerance, lack of treatment efficacy or loss of response	The population addressed in the submission is in line with the final scope. TNFis are typically used as the first biologic treatment in patients who are intolerant or have had an inadequate response, or loss of response to CvT.1 As a result, exposure to TNFis forms the basis for clinical decision-making, with treatment options differing in two distinct sub-populations: TNFi-naïve and TNFi-experienced. This is reflected in the NICE restriction on the use of ustekinumab and is in line with the current use of other biologic treatments in UK clinical practice.[1]	The ERG considered the overall population included in the company scope to be broadly appropriate. While the ERG agreed with the company's decision to stratify its analyses by subpopulations related to treatment experience, it considered the stratification to be inconsistent with the NICE scope in that TNFi experience does not provide an absolute distinction between the first and second line following CvT.
Intervention	Ozanimod	Ozanimod[a]	N/A	The intervention in the company's main trial, TRUENORTH,[27,28] matches the scope and licence for ozanimod. The company's phase 2 trial compared the licensed dose of 1 mg daily with a lower dose; the ERG appraisal of this trial is restricted to the licensed dose.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Comparator (s)	Current clinical management including: • TNFi (infliximab, adalimumab and golimumab) • Vedolizumab • Ustekinumab • Tofacitinib • Conventional therapies (aminosalicylates, oral corticosteroids and/or immunomodulators), without biological treatments	The submission population has been split into two distinct subpopulations: TNFi-naïve and TNFi-experienced. The relevant comparators differ in these two populations: •TNFi-naive: o Infliximab (and associated biosimilars) o Adalimumab (and associated biosimilars) o Vedolizumab •TNFi-experienced: o Vedolizumab o Ustekinumab	The SmPC for ozanimod states that patients must have failed CvT or a biologic. As biologics are only offered after failure on CvT in clinical practice, CvT is not viewed as a relevant comparator to ozanimod in either population. **TNFi-naïve:** • Following failure with CvT the majority of patients are initially treated with TNFis • As a result, whilst the NICE recommendation for vedolizumab and tofacitinib do not restrict their use in patients who have failed, cannot tolerate or are unsuitable for TNFis, neither tofacitinib nor vedolizumab are typically used first line in TNFi-naïve patients. This was supported by feedback received from clinical consultation conducted as part of this appraisal • TNFis are not suitable for all patients and vedolizumab may be used in a small proportion of TNFi-naïve patients who are contraindicated to TNFis or have specific safety concerns surrounding their use • TNFis and vedolizumab have therefore been considered as relevant comparators in the TNFi-naïve population	The ERG noted that the comparators included in the submission were not consistent with the NICE final scope. The ERG considered the exclusion of tofacitinib as a comparator from both the TNFi-naïve and –experienced subgroups, as an outstanding area of uncertainty, and misaligned with UK clinical practice. The ERG further noted the exclusion of TNFis in the TNFi-experienced subgroup, though clinical advice to the ERG indicated that within-class treatment switching does occur if TNFi failure is due to immunogenicity. The exclusion of adalimumab was of particular concern as it was included as a comparator in the biologic failure subgroup in TA633,[20] indicating that one TNFi may be prescribed following the failure of another. This was confirmed by clinical advice to the ERG. The company excluded CvT as a comparator. Based on clinical advice to the ERG, this was considered to be an appropriate exclusion for the target population. The ERG accepted that the exclusion of golimumab from the company's scope was an omission made in error, due to clarification

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
			TNFi-experienced: • In line with the NICE final scope both ustekinumab and vedolizumab were considered relevant comparators in the TNFi-experienced populations • Neither tofacitinib or TNFis were considered relevant comparators in the TNFi-experienced population • Tofacitinib was not viewed as a relevant comparator as, in line with the opinion of clinicians consulted in TA633,[20] feedback from clinical consultation received as part of this appraisal noted that whilst tofacitinib may be effective for some patients, concerns regarding its safety profile mean it is not typically used as a first line treatment option in TNFi-experienced patients. There has been no downgrading in the EMA warnings and restrictions associated with tofacitinib since the ustekinumab submission.2 The restricted use of tofacitinib combined with concerns of its safety profile negates it as a standard comparator to ozanimod in this population (Section B.1.3.4) • TNFis were not considered relevant comparators in the TNFi-experienced population as TNFi switching is no longer routine clinical practice. As a result, receiving a second TNFi is only clinically relevant in a small	and its inclusion in both the NMA and as a comparator in the economics.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
			proportion of TNFi-experienced patients. The exclusion of TNFis is in line with the accepted assumption in TA633.[20]	
Outcomes	Outcome measures include: • Mortality • Measures of disease activity • Rates of and duration of response, relapse and remission • Rates of hospitalisation • Rates of surgical intervention • Endoscopic healing • Mucosal healing (combined endoscopic and histological healing) • Corticosteroid-free remission • Adverse effects of treatment • Health-related quality of life	Outcome measures include: • Measures of disease activity; change in the 3-component Mayo score • Rates of and duration of response, relapse and remission • Endoscopic healing • Mucosal healing (combined endoscopic and histological healing) • Corticosteroid-free remission • Adverse effects of treatment • Health-related quality of life	Mortality, rates of hospitalisation and rates of surgical intervention were not primary or secondary endpoints in TRUENORTH.[27,28] Data were therefore only collected on these events when assessing adverse events.	The outcomes reported by the company for the trial TRUENORTH[27,28] are relevant to the NICE scope, and clinically meaningful for evaluating the efficacy of treatments for UC. The ERG noted the omission of mortality, rates of hospitalisation and rates of surgical intervention as primary or secondary outcomes from the company scope. Clinical advice to the ERG indicated that mortality and rates of hospitalisation are broadly invariant with respect treatment with biologics or small molecules. Based on further clinician input, rates of surgery are likely similarly unchanged, though uncertainty remains as to whether the use of these treatments may result in a reduction in surgery. The ERG considered that this outcome could have been included in the NMA and used subsequently in the economic modelling, given its importance in the treatment pathway and disease course. However, the ERG acknowledges that rates of surgery would likely be little changed between treatments and are unlikely to impact the base case ICER, given that the key drivers of

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
				cost-effectiveness relate to treatment effect which, to some extent, incorporate surgical outcomes. The economic model captured treatment discontinuation, disutility and costs associated with serious adverse events only. HRQoL data were included in the economic model. The ERG noted that HRQoL data were collected in the pivotal TRUENORTH[27,28] trial using the EQ-5D-5L instrument; however, these data were not used in the company's base case. Instead, the company used published literature (from Woehl et al.[29] and Arsenau et al.[30]) and assumption to derive health state utility values.
Economic analysis	• The cost-effectiveness of treatments is expressed in terms of incremental cost per quality-adjusted life year • The time horizon for estimating cost-effectiveness was set at a lifetime horizon to sufficiently reflect any differences in costs or outcomes between the technologies being compared	As per final scope and NICE reference case	In line with the NICE final scope	The company submitted a cost utility analysis which used ICERs and QALYs as appropriate. A lifetime horizon was used in the base case. The ERG considered this to be reasonable (see Section 4.2.5). Costs were considered from an NHS and Personal Social Services perspective, in line with NICE guidance. Overall, the ERG considered that the economic analysis provided by the company was aligned with NICE's preferred reference case with respect to time horizon, perspective and outcomes.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
	• Costs are considered from a NHS and Personal Social Services perspective • The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account			
Subgroups	If the evidence allows the following subgroups will be considered: • People who have been previously treated with one or more biologic • People who have not received prior biologic therapy	Clinical consultation conducted as part of this appraisal indicated that exposure to TNFis forms the basis for clinical decision-making, with treatment options differing in two distinct sub-populations: • TNFi-naïve • TNFi-experienced	• Economic analyses were conducted for ozanimod for sub-populations based on prior TNFi exposure owing to the relevant comparators differing between these sub-populations. These analyses informed the base case cost-effectiveness analysis for comparisons versus infliximab, adalimumab, golimumab and vedolizumab (in TNFi-naïve patents) and vedolizumab and ustekinumab (in TNFi-experienced patients) • Subgroup analyses were informed by the Network Meta-analysis (NMA). The efficacy of ozanimod in the NMA was based on the subgroups of TRUENORTH stratified by TNFi exposure.	In the economic analysis, results have been presented for two distinct subgroups i.e., TNFi-naïve and -experienced patients. The ERG noted that final scope issued by NICE stated that subgroups should be stratified according to those who have been treated previously with biologics and those who have not received biologic treatment. The ERG noted that the previous UC appraisal for ustekinumab (TA633) was for a treatment licensed for patients with moderately to severely active UC who had an inadequate response or lost response to previous biologic therapy.
Special considerations including	None	The company did not identify any equity or equality concerns in the scope	N/A	The ERG agreed that there are no equity or equality concerns to be considered in this appraisal.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
issues related to equity or equality				

Abbreviations: cPAS, confidential patient access scheme; CvT, conventional therapy; EMA, European Medicines Agency; EQ-5D-5L, European Quality of Life Five Dimension Five Level; ERG, Evidence Review Group; HRQoL, health-related quality of life; ICER, incremental cost-effectiveness ratio; JAK, Janus kinase; N/A, not applicable; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NMA, network meta-analysis; SmPC, Summary of Product Characteristics; TA, technology appraisal; TDM, therapeutic drug monitoring; TNFi(s), tumour necrosis factor inhibitor(s); UC, ulcerative colitis

Note: [a] Ozanimod presents in three distinct capsule strengths each with two reportable weights (ozanimod hydrochloride 0.25 mg,0.50 mg, and 1.0 mg, which are equivalent to ozanimod 0.23 mg, 0.46 mg, and 0.92 mg, respectively).

3. CLINICAL EFFECTIVENESS

3.1. Critique of the methods of review(s)

The Company undertook a systematic literature review (SLR) to identify randomised controlled trials (RCTs) providing evidence for ozanimod (summarised in Section 3.2) and comparators to ozanimod. These were used to inform their indirect treatment comparison (Sections 3.3 and 3.4) in people with moderately to severely active UC. An additional SLR was conducted to identify any non-randomised trials of ozanimod, but yielded no results. An overview of the methods used in the SLRs is provided in Table 8 below.

Table 8: Summary of ERG'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	ERG assessment of robustness of methods
Searches	Appendix D.2.1.	The searches of bibliographic databases and grey literature sources are considered broadly appropriate; however, the ERG noted that specific searches for adverse reactions were not conducted. The search methods for the additional SLR to identify non-randomised trials were provided in response to clarification question A1.
Inclusion criteria	Appendix D.2.2., Table 7 (pages 50-51)	The inclusion criteria for the clinical effectiveness review are considered broadly appropriate to the decision problem. The ERG noted that the subgroup 'biologic treatment-failure and biological treatment non-failure with and without prior corticosteroid use' is fully aligned with the population detailed in the NICE scope, but not the company scope defining subgroups by TFNi experience; as highlighted in Table 7. The ERG further noted the inclusion of certolizumab as a comparator, though this treatment is listed in neither the NICE scope nor the decision problem addressed by the CS. The ERG accepted the company's clarification that certolizumab was included in error. The ERG disagreed with the company's decision to exclude phase 4 trials from the NMA, though it notes that such trial data are not currently available and therefore did not investigate this further.
Screening	Appendix D.2.2. (page 49)	Screening was conducted to appropriate standards to minimise selection bias, with duplicate screening at title/abstract and full-text stages. Arbitration by a third reviewer is also described, though the ERG noted that it was

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
		not explicitly stated whether this was done at both screening stages.
Data extraction	Appendix D.2.3. (page 51)	Data extraction was conducted to appropriate standards to minimise selection bias, with single reviewer extractions validated by a second reviewer. Though data extraction was not done independently and in duplicate, the ERG noted that data validation by a second reviewer is permissible with the AMSTAR 2 critical appraisal tool,[31] and further concluded that arbitration conducted by a third reviewer, if necessary, would minimise potential error or bias.
Tool for quality assessment of included study or studies	Document B, Section B.2.4., Table 19 (page 60); Appendix D.2.3. (page 51), Appendix D.4.4 (page 129), Appendix D.6. (pages 169-170)	The risk of bias assessment of TRUENORTH[27,28] in Document B of the CS was reported according to the Centre for Reviews and Dissemination (2009)[32] tool. The tool was also used to assess the risk of bias of all RCTs included in the company's NMA. The ERG considered this method appropriate, though it noted that the updated Cochrane risk of bias 2 tool[33] is generally preferred. No risk of bias assessment was reported for the long-term trial extension to TRUENORTH. The ERG considered this acceptable, given the ongoing nature of this trial.
Evidence synthesis	Appendix D.4.1. (pages 122-123), Appendix D.4.2. (pages 124-126), Appendix 4.3. (pages 127-129)	No synthesis of trials investigating ozanimod was conducted, as there is only one trial per comparison available. The ERG considered this reasonable. The company conducted several NMAs to evaluate the comparative efficacy of ozanimod with other available treatments within the TNFi-naïve and –experienced subgroups; these were further stratified by induction and maintenance phases for each subgroup. The ERG considered that further outcomes, particularly adverse events or treatment discontinuations, could have been evaluated in the NMAs, however, the company did not report their feasibility assessment with regards to outcomes and therefore it is not possible to determine if these outcomes were considered but found not feasible for analysis. The methods used in the NMAs were appropriate, though the ERG highlighted concerns about heterogeneity in the networks and the paucity of evidence, which both contributed to uncertainty in the results.

Abbreviations: AMSTAR, A MeaSurement Tool to Assess systematic Reviews; CRD, Centre for Reviews and Dissemination; CS, Company submission; ERG, Evidence Review Group; NMA, network meta-analysis; TNFi, tumour necrosis factor inhibitor

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

The company presented evidence for ozanimod from one Phase 3 RCT (TRUENORTH) including two cohorts: one placebo-controlled cohort and a second enrichment cohort; with responders from both cohorts re-randomised following an induction period. Further evidence came from a key supporting phase 2 placebo-controlled dose-ranging RCT in participants with endoscopically-confirmed UC (TOUCHSTONE). An overview of the methods used in these studies is presented in Sections 3.2.1 to 3.2.5.

3.2.1. Study design

The company's primary evidence for ozanimod is derived from TRUENORTH,[27,28] a multicenter, phase 3 study with a 10-week induction phase followed by a 42-week maintenance phase. The trial enrolled a total of 1,012 participants: some had no prior experience to TNFi; others had been treated with TNFi before. Eligible participants were either randomised in a 1:2 ratio to placebo or 1 mg ozanimod (called 'cohort 1') or included in an open-label enrichment cohort which was also allocated 1 mg ozanimod ('cohort 2') during the induction phase. Following induction, responders to ozanimod from both cohorts were re-randomised to receive placebo or 1 mg ozanimod during the maintenance phase, while responders to placebo continued placebo in the maintenance phase. Induction non-responders in both arms, as well as those who had relapsed during the maintenance phase, had the option of entering the open-label extension (OLE) trial. The trial measured a broad range of clinical efficacy, quality of life and safety outcomes up to 52 weeks and the ERG considered the large trial to be well conducted, though some methodological concerns that could bias results are described in Section 3.2.4.3. Despite these concerns, the ERG agreed with the company's decision to use data from this trial as the primary clinical effectiveness evidence.

The key supporting trial, TOUCHSTONE,[34] is a multicenter, phase 2 dose-finding study with an 8-week induction and 24-week maintenance phase. The study enrolled 199 participants who were randomised to receive 0.5 mg ozanimod, 1 mg ozanimod or placebo. Participants with clinical improvement during the induction phase continued their blinded regimen during maintenance; induction non-responders and those who relapsed during the maintenance phase had the option of entering the OLE trial. The ERG agreed that data from this trial are suitable as supporting evidence, given that the trial was well conducted and reported on outcomes within the NICE scope of this appraisal.

Following the completion of either TOUCHSTONE or TRUENORTH, participants had the option of continuing in the single-arm 1 mg ozanimod OLE trial. The study had, at the time of writing, reached its primary completion date and is expected to report maximal follow-up to six years, with its primary outcomes related to the safety of ozanimod. The ERG is of the opinion that the long-term safety evidence from this trial would reduce the uncertainty around the safety of ozanimod for moderately to severely active UC. An overview of the trial designs in provided in Table 9.

Table 9: Overview of ozanimod trial designs

Study name and acronym	Study design	Phase	Intervention / Comparator	Study objectives	Population
TRUENORTH (NCT02435992)	Multicentre, placebo-controlled study. 1-week dose titration within a 10-week induction. Induction period had 2 cohorts, one randomised and double blind and one open label enrichment cohort. Responders to ozanimod re-randomised to 42-week maintenance period.	3	1 mg ozanimod hydrochloride daily / placebo	Safety and efficacy	N=1012. Adults aged 18 to 75 with moderately to severely active UC. (N=526 for maintenance period)
TOUCHSTONE/ TRUENORTH OLE – Ongoing (NCT02531126)	Multicenter, single group assignment, OLE.	3	1 mg ozanimod	Safety and efficacy	N=878. Adults aged 18 to 75 who had participated in either TRUENORTH or TOUCHSTONE.
TOUCHSTONE (NCT01647516)	Multicentre, randomised, double-blind, placebo-controlled study. 1-week dose titration within a 9-week induction. Responders re-randomised to	2	1 mg ozanimod / 0.5 mg ozanimod / placebo	Safety and efficacy	N=199. Adults ages 18 to 75 with moderately to severely active UC.

Study name and acronym	Study design	Phase	Intervention / Comparator	Study objectives	Population
	24-week maintenance period.				

Abbreviations: OLE, open-label extension; UC, ulcerative colitis

3.2.2. Trial populations

3.2.2.1. Eligibility criteria

Key inclusion and exclusion criteria used in the two included trials and their open-label extension are summarised in Table 10 below. Potential participants were identified through endoscopically confirmed UC of moderate to severe activity, defined by a Mayo score of 6 to 12. Participants in TRUENORTH[27,28] were additionally required to receive aminosalicylate or corticosteroids and could have had prior treatment with immunosuppressants, though the use of these needed to be stopped prior to randomisation. The ERG considered the age, definition of the condition and other inclusion criteria to be appropriate for the target population.

The TRUENORTH trial[27,28] excluded potential participants with a physician-judged likelihood of receiving colectomy or ileostomy within 12 weeks of baseline, recent evidence of serious UC symptoms, diagnosis of Crohn's disease (CD) or other types of colitis, cardiovascular (CV) conditions, or a history of certain eye conditions; as well as excluding participants who are pregnant or lactating. Potential participants in TOUCHSTONE[34] were excluded for current use of TNFis. As there were no explicit exclusion criteria related to TNFi experience, the ERG agreed that both TNFi-naïve and -experienced participants would be included in the trial populations and that data from these trials align with the proposed positioning of ozanimod as a first- or second-line treatment.

For the open-label extension, all participants who participated in either TOUCHSTONE[34] or TRUENORTH[27,28] were eligible for inclusion. For this long-term extension, participants were excluded if they were treated with breast cancer resistance protein inhibitors, had clinically relevant CV conditions, or had liver function impairment. The ERG considered these exclusions appropriate given the long-term safety concerns of S1P receptor modulators, but noted that this would limit the generalisability of any conclusions on the safety of ozanimod, though it acknowledged that participants with similar conditions would likely not be prescribed ozanimod in UK practice.

Table 10: Eligibility for the included trials

Study	Inclusion criteria	Exclusion criteria
TRUENORTH (NCT02435992)	Aged 18 to 75 years (at screening for Cohort 1 and 2) UC confirmed on endoscopy Moderately to severely active UC (Mayo score 6-12) Currently receiving treatment with aminosalisylate, prednisone, or budesonide Can be receiving azathioprine, mercaptopurine, or methotrexate, but treatment will be stopped prior to randomisation	Physician judgment that the patient is likely to require colectomy or ileostomy within 12 weeks of baseline. Current or recent (within 3 months) evidence of fulminant colitis, toxic megacolon, or bowel perforation. Diagnosis of CD, indeterminate colitis, or the presence of fistula consistent with CD or microscopic colitis, radiation colitis, or ischemic colitis Clinically relevant cardiovascular conditions or other relevant diseases that could impact the implementation or interpretation of the trial, or put the patient at risk History of uveitis or unknown macular edema Pregnancy, lactation, or a positive serum beta-human chorionic gonadotropin (β-hCG) measured during screening
TOUCHSTONE (NCT01647516)	18 Years to 75 Years UC confirmed on endoscopy Moderately to severely active UC (Mayo score 6-12)	Current use of anti-TNF agents
OLE (NCT02531126)	Aged 18 to 75 years Previously participated in a trial of ozanimod and meets the criteria for participation in the open-label extension as outlined in the prior trial (i.e. non-responders after induction or relapse/completion of maintenance phase)	Receiving treatment with breast cancer resistance protein inhibitors Clinically relevant cardiovascular conditions Liver function impairment

Abbreviations: β-hCG, beta-human chorionic gonadotropin; CD, Crohn's disease; OLE, open-label extension; TNF, tumour necrosis factor; UC, ulcerative colitis

3.2.2.2. Baseline characteristics

The baseline characteristics of the participants in the TOUCHSTONE[34] and TRUENORTH[27,28] trials are presented in Table 11, alongside comparative characteristics from a cross-sectional, retrospective UK cohort dataset presented by King et al. (2020).[35]

Table 11: Baseline characteristics of the intention-to-treat populations of the included trials at induction, and their comparability with a cross-sectional and retrospective UK cohort study[35]

Characteristic	TRUENORTH			TOUCHSTONE			King et al.[35]
	Ozanimod (cohort 1)	Placebo	Ozanimod (cohort 2)	Placebo	Ozanimod 0.5 mg	Ozanimod 1 mg	
Mean age (years) (SD)	41.4 (13.54)	41.9 (13.64)	42.1 (13.72)	41.9 (12.3)	38.8 (12.1)	41.8 (11.0)	51 (37-65)[a]
Female	42.9%	33.8%	41.7%	46%	51%	28%	50.1%
White race	[redacted]	[redacted]	[redacted]	94%	91%	93%	
Mean weight (kg) (SD)	[redacted]	[redacted]	[redacted]	72.6 (14.9)	72.3 (16.9)	77.4 (16.3)	
BMI (kg/m^2) (SD)	25.40 (5.492)	25.11 (4.477)	25.88 (5.796)	NR	NR	NR	<25 – 38.49% 25-30 – 28.21% >30 – 14.55% Unknown – 18.76%
Tobacco/nicotine use	[redacted]	[redacted]	[redacted]	Current – 5%	Current – 6%	Current – 6%	Current – 12.33% Former – 26.36% Never – 54.22% Unknown – 7.08%
Region	[redacted]	[redacted]	[redacted]	NR	NR	NR	

Characteristic	TRUENORTH			TOUCHSTONE			King et al.[35]
4-component Mayo score	8.9 (1.47)	8.9 (1.35)	9.1 (1.49)	8.6 (1.5)	8.3 (1.5)	8.5 (1.6)	
Median C-reactive protein (mg/L) (range)	4.0 ([redacted])	5.0 ([redacted])	5.0 ([redacted])	4.9 (0.20-141.4)	3.9 (0.10-131.2)	4.3 (0.10-82.5)	
Median faecal calprotectin (µg/g) (range)	1079.48 ([redacted])	1349.79 ([redacted])	1259.85 ([redacted])	1272 (30-8380)	1477 (66-11,108)	1238 (10-10,511)	
Median lactoferrin (µg/g) (range)	NR	NR	NR	29.0 (1.4-1049)	30.6 (1.4-483)	29.9 (1.4-586)	
Disease extent	Left side of colon – 62.5% Extensive – 37.5%	Left side of colon – 62.0% Extensive – 38.0%	Left side of colon – 64.6% Extensive – 35.4 %	Left side of colon – 63% Extensive – 37%	Left side of colon – 63% Extensive – 37%	Left side of colon – 61% Extensive – 39%	
Concomitant medication	Glucocorticoid – 27.7% Aminosalicylate – 87.2%	Glucocorticoid – 32.4% Aminosalicylate – 84.3%	Glucocorticoid – 33.8% Aminosalicylate – 85.8%	Glucocorticoid – 37% Aminosalicylate – 88%	Glucocorticoid – 34% Aminosalicylate – 82%	Glucocorticoid – 40% Aminosalicylate – 79%	
Previous medication	TNFi – 30.3%	TNFi – 30.1%	TNFi – 43.4%	Immunosuppressant – 26% TNFi – 15%	Immunosuppressant – 37% TNFi – 20%	Immunosuppressant – 33% TNFi – 19%	
Mean age at onset/diagnosis (years) (SD)	[redacted]	[redacted]	[redacted]	35.8 (13.0)	33.1 (11.3)	35.2 (12.1)	
Mean years since diagnosis (SD)	6.9 (6.61)	6.8 (7.04)	7.91 (7.365)	6.1 (5.5)	5.9 (5.4)	6.7 (6.8)	

Abbreviations: BMI, body mass index; NR, not reported; SD, standard deviation; TNFi, tumour necrosis factor inhibitor

Note: [a] King et al. only reported the median age with interquartile range, as shown

Comparability of trial arms

As shown in Table 11, the baseline characteristics of the participants included in the ITT populations of the TRUENORTH[34] and TOUCHSTONE[34] trials were balanced across trial arms. The ERG noted that randomisation had been stratified by prior corticosteroid use and prior TNFi exposure. The company did not provide baseline characteristics per trial arm for the corticosteroid use stratum; these were, however, reported separately by TNFi experience in the CS (Document B, Table 15, p.54). Demographic and anthropometric characteristics between the placebo and ozanimod trial arms were comparable, though the ERG noted slightly higher proportions of male participants in the placebo arms of both the TNFi-naïve and –experienced strata. The ERG further noted a higher proportion of participants from Europe, as well as lower proportions of participants from other regions and with extensive disease in the ozanimod Cohort 2, when compared with the ozanimod Cohort 1 and placebo arms.

Relevance of trial populations to the target population

The overall characteristics of participants in the trials appear broadly comparable with those of the UK cohort dataset, with the only exceptions being that the average age of participants in the UK dataset is approximately 10 years older than in the ozanimod trials, and that there is a smaller proportion of current smokers included in the trials. The comparative data available for the relevant population in the UK are limited, however, and no comparisons are possible for a number of baseline characteristics; in particular across the range of biomarkers reported in the ozanimod trials. The ERG acknowledges that such unknown imbalances in respect of the UK population may exist in the trial populations, but considered the comparability between demographic and anthropometric characteristics to be reassuring. In addition, consultation with clinical experts indicated that the populations in TRUENORTH[27,28] and TOUCHSTONE[34] broadly reflect the characteristics of people with moderate to severe UC in the UK.

No comparative characteristics for the TNFi-naïve and –experienced strata could be found in published literature. The generalisability of evidence from these subgroups to the corresponding UK populations is an area of uncertainty in this appraisal.

3.2.3. Intervention characteristics

The characteristics of interventions delivered during the TOUCHSTONE[34] and TRUENORTH[27,28] trials, as well as their open-label extension, are summarised in Table 12. Ozanimod is delivered through oral administration of 1 mg (0.93 mg ozanimod hydrochloride)

capsules once a day, following an up-titration regimen of 0.25 mg for the first four days, 0.5 mg for Days 5 to 7 and 1 mg thereafter. The company did not provide an explicit rationale for the up-titration of ozanimod in the CS, but the ERG noted that this approach is identical to treatment described in patients with multiple sclerosis, with the rationale being attenuation of first-dose heart rate and atrioventricular conduction effects.[36] Furthermore, up-titration seems to be associated with the use of S1P modulators in general. As a result, the ERG considered this step to be appropriate.

A dose of 1 mg daily was selected following the completion of the phase 2 TOUCHSTONE[34] dose-finding trial, which also evaluated a lower maintenance dose (0.5 mg daily) of ozanimod and found slightly higher occurrences of clinical remission and clinical response and lowered lymphocyte counts with the higher dose.[34] No reductions or increases in dose, with the exception of up-titration, were permitted during TRUENORTH or indicated in the license for ozanimod.

The ERG noted that the use of concomitant treatment in TRUENORTH[27,28] (Document B, Table 14, p.52) and TOUCHSTONE[34] (Appendix L.1.3., Table 73, p.297) was balanced between the ozanimod cohort 1, ozanimod cohort 2 and placebo arms for TRUENORTH as well as the between the ozanimod 0.5 mg, ozanimod 1 mg and placebo arms for TOUCHSTONE[34] at induction. The most commonly used concomitant medication was aminosalicylates, followed by glucocorticoids and immunomodulators (reported for TRUENORTH only). The ERG further noted that the placebo responders arm during the maintenance phase of TRUENORTH comprised a far larger proportion of participants on aminosalicylates and lower proportion of participants on immunomodulators compared to those who has received ozanimod during induction: this was considered to be a function of the type of CvTs placebo responders were receiving at the time of response.

Table 12: Intervention characteristics of the included trials

Trial		Treatment
TRUENORTH induction (Cohort 1)	Ozanimod	Up-titration of 0.25 mg for Days 1 to 4, 0.5 mg for Days 5 to 7 and 1 mg thereafter with daily double-blinded oral administration 10 weeks
	Placebo	Matched double-blind oral placebo administered daily 10 weeks
TRUENORTH induction (Cohort 2)	Ozanimod	Up-titration of 0.25 mg for Days 1 to 4, 0.5 mg for Days 5 to 7 and 1 mg thereafter with daily double-blinded oral administration 10 weeks
TRUENORTH maintenance (re-randomised)	Ozanimod	Dose of 1 mg, daily double-blinded oral administration 42 weeks, up to study duration of 52 weeks
	Placebo	Matched double-blind oral placebo administered daily 42 weeks, up to study duration of 52 weeks
TOUCHSTONE/ TRUENORTH OLE		Dose of 1 mg, daily open-label oral administration Up to 6 years, or upon discontinuation from the sponsor
TOUCHSTONE	Ozanimod 0.5 mg	Up-titration of 0.25 mg for Days 1 to 4, and 0.5 mg from Day 5 onwards with daily double-blinded oral administration 32 weeks
	Ozanimod 1 mg	Up-titration of 0.25 mg for Days 1 to 4, 0.5 mg for Days 5 to 7 and 1 mg thereafter with daily double-blinded oral administration 32 weeks
	Placebo	Matched double-blind oral placebo administered daily 32 weeks

Abbreviation: OLE, open-label extension

3.2.4. Clinical effectiveness results

An overview of the clinical outcomes specified by NICE i.e., whether they were reported in the trials, how they were defined and how they were measured is provided in Section 3.2.4.1, along with limitations of these means of ascertainment captured where necessary.

3.2.4.1. NICE-scoped outcomes

Mortality

Mortality was not assessed as an outcome in TRUENORTH[27,28] or TOUCHSTONE.[34] The company did indicate that mortality is captured to a certain extent in TRUENORTH through its

reporting of adverse events. Though the ERG did not consider adverse events to be sufficiently specific to make any conclusions relating to the effect of ozanimod on mortality in the population of interest, clinical advice to the ERG did confirm that mortality is broadly invariant with respect to treatment with biologics or small molecules. As a result, the ERG did not consider the omission of mortality from the company submission to be highly problematic, though it noted the uncertainty around the effect of ozanimod on this outcome.

Measures of disease activity

Both TOUCHSTONE[34] and TRUENORTH[27,28] report disease activity. TRUENORTH measures disease activity through the three-component Mayo score, consisting of three sub-scores; rectal bleeding, stool frequency and mucosal appearance through endoscopy. By comparison, TOUCHSTONE reports disease activity using the 4-component score however the CS refers to the 3-component score. The primary outcomes of both studies included patients achieving clinical remission according to their three-component Mayo score. In addition, secondary outcomes included change in Mayo score and clinical response according to this score at weeks 8 and 32 in TOUCHSTONE; and weeks 10 and 52 in TRUENORTH. The company indicated that the Mayo scoring system is the most widely used, the ERG noted mention of the Truelove and Witts' severity index system and the UC symptom score (UCSS). Clinical advice to the ERG confirmed that the use of the Mayo scoring system is broadly appropriate.

Rates of and duration of response, relapse and remission

Both the TOUCHSTONE[34] and TRUENORTH[27,28] trials reported the number of patients achieving remission or clinical response to treatment. TOUCHSTONE defined clinical remission as a four-component Mayo score <2, with none of the individual sub-scores >1. This was recorded as a primary endpoint at week 8 and a secondary endpoint at week 32. TRUENORTH defined remission according to both the overall and the sub-scores of the three- and four-component Mayo scores. Clinical remission per the three-component Mayo score was defined as a rectal bleeding sub-score (RBS) of 0, with both the stool frequency sub-score (SFS) and endoscopy sub-score ≤1; clinical remission per the four-component Mayo score was defined the same as for TOUCHSTONE. The trial reported rates of remission at both 10 and 52 weeks as primary outcomes. It also reported the number of patients who remained in remission from week 10 to week 52. The TRUENORTH study also included a further group of those in 'durable clinical remission' at the 52-week time point.

Clinical response was defined in TRUENORTH[27,28] using both the overall and constituent sub-scores of both the three- and four-component Mayo score. The four-component definition was a reduction from baseline in the overall score of ≥3 points and ≥30%, and a reduction from baseline in the RBS of ≥1 point or an absolute RBS of ≤1 point. Similarly, the three-component definition was a reduction from baseline in the overall score of ≥2 and ≥35%, a reduction from baseline in RBS of ≥1 point and an absolute RBS of ≤1 point. In TOUCHSTONE,[34] clinical response was defined as a decrease in Mayo score of ≥3 points and ≥30% and a decrease in RBS of ≥1 point or an absolute RBS of ≤1.

Rates of hospitalisation

The company reported that rates of hospitalisation were not assessed as an outcome in TRUENORTH[27,28] or TOUCHSTONE,[34], though the ERG noted that rates of hospitalisation were listed as an outcome of TRUENORTH in the CS (Document B, Section B.2.2, Table 8). Instead of reporting this outcome, the company indicated that hospitalisation is captured to a certain extent in TRUENORTH through its reporting of adverse events. Though the ERG did not consider adverse events to be sufficiently specific to make any conclusions relating to the effect of ozanimod on hospitalisation in the population of interest, clinical advice to the ERG did confirm that this outcome is broadly invariant with respect treatment with biologics or small molecules. As a result, the ERG did not consider the omission of hospitalisation rates from the company submission to be highly problematic, though it noted the uncertainty around the effect of ozanimod on this outcome.

Rates of surgical intervention

Surgical intervention rates were not assessed as an outcome in TRUENORTH[27,28] or TOUCHSTONE.[34] The company did indicate that surgeries are captured to a certain extent in TRUENORTH through its reporting of adverse events. Though the ERG did not consider adverse events to be sufficiently specific to make any conclusions relating to the effect of ozanimod on surgery rates in patients with moderately to severely active UC, clinical advice to the ERG did indicate that rates of surgery are likely unchanged by treatment with biologics and small molecules. However, some uncertainty remains as to whether the use of these treatments may result in a reduction in surgery. The ERG considered that this outcome could have been included in the NMA and used subsequently in the economic modelling, given its importance in the treatment pathway and disease course.

Endoscopic healing

TOUCHSTONE[34] and TRUENORTH[27,28] both report endoscopic findings in addition to the endoscopic sub-score within the four-component Mayo score. However, the ERG noted that only TRUENORTH pre-specified the percentage of patients with 'endoscopic improvement', also defined as a Mayo endoscopy sub-score of ≤1, at the end of induction and maintenance phases (10 and 52 weeks respectively) as an outcome. The ERG further noted an inconsistency in TOUCHSTONE, where an endoscopy sub-score ≤1 at the end of the induction and maintenance phases (weeks 8 and 32 respectively) was also pre-specified as an outcome, however, this was termed 'mucosal healing'.

Mucosal healing (combined endoscopic and histological healing)

The ERG noted that mucosal healing was defined in TRUENORTH as a combination of the endoscopic healing outcome as well as histological healing, defined as a Geboes score <2.0. The latter is achieved when there are no neutrophils in the epithelial crypts or lamina propria and none of the following: increased eosinophils; crypt destruction; or erosions, ulceration or granulation of the tissue. TRUENORTH[27,28] reported the percentage of patients with mucosal healing at weeks 10 and 52, which was defined as 'endoscopic improvement with histological remission'; this also included a Mayo endoscopy sub-score of ≤1 and Geboes score <2.0. The TOUCHSTONE[34] trial, however, pre-specified endoscopy sub-scores as an outcome, as described in the section above, and called this 'mucosal healing'. In addition, this trial pre-specified 'histological remission', defined as a Geboes score <2.0, suggesting that TOUCHSTONE did not consider histological remission to be a component of mucosal healing. The ERG noted the company's acknowledgement of the stricter definition of mucosal healing in TRUENORTH when compared to other trials in UC.

Corticosteroid-free remission

TRUENORTH[27,28] reported the percentage of patients in corticosteroid-free remission. This was defined as those who had not received corticosteroids more than 12 weeks at week 52 of the trial. The company indicated in the CS that relapse within 12 weeks of corticosteroid discontinuation demonstrates steroid-dependent remission.

Adverse effects of treatment

TOUCHSTONE[34] separately reported the number of patients with treatment emergent adverse events (TEAEs) in the induction period and the maintenance period. A TEAE was classed as any event beginning on or after the first dose or an ongoing event that became more severe after the first dose, or up to 90 days after the last dose. An adverse event (AE) was described as serious if it resulted in death; was life threatening; required hospitalisation or elongation of a hospital stay; caused persistent disability/incapacity; was a congenital anomaly; or constituted an important medical event. The severity of AEs was assessed by the investigator according to their impact of patients' normal activities.

TRUENORTH[27,28] also reported the incidence, severity and relationship between the following TEAEs, serious AEs, TEAEs leading to discontinuation of ozanimod and TEAEs of special interest. In addition, changes from baseline in clinical laboratory measures, vital signs, ECG and pulmonary function tests were measured.

Health-related quality of life

Change in health-related quality of life (HRQoL), from baseline to week 10, was assessed in the TRUENORTH[27,28] study using both the SF-36 and EQ-5D five-level (5L) version, using both a summary index score and the patient's self-rated health status using a graduated visual analogue scale (VAS). The company reported that EQ-5D-5L data were cross-walked to EQ-5D-3L index scores using the algorithm included in van Hout et al. (2012);[37] the weighted average across treatment arms was used to inform health states. The ERG noted that this is the approach preferred by NICE.

3.2.4.2. Trial outcomes

Table 13 and Table 14 list the outcomes measured in the two trials providing the primary and key supporting evidence (TRUENORTH[27,28] and TOUCHSTONE,[34] respectively). Outcomes corresponding to NICE-scoped outcomes are also indicated.

Table 13: Outcomes per treatment phase reported in TRUENORTH

Outcome	NICE-scoped
Proportion in clinical remission; three- and four-component Mayo score (induction)	✓
Proportion with clinical response; three- and four-component Mayo score (induction)	✓
Proportion with endoscopic improvement (induction)	✓ (endoscopic healing)
Proportion with mucosal healing (induction)	✓
Changes in three-, four- and partial Mayo scores (induction)	✓ (measures of disease activity)
Proportion in histologic remission (induction)	× (only as part of mucosal healing)
Proportion with clinical response, clinical remission or endoscopic improvement in patients with prior TNFi experience (induction)	✓
Change in SF-36 and EQ-5D (induction)	✓
Proportion in clinical remission; three- and four-component Mayo score (maintenance)	✓
Proportion with clinical response; three- and four-component Mayo score (maintenance)	✓
Proportion with endoscopic improvement (maintenance)	✓ (endoscopic healing)
Proportion with maintenance of remission (maintenance)	×
Proportion with corticosteroid-free remission (maintenance)	✓
Proportion with mucosal healing (maintenance)	✓
Proportion with durable clinical remission (maintenance)	×
Changes in three-, four- and partial Mayo scores (maintenance)	✓ (measures of disease activity)
Proportion in histologic remission (maintenance)	× (only as part of mucosal healing)
Proportion with clinical response, clinical remission or endoscopic improvement in patients with prior TNFi experience (maintenance)	✓
Change in SF-36 and EQ-5D (maintenance)	✓
Health resource utilisation (maintenance)	×
Work productivity (maintenance)	×

Abbreviations: EQ-5D, European Quality of Life Five Dimension; NICE, National Institute for Health and Care Excellence; SF-36, 36-item Short Form Health Survey

Table 14: Outcomes per treatment phase reported in TOUCHSTONE

Outcome	NICE-scoped
Proportion with clinical response (induction)	✓
Changes in Mayo scores (induction)	✓ (measures of disease activity)
Proportion with mucosal healing (induction)	✓ (different definition to TRUENORTH)
Proportion with TEAE (induction)	✓
Proportion with clinical response (maintenance)	✓
Proportion in clinical remission (maintenance)	✓
Changes in Mayo scores (maintenance)	✓ (measures of disease activity)
Proportion with mucosal healing (maintenance)	✓ (different definition to TRUENORTH)
Proportion with durable clinical remission (maintenance)	×
Proportion with TEAE (maintenance)	✓

Abbreviation: NICE, National Institute for Health and Care Excellence; TEAE, treatment-emergent adverse event

3.2.4.3. 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.2.3., p.51). The critical appraisal of published evidence from the key supporting trial, i.e. the TOUCHSTONE (Sandborn et al. 2016)[34] study, according to the University of York CRD[32] criteria was reported in Appendix D.6. (p.169). Published and unpublished evidence from the pivotal TRUENORTH study (Sandborn et al. 2021[28] and TRUENORTH CSR,[27] respectively) was also critically appraised using the University of York CRD criteria; the results of this appraisal were reported in Document B, Section B.2.4., Table 19 (p.60). No risk of bias assessment was reported for the long-term trial extension to TRUENORTH, but the ERG considered this acceptable, given the ongoing nature of this trial.

As noted in Table 8, the ERG considered the CRD criteria to be appropriate for the appraisal of these studies, though it noted that the Cochrane risk of bias tool was updated in 2019[33] and that Cochrane risk of bias 2 is generally the preferred tool for appraising risk of bias in RCTs.

TOUCHSTONE

The company appraised this trial as having no methodological concerns, aside from some uncertainty around blinding of providers, participants and outcome assessors. Assessments were made at the trial level and not the individual outcome level. The ERG agreed broadly with

the assessments of the company according to the domains of the tool, though some domains were less obvious than others. The ERG noted that there was no explicit description of allocation concealment and considered 'Unclear' to be a more appropriate judgment than 'Yes' for this domain. While the ERG agreed that participants in the two groups of interest, placebo and 1 mg ozanimod, were similar in terms of prognostic factors it did note some considerable differences; with more men included in the 1 mg ozanimod arm, a lower lactoferrin range on average in the 1 mg ozanimod group, and a lower proportion of participants in the placebo arm with previous medication use. These differences were not large or numerous enough to consider the randomisation and balancing of known and unknown prognostic factors to have failed. The ERG also noted that NCT01647516 does not list the change in Mayo score at week 32 as an outcome, even though it is listed and reported in the trial publication (Sandborn et al 2016);[34] it did take cognisance that this was an exploratory secondary outcome. Finally, it was noted that while an intention-to-treat analysis was conducted for the primary analysis this was not done using an imputation technique, but rather an assumption of non-response in participants with missing data. While this is a conservative assumption which would bias results to the null, the ERG considered 'Unclear' to be a more appropriate response for this domain.

TRUENORTH

The company appraised this trial as having no methodological concerns, with the assessment made at the trial level. No differential judgments were made by outcomes. The ERG agreed broadly with the assessments of the company according to the domains of the tool, though some domains were less obvious than others. While the ERG agreed that participants in the two groups of interest, placebo and 1 mg ozanimod in cohort 1, were similar in terms of prognostic factors it did note considerable differences; less men were included, median faecal calprotectin was lower on average, and the range of C-reactive protein (CRP) was lower on average in the ozanimod arm. These differences were not large or numerous enough to consider the randomisation and balancing of known and unknown prognostic factors to have failed. Furthermore, the ERG noted a discrepancy in the manner in which blinding was assessed between this trial and the TOUCHSTONE[34] trial. Given that blinding was reported similarly in the two trial publications (Sandborn et al 2016[34] and Sandborn et al 2021[28]) as well as the corresponding trial registries (NCT01647516 and NCT02435992, respectively), combined with Cohort 2 receiving ozanimod open-label, the ERG considered 'High' to be a more appropriate judgment for this domain. The receipt of open-label ozanimod in Cohort 2 is of particular concern for the induction phase, with participants self-reporting QoL and stool frequency

outcomes whilst unblinded; the ERG noted that such participants accounted for approximately 46% of participants treated with ozanimod in the ITT population. The ERG did not consider the company's judgment on imbalances in dropouts between groups and consequent analytical approach to be fully appropriate: it noted high and differential attrition between the ozanimod and placebo arms; indicating that the value of the outcome was unlikely to be independent of the missingness. While this may not have been unexpected, this may have influenced the validity of the non-response assumption for missing participants in the intention-to-treat analysis. Given that this assumption if conservative, and that sensitivity analyses using multiple imputation indicated that the results of primary analyses are robust, the ERG considers 'Unclear' to be a more reasonable judgment for the domain dealing with analytical approach.

TOUCHSTONE/TRUENORTH OLE

The company did not conduct a quality assessment for the OLE; it also did not comment on potential sources of bias present in this study. The ERG considered the study to be at high risk of bias. As the study did not have a control group, it is not possible to determine whether any observed changes are due to treatment with ozanimod, or natural disease progression over time. Furthermore, the open-label design may have resulted in ascertainment bias, with self-reported sub-scores of the Mayo score (e.g., stool frequency and rectal bleeding) particularly prone to this over-estimation of treatment effect.

3.2.5. Description and critique of the results of the studies

3.2.5.1. Clinical effectiveness results

The primary goal of treatment for UC is to induce remission. During periods of remission, patients' symptoms are minimal and the inflammation of the colon is reduced. In turn, this improves long-term prognoses by reducing the likelihood of developing complications such as colorectal cancer. In addition to the requirement of inducing remission, UC drugs must be able to maintain this state for as long as possible without relapse, ideally without the need for concomitant corticosteroids. This is particularly important since treatment options for UC are limited, with the result that each failed line of treatment takes a patient closer to a surgical last resort. The direct effect of these sub-scores on patients can also be identified through measures of quality of life, this is particularly pertinent given that UC is a chronic disease without a known cure.

Measures of disease activity

The company presented results from TRUENORTH[27,28] for the change from baseline in three-component Mayo score, and reported a greater reduction in the ozanimod group than in the placebo group at week 10 (LS mean (SE) change from baseline [redacted] for ozanimod and [redacted] for placebo, [redacted]. The company also reported a greater reduction in the three-component Mayo score in patients treated with ozanimod compared to those treated with placebo during maintenance at the 52-week time point (LS mean (SE) change from baseline [redacted] for ozanimod and [redacted] for placebo, [redacted].

In TOUCHSTONE[34], a significantly greater reduction in the three-component Mayo score was reported in the 1 mg ozanimod group when compared to placebo following induction up to week 10 (mean (SD) change from baseline -3.4 (2.79) for 1 mg ozanimod and -2.0 (2.52) for placebo, p=0.0042). A significantly greater reduction was also observed in the 1 mg ozanimod group when compared to placebo after maintenance at the 32-week time point (mean (SD) change from baseline -3.4 (2.93) for 1 mg ozanimod and -1.6 (2.72) for placebo, p=0.0004).

Clinical remission

Achievement of clinical remission with the three-component Mayo score (as defined in Section 3.2.4.1) was the primary endpoint during both the induction and maintenance phases of the TRUENORTH[27,28] study. During the induction phase, a significantly greater proportion achieved clinical remission at week 10 in the ozanimod arm versus placebo (18.4% vs. 6.0%, p<0.0001; OR (95% Wald confidence interval [CI]) 3.59 (1.94 to 6.63)). This was also reflected in the results from the maintenance phase at 52 weeks (37.0% vs. 18.5%, p<0.0001; OR (95% Wald CI) 2.76 (1.767 to 4.294)).

During the maintenance phase, the company provided further characterisation of patients' remission states with secondary endpoints measuring maintenance of remission and durable clinical remission. Maintenance of clinical remission, defined as the proportion of patients in clinical remission at the end of the maintenance period (52-week timepoint) in the subset of patients in clinical remission at the end of the induction period (10-week time point), was significantly higher in those receiving ozanimod when compared with placebo (51.9% vs. 29.3%, p=0.0025; OR (95% Wald CI) 2.88 (1.45 to 5.74)). Similarly, durable clinical remission, defined as those achieving remission at the end of both induction and maintenance periods, was

significantly higher in those receiving ozanimod than in the placebo arm (17.8% vs. 9.7%, p=0.003; OR (95% Wald CI) 2.65 (1,39 to 5.06)).

By comparison, the rates of clinical remission (defined as a Mayo score ≤2, with no subscore >1) at week eight were also greater in the ozanimod arm of the TOUCHSTONE[34] study (16.0% vs. 6.0%, p=0.048). Though the ERG noted a slight discrepancy in the reporting of the p-value presented in the CS appendices (Appendix L.3.2., p.302 and Table 80, p.303) in the maintenance phase at week 32, a significantly greater proportion (by either value) of those in the 1 mg ozanimod arm also achieved clinical remission than in the placebo arm (21% vs. 6%, p=0.01).

Clinical response

Clinical response in TRUENORTH[27,28] is presented by the company according to the three-component Mayo score. At week 10, a significantly greater proportion of patients receiving ozanimod achieved clinical response than those receiving placebo (47.8% vs. 25.9%, p<0.0001; OR (95% Wald CI) 2.67 (1.86 to 3.84)). This was also reflected in the maintenance phase at week 52 (60.0% vs. 41.0%, p<0.0001; OR (95% Wald CI) 2.27 (1.542 to 3.33)).

At the end of the induction phase of the TOUCHSTONE [34] study, there was also a significantly higher proportion of patients in the 1 mg ozanimod arm achieving clinical response than in the placebo arm (57% vs. 37%, p=0.02). At the end of the maintenance phase, there remained a greater proportion of the 1 mg ozanimod group with clinical response compared to placebo (51% vs. 20%, p<0.001) at week 32.

Hospitalisation

Though the company reported that rates of hospitalisation were not assessed as an outcome in TRUENORTH[27,28], the ERG noted that rates of hospitalisation were listed as an outcome of TRUENORTH and that some information on hospitalisations was provided in the CS (Document B, p.72). A very low overall rate of hospitalisations was reported with no accompanying test for significance between rates for ozanimod versus placebo ([redacted] vs. [redacted]).

Endoscopic improvement

While included in the overall four-component Mayo score used to quantify remission rates, endoscopic improvement was also reported as a separate secondary endpoint. It is unclear how endoscopic improvement is defined in the CS, however, endoscopic healing is defined in the

TRUENORTH[27,28] study as an endoscopic sub-score ≤1. The ERG considered that the terms 'improvement' and 'healing' may have been used interchangeably in the CS.

The company presented endoscopic improvement data as a secondary endpoint in both the induction and maintenance phases of TRUENORTH[27,28]. Endoscopic improvement was significantly greater in the ozanimod arm than in the placebo arm at week 10 of the induction phase (27.3% vs. 11.6%, p<0.001; OR (95% Wald CI) 2.88 (1.80 to 4.60)) and week 52 of the maintenance phase (45.7% vs. 26.4%, p<0.001; OR (95% Wald CI) 2.48 (1.65 to 3.72)).

Mucosal healing

Mucosal healing was also reported as secondary endpoint for both 10- and 52-week time points for the induction and maintenance phases. Mucosal healing can be defined as a lack of endoscopic or histological activity, or a combination of these. The CS, as in the TRUENORTH[27,28] study, defined mucosal healing as a Mayo endoscopy sub-score ≤1 and a Geboes index score <2.0. Again, this is presented for both the induction and maintenance phases. At week 10 of the induction phase, a significantly greater proportion of those in the ozanimod arm showed mucosal healing than those in the placebo arm (12.6% vs. 3.7%, p<0.001; OR (95% Wald CI) 3.77 (1.76 to 8.07)). This was similar at week 52 of the maintenance phase (29.6% vs. 14.1%, p<0.001; OR (95% Wald CI) 2.64 (1.64 to 4.26)).

Mucosal healing was defined differently in TOUCHSTONE[34], when compared to the stricter definition in TRUENORTH[27,28], as a Mayo endoscopy subscore of ≤1. This was also greater in the 1 mg ozanimod arm of the study compared to placebo at both the 8-week induction phase (34% vs. 12%, p=0.002) and 32-week maintenance phase (33% vs. 12%, p=0.005) time points. TOUCHSTONE[34] also reported histological remission separately, and defined this as a Geboes score of <2.0. Rates of histological remission were not significantly different between the 1 mg ozanimod and placebo arms at week eight (22% vs. 11%, p=0.07). Following maintenance at the 32-week time point, however, a statistically significantly greater proportion of the 1 mg ozanimod arm had achieved histological remission (31% vs. 8%, p<0.001).

Corticosteroid-free remission

During the maintenance phase, the company provided further characterisation of patients' remission states with a secondary endpoint measuring corticosteroid-free remission. A significantly greater proportion of patients treated with ozanimod achieved corticosteroid-free remission, defined as having been in remission without the need for corticosteroids for the prior

≥12 weeks during the maintenance phase, than those receiving placebo at week 52 (31.7% vs. 16.7%, p<0.001; OR (95% Wald CI) 2.56 (1.60 to 4.09)). The company considered this 12-week threshold to be clinically meaningful since relapse within 12 weeks is considered to be an indicator of steroid dependence in UC patients.

Adverse effects

The safety population reported in the CS includes all patients who received at least one dose of ozanimod. Treatment emergent adverse events (TEAEs) were defined as any AE with onset or worsening on or after the date of the first dose. TEAEs which occurred beyond the 90-day follow-up period were excluded. The company provide an overview of adverse events in Table 37 of document B of the CS.

Rates of treatment emergent adverse events (TEAEs) were similar between the ozanimod and placebo arms of cohort 1 (40.1% vs. 38% respectively). Rates of severe TEAEs, serious TEAEs, related serious TEAEs and those leading to interruption or discontinuation were also similar during the induction phase. However, in the maintenance period specifically, rates of TEAEs were higher in the ozanimod arm than the placebo arm (49.1% vs. 36.6% respectively). TEAEs suspected to be related to treatment were also higher in the ozanimod arm than placebo arm during the maintenance phase ([redacted] respectively)

In addition, a health resource utilisation questionnaire was used to collect data on hospitalisations, doctor visits and emergency room visits during both the induction and maintenance phases. During the induction phase, hospitalisation rates were [redacted] and [redacted] for ozanimod and placebo respectively. During the maintenance phase, hospitalisation rates were similarly low at [redacted] and [redacted] for ozanimod and placebo respectively. Conversely, in those re-randomised to ozanimod, rates of serious TEAEs and TEAEs leading to discontinuation were slightly lower than in those re-randomised to placebo.

One death occurred during the induction period in cohort 2. However, this was considered unrelated to ozanimod.

TOUCHSTONE also reported AE data for all three arms. The ERG noted the proportion of patients affected by AEs in the placebo, 0.5 mg ozanimod and 1 mg ozanimod arms were similar (40% vs. 40% vs. 39%, respectively). Serious AEs also occurred in similar proportion across the three arms (9% vs. 2% vs. 4% respectively. The most common AEs were UC flares, [redacted]

The ERG noted that the company only included serious infection AEs in the model; the company justified this approach by citing its high associated cost. This approach was accepted in TA633,[20] therefore the ERG considered it broadly appropriate. The TRUENORTH CSR[27] reported the incidence rates of serious infections during the induction phase as [redacted] and [redacted] for ozanimod and placebo, respectively. During the maintenance phase, those re-randomized to ozanimod had incidence rates of [redacted] compared to [redacted] in those re-randomised to placebo. The timeframe of reporting these results, specifically for the maintenance phase, was not clear to the ERG, i.e. it was not clear whether these results were annualised. As a result, the ERG was not able to validate the two-week cycle probability of serious infections as reported in the CS (Document B, Section B.3.3.9, Table 50). Furthermore, the ERG noted the data provided within the CSR are limited: Tables 14.3.2.1A and 14.3.2.1B were cited, but neither were made available to the ERG.

Health-related quality of life

Health related quality of life was presented for both the induction and maintenance phases through the EQ-5D-5L and the SF-36 in TRUENORTH[27,28]. However, the reporting of both the overall and component sub-scores was incomprehensive. The elements of the EQ-5D and SF-36 that were presented in the CS are described below.

For the induction phase, the physical component summary (PCS) score of the SF-36 was significantly improved in those treated with ozanimod compared to placebo [redacted], with a significantly greater proportion of patients treated with ozanimod compared to placebo achieved a minimally clinically important difference (MCID) for this score ([redacted] vs. [redacted], [redacted]). The mental composite summary score (MCS) score, however, showed no significant difference ([redacted] vs.[redacted] [redacted]) between ozanimod and placebo, respectively, and no difference in the proportion of patients who achieved MCID. The company reported that there were certain domains of the MCS score which showed significant improvement in the ozanimod group compared to placebo, including vitality ([redacted]), social functioning ([redacted]) and mental health ([redacted]). Scores on the SF-36 global health were also significantly improved with ozanimod compared to placebo ([redacted]), as were health utility scores ([redacted]). In terms of the EQ-5D summary index score, those in the ozanimod arms had a significantly greater mean change from baseline than those receiving the placebo ([redacted] vs.[redacted], [redacted]). Similarly, the mean change from baseline in the VAS, representing the self-reported health status, was significantly greater in the ozanimod arm than the placebo arm ([redacted] vs. [redacted], [redacted]).

For the maintenance phase, the company reported that SF-36 scores generally improved for patients randomised to both ozanimod and placebo. The PCS score, however, was most significantly improved in those receiving ozanimod compared to those receiving placebo ([redacted]). In addition, a significantly greater proportion of patients treated with ozanimod compared to placebo achieved MCID for this score ([redacted] vs. [redacted], [redacted]). There was, however, no difference in the proportion of patients achieving MCID in the SF-36 MCS scores at the 52-week time point. Also at week 52, the company reported that there was no significant difference between the EQ-5D summary index between placebo and ozanimod groups, though the ERG noted an inconsistency in reporting: [redacted] in text (Document B, p.70) versus [redacted] (Document B, Figure 19, p.71). VAS scores in those receiving ozanimod relative to those receiving placebo were significantly improved ([redacted] vs. [redacted], [redacted]).

Subgroup analyses

The CS presents two subgroups, defined as those that are TNFi-naïve and those that are TNFi-experienced.

During the induction phase, rates of clinical remission were higher in the ozanimod groups compared to placebo for both the TNFi-naïve ([redacted] vs. [redacted], respectively; [redacted] and TNFi-experienced ([redacted] vs. [redacted], respectively; [redacted] subgroups. These results were reflected in the maintenance phase at week 52 with higher rates of remission in the ozanimod groups compared to placebo for both the TNFi-naïve ([redacted] vs. [redacted], respectively; [redacted] and TNFi-experienced ([redacted] vs. [redacted], respectively; [redacted] subgroups. These results broadly reflect those from the overall population. The ERG noted and agreed with the company's position that lower rates of remission in the TNFi-experienced subgroup is explained by the fact that those with prior TNFi exposure are more challenging to treat.

This pattern also extended into the secondary outcomes, as shown in Table 15 below.

Table 15: Efficacy outcomes by TNFi-exposure subgroup and treatment arm

	TNFi-naïve		TNFi-experienced	
	Ozanimod	**Placebo**	**Ozanimod**	**Placebo**
Induction				
Clinical remission	[redacted]	[redacted]	[redacted]	[redacted]
Clinical Response	[redacted]	[redacted]	[redacted]	[redacted]
Endoscopic Improvement**	[redacted]	[redacted]	[redacted]	[redacted]
Mucosal Healing	[redacted]	[redacted]	[redacted]	[redacted]
Maintenance				
Clinical remission	[redacted]	[redacted]	[redacted]	[redacted]
Clinical Response	[redacted]	[redacted]	[redacted]	[redacted]
Endoscopic Improvement	[redacted]	[redacted]	[redacted]	[redacted]
Mucosal Healing	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviation: TNFi, tumour necrosis factor inhibitor

Note: * Significant difference at the 5% level between ozanimod and placebo; ** Different value for endoscopic improvement in TNFi-naïve placebo arm during induction in text ([redacted]; Document B, p.74) and Figure 22 ([redacted]; Document B, p.75)

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

3.3.1. Search strategy

Two search strategies were used, one to identify RCTs and another to identify non-randomised trials, evaluating the efficacy and safety of ozanimod and its comparators for moderately to severely active UC for the company submission. The methods of these searches are described in Section 3.1.

3.3.2. Feasibility assessment

The company conducted a feasibility assessment for the NMA included in its submission, as described in CS Section 2.8.3 (Document B, pp.86-95). The company motivated its subgrouping of populations based on TNFi experience for the NMA by citing differential efficacy of treatment between first and second line biologics, with the former resulting in higher response rates and fewer patients requiring dose escalations than the latter. The ERG noted that subgrouping

based on degree of biologic experience is in line with the approach required in the NICE scope, with differences in efficacy between patients treated with first and second line treatments highly biologically plausible.

The ERG did not, however, consider the company's justification based on lower efficacy of treatments in the second line to be necessarily appropriate: literature cited included clinician surveys[38] and small studies (including less than 100 patients each) in participants with CD,[39-41] all describing comparative efficacy of TNFis (all considered first line according to the company decision problem). The ERG did not find these studies to be generalisable to the target population and considered that their publication dates reflect a clinical treatment landscape that is different to the current context. In contrast, a recent NMA[42] conducted in patients with moderately to severely active UC, and including most of the trials included in the company's NMA, generally found higher rates of clinical remission and endoscopic improvement for second-line treatments (ustekinumab and tofacitinib) versus placebo, when compared to first-line treatments (vedolizumab and TNFis) versus placebo. This was especially true for patients with prior TNFi exposure, in whom second-line treatments are typically used.[42]

3.3.2.1. Trial design

Maintenance trial design

The company submission also detailed the management of differences in trial design for the management phase; differentiated by the use of a 'treat-through' approach, i.e. once-off randomisation to treatment or placebo at baseline, or a 're-randomised' approach, i.e. randomising responders to a treatment again following an induction period during which participants were randomised to treatment or placebo. In addition to the information described in the CS (Document B, pp.86-88 and pp.97-98), further details on this approach are also provided in the appendices (Appendix D.4.1.) and in the response to clarification question A18.

Briefly, the approach allowed a comparison of like with like in terms of remitters in the maintenance phase who had responded by the end of an induction phase; this was achieved by estimating the number of clinical remitters among induction responders by applying a responders-to-remitters ratio from a comparable treatment arm from a similar trial investigating the same treatment or treatment class. For example, the responders-to-remitters ratio for TNFi-naïve participants of ULTRA 2[43] (comparing adalimumab with placebo) was applied to the total

number of remitters in ACT1[44] (comparing infliximab with placebo in a TNFi-naïve population) to estimate the number of remitters in ACT1 who had been induction responders.

The ERG considered it sensible to account for this source of heterogeneity, and noted the precedent for an approach to re-calculating data from different designs to allow comparisons in TA547[21] and TA633.[20] In TA633, the appraisal committee preferred the converse approach (i.e. recalculating re-randomised trials to approximate treat-through trials). The ERG acknowledged, however, the considerable uncertainty in recalculating re-randomised trials to approximate treat-through trials as reconfiguring these numbers to mimic the results of re-randomised trials may have biased relative measures of effectiveness. In particular, this approach assumes that there are no systematic differences between the baseline characteristics of induction responders and non-responders. If this is not the case, potential imbalances in treatment effect modifiers may have biased the results to an unknown extent, though the ERG accepted that the results of sensitivity analyses excluding treat-through trial designs (Appendix D.4.5.3.) demonstrated very little difference in point estimates, therefore the ERG did not consider the inclusion of data from these trials to be inappropriate. The assessment of the differences between the NMA base case and sensitivity analyses excluding these trials is described in more detail in Section 3.5.2.

Time point of assessment

The ERG considered the company's decision to restrict time points of assessment for the induction and maintenance phases of treatment within each subgroup to be a sensible approach in dealing with heterogeneity introduced by varying time points between trials. It was noted, however, that no clinical basis was provided for the selected restrictions, i.e. 6 to 14 weeks for induction and 52 to 60 weeks for maintenance. The ERG considered that the choice of these time points may have been led by maximal available data rather than clinical guidance, particularly in the induction phase; where only the trial by Sands et al. (2001)[45] was excluded.

3.3.2.2. Eligibility criteria

The company described trial eligibility criteria as a potential source of heterogeneity, a position with which the ERG concurred. The company report comparable inclusion criteria across trials for age, time since diagnosis, Mayo score and endoscopic sub-score and prior experience with CvT. Though the ERG acknowledged that these are important potential sources of heterogeneity that have been addressed in the company submission, it did note that UC SUCCESS[46] included participants aged 21 years and older while the Suzuki[47] trial listed 15

years and older as an age inclusion criterion. In addition, a study by Macaluso et al. (2018),[48] investigating factors affecting clinical and endoscopic outcomes in placebo arms of trials for UC treatments, indicated that concomitant steroids use, no prior TNFi experience, endoscopic central reading and duration of disease at baseline all affected these outcomes differentially. While the ERG considered TNFi experience to have been addressed through the stratification of analyses by prior TNFi exposure, other factors listed were not addressed through subgrouping or explored in sensitivity analysis.

The ERG noted the exclusion of trials conducted in exclusively Asian participants in a sensitivity analysis. The rationale for this was not clear or well-described in the CS, with the ERG noting no reported difference in pharmacokinetics between Japanese and Caucasian patients according to the SmPC. In addition, while inflammatory bowel disease (IBD) is modelled to be an emerging epidemic in Asia[49] this phenomenon is hypothesised to be related to changes in lifestyle, particularly the westernisation of diet,[50] rather than physiological differences in response. The ERG further considered that Asian patients are treated in the NHS and that the NICE scope did not exclude this population, therefore it agreed with the company's decision to include these trials in the base case NMA.

3.3.2.3. Subgroup definitions

TNFi versus biologic experience

The company's decision to separate participants into two mutually exclusive subgroups based on whether they were TNFi naïve or experienced, was considered a departure from the NICE scope; indicating subgrouping based on experience with biologics. The ERG noted that, of the trials included in the NMAs that reported on TNFi-experienced participants or a mix of TNFi-naïve and -experienced participants, only half listed TNFi experience as an explicit criterion. Four trials included participants with experience of 'biologics' or 'investigational' treatments, with two trials explicitly including those experienced with tofacitinib and vedolizumab (Motoya[51] and UNIFI,[52] respectively). The company justified this approach by indicating the following:

- TNFis are used almost exclusively in the first line.
- With the exception of UNIFI, across all trials included in the NMA, including TRUENORTH, subgroups were stratified by TNFi experience rather than biologic experience. This terminology is therefore a more accurate classification of the subgroups in which efficacy results are available.

- This approach is in line with a previous NMA in UC (TA547).

With respect to the first of these points, clinical advice to the ERG presented a more complex situation in UK clinical practice, as described in Section 2.2.1.

The ERG considered the heterogeneity in the TNFi-experienced subgroup to be a source of uncertainty in respect of the results of the NMA. The company reported that an overwhelming majority of participants (98.8%) in the UNIFI trial in the 'biologic failure' subgroup had experienced failure with at least one TNFi and justified their approach based on these numbers. The ERG could not find similar proportions for the study by Motoya et al. (2019)[51] (including biologic experience and describing required tofacitinib clearance period) or A3921063[53] (stipulating clearance periods for 'investigational' treatments). Furthermore, the ERG noted that, though the company described subgroups as being defined by TNFi experience, TOUCHSTONE[34] also specified clearance periods for 'biologic' or 'investigational' experience. As a result, the ERG is of the opinion that subgrouping by prior TNFi experience may limit generalisability of the results to the NICE scope, but is in line with the method of stratification used by the majority of the trials included in the NMA.

TNFi experience versus failure

The ERG accepted the company's explanation of heterogeneity within the TNFi-experienced subgroup in respect of failure, intolerance or inadequate response – particularly as Table 26 of the CS (Document B, p.91) showed that this information was not available in at least half of the trials. While the uncertainty caused by this heterogeneity is noted, the ERG further considered that this was not something the company could address as exclusion of these trials would result in sparse NMA networks. Finally, the company's approach is in line with the NICE scope, which specified inadequate response, loss of response or intolerance to biologic therapy.

3.3.2.4. Baseline characteristics

The company provided baseline characteristics of participants entering the induction and maintenance phases of trials included in the NMA in appendices to the CS (Appendix D.4., Tables 13 and 14, pp.113-121; though the ERG noted an erroneous reference to Appendix D.4.1.). These characteristics were reported as 'broadly similar' by the company, though the ERG noted large variations in C-reactive protein (CRP) levels and the proportion of patients with extensive disease, as well as some variation in concomitant steroid use and years since diagnosis. While the company also acknowledged this variation in the CS, it indicated that this

heterogeneity in baseline characteristics had been accepted in TA633[20] and TA547.[21] The ERG noted this precedent, and agreed to an extent that there are no alternative approaches which would improve the certainty around the estimates generated by the NMA, given that it is not known whether these characteristics are effect modifiers for UC treatment.

The ERG did, however, consider that excluding trials with outlier values for potential effect modifiers in sensitivity analyses, or running meta-regressions based on the values of such characteristics, could provide additional certainty around effect estimates generated by the NMA, but found there to be a paucity of published literature on the identification and cut-off levels of effect modifiers for UC.

3.3.2.5. Outcomes

The company presented the availability of outcome data by TNFi subgroups for the induction and maintenance phases by trials included in the NMA in the CS (Tables 27 and 28, respectively; pp.92-93), though the data itself were not presented. In cases where studies only reported clinical remission, but not clinical response, the company reported leveraging an ordinal response-remission NMA to retain studies – this approach is described in further detail in Section B.2.8.4 of the CS.

The ERG also noted the variation reported by the company in respect of the measurement of outcomes. This was considered to introduce some heterogeneity through the use of a three- or four-component Mayo score; the endoscopy sub-score of this measurement introduced further heterogeneity through local or central readings, with the latter providing greater objectivity than the former. The ERG further noted that the company restricted outcome measurement to the Mayo score, with trials using the UC Symptom and modified Truelove and Witts scoring systems excluded from the NMAs. This was considered to be an appropriate step in managing heterogeneity in outcome measurement: while this could have been further managed by restricting to the three- or four-component Mayo score, or to endoscopy scores read centrally, the ERG appreciated that this may have resulted in very sparse networks and highly imprecise effect estimates. Furthermore, a sensitivity analysis including three-component Mayo score data instead of four-component data from TRUENORTH[27,28] showed very little change in the effect estimates of the NMAs (comparing Appendix D.4.5.4. and base case league tables provided by the company in its clarification response to question A13).

The ERG did, however, consider the use of outcome data from an unweighted average of the placebo arms to be contrary to NICE guidance on this topic, which suggests using the evidence

sources that are generalisable to the decision problem to inform the baseline model (Dias et al. 2013).[1] It considered the approach taken by the company to increase the uncertainty and considerably reduce the generalisability of the findings of the NMA, and therefore recommends that the use of placebo rates from a single, generalisable trial (or possibly multiple generalizable sources of evidence) would yield results that are more aligned with the NICE scope.

3.3.3. Study selection criteria

The selection criteria used by the company are described in the CS appendices, with specific selection criteria presented in Appendix D.2.2. (Table 7, p.50-51). The ERG considered these criteria to be broadly appropriate, and noted specifically the inclusion of tofacitinib as a comparator treatment, resulting from the company's decision to include all treatments specified in the NICE final scope in the NMA.

As discussed in Section 3.3.2, the ERG considered the stratification of the company's analyses by prior TNFi experience to be a departure from NICE scope which may limit the generalisability, both to populations which are naïve to and have experience of biologics, but is in line with the method of stratification used by the majority of the trials included in the NMA.

The company chose to exclude phase 4 trials from the submission, which the ERG did not consider appropriate as such evidence could have been used to inform other links in the networks. In particular, real-world evidence for ozanimod could provide additional insights into the long-term safety and efficacy of the treatment. To determine the potential impact of this exclusion, the ERG conducted a search for phase 4 trial evidence for ozanimod and its comparators in the submission. The results of this search yield are reported in Section 3.5.1.

Following the completion of its screening, the company imposed additional exclusion criteria as part of its feasibility assessment. While this typically considered to introduce potential bias, the ERG agreed with the company's exclusions based on unlicensed doses and ineligible comparisons, as well as trials with substantially different follow-up time points. Notably, tofacitinib was not excluded from the search yield at this stage and was included in the company's NMAs. Furthermore, the ERG noted the absence of UC-SUCCESS[46] from the TNFi-naïve induction phase evidence network, even though it had not been excluded for any reasons stated in the feasibility assessment. The ERG did not consider the exclusion of this ostensibly eligible source of evidence to be appropriate and regarded it as decreasing the confidence in the results of the NMA.

3.3.4. Included studies

The flow of studies identified for the NMAs was reported clearly in a PRISMA diagram (Appendix D.3., Figure 1). The company reported including 28 trials in the qualitative synthesis, of which 22 trials were included in the quantitative synthesis, i.e. NMA. This discrepancy resulted from three trials, namely OASIS (etrasimod),[54] HICKORY[55] and EUCALYPTUS[56] (both etrolizumab), being excluded due to the treatments of interest not having FDA or European Medicines Agency (EMA) approval at the time of the appraisal. The ERG noted that this approach was in line with TA547[21] and TA633[20] and considered the exclusions appropriate. It was not clear to the ERG at which stage the PURSUIT-IV trial,[57] investigating intravenous golimumab, was excluded. This trial was reportedly excluded on the basis of the treatment not having FDA or EMA approval, bringing the total number of trials excluded for this reason to four. Along with the trials reported in Table 24 of the CS (Document B, pp.84-86), this increases the number of trials that should have been included in the qualitative synthesis to 29.

A further three trials were excluded due to one comparing an approved and unapproved dose of adalimumab (SERENE-UC)[58] and the remaining two not using the Mayo clinic score to assess outcomes: the study by Probert et al. (2003)[59] used the UC Symptom scoring system; the study by Sands et al. (2001)[45] used the modified Truelove and Witts scoring system (Sands). The ERG agreed with the exclusion of the SERENE-UC[58] trial as no placebo arm was included in the study, and therefore no eligible comparison was available to inform links in the networks. Very little information was available in the CS, primary studies or in the published literature on what the modified Truelove and Witts and UC Symptom scoring system comprised. The ERG considered that the inclusion of these trials would have exacerbated outcome-related heterogeneity in the NMAs and found the exclusions broadly appropriate.

For the induction phase of treatments, a total of 18 trials reported at least one outcome in the TNFi-naïve and/or TNFi-experienced subgroup. For the TNFi-naïve subgroup, 15 trials reported on clinical response and 14 on clinical remission. For the TNFi-experienced subgroup, eight trials reported on clinical response and seven on clinical remission. For the maintenance phase of treatments, a total of 12 trials reported at least one outcome in at least one of the subgroups related to TNFi experience. Of the trials reporting on TNFi-naïve populations, 10 reported on clinical response and 12 on clinical remission. Of the trials reporting outcomes for TNFi-experienced populations, six reported on clinical response and eight on clinical remission.

The majority of trials (n=20) included in the NMA had a placebo arm, though many trials included multiple arms investigating different doses of the same treatment (n=12). A total of three trials included a head-to-head comparison between active treatments, namely UC-SUCCESS[46] (comparing azathioprine and infliximab), VARSITY[60] (comparison vedolizumab and adalimumab) and GEMINI 1[61] (comparing infliximab with various arms treated with vedolizumab). Of these head-to-head trials, only GEMINI 1 was placebo-controlled.

The number of included RCTs for each comparator treatment were as follows: adalimumab, n=4; azathioprine, n=1 (not a relevant comparator in this appraisal); golimumab, n=3; infliximab, n=6; ozanimod, n=2; tofacitinib, n=4; ustekinumab, n=1; vedolizumab, n=4; and placebo, n=20.

Trials included in the NMA were conducted between dates ranging from 2002 to 2021, according to the trial registries of these studies. The trials were conducted across a range of geographic locations and healthcare settings, with the majority conducted in multiple countries (n=17). Four trials were conducted in Japan only, and one trial was conducted in China. Time points of assessment following the induction phase ranged from six to 14 weeks; follow-up after induction and maintenance ranged from 30 to 60 weeks.

The eligibility criteria of the trials in the company NMA are reported in the appendices to the CS (Appendix D.3.1., Table 9, pp.56-62). These criteria showed very little between-trial variation in diagnostic criteria, with only UC-SUCCESS[46] not specifying an endoscopic sub-score of the Mayo score and TRUENORTH[27,28] specifying additional criteria for the rectal bleeding and stool frequency sub-scores. A number of trials did not specify age as an eligibility criterion, indicating the possibility of including paediatric patients. The ERG considered this to be unlikely, given that UC typically does not present before 15 years of age, and was of the opinion that the heterogeneity introduced and departure from the NICE-scoped population by including a few paediatric patients would be negligible. There was some heterogeneity in the trials reporting age inclusion criteria: most specified participants aged 18 or older, with some indicating an upper age limit; the studies by Motoya et al. (2019)[51] and Suzuki et al. (2014)[47] recruited patients from 15 years, while UC-SUCCESS[46] amended its minimum age criteria from 18 years to 21 years. The ERG did not consider this variation in age inclusion to meaningfully affect heterogeneity, given the high background heterogeneity between the included trials. Finally, previous experience with CvT and active treatment was stipulated for a number of trials – the ERG was of the opinion that this heterogeneity was address through the subgrouping of the overall

population by TNFi experience, though it did not consider the approach to be exactly aligned to the NICE scope, as discussed in Section 3.3.2.3.

3.3.5. Quality assessment of studies included in indirect treatment comparison

The company reported using the University of York CRD[32] criteria for assessing risk of bias for the trials included in the NMA. The ERG noted that the domains used in the assessment of the trials were appropriate for the CRD criteria. The judgments were summarised in Appendix D.6. (Table 15, pp.169-170). Overall, the company assessed most studies included in the NMA to have had appropriate randomisation and, to a somewhat lesser extent, adequate concealment of treatment allocation. The ERG noted that all trials at unclear risk of bias for appropriate randomisation were also at unclear risk for allocation concealment, representing a serious potential risk of baseline imbalance for confounders and effect modifiers. A number of studies, some of those with potential risk of selection bias, were considered to have an unclear risk of bias related to baseline imbalance for prognostic factors. These factors were identified by the ERG as potential factors that could increase the uncertainty in the NMA estimates. The blinding of care providers, participants and outcome assessors was mostly assessed as unclear across trials; the ERG considered that this may have systematically biased results in favour of active treatments, particularly given that outcomes of interest were at least partially self-reported, i.e. stool frequency and rectal bleeding as part of the Mayo score. Studies included in the NMAs were generally considered not to have quality issues related to unexpected imbalance in attrition between trial arms, or for selective outcome reporting; most trials were also judged as having conducted intention-to-treat analyses, and for doing so appropriately.

The ERG noted, however, that the company did not provide justifications for their quality assessments, which made it difficult to determine whether these were reasonable. The company did report that these assessments were made by a single reviewer, with validation by a second reviewer and, where necessary, resolution by a third reviewer. This was considered to be appropriate. Within the timeframe of the appraisal it was not possible for the ERG to conduct independent assessments of the quality of trials included in the NMA. The ERG did, however, compare these judgments with a comprehensive assessment of quality appraisal done in TA633.[20] This assessment suggested that all trials were considered to be a low risk of bias ('Yes' according to CRD criteria) for the randomisation and allocation concealment domains – this is contrary to the assessment in this appraisal, with the SERENE-UC,[62] Probert et al. (2003)[59] and Sands[45] trials considered 'Unclear' for these domains. These assessments in

terms of randomisation were considered by the ERG to be appropriate, as very little information on SERENE-UC was available in the public domain at the time of submission; the studies by Probert[59] and Sands[45] did not provide clear methodology around randomisation. The ERG noted that none of these trials were included in the NMAs following the feasibility assessment. In terms of allocation concealment, the ERG noted that PURSUIT J[63] and UNIFI[52] were additionally discrepant from the assessment in TA633[20] as they were also considered at 'Unclear' risk. The ERG also agreed with these assessments in the current appraisal, as all five trials reported insufficient information to enable an assessment.

TA633[20] acknowledged heterogeneity between trials for balance in prognostic factors, but broadly considered an assessment of low risk ('Yes') to be appropriate, though it identified the Study A3921063 and ACT 1 trial to have the greatest within-trial variability. This was broadly reflected in the assessments of the current submission, with Study A3921063[53] as well as ACT1[44] judged as having unclear risk. The ERG noted that ACT 2,[44] SERENE-UC,[62] Sands, Motoya,[51] OCTAVE 2[64] and OCTAVE SUSTAIN[64] and were additionally judged as having unclear risk, but could not verify this as the prognostic factors used in the assessment were not clearly identified in the CS.

The largest discrepancy between the assessment done in TA633[20] and the current submission is with regards to blinding, with the former indicating most trials were at low risk of bias while the latter assessed most trials as posing an unclear risk of bias. In addition, TA633[20] indicated that 'No' for all trials except ULTRA 1 is an appropriate assessment for imbalance in dropouts – the ERG noted that several trials were judged as 'Unclear' or 'High' in this submission, with ULTRA 1[65] not among these. The assessment of intention-to-treat analyses also yielded some discrepancies between the assessment in TA633 and the current appraisal – the OCTAVE trials are assessed in the CS as having conducted appropriate analyses, in line with opinion in TA633; on the other hand, the CS judged Probert[59] and PURSUIT-M[66] as 'Unclear' and 'Yes' for this domain while TA633[20] considered that intention-to-treat analyses were not reported for these trials. The ERG considered that this may be due to systematic differences in assessing these domains, with assessment of attrition and appropriate analysis further exacerbated by uncertainty around the time of assessment (induction versus maintenance), and noted this as an area of potential uncertainty.

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

The following sections contain an appraisal of the company's NMA methods and results, as conducted by the ERG. The ERG considered that the model applied by the company followed recommended practice and that logical steps had been taken to address some potential effect modifiers (primarily prior TNFi treatment and differences in trial design). Overall, the ERG considered that the company could have selected a more generalisable approach to assessing baseline risk in the placebo arm (in this context, the probability of being in non-response and non-remission, under placebo) of the NMAs, thereby providing effect estimates that are more applicable to the UK context. As a result, the ERG conducted scenario analyses using more UK-appropriate alternate values for the placebo arm of the NMAs. The ERG was also of the opinion that the placebo arms of trials included in the NMAs showed considerable variability in baseline characteristics, and may differ in respect of potential effect modifiers. Furthermore, given the heterogeneity in the evidence base used to conduct the NMAs, the ERG considered RE modelling to be a more appropriate choice; an approach that was not applied to all company analyses. Though the ERG noted the company's assertion that RE modelling was not done throughout due to failure of the model to converge, it considered that RE modelling could have been conducted using an informative prior distribution from a relevant context, e.g. those reported in Turner et al. (2012)[67] or Turner et al. (2015).[68]

3.4.1. Summary of analyses undertaken

The company carried out four NMAs representing the combinations of TNFi-naïve and -experienced during induction and maintenance periods. The NMA models were based on a multinomial model with probit link, described further in NICE TSD2,[69] and the analysis was carried out in a Bayesian framework using JAGS. The company explained that the underlying JAGS code was 'in line with' the WinBUGS code presented in Example 6 of the appendix to NICE TSD 2.[69]

The company favoured the default use of RE models, but selected FE models in three of its four settings (induction in the TNFi-experienced subgroup, and maintenance phases for both TNFi experience subgroups), following an assessment of convergence with the Gelman-Rubin convergence statistic. With RE models they found non-convergence, or overly large variances in estimates, in three settings, so FE models were applied instead.

Doses were pooled in analyses, that is, grouped as the same treatment where individual doses of the same active agent had the same method of administration (summarised in Document B, Table 31). A sensitivity analysis was carried out without pooling of doses; the results of which were broadly comparable to the base case (see Section 3.4.4.5).

The trials included were both ‘treat-through’ and ‘re-randomised’ in design. The company used a procedure to make the treat-through trials emulate re-randomised trials (see 3.4.2.3). A sensitivity analysis was carried out in which treat-through trials were excluded (see Section 3.4.4.5) and, given the potential uncertainty introduced by this approach, as highlighted in Section 3.3.2.1, the ERG undertook a head-to-head comparison of the results of the base case and sensitivity analysis. The results of this comparison demonstrated very little difference between the two approaches, as described in Section 3.5.2; the ERG therefore considered the approach to be appropriate.

The company did not include basic results in the CS, i.e. numbers or proportions partially responding or remitting by trial arm. These were supplied in Appendix 2 of the company’s clarification response, and are reproduced with reformatting in Table 16 to Table 19 below. Estimates of the proportions responding/remitting under the company’s NMAs are given in Appendix 3.3 of the company’s clarification response. Complete sets of pairwise estimates of odds ratio from the NMAs are provided in Figures 3 to 10 of the company’s clarification response.

Table 16: Company's base case NMA inputs for TNFi-naïve subgroups during induction; provided during clarification

Trial name	Induction period (weeks)	Treatments	Clinical response			Clinical remission		
			n	N	%	n	N	%
ACT 1	8	Infliximab Pooled	159	243	65.4%	86	243	35.4%
		Placebo	45	121	37.2%	18	121	14.9%
ACT 2	8	Infliximab Pooled	161	241	66.8%	74	241	30.7%
		Placebo	36	123	29.3%	7	123	5.7%
GEMINI 1	6	Vedolizumab 300 mg IV	69	130	53.1%	30	130	23.1%
		Placebo	20	76	26.3%	5	76	6.6%
Jiang 2015	8	Infliximab Pooled	32	41	78.0%	22	41	53.7%
		Placebo	15	41	36.6%	9	41	22.0%
Kobayashi 2016	8	Infliximab Pooled	57	104	54.8%	21	104	20.2%
		Placebo	37	104	35.6%	11	104	10.6%
Motoya 2019	10	Vedolizumab 300 mg IV	42	79	53.2%	22	79	27.8%
		Placebo	15	41	36.6%	6	41	14.6%
OCTAVE 1 + 2	8	Placebo	43	110	39.1%	13	110	11.8%
		Tofacitinib 10 mg BID	284	440	64.5%	106	440	24.1%
PURSUIT-SC	6	Placebo	89	292	30.5%	20	292	6.8%
		Golimumab 200/100 mg SC	147	294	50.0%	52	294	17.7%
Study A3921063	8	Placebo	15	33	45.5%	NA	NA	NA
		Tofacitinib 10 mg BID	14	23	60.9%	NA	NA	NA
Suzuki 2014	8	Adalimumab 160/80/40 mg Q2W	45	90	50.0%	9	90	10.0%
		Placebo	34	96	35.4%	11	96	11.5%
TRUE NORTH	10	Ozanimod 1 mg QD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Trial name	Induction period (weeks)	Treatments	Clinical response			Clinical remission		
			n	N	%	n	N	%
		Placebo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ULTRA 1	8	Adalimumab 160/80/40 mg Q2W	71	130	54.6%	24	130	18.5%
		Placebo	58	130	44.6%	12	130	9.2%
ULTRA 2	8	Adalimumab 160/80/40 mg Q2W	89	150	59.3%	32	150	21.3%
		Placebo	56	145	38.6%	16	145	11.0%
UNIFI	8	Placebo	56	158	35.4%	15	158	9.5%
		Ustekinumab Pooled	194	312	62.2%	60	312	19.2%
VARSITY	14	Adalimumab 160/80/40 mg Q2W	151	305	49.5%	72	305	23.6%
		Vedolizumab 300 mg IV	213	304	70.1%	84	304	27.6%

Abbreviations: BID, twice a day; IV, intravenous; NMA, network meta-analysis; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Table 17: Company's base case NMA inputs for TNFi-naïve subgroups during maintenance; provided during clarification

Trial name	Maintenance period (weeks)	Treatments	Clinical response			Clinical remission		
			n	N	%	n	N	%
ACT 1	54	Infliximab pooled	92	159	57.9%	53	159	33.3%
		Placebo	17	45	37.8%	10	45	22.2%
GEMINI 1	52	Placebo	21	79	26.6%	15	79	19.0%
		Vedolizumab pooled	88	145	60.7%	68	145	46.9%
Motoya 2019	60	Placebo	10	28	35.7%	10	28	35.7%
		Vedolizumab pooled	16	24	66.7%	13	24	54.2%
	52	Placebo	27	109	24.8%	12	109	11.0%

Trial name	Maintenance period (weeks)	Treatments	Clinical response			Clinical remission		
			n	N	%	n	N	%
OCTAVE SUSTAIN		Tofacitinib pooled	132	219	60.3%	94	219	42.9%
PURSUIT-J	54	Golimumab pooled	18	32	56.3%	16	32	50.0%
		Placebo	6	31	19.4%	2	31	6.5%
PURSUIT-M	54	Golimumab pooled	146	302	48.3%	101	302	33.4%
		Placebo	48	154	31.2%	34	154	22.1%
Suzuki 2014	52	Adalimumab 40 mg Q2W	50	82	61.0%	38	82	46.3%
		Placebo	12	34	35.3%	8	34	23.5%
TRUE NORTH	52	Ozanimod 1 mg QD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Placebo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ULTRA 2	52	Adalimumab 40 mg Q2W	44	89	49.4%	34	89	38.2%
		Placebo	24	56	42.9%	15	56	26.8%
UNIFI	52	Placebo	44	87	50.6%	27	87	31.0%
		Ustekinumab pooled	144	187	77.0%	91	187	48.7%
VISIBLE 1	52	Placebo	NA	NA	NA	7	37	18.9%
		Vedolizumab 108 mg Q2W SC	NA	NA	NA	36	67	53.7%
		Vedolizumab pooled	NA	NA	NA	17	32	53.1%

Abbreviations: NMA, network meta-analysis; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Table 18: Company's base case NMA inputs for TNFi-experienced subgroups during induction; provided during clarification

Trial name	Induction period (weeks)	Treatments	Clinical response			Clinical remission		
			n	N	%	n	N	%
GEMINI 1	6	Vedolizumab 300 mg IV	32	82	39.0%	8	82	9.8%
		Placebo	13	63	20.6%	2	63	3.2%
Motoya 2019	10	Vedolizumab 300 mg IV	23	85	27.1%	8	85	9.4%
		Placebo	12	41	29.3%	4	41	9.8%
OCTAVE 1 + 2	8	Placebo	29	124	23.4%	1	124	0.8%
		Tofacitinib 10 mg BID	237	465	51.0%	53	465	11.4%
Study A3921063	8	Placebo	5	15	33.3%	NA	NA	NA
		Tofacitinib 10 mg BID	6	10	60.0%	NA	NA	NA
TRUE NORTH	10	Ozanimod 1 mg QD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Placebo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ULTRA 2	8	Adalimumab 160/80/40 mg Q2W	36	98	36.7%	9	98	9.2%
		Placebo	29	101	28.7%	7	101	6.9%
UNIFI	8	Placebo	44	161	27.3%	2	161	1.2%
		Ustekinumab Pooled	169	330	51.2%	40	330	12.1%
VARSITY	14	Adalimumab 160/80/40 mg Q2W	26	81	32.1%	10	81	12.3%
		Vedolizumab 300 mg IV	44	79	55.7%	18	79	22.8%

Abbreviations: BID, twice a day; IV, intravenous; NMA, network meta-analysis; Q2W, every two weeks; QD, once a day; TNFi, tumour necrosis factor inhibitor

Table 19: Company's base case NMA inputs for TNFi-experienced subgroups during maintenance; provided during clarification

Trial name	Maintenance period (weeks)	Treatments	Clinical response			Clinical remission		
			n	N	%	n	N	%
GEMINI 1	52	Placebo	6	38	15.8%	2	38	5.3%
		Vedolizumab pooled	37	83	44.6%	30	83	36.1%
Motoya 2019	60	Placebo	5	14	35.7%	3	14	21.4%
		Vedolizumab pooled	11	17	64.7%	10	17	58.8%
OCTAVE SUSTAIN	52	Placebo	13	89	14.6%	10	89	11.2%
		Tofacitinib pooled	92	176	52.3%	54	176	30.7%
TRUENORTH	52	Ozanimod 1 mg QD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Placebo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ULTRA 2	52	Adalimumab 40 mg	15	36	41.7%	10	36	27.8%
		Placebo	6	29	20.7%	2	29	6.9%
UNIFI	52	Placebo	34	88	38.6%	15	88	17.0%
		Ustekinumab pooled	98	161	60.9%	52	161	32.3%
VISIBLE 1	52	Placebo	NA	NA	NA	1	19	5.3%
		Vedolizumab 108 mg Q2W SC	NA	NA	NA	13	39	33.3%
		Vedolizumab pooled	NA	NA	NA	6	22	27.3%

Abbreviations: NMA, network meta-analysis; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

The same trials supplied data for both the baseline risk estimates and the relative risk model. The baseline risk model made use of all placebo arms from these trials.

3.4.2. Critique of assumptions used in the indirect treatment comparison

3.4.2.1. General methodology

The company made use of a multinomial model with probit link for the counts of trial participants exceeding a set of thresholds. This modelling approach followed recommended practices (see TSD2[69]) and has precedent in UC submissions (e.g. TA547[21]).

The multinomial with probit modelling approach allows trials to utilise alternative thresholds within the pooled analysis. The set of thresholds across trials represent varying definitions of trial response and remission (see Document B, Tables 29 and 30). Although the inclusion criteria restricted trial outcomes to those involving reductions in Mayo scores, between-trial variation remained, primarily in the extent of Mayo score reduction constituting a response category. The modelling also accounts for inherent correlation between outcomes (i.e. between counts in different response categories), which has an advantage over a previous UC submission (TA633[20]) in which separate NMAs were carried out for remission and response, thereby disregarding the correlation.

The ERG noted that company network diagrams (Document B, Figures 31, 34, 37 and 40) are star-shaped, or nearly so, with a central node representing placebo. The networks have very few loops, indicating a lack of 'indirect' evidence in the networks. There are also few replicates of trials between the nodes, so difficulties in estimating heterogeneity precisely (if at all) may be anticipated. The Bayesian approach to modelling, as used by the company, seems apt if prior information can be justified and utilised, though this was not the case.

The company analysis took steps to account for heterogeneity, including:

- conducting separate NMAs for TNFi-experienced/TNFi-naïve combined with induction/maintenance phases,
- processing of 'treat-through' trials to emulate 're-randomised' trials (see Section 3.4.2.3 below) to remove or at least modulate a major heterogeneity from trial design,
- restriction of trials to those reporting outcomes determined using Mayo scores only, and

- restricting the time points of assessment in trials to 6 to 14 weeks for induction, and 52 to 60 weeks for maintenance.

The company pooled information of 'doses of the same active agent that had the same method of administration' for the base case (Document B, section B.2.8.4). Pooling of doses is contrary to Dias et al. (2018)[70] recommendation 'The default network meta-analysis … treats every intervention, every dose and every treatment combination a separate treatment' [p15]. On the other hand, the doses were at licensed levels and the ERG was advised by clinicians that the doses matched clinical practice (i.e. not implausibly low/high).

3.4.2.2. Choice of model (fixed effect or random effects)

The ERG noted that the underlying trials informing the NMA displayed considerable heterogeneity in respect of setting, inclusion criteria and baseline characteristics. As such, the ERG considered the RE model to be more appropriate than the FE model, as the latter assumes no between-trial heterogeneity is present. Heterogeneity in UC trials is known to be high, based on published literature (Macaluso et al. 2018).[48] It is also evident in the decision problem (variation in TNFi experience, see Table 7) and data (see baseline characteristics, CS Appendix, Tables 13 and 14; and outcome data, Table 16-Table 19 of Section 3.4.1). RE models are, therefore, better suited to NMAs of this condition due to the highly heterogeneous nature of studies; the company acknowledged this in the CS and reported that these were favoured over FE by default (Document B, Section B.2.8.4,p.96).

The ERG noted, however, that when attempting to fit RE models, the company reported model non-convergence in both TNFi-naïve and TNFi-experienced maintenance settings; and highly uncertain posterior standard deviation of between-trial variation within treatments in the setting for TNFi-experienced patients in the induction phase. As a result, the FE model was selected and applied by the company in all but the setting for TNFi-naïve patients in the induction phase.

The ERG noted that the aforementioned problems can often be remedied in RE models by using a more informative prior distribution on the variance parameter. The company did not report trying this remedy. The ERG acknowledged that in comparison to a FE model the likely results under the more defensible RE model would be similar point estimates, but an increase in the width of the credible intervals.

3.4.2.3. Choice of trial design ('treat-through' versus 're-randomised')

The company attempted to reduce a major potential source of heterogeneity in study design by carrying out a procedure to translate 'treat-through' trials to 're-randomised'. The ERG agreed with the principle but found the explanation poor in CS – it was clarified in response to questions A17 & A18. This process or similar appears to have been applied previously in e.g. TA547,[21] while the reverse approach (translating re-randomised to treat-through) was preferred in e.g. TA633.[20]

The procedure involved assuming those responding at the end of induction then enter the maintenance phase (as would happen in a 're-randomising' trial). This is discussed in more detail in Section 3.3.2.1 of this document. The assumptions are that the ratio of response during maintenance and response during induction is similar between similar trials, and that there is no delayed response, i.e. no participants respond after induction and then enter the maintenance phase. The ERG considered that there may be problems with this approach: there is evidence that these assumptions may be violated, and further there may be systematic differences between 're-randomised' and 'treat-through' trials due to differences in the trial process following randomisation. However, a sensitivity analysis conducted by the company indicated that the modification of treat-through trials is not very influential (see Section 3.5.2), and therefore the ERG did not investigate further.

3.4.2.4. Baseline risk

The sources for the baseline risk (the probability in the placebo arm of remaining in UC / not in response or remission) in the CS are the placebo arms, where they exist, of exactly the same set of trials used to estimate relative treatment effects. This runs contrary to TSD5,[71] which recommends separate modelling of relative treatment effects and baseline effects, and potentially separate sources of evidence, with the latter not restricted to randomized trials.

The studies informing the baseline model in the CS have not undergone a separate search process oriented to the baseline setting ("Investigators should identify evidence sources to inform the baseline model based on a protocol-driven systematic search ... " Dias et al. (2018)[70] (p.157)).

Furthermore, there has been no filtering of these studies towards the baseline setting. The ERG noted there is high variability in baseline characteristics across placebo arms of included trials (Appendix D.4, Tables 13 and 14). High heterogeneity in baseline variables may weaken

external validity: it may be that only a subset of these trials will match the decision problem population, or even none at all (in which case other sources of information e.g. registry data would be essential).

The ERG recommends selecting sources closest to UK clinical practice, with most appropriate choices for factors identified as important determinants of outcome, i.e. concomitant steroid use, duration of disease, prior TNFi, endoscopic central reading (Macaluso et al. 2018).[48] Clinical advice to the ERG confirmed that these factors are important to consider, in addition to severity of disease defined by a modified Mayo score of 9 or 10; endoscopy comprises one of the parts of this score. For its revised base case analysis, the ERG chose the following sources as best representing baseline risk: for the induction phase, PURSUIT SC[72] was selected for the TNFi-naïve subgroup, as it includes a similar gender split and roughly the same age as a large UK cohort[35] and OCTAVE 2[64] was selected for the TNFi-experienced subgroup, for the same reason; for the maintenance phase, PURSUIT M[66] was selected for the TNFi-naïve subgroup, and GEMINI1[61] for the TNFi-experienced subgroup, as the placebo arms of these trials still matched the age and gender split of the UK cohort most closely. In addition, these trials were all conducted in populations not exclusively including Asian participants; all trials were also assessed as having low risk of selection bias and were considered balanced in terms of prognostic factors at baseline (Appendix D.6., Table 15, pp.169-170). In these four studies, trial-level placebo arm average age since diagnosis was between 6.0 and 7.8 years, and average concomitant steroid use was between 42.9% and 57%. The characteristics of participants in the placebo arms of the selected trials are summarised in Table 20.

The company highlighted that remission or response data for the ERG's selected baseline trial for TNFi-experienced participants during induction (OCTAVE 2) were only available when pooled with results from the OCTAVE 1 trial. The baseline values for OCTAVE 1 are therefore also supplied in Table 20, where it can be seen that there is generally good correspondence, but that compared with OCTAVE 2 the percentage of males is about 13% higher, and the percentage with TNFi exposure about 5% less. The need to pool the ERG's selected trial (OCTAVE 2) with a similar trial (OCTAVE 1) is a limitation of the ERG's exploratory analysis.

Table 20: Baseline characteristics of participants in the placebo arms of trials selected for the ERG's placebo baseline risk NMA scenario

Characteristic (mean, unless otherwise specified)	TNFi-naïve (induction phase)	TNFi-experienced (induction phase)		TNFi-naïve (maintenance phase)	TNFi-experienced (maintenance phase)
	PURSUIT SC	OCTAVE 1	OCTAVE 2	PURSUIT M	GEMINI 1
Age, years (mean)	39.0	41.8	40.4	40.2	40.3
Male (%)	52.9	63.1	49.1	48.1	54.8
CRP (mg/L) (mean)	10.7	4.7	5.0[a]	9.6	NR
Years since diagnosis (mean)	6.0	6.0	6.2[a]	6.9	7.8
Mayo score (mean)	8.3	9.1	8.9	8.3	8.4
Left-sided disease (%)	57.0	30.3	35.1	NR	42.1
Extensive disease (%)	43.0	54.1	50.5	NR	13.5
Concomitant steroid use (%)	42.9	47.5	49.1	56.4	57
Biologic (TNFi) exposure (%)	NA	53.3	58.0	NA	37
Prior TNFi failure (%)	NA	52.5	53.6	NA	30.2

Abbreviations: CRP, C-reactive protein; NMA, network meta-analysis; TNFi, tumour necrosis factor inhibitor

Note: [a] median values

The ERG acknowledges the limitations of using these trials for placebo baseline risk, given the unsystematic selection of these based on limited information related to demographics, settings and methodological quality. This approach was selected due to time constraints within the appraisal and should therefore be seen as an attempt at improving the generalisability of results, vis-à-vis that of an unweighted average of all placebo arms, albeit with its own uncertainty. Clinical advice to the ERG confirmed that the baseline risk values used from the selected trials are broadly acceptable and representative of the relevant population by TNFi experience, as well as the treatment phase, but cautioned that no patients with proctitis were explicitly included. These patients are estimated by clinical experts to the ERG as representing approximately 20% of patients treated in the Royal Devon and Exeter NHS Foundation Trust. As a result, clinical advice to the ERG indicated that participants included in trials generally have more severe disease and increased use of steroids when compared to the 'general' population of patients with moderately to severely active UC in the UK. The ERG noted this caution and

considered that baseline placebo risk from the selected trials may not be fully generalisable to the target population, and recommends that a proper protocol-driven systematic review procedure as described in NICE guidance (Dias et al 2013)[1] is followed in respect of estimating baseline placebo risk, including non-RCT sources where available.

The baseline data (response or remission proportion; remission proportion in the placebo arm) for the selected trials are shown in Table 21 along with the estimates from the ERG's updated NMA. These estimated proportions are smaller than those used in the company base case where all trials with placebo arms were used. More information on the results of the ERG's updated NMA is given in Section 3.5.3.1.

Table 21: Comparison of response/remission proportions from data with estimates from NMA with updated baseline selection

Setting	Trial supplying baseline risk	Data source table[a]	Response or remission	NMA estimate	Remission	NMA estimate
Induction/ naïve	PURSUIT SC	31	89/292 (0.30)	0.30	20/292 (0.07)	0.07
Induction/ experienced	OCTAVE 1 + 2	32	29/124 (0.23)	0.22	1/124 (0.008)	0.03
Maintenance/ naïve	PURSUIT M	33	48/154 (0.31)	0.31	34/154 (0.22)	0.17
Maintenance/ experienced	GEMINI 1	34	6/38 (0.16)	0.15	2/38 (0.05)	0.06

Abbreviation: NMA, network meta-analysis

Note: [a] Clarification response, Appendix 2

3.4.2.5. Effect modification

The NMAs are to some extent protected from bias with respect to select potential effect modifiers by the company's approach. In the case of trial design (re-randomised versus treat-through), the company made a statistical adjustment to the treat-through trials (though the ERG notes some issues with this process). In the case of prior TNFi, the NMAs are conditioned on this factor i.e. they are analysed separately and prior TNFi is held fixed within each analysis (though the division between these levels (naïve/ experienced) may be somewhat blurred, see 3.3.2.3). Some other measures taken by the company are listed in Section 3.4.2.1.

The CS recorded baseline characteristics in Appendices Tables 13 and 14. This includes information on known prognostic or effect modifying factors mentioned by clinicians advising the

ERG, such as steroid use, duration of disease and age. There is wide variation between trials in some characteristics that may be effect modifiers, for example, extent of disease is plausibly related to treatment effect. A comment on the variation was requested in clarification A8. Placebo arm extensive disease in the induction phase ranges from 7.1% (VISIBLE1[73]) to 80.8% (Kobayashi et al. 2016[74]). In response, the company indicated that the range was reduced (7.1% to 56.2%) when excluding trials which only recruited Asian participants; the ERG noted that this remains a considerable range, a position clinical advice to the ERG confirmed. In relation to placebo arm CRP (mg/L) ranges from 3.2 (ULTRA1[65]) to 35.1 (Jiang et al. 2015[75]) in the induction phase, the company responded that clinicians in a previous appraisal described CRP measurements as 'non-specific' inflammatory marker. Clinical advice to the ERG confirmed that CRP has limited utility as marker outside of the acute severe UC context not covered in this appraisal, however, clinical experts noted that CRP levels <5 mg/L in the UNIFI[52] trial were predictive of response. Duration of disease appears to be fairly consistent when reported for the maintenance phase (ranging from 5.4 to 8.7 years) but much less so in the induction phase (ranging from 3.8 to 14.6 years). Concomitant steroid use is also relatively consistent during maintenance (ranging from 28% to 58%) but not the induction phase (ranging from 27% to 100%) (Document B, p.91; full trial-level data in CS Appendix D.4, Tables 13 and 14).

The ERG carried out random-effects meta-analyses of the response (no remission) and remission proportions in the placebo arms (data supplied in Clarification Response, Tables 31 to 34). The results are shown in Figure 1 to Figure 8, and the estimated I^2 from these analyses in Table 22.

Table 22: Estimates of heterogeneity (I^2) from placebo arms of the trials included in the CS, as calculated by ERG

	Estimate of I^2 (%) for remission	**Estimate of I^2 (%) for response (no remission)**
Induction/TNFi-naïve	44.5	28.8
Induction/TNFi-experienced	56.8	27.9
Maintenance/TNFi-naïve	50.6	65.0
Maintenance/TNFi-experienced	3.3	65.3

Abbreviation: TNFi, tumour necrosis factor inhibitor

Heterogeneity is substantial in many settings, most notably in the maintenance setting for response no remission. It is more moderate in the induction setting for response no remission,

and very low in the maintenance/experienced setting for remission. Unexplained heterogeneity in outcomes between trials might signal the influence of effect modifiers (known or unknown) and therefore potential bias in the NMA. Overall, there appears to be some reduction in heterogeneity compared to an earlier analysis of placebo arm outcomes in UC (Macaluso et al. 2018),[48] which could be attributed to the measures taken by the company (e.g. restricting time point of assessment, conditioning on TNFi experience, etc.)

Figure 1 : Placebo-arm, trial-specific response (no remission) proportions for the TNFi-naïve during induction NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 31

Figure 2: Placebo-arm, trial-specific response (no remission) proportions for the TNFi-experienced during induction NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 32

Figure 3: Placebo-arm, trial-specific response (no remission) proportions for the TNFi-naïve during maintenance NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 33

Figure 4: Placebo-arm, trial-specific response (no remission) proportions for the TNFi-experienced during maintenance NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 34

Figure 5: Placebo-arm, trial specific remission proportions for the TNFi-naïve during induction NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 31

Figure 6: Placebo-arm, trial specific remission proportions for the TNFi-experienced during induction NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 32

Figure 7: Placebo-arm, trial specific remission proportions for the TNFi-naïve during maintenance NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 33

Figure 8: Placebo-arm, trial specific remission proportions for the TNFi-experienced during maintenance NMA setting

Abbreviation: RE, random effects

Note: Data obtained from the company's clarification response, Appendix 2, Tables 34

3.4.3. Relevance to the target population

The ERG considered the company's analyses to be broadly appropriate for the populations of interest, though it raised concerns about the generalisability of results stratified by prior TNFi experience, given the reality of the current treatment pathway is more complex. As such, the ERG is of the opinion that analyses stratified by biologic experience would have been more appropriate. The ERG also did not agree with the exclusion of tofacitinib, given its prominent role in the UK treatment landscape, as per clinical advice to the ERG.

There is a lack of published literature on effect modifiers in UC; as a result, the ERG considered that high variability in baseline characteristics of placebo arms of included trials may have represented imbalances in unknown effect modifiers, though it is known that treatment efficacy

varies widely between individuals based on demographic, medication use and clinical characteristics[48]. Furthermore, in line with NICE guidance[1], the ERG was of the opinion that an unweighted average of outcomes reported in the placebo arms of trials included in the NMA, the approach confirmed by the company in clarification response A15, was not appropriate. The ERG considered the use of placebo outcomes from studies that are highly generalisable to the UK-specific context and are identified through a proper, protocol-driven systematic review to be the most appropriate. However, given time constraints in this appraisal, the ERG selected placebo baseline risk from a single trial per TNFi experience and treatment setting dyad from the trials included in the NMA which matched the relevant dyad most closely for its base case.

3.4.4. Results of the indirect treatment comparison

The results of the company's base case NMAs are provided in the following sections, according to subgroups by prior TNFi experience and stratification by the induction and maintenance phases of the treatment.

3.4.4.1. TNFi-naïve participants (induction phase)

A summary of the results of the company's base case NMA for TNFi-naïve participants during the induction phase, comparing ozanimod to comparators and placebo, is presented in Table 23. Rank data taken from league tables provided during clarification indicated that ozanimod ranked 3rd (of 8) for both clinical response and clinical remission in this subgroup during induction; the ERG interpreted this with caution as no confidence intervals were reported around these ranks.

Furthermore, the results suggested that ozanimod was more likely to achieve clinical response and clinical remission during treatment induction of TNFi-naïve participants when compared to pooled doses of ustekinumab, tofacitinib, golimumab, adalimumab and placebo; the ERG noted that this effect was only statistically significant versus placebo. Ozanimod was out-performed by pooled doses of infliximab and vedolizumab for both outcomes, though the ERG noted that these effects were also non-significant.

Table 23: NMA outcomes for ozanimod 1 mg QD versus comparators in TNFi-naïve participants during the induction phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Infliximab	Pooled	[redacted]	[redacted]
Vedolizumab	300 mg IV	[redacted]	[redacted]
Ustekinumab	Pooled	[redacted]	[redacted]
Tofacitinib	10 mg BID	[redacted]	[redacted]
Golimumab	200/100 mg SC	[redacted]	[redacted]
Adalimumab	160/80/40 mg Q2W	[redacted]	[redacted]
Placebo	-	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; IV, intravenous; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Notes: [a] random effects NMA. Comparators serve as the reference group; darker coloured cells represent statistically significant differences: blue cells favour ozanimod, grey cells favour the comparator.

Due to the wide credible intervals reported around the relative effect estimates for comparators, the ERG considered all comparators versus placebo as effects were expected to be more precise as a result of the weight and proximity of evidence relative to placebo. This comparison reflected ozanimod versus other treatments: pooled doses of infliximab and vedolizumab resulted in greater relative effect against placebo when compared to ozanimod, all other treatments had smaller relative effects against placebo; all results against placebo were statistically significant. The results of this comparison are reported in Table 24.

Table 24: NMA outcomes for comparators versus placebo in TNFi-naïve participants during the induction phase

Treatment	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Infliximab	Pooled	[redacted]	[redacted]
Vedolizumab	300 mg IV	[redacted]	[redacted]
Ozanimod	1 mg QD	[redacted]	[redacted]
Ustekinumab	Pooled	[redacted]	[redacted]
Tofacitinib	10 mg BID	[redacted]	[redacted]
Golimumab	200/100 mg SC	[redacted]	[redacted]
Adalimumab	160/80/40 mg Q2W	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; IV, intravenous; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Notes: [a] random effects NMA. Placebo serves as the reference group; darker coloured cells represent statistically significant differences: blue cells favour treatment, grey cells favour placebo.

3.4.4.2. TNFi-naïve participants (maintenance phase)

A summary of the results of the company's base case NMA for TNFi-naïve participants during the maintenance phase, comparing ozanimod to comparators and placebo, is presented in Table 25. Rank data taken from league tables provided during clarification indicated that ozanimod ranked 7th (of 9) for both clinical response and clinical remission in this subgroup during maintenance; the ERG interpreted this with caution as no confidence intervals were reported around these ranks.

The results suggested that ozanimod was out-performed by all treatments, with the exception of adalimumab and placebo, in terms of clinical response. Notably, pooled doses of vedolizumab as well as vedolizumab 108 mg resulted in significantly higher clinical response than ozanimod. Ozanimod was also out-performed by all treatments except adalimumab and placebo for clinical remission as an outcome. The ERG noted that pooled doses of vedolizumab and pooled doses of tofacitinib resulted in significantly higher clinical remission than ozanimod. The results of comparisons with all other active treatments for the two outcomes were statistically non-significant.

Table 25: NMA outcomes for ozanimod 1 mg QD versus comparators in TNFi-naïve participants during the maintenance phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Adalimumab	40 mg Q2W	[redacted]	[redacted]
Infliximab	Pooled	[redacted]	[redacted]
Golimumab	Pooled	[redacted]	[redacted]
Ustekinumab	Pooled	[redacted]	[redacted]
Vedolizumab	Pooled	[redacted]	[redacted]
Vedolizumab	108 mg Q2W SC	[redacted]	[redacted]
Tofacitinib	Pooled	[redacted]	[redacted]
Placebo	-	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Notes: [a] fixed effects NMA. Comparators serve as the reference group; darker coloured cells represent statistically significant differences: blue cells favour ozanimod, grey cells favour the comparator.

Due to the large number of non-significant relative effect estimates for comparators, the ERG considered all comparators versus placebo as effects were expected to be more precise as a

result of the weight and proximity of evidence relative to placebo. All treatments, with the exception of adalimumab, resulted in greater relative effect against placebo when compared to ozanimod; all results against placebo were statistically significant. The results of this comparison are reported in Table 26.

Table 26: NMA outcomes for comparators versus placebo in TNFi-naïve participants during the maintenance phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Adalimumab	40 mg Q2W	[redacted]	[redacted]
Ozanimod	1 mg QD	[redacted]	[redacted]
Infliximab	Pooled	[redacted]	[redacted]
Golimumab	Pooled	[redacted]	[redacted]
Ustekinumab	Pooled	[redacted]	[redacted]
Vedolizumab	Pooled	[redacted]	[redacted]
Vedolizumab	108 mg Q2W SC	[redacted]	[redacted]
Tofacitinib	Pooled	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Notes: [a] fixed effects NMA. Placebo serves as the reference group; darker coloured cells represent statistically significant differences: blue cells favour treatment, grey cells favour placebo.

3.4.4.3. TNFi-experienced participants (induction phase)

A summary of the results of the company's base case NMA for TNFi-experienced participants during the maintenance phase, comparing ozanimod to comparators and placebo, is presented in Table 27. Rank data taken from league tables provided during clarification indicated that ozanimod ranked 2nd (of 6) for both clinical response and clinical remission in this subgroup during induction; the ERG interpreted this with caution as no confidence intervals were reported around these ranks.

Results suggested that ozanimod was more likely to achieve clinical response and clinical remission during treatment induction of TNFi-experienced participants when compared to pooled doses of ustekinumab, vedolizumab, adalimumab and placebo; the ERG noted that this effect was only statistically significant versus adalimumab and placebo. Ozanimod was out-performed by tofacitinib for both outcomes, though the ERG noted that these effects were non-significant.

Table 27: NMA outcomes for ozanimod 1 mg QD versus comparators in TNFi-experienced participants during the induction phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Adalimumab	160/80/40 mg Q2W	[redacted]	[redacted]
Vedolizumab	300 mg IV	[redacted]	[redacted]
Ustekinumab	Pooled	[redacted]	[redacted]
Tofacitinib	10 mg BID	[redacted]	[redacted]
Placebo	-	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; IV, intravenous; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; TNFi, tumour necrosis factor inhibitor

Notes: [a] fixed effects NMA. Comparators serve as the reference group; darker coloured cells represent statistically significant differences: blue cells favour ozanimod, grey cells favour the comparator.

Due to a number of non-significant relative effect estimates for comparators and wide credible intervals for adalimumab, the ERG considered all comparators versus placebo as effects were expected to be more precise as a result of the weight and proximity of evidence relative to placebo. All treatments resulted in greater relative effect against placebo when compared to ozanimod; all results against placebo were statistically significant with the exception of adalimumab. The results of this comparison are reported in Table 28.

Table 28: NMA outcomes for comparators versus placebo in TNFi-experienced participants during the induction phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Adalimumab	160/80/40 mg Q2W	[redacted]	[redacted]
Vedolizumab	300 mg IV	[redacted]	[redacted]
Ustekinumab	Pooled	[redacted]	[redacted]
Ozanimod	1 mg QD	[redacted]	[redacted]
Tofacitinib	10 mg BID	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; IV, intravenous; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; TNFi, tumour necrosis factor inhibitor

Notes: [a] fixed effects NMA. Placebo serves as the reference group; darker coloured cells represent statistically significant differences: blue cells favour treatment, grey cells favour placebo.

3.4.4.4. TNFi-experienced participants (maintenance phase)

The results of the company's base case NMA for TNFi-experienced participants during the maintenance phase, comparing ozanimod to comparators and placebo, are summarised in Table 29. Rank data taken from league tables provided during clarification indicated that

ozanimod ranked 4th (of 7) for both clinical response and clinical remission in this subgroup during maintenance; the ERG interpreted this with caution as no confidence intervals were reported around these ranks.

Results suggested that ozanimod was more likely to achieve clinical response and clinical remission during treatment maintenance of TNFi-experienced participants when compared to pooled doses of ustekinumab, adalimumab and placebo; the ERG noted that this effect was only statistically significant versus placebo. Ozanimod was out-performed by pooled doses of tofacitinib, pooled doses of vedolizumab and vedolizumab 108 mg for both outcomes, though the ERG noted that these effects were non-significant.

Table 29: NMA outcomes for ozanimod 1 mg QD versus comparators in TNFi-experienced participants during the maintenance phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Ustekinumab	Pooled	[redacted]	[redacted]
Adalimumab	40 mg Q2W	[redacted]	[redacted]
Vedolizumab	Pooled	[redacted]	[redacted]
Tofacitinib	Pooled	[redacted]	[redacted]
Vedolizumab	108 mg Q2W SC	[redacted]	[redacted]
Placebo	-	[redacted]	[redacted]

Abbreviations: CrI, credible interval; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Notes: [a] fixed effects NMA. Comparators serve as the reference group; darker coloured cells represent statistically significant differences: blue cells favour ozanimod, grey cells favour the comparator.

Due to the large number of non-significant relative effect estimates for comparators, the ERG considered all comparators versus placebo as effects were expected to be more precise as a result of the weight and proximity of evidence relative to placebo. Pooled doses of vedolizumab as well as tofacitinib and vedolizumab 108 mg resulted in greater relative effect against placebo when compared to ozanimod, with pooled doses of ustekinumab and adalimumab resulting in smaller relative effects against placebo when compared to ozanimod. The ERG noted that all results against placebo were statistically significant, but that large imprecision was still observed. The results of this comparison are reported in Table 30.

Table 30: NMA outcomes for comparators versus placebo in TNFi-experienced participants during the maintenance phase

Comparator	Dose	Clinical response; OR (95% CrI) [a]	Clinical remission; OR (95% CrI) [a]
Ustekinumab	Pooled	[redacted]	[redacted]
Adalimumab	40 mg Q2W	[redacted]	[redacted]
Ozanimod	1 mg QD	[redacted]	[redacted]
Vedolizumab	Pooled	[redacted]	[redacted]
Tofacitinib	Pooled	[redacted]	[redacted]
Vedolizumab	108 mg Q2W SC	[redacted]	[redacted]

Abbreviations: CrI, credible interval; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Notes: [a] fixed effects NMA. Placebo serves as the reference group; darker coloured cells represent statistically significant differences: blue cells favour treatment, grey cells favour placebo.

3.4.4.5. Sensitivity analyses

Sensitivity analyses reported by the company comprised an assumption of not pooling doses of the same active treatment if it had the same method of administration, the exclusion of trials with a treat-though design, using the three-component Mayo score instead of the four-component Mayo score in the TRUENORTH trial, the exclusion of trials conducted exclusively in Asian populations, and the inclusion of TOUCHSTONE[34] in the TNFi-naïve analysis.

The ERG generally agreed with the company's assessment that the results of sensitivity analyses demonstrated robustness of the base case NMA for the factors that were explored, but noted that wide confidence intervals in the base case, in particular for the TNFi-experienced subgroup, made it difficult to identify meaningful differences across analyses. Of note, the sensitivity analyses using three-component Mayo scores for TRUENORTH indicated large shifts in the point estimates for ozanimod relative to placebo, particularly in the TNFi-experienced subgroup. The ERG also noted higher point estimates, on average, for vedolizumab, infliximab, golimumab and adalimumab when Asian trials were excluded – this was considered to be an expected effect given the investigation of these treatments by the five trials including Asian participants only, i.e. Motoya,[51] Kobayashi,[74] Jiang,[75] PURSUIT J[63] and Suzuki.[47]

The ERG also did not consider all uncertainties in the NMA to have been addressed by the sensitivity analyses, though it did note an exploration of the effect of re-calculated data from treat-through trials, which was identified as a source of considerable uncertainty. It conducted a

thorough comparative assessment of the sensitivity analysis excluding data from treat-through trial designs, reported in Section 3.5.2.

3.4.5. Conclusions on the indirect treatment comparison

The company carried out NMAs in four settings combining trial phase (maintenance/induction) with prior TNFi treatment (experienced/naïve) using a modelling approach (multinomial with probit link) that was very appropriate in the ERG's view. A further strength of the submission was in the form of measures the company took to counter or reduce potential effect modification from factors including trial phase, prior TNFi treatment, trial design (re-randomised/treat-through), outcome definition and timepoint of assessment. However, evidence of further variation in potential effect modifiers is apparent in baseline characteristics (e.g. extensive disease, see Section 3.4.2.5) and placebo arm responses (see Figure 1 through Figure 8). The company also mainly used FE models when RE models are more appropriate (given heterogeneity), as the latter would not converge – the ERG recommends using informative priors as a potential remedy as FE models do not account adequately for the observed between-trial heterogeneity. Furthermore, the ERG considered the company's approach in using all available trials to inform the baseline placebo risk to be a limitation, and recommends instead a systematic review to select the most appropriate sources for the UK context and rerun the NMAs.

Results of the NMA indicated that ozanimod was a ranked in the top three treatments for the induction phase in both TNFi-naïve and -experienced patients, though only placebo was significantly outperformed in both subgroups; adalimumab was additionally significantly outperformed in the TNFi-experienced subgroup during the induction phase. Comparison with all other active treatments yielded non-significant results; in particular, the ERG noted that tofacitinib was non-significantly outperformed by ozanimod in the TNFi-naïve subgroup, and non-significantly outperformed ozanimod in the TNFi-experienced subgroup. These results indicate that ozanimod is a moderately effective treatment, that considerable uncertainty exists around its relative treatment effect and that tofacitinib is a comparable treatment in terms of efficacy in the induction phase (approximately 6 to 14 weeks), regardless of prior TNFi experience.

In the maintenance phase, results of the NMA show that ozanimod was ranked in the lowest three for the TNFi-naïve subgroup and middle of the range for the TNFi-experienced subgroup. In the former, ozanimod only significantly outperformed placebo and was significantly

outperformed by pooled doses of vedolizumab (for clinical response and remission), vedolizumab 108 mg (for clinical response only) and tofacitinib (for clinical remission only). Comparisons with all other active treatments were non-significantly in favour of the comparator. These results indicate that ozanimod may be a less efficacious treatment for the maintenance of TNFi-naïve patients, though considerable uncertainty exists about its relative treatment effect. The ERG also noted that results suggested tofacitinib may be a more efficacious treatment compared to ozanimod in this setting. In the TNFi-experienced subgroup, ozanimod significantly outperformed placebo and non-significantly outperformed ustekinumab and adalimumab, with all other comparators non-significantly outperforming ozanimod. The results suggest that considerable uncertainty exists around the relative efficacy of ozanimod against comparators for the maintenance of TNFi-experienced patients.

3.5. Additional work on clinical effectiveness undertaken by the ERG

3.5.1. Additional searches

As described in Sections 3.1 and 3.3.3, the ERG did not consider the company's exclusion of evidence from phase 4 studies from the submission to be appropriate. Consequently, the ERG carried out additional searches for phase 4 trials reporting on ozanimod and its comparators for moderately to severely active UC. Searches for phase 4 studies of ozanimod were conducted in Ovid MEDLINE, Ovid Embase, clinicaltrials.gov, the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) and the European Union (EU) Clinical Trials Register; Ovid MEDLINE and Embase were searched from 1946 and 1974, respectively (Appendix A). These searches yielded no results, with only one trial investigating ozanimod for multiple sclerosis identified.

Searches for Phase 4 studies of relevant comparators were conducted in Ovid MEDLINE and Ovid Embase, from 1946 and 1974, respectively (Appendix A). The searches yielded 28 potentially eligible records, though screening of this yield identified no studies reporting real-world evidence for adalimumab, tofacitinib or ustekinumab, filgotinib, etrasimod or etrolizumab. The ERG noted that filgotinib is currently under appraisal by NICE for moderately to severely active UC, and that etrasimod and etrolizumab do not currently have FDA or EMA approval – it therefore considered that these were not comparators of interest for this appraisal and that it would not have been possible for evidence on these treatments to inform links in NMA networks.

A total of 13 publications reporting results of phase 4 trials investigating golimumab were identified. Nine records, all related to the GO-COLITIS trial (NCT02092285), reported results for TNFi-naïve populations; four records reported results for participants with mixed TNFi experience, with two of these related to the GORE-UC trial,[76] one related to GO-LEVEL (NCT03124121)[77] and another related to a trial by Yu et al. (2021).[78] These trials were all considered ineligible as they described single-group assignment to golimumab.

Ten publications reporting phase 4 results for infliximab were identified; six were related to the NOR-SWITCH trial (n=4; NCT02148640) and its open label extension (n=2), two related to the NOR-DRUM trial (NCT03074656), one to the SECURE[79] trial and one to the trial by Park et al. (2015).[80] All trials were assumed to describe results in participants with mixed TNFi experience, though none stated this explicitly. The trials were all considered ineligible by the ERG: NOR-SWITCH and the SECURE trial both compared infliximab with a biosimilar; NOR-DRUM compared infliximab plus standard of care with infliximab plus TDM and the trial by Park et al. (2015)[80] described single-group assignment to an infliximab biosimilar.

Five publications reporting phase 4 results for vedolizumab were obtained during the searches. Two publications were related to a trial by Coletta et al. (2020)[81] (Eudract number 2015-003270-32) and one each to the trials by Danese et al. (2021)[82] (NCT02743806), Osterman et al. (2020)[83] and Vermeire et al. (2020).[84] As for infliximab, all trials were assumed to describe results in participants with mixed TNFi experience, though this was not explicit. All trials were considered ineligible as the trial by Coletta et al. (2020)[81] and Danese et al. (2021)[82] described single-group assignment, the trial by Vermeire[84] investigated single-group de-escalation of vedolizumab dosing and the trial by Osterman et al. (2020)[83] compared serum vedolizumab concentration in responders and non-responders.

As such, the ERG did not consider that the exclusion of phase 4 evidence from the CS meaningfully changed the results of the NMA or conclusion of the submission, though the methodological bias this approach could introduce is reiterated.

3.5.2. Validation of robustness of NMAs including treat-through trial data

As described in Section 3.3.2, the ERG considered the combination of re-calculated treat-through trial data with data from re-randomised trials to be a potential source of heterogeneity and bias in the base case NMA. Therefore, the sensitivity analysis of each comparator relative

to placebo was compared with the corresponding base case result; these are summarised in Table 31 and Table 32.

Table 31: Comparison of treatment effect relative to placebo between the NMA base case and sensitivity analysis excluding treat-through trial data: TNFi-naïve population in the maintenance phase

Comparator vs. placebo	Clinical response estimate; OR (95% CrI)		Clinical remission estimate; OR (95% CrI)	
	Base case [a]	Sensitivity analysis [a]	Base case [a]	Sensitivity analysis [a]
Tofacitinib 10 mg BID	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab 108 mg Q2W SC	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab pooled	[redacted]	[redacted]	[redacted]	[redacted]
Ustekinumab pooled	[redacted]	[redacted]	[redacted]	[redacted]
Golimumab pooled	[redacted]	[redacted]	[redacted]	[redacted]
Infliximab pooled	[redacted]	[redacted]	[redacted]	[redacted]
Ozanimod 1 mg QD	[redacted]	[redacted]	[redacted]	[redacted]
Adalimumab 40 mg Q2W	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BID, twice a day; CrI, credible interval; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor; vs., versus

Notes: [a] fixed effects NMA. Blue cells represent statistically significant results

Table 32: Comparison of treatment effect relative to placebo between the NMA base case and sensitivity analysis excluding treat-through trial data: TNFi-experienced population in the maintenance phase

Comparator vs. placebo	Clinical response estimate; OR (95% CrI)		Clinical remission estimate; OR (95% CrI)	
	Base case [a]	Sensitivity analysis [a]	Base case [a]	Sensitivity analysis [a]
Vedolizumab 108 mg Q2W SC	[redacted]	[redacted]	[redacted]	[redacted]
Tofacitinib pooled	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab pooled	[redacted]	[redacted]	[redacted]	[redacted]
Ozanimod 1 mg QD	[redacted]	[redacted]	[redacted]	[redacted]
Adalimumab 40 mg	[redacted]	[redacted]	[redacted]	[redacted]

Comparator vs. placebo	Clinical response estimate; OR (95% CrI)		Clinical remission estimate; OR (95% CrI)	
	Base case [a]	Sensitivity analysis [a]	Base case [a]	Sensitivity analysis [a]
Ustekinumab pooled	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: CrI, credible interval; NMA, network meta-analysis; OR, odds ratio; Q2W, every two weeks; QD, once a day; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor; vs., versus

Notes: [a] fixed effects NMA; blue cells represent statistically significant results

The ERG considered the comparison between the NMA base case and the sensitivity analysis excluding treat-through data to indicate that re-calculated treat-through data did not meaningfully bias the relative point estimates, with the majority of relative risks differing by less than 0.2 between the base case and sensitivity analysis in both subgroups related to prior TNFi experience. The only exception was for comparisons of golimumab with placebo in the TNFi-naïve subgroup, where base case estimates were 0.6 lower compared to estimates derived from sensitivity analyses.

Furthermore, the ERG noted that the 95% credible intervals around relative effect estimates were considerably wider for sensitivity analyses conducted in the TNFi-naïve subgroup, and fewer results were nominally significant as a result. The ERG found this plausible and likely attributable to the fact that three of the four treat-through trials, i.e. ACT1,[44] ULTRA2[43] and Suzuki (2014),[47] were conducted in exclusively TNFi-naïve populations; their exclusions therefore resulting in a very sparse network and considerable imprecision around the point estimate. Given that only the VARSITY[60] trial reported treat-through data for the TNFi-experienced population, the ERG noted that base case precision and nominal significance was generally retained.

3.5.3. Validation of company NMAs

The company supplied JAGS model code in CS Appendix D.4.3, but no accompanying data (either coded or within the CS). In clarification question A6 the ERG requested code for data setup and execution of the NMA. Although in response the company supplied data and setup information, this was not provided in an executable form. It was therefore necessary for the ERG to finalise coding itself, including reconfiguring response data and initial values provided as a printout or list. The procedure to obtain stable estimates as described in the CS (Document B, p.97) was not fully specified and the ERG could not implement it. In addition, the ERG noted that the code supplied as the RE model (company's clarification response, Figure 12) was for

the FE case. As a result, implementation of JAGS code by the ERG was time-consuming, resulting in limited latitude for revisions of the base case NMA or further NMA scenarios. In spite of this, coding was largely successful, and the ERG were able to closely, but not identically, replicate the NMA results provided by the company.

3.5.3.1. ERG base case NMA using alternative baseline placebo risk values

As discussed in Section 3.4.2.4, the ERG considered the use of a more generalisable study to the UK context to inform the placebo baseline risk in the NMA to be a more appropriate approach than the unweighted average of all placebo arms used by the company. The ERG included the NMA estimates generated using these values in its revised base case, though the limitations associated with this approach are acknowledged in Section 3.4.2.4. Time constraints precluded the ideal approach, i.e. using highly generalisable studies identified through a protocol-driven systematic review. The trials selected for the ERG base case NMA, as well as the baseline characteristics of participants in the placebo arm of each trial, are summarised in Table 20.

For its base case, the ERG modified the company JAGS code to select the placebo arm of a single trial in each setting (whereas the company's code averaged all placebo arms) as data for estimation of baseline risk. The process for estimation of relative effects was unchanged and estimates of OR were similar between the ERG and company base case. The pattern of convergence was similar to that reported by the company, and the same model choices were applied, namely RE for the induction setting in TNFi-naïve participants, and FE otherwise. Estimated probabilities of being in each response category by treatment and setting are affected; the results of the ERG base case NMAs, and comparative ERG-replicated company NMA results, are summarised in Figure 9 and Table 33.

Figure 9: Visualisation of results of ERG base case NMAs with revised placebo baseline risk compared to ERG-replicated results of company base case NMAs

Abbreviations: ADA, adalimumab; CS, company submission; ERG, evidence review group; GOL, golimumab; IFX, infliximab; ind/naïve, TNFi-naïve subgroup during induction; ind/exp, TNFi-experienced subgroup during induction; maint/naïve, TNFi-naïve subgroup during maintenance; maint/exp, TNFi-experienced during maintenance; no_resp, no response; OZA, ozanimod; partial_resp, partial response (response no remission); PBO, placebo; TOF, tofacitinib; UST, ustekinumab; VEDO, vedolizumab; VEDO 108, vedolizumab 108 mg Q2W SC

Table 33: Numerical results of ERG base case NMAs with revised placebo baseline risk compared to ERG-replicated results of company base case NMAs

Treatment	Setting (TNFi experience/treatment phase)	Response		Remission		No response	
		Company	ERG	Company	ERG	Company	ERG
PBO	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PBO	Experienced/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PBO	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PBO	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OZA	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OZA	Experienced/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OZA	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OZA	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	Experienced/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFX	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IFX	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF	Experienced/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
UST	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
UST	Experienced/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
UST	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
UST	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEDO	Naïve/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEDO	Experienced/induction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEDO	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEDO	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEDO 108	Naïve/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEDO 108	Experienced/maintenance	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: ADA, adalimumab; ERG, evidence review group; GOL, golimumab; IFX, infliximab; OZA, ozanimod; PBO, placebo; TOF, tofacitinib; UST, ustekinumab; VEDO, vedolizumab; VEDO 108, vedolizumab 108 mg Q2W SC

The comparison of results from the ERG base case NMAs with results from the company NMAs indicate that the company approach estimated a higher response for most treatments, with the 'no response' outcome higher for placebo and almost all active treatments in the ERG NMAs. The ERG noted that these findings are indirectly validated by clinical advice to the ERG, which suggested that remission and response in the placebo arms of trials included in the company NMA are higher than expected. The ERG concluded that using placebo baseline risks from more generalisable studies represented a more conservative base case.

3.5.3.2. Re-estimation of company NMA using random effects modelling

After closely replicating the NMA results provided by the company as described in Section 3.5.3.1, the ERG attempted to re-run the NMA using RE models with alternative baseline placebo values for the three analyses where FE were used (see Section 3.4.2.2). Clear non-convergence, broadly in line with what was observed in the company base case, was detected for the odds ratios of these analyses. As described in Section 3.4.2.2, the ERG considered that the use of an appropriate informative prior distribution, e.g. those reported in Turner et al. (2012)[67] or Turner et al. (2015),[68] could resolve this problem.

3.6. Conclusions of the clinical effectiveness section

Based on the evidence presented in the CS for the pivotal TRUENORTH[27,28] trial and the supplementary TOUCHSTONE.[34] trial, as summarised in Section 3.2.5.1, the ERG concluded that ozanimod has a significant effect on the outcomes of clinical remission and clinical response in both the induction and maintenance phases in the overall population, when compared to placebo. Furthermore, ozanimod resulted in significant improvements in other categories of remission (maintenance of remission, durable remission and corticosteroid-free remission), endoscopic healing, mucosal healing and measures of disease activity compared to placebo. The results of the effect of ozanimod on HRQoL were more variable, with some domains showing no significant change compared to placebo. Furthermore, the ERG noted that the proportion of various adverse events were higher for ozanimod compared to placebo, though no formal tests of significance are reported. As a result, the ERG concluded that ozanimod is an effective treatment compared to placebo, though its effect on quality of life and its safety are uncertain. These results were mostly reflected in the clinical results reported for subgroup analyses by TNFi experience, though the ERG noted that effects were smaller in the TNFi-experienced subgroup. The company posited that this was due to TNFi-experienced patients being more difficult to treat, a position the ERG agreed with.

As discussed in Section 3.4.5, the results of the indirect treatment comparison showed that ozanimod is a moderately effective treatment during the induction phase of treatment, regardless of TNFi experience, but that considerable uncertainty exists around its relative treatment effect. The ERG further noted that tofacitinib is a comparable treatment to ozanimod in terms of efficacy in the induction phase. The results of the indirect treatment comparison for the maintenance phase further indicated that ozanimod may be a less efficacious treatment for the maintenance of TNFi-naïve patients, though considerable uncertainty exists about its relative treatment effect, and that tofacitinib may be a more efficacious treatment compared to ozanimod in this setting. Furthermore, in the TNFi-experienced subgroup, results suggested that considerable uncertainty exists around the relative efficacy of ozanimod against comparators for the maintenance of patients with prior TNFi experience.

4. COST-EFFECTIVENESS

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

The company undertook a SLR to identify evidence for outcomes relevant to the cost-effectiveness, HRQoL, healthcare resource use (HCRU) and cost of ozanimod and comparator treatments for the treatment of moderate to severe UC.

Table 34. Summary of ERG'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	ERG assessment of robustness of methods
Searches	Appendix G.2.	The ERG noted the following limitations: a recognised filter for identifying cost-effectiveness studies was not used; database searches applied limits that excluded conference abstracts from search results, however, hand searching and database searches of known conference proceedings may have mitigated this issue. Missing search strategies were provided in response to clarification question B1.
Inclusion criteria	Appendix G.3	The ERG notes that studies reporting 'primarily clinical outcomes' are excluded. It is not clear whether this may involve the exclusion of studies reporting both clinical and cost-effectiveness evidence. Despite this, the ERG considered the inclusion criteria to be broadly appropriate to encompass the cost-effectiveness evidence for all the relevant comparators to this appraisal.
Screening	Appendix G.3	Title and abstract screening was conducted by two reviewers, with a third available to resolve discrepancies. The same procedure was followed at full-text screening. Where studies gave insufficient information at full-text screening, they were excluded. It is not clear how this was applied and whether this may have led to relevant studies being excluded. In addition, updates were made to the SLR during screening. While the impact of these factors is unclear, the ERG considers

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
		the screening methods to be broadly appropriate.
Data extraction	Appendix G.4	It is unclear why detailed data extractions were done in US studies while those from other countries underwent less extensive extraction. Despite this, the ERG considers the data extraction to be broadly appropriate.
QA of included studies	Appendix G.4	QA was completed using the Drummond checklist, as recommended by NICE. Therefore the ERG considers the QA to be appropriate.

Abbreviations: CS, Company Submission; ERG, Evidence Review Group; HRQoL, health-related quality of life; NICE, National Institute for Health and Care Excellence; QA, quality assessment

Table 35. Summary of ERG'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	ERG assessment of robustness of methods
Searches	Appendix H.2.	The ERG noted the following limitations: a recognised filter for identifying health-related quality of life studies was not used; database searches applied limits that excluded conference abstracts from search results, however, hand searching and database searches of known conference proceedings may have mitigated this issue. Missing search strategies were provided in response to clarification question B1.
Inclusion criteria	Appendix H.3.	Studies with <50 patients of interest are excluded though it is unclear why this is the case and whether this may have excluded relevant studies. Besides this, the ERG considers the inclusion criteria to be broadly appropriate to capture HRQoL studies relevant to UC.
Screening	Appendix H.3.	Title and abstract screening was conducted by two reviewers, with a third available to resolve discrepancies. The same procedure was followed at full-text screening. Where studies gave insufficient information at full-text

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
		screening, they were excluded. It is not clear how this was applied and whether this may have led to relevant studies being excluded. In addition, updates were made to the SLR during screening. While the impact of these factors is unclear, the ERG considers the screening methods to be broadly appropriate.
Data extraction	Appendix H.4.	Data from all studies was extracted in detail regardless of geographic region. The ERG considers the methods to be appropriate to have extracted all relevant data.
QA of included studies	Appendix H.4.	Methods of QA of the included studies are not described.

Abbreviations: CS, Company Submission; ERG, Evidence Review Group; HRQoL, health-related quality of life; QA, quality assessment; SLR, systematic literature review; UC, ulcerative colitis

Table 36. Summary of ERG'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	ERG assessment of robustness of methods
Searches	Appendix G.2.	Searches for cost and resource use studies and cost-effectiveness studies were conducted in a single SLR. The ERG noted the following limitation: database searches applied limits that excluded conference abstracts from search results. The ERG conducted additional searches of Ovid MEDLINE (reported in Appendix A) to retrieve studies that may have been missed. Following screening of additional search results the ERG was satisfied all relevant studies had been identified. Missing search strategies for bibliographic database searches were provided in response to clarification question B1.
Inclusion criteria	Appendix G.3.	The ERG notes that studies reporting 'primarily clinical outcomes' are excluded. It is not clear whether this may involve the exclusion of studies reporting both clinical and resource use and cost evidence. Despite this, the ERG considered the inclusion criteria to

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
		be broadly appropriate to encompass the resource use and cost evidence for all the relevant comparators to this appraisal.
Screening	Appendix G.3.	Title and abstract screening was conducted by two reviewers, with a third available to resolve discrepancies. The same procedure was followed at full-text screening. Where studies gave insufficient information at full-text screening, they were excluded. It is not clear how this was applied and whether this may have led to relevant studies being excluded. In addition, updates were made to the SLR during screening. While the impact of these factors is unclear, the ERG considers the screening methods to be broadly appropriate.
Data extraction	Appendix G.4.	It is unclear why only US and UK studies were extracted in detail, particularly for resource use data. Besides this, the data extraction is considered acceptable by the ERG.
QA of included studies	Appendix G.4.	The description of the methods for QA of resource use and costs is not completely clear, however the CS states that each study was compared to the NICE reference case, suggesting that QA will have been conducted appropriately.

Abbreviations: CS, Company Submission; ERG, Evidence Review Group; HRQoL, health-related quality of life; NICE, National Institute for Health and Care Excellence; QA, quality assessment; SLR, systematic literature review

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

4.2.1. NICE reference case checklist

The NICE reference case checklist with regards to aspects of the appraisal, as well as the ERG's comment on the company's submission, is summarised in Table 37.

Table 37: NICE reference case checklist

Attribute	Reference case	ERG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	QALYs were estimated for patients. Carer disutility was not included in the analysis.
Perspective on costs	NHS and PSS	NHS and PSS as appropriate.
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	The company submitted a cost utility analysis and presented both pairwise results and a fully incremental analysis.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	A lifetime horizon was used in the base case analysis. The ERG considered this to be appropriate.
Synthesis of evidence on health effects	Based on systematic review	Clinical data used in the economic model for both the treatment naïve and treatment experienced subgroups were primarily derived from the induction and maintenance NMAs conducted by the company. For the extended induction scenario analysis, clinical efficacy data were based on individual trial arms.
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.	QALYs were used as appropriate.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Utility values used in the base case were derived from published literature (Woehl et al.[29] and Arsenau et al.[30]). The ERG noted that QoL data were collected in the TrueNorth study using the EQ-5D-5L (which were cross-walked to EQ-5D-3L). These values were used in a company scenario analysis.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	The paper by Dolan (1997)[85] was used and was considered to be appropriate.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	There were no equity concerns.

Attribute	Reference case	ERG comment on company's submission
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	NHS reference costs, previous NICE appraisals and published literature were used to estimate costs and resource use.
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Costs and outcomes were discounted at 3.5% as appropriate.

Key: EQ-5D, European Quality of Life Five Dimension; HRQoL: health-related quality of life; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; PSS, Personal Social Services; QALY, quality-adjusted life year; QoL, quality of life; TA, technology appraisal

4.2.2. Model structure

The company submitted a de novo hybrid decision analytic model which consisted of two parts (active treatment and post active treatment). The ERG noted that the model is a Markov-based model that utilised tunnel states within the Markov cohort trace for the induction period. Based on a review of previous UC appraisals i.e. ustekinumab TA633[20] and tofacitinib TA547[21], models were characterised by the use of decision trees (for the induction period) and a Markov (for the maintenance/post active treatment periods). Conceptually, a decision tree can be modelled within a Markov model framework as a set of tunnel states. In addition, the company stated (p141 of the CS) that the use of tunnel states has the added benefit *'of allowing patients to enter the maintenance phase at any cycle, therefore enabling the variable length of induction periods between treatments'*. Furthermore, the use of tunnel states was stated to capture '*the effective decision tree at the end of the induction period'* thereby determining the initial health state distribution for the maintenance period. As such, the tunnel state approach is equivalent to a decision tree and, as it explicitly includes a time dimension, allows for more accurate modelling of the varying induction lengths of different treatments. For completeness, the company provided additional rationale for the use of tunnel states upon request from the ERG (see as B5 in the clarifaction document). The ERG considered the company's justifcation to be reasonable and that the approach did not appear to introduce any bias into the analysis.

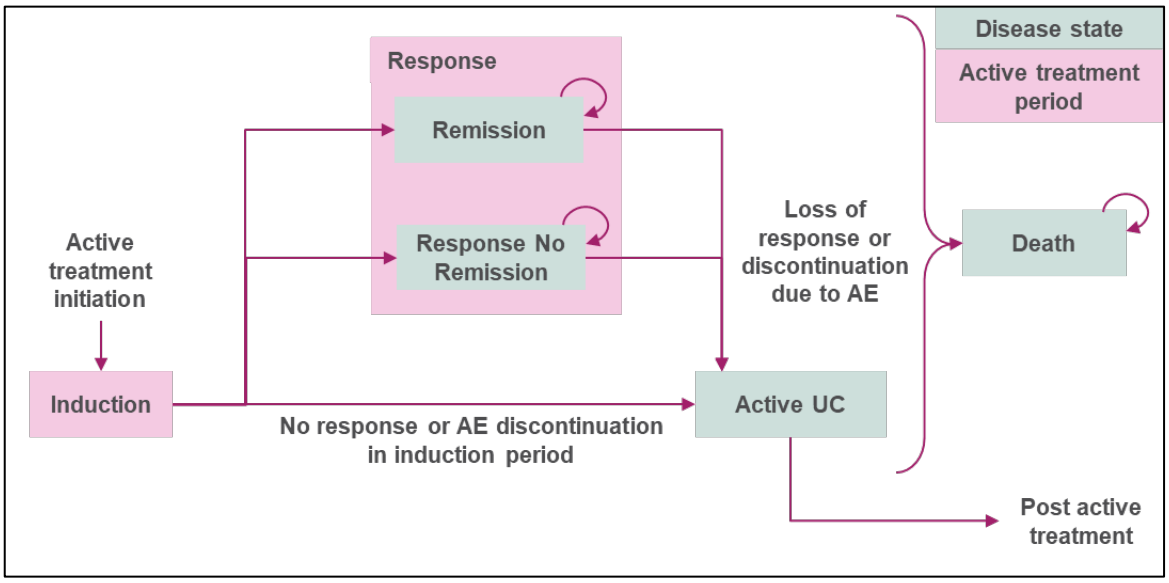
4.2.2.1. Active treatment period

The active treatment portion of the model consisted of both an induction and maintenance phase. All patients entered the model by initiating active treatment and progressed through a series of tunnel states (which reflected the specific induction period for each treatment). Once patients reached the final induction tunnel state, they were distributed into one of three health

states, 'Remission', 'Response (No remission)' and 'Active UC' (Figure 10). The probability of transitioning between key health states was derived from the NMAs outlined in Section 3.4.4 . See Section 4.2.6 for further discussion.

During induction patients could discontinue treatment due to serious adverse events or be absorbed by the 'Death' state. Patients distributed into the 'Remission', 'Response (No remission)' health states were assumed to remain in these states until they lose their initial response, discontinue due to adverse events or die. This is active treatment maintenance phase.

Figure 10: Active treatment model structure (without extended induction)



Abbreviation: AE, adverse events

4.2.2.2. Post active treatment phase

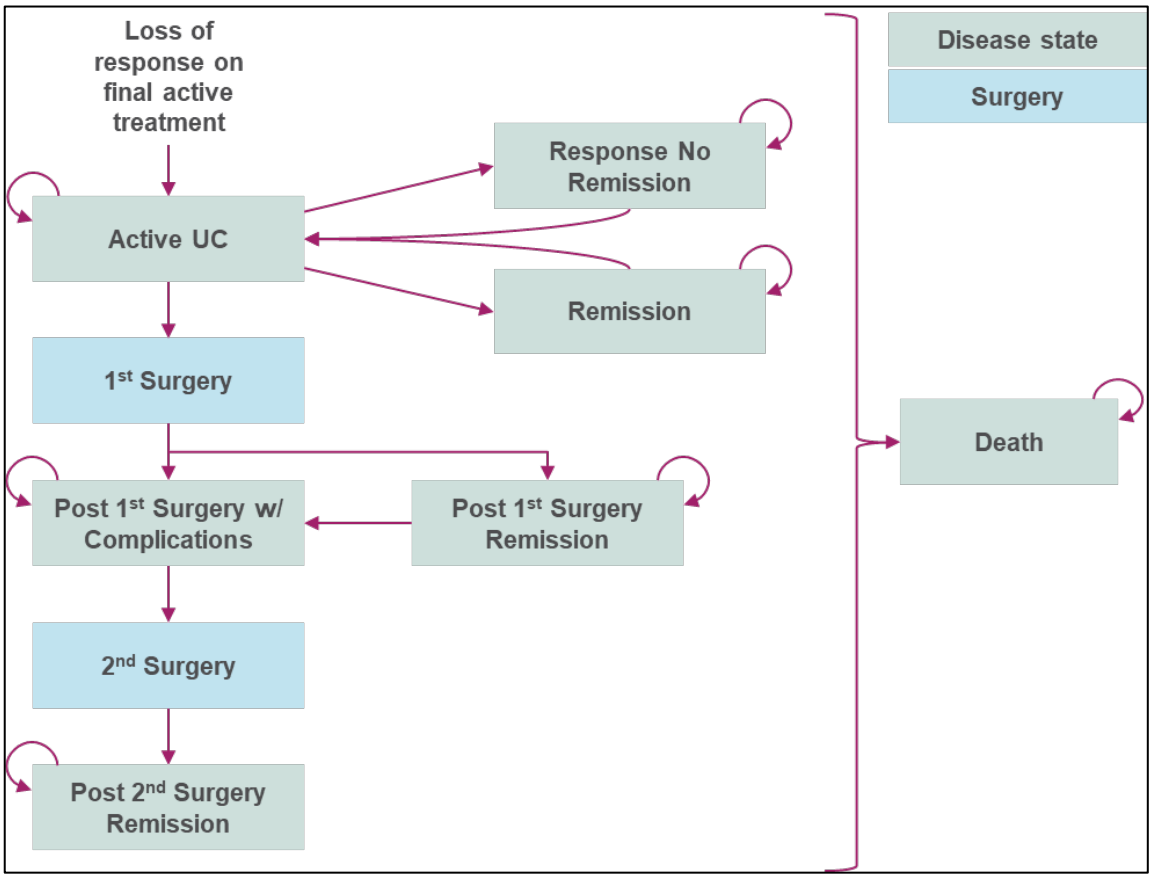
Patients that discontinued active treatment due to AEs, loss of response, or failure to achieve response entered the post active treatment component of the model and were assumed to initially enter the 'Active UC' health state. This part of the model consisted of 9 health states including 'Remission', 'Response (No remission)', 'Active UC', '1st Surgery', 'Post 1st Surgery Remission', 'Post 1st Surgery Complications', '2nd Surgery', 'Post 2nd Surgery Remission' and 'Death'**.** The ERG confirmed that these health states have been used and accepted in previous ustekinumab TA633[20] and are considered to accurately reflect the nature of the condition and key clinical events. In TA633,[20] the ERG had highlighted that a major limitation with the model structure was the non-inclusion of remission and response health states in the post active

treatment phase as not all patients follow an active form of the disease. This limitation has been addressed in the current model structure submitted by the company, as it considers remission and response health states following the 'Active UC' health state. In addition, the model allows for the spontaneous response and remission as well in the 'Active UC' health state.

The ERG noted that the modelled spontanous remission rate of 0.5% per model cycle (12% per year), was not based on clinical data, but was an arbitrary value chosen by the company to align with NICE committee preferences in TA633.[20] In TA633, the NICE committee stated that *'there is likely to be a small number of people who improve without treatment'* and generally preferred a low spontaneous remission rate, closer to the company's original modelled estimate of 0%. In order to validate the spontaneous remission rate, the ERG sought clinical input. Based on clinical expert responses, spontaneous remission was considered plausible for patients who no longer received active treatment. The rate of spontaneous remission in clinical practice, was somewhat variable i.e. between 5% to 30% per flare of active disease. In order to explore uncertainty, the company conducted scenario analyses in which 0% and 1% rates of spontaneous response were tested. The ERG has additionally conducted a scenario analysis which used a higher rate of spontaneous response compared to the company's base case estimate (reflective of 0.75% per model cycle). It should be noted that this analysis may lead to implausibly high spontaenous remission rates over the modelled time horizon and therefore should be interpreted with caution. See Section 6.1.5 for results.

Finally, it should be noted that in the post active model component, patients were assumed to receive best supportive care, comprising components of conventional therapy (See Section 4.2.8.3 for the list of CvT treatments provided).

Figure 11. Post active treatment model structure



Abbreviation: UC, ulcerative colitis

4.2.2.3. Subsequent treatments/Treatment sequencing

The base case analysis did not include subsequent treatments. Although the company acknowledged that it is possible for patients in the TNFi-naïve and TNFi-experienced subgroups to receive subsequent treatment (based on clinical expert feedback), subsequent treatments were not considered in the base case due to the lack of robust efficacy data available and uncertainty surrounding the treatments patients are likely to receive. The ERG noted that the exclusion of subsequent treatments is in line with TA633[20] (although the impact of including subsequent treatment was tested in a scenario analysis by the ERG). In TA547,[21] the base case did not consider subsequent treatments, however the model was flexible enough to allow the ERG to conduct a scenario analyses in the TNFi-naïve subgroup whereby various treatment sequencing strategies were explored including within class switching amongst TNFis as well as step up and step down approaches (outside class). The ERG noted in this appraisal that the results were not especially sensitive to this analysis. Overall, the ERG considered the impact of

treatment sequencing to have a moderate impact on costs, but minor impact on QALYs gained, as outlined in a published study by Wu et al. (2018)[86], which assessed the cost effectiveness of TNFis for the treatment of UC (from a UK and Chinese perspective). In this study, which compared 14 different treatment sequencing strategies, the total QALY gained for each strategy were largely similar (ranging between 10.49 to 12.37).

Within the current appraisal for ozanimod, the company provided limited scenario analysis which allowed for subsequent treatment usage in the TNFi-naïve subgroup (this was not conducted for the TNF-experienced subgroup as the company stated that there was a lack of available data to inform efficacy and clinical opinion to the company noted that treatments provided after failure on multiple biologics were likely to be patient dependent and variable). The scenario analysis allowed for the modelling of either vedolizumab or ustekinumab as plausible subsequent treatment options (after having received ozanimod or TNFis first line). The ERG noted that this did not have a meaningful impact on base case cost-effectiveness results. For completeness, the ERG considered undertaking additional scenario analyses using alternative subsequent treatments/sequencing options, however the model lacked the flexibility to conduct this. The ERG acknowledged that the exclusion of subsequent treatments from the base case analysis is an area of structural uncertainty, as it was not possible to adequately test via scenario analysis. Given the small differences in costs between treatments, incorporating treatment sequencing may have considerable impact on the base case results.

4.2.2.4. Extended induction

The ERG noted that extended induction was not considered as part of the company's base case analysis on the basis that it is not standard clinical practice in the UK for all treatments and further noted limitations associated with using trial data to inform the patient distributions into the health states (see p.154 of the CS). However, the company did provide a scenario analysis which explored the impact of including extended induction based on the SmPC for each treatment (See Section 5.2.3). Clinical input to the ERG, confirmed that extended induction does occur in clinical practice, albeit it is treatment dependent and therefore highly variable, i.e. if patients do not respond to treatment by the end of induction as per the SmPC, they can receive further treatment up to a total of 12 to 16 weeks before switching. The ERG noted that a delayed response phase was included in the TA633[20] model and was generally accepted, however, the ERG broadly agreed with the company's approach to exclude extended induction in the current appraisal, as the assumption that all treatments will require extended induction is highly unlikely. Furthermore, the ERG noted that for the scenario analysis which included

extended induction, patient distribution into the Remission, Response (No remission) and Active UC health states were derived directly from individual clinical trials (as opposed to the NMAs).

4.2.3. Population

For the economic analysis, the company submitted two distinct subgroup analyses for TNFi-naïve and TNFi- experienced patients, which was considered to be broadly consistent with the decision problem in Table 7. Modelled patient characteristics were taken from the TrueNorth[27,28] study for both subgroups, which was a multicentered, international study. Clinical expert opinion to the ERG confirmed that mean weight and mean age were generally representative of patients in the UK (albeit there may be more of an equal distribution of male and female patients in both subgroups). Patient characteristics were also found to be broadly similar to those used in previous NICE TAs for moderately to severe UC including ustekinumab TA633[20] and tofacitinib TA547.[21] Overall, the ERG considered the modelled patient characteristics presented in Table 38 to be appropriate.

Table 38: Modelled patient characteristics

Characteristic	Population	
	TNFi-naïve	TFNi-experienced
Mean weight, kg	[redacted]	[redacted]
Proportion of female, %	[redacted]	[redacted]
Mean age, years	[redacted]	[redacted]

Abbreviations: TNFi, tumour necrosis factor inhibitor

4.2.4. Interventions and comparators

In the TNFi-naïve subgroup the primary comparators were TNFis including infliximab (biosimilar), adalimumab (biosimilar), golimumab and the biologic treatment vedolizumab. The company stated that ustekinumab was not considered as a relevant comparator within this subgroup, given that NICE guidance states that ustekinumab is restricted to patients who have failed CvT or a biologic AND who have failed a TNFi or for whom a TNFi cannot be tolerated or is unsuitable. The ERG acknowledged that this restriction is in place for ustekinumab and that the company's rationale to exclude ustekinumab from this subgroup seemed reasonable. Furthermore, clinical opinion to the ERG noted that ustekinumab is not used as a first line treatment.

In the TNFi-experienced subgroup, comparators were ustekinumab and vedolizumab. The company stated that TNFis were not appropriate comparators for this subgroup, as TNFi switching is no longer routine clinical practice. Based on clinical expert opinion to the ERG, treatment switching amongst TNFis may occur in UK clinical practice. One expert stated that further TNFis are used, though the choice of subsequent TNFi is dependent largely on why patients did not respond to initial treatment. For example if the patient failed due to immunogenicity, a second TNFi would be tried. A second clinical expert noted that switching is uncommon, however patients could switch to adalimumab if they do not respond to infliximab. The ERG also noted that adalimumab was included as a relevant comparator in TA633.[20] The ERG considered undertaking a scenario analysis which included adalimumab as a relevant comparator in the TNFi-experienced subgroup, however the model does not allow a flexible selection of comparators interchangeably between the subgroups and so it was not possible conduct this analysis.

The ERG noted that the company excluded tofacitinib as a comparator from both the TNFi-naïve and TNFi- experienced subgroups stating that there are significant safety concerns associated with treatment. The ERG did not consider the company's rationale to be sufficient, given that tofacitinib (TA547)[21] has been recommended for use by NICE as a viable treatment option. Furthermore, based on clinical opinion to the ERG, tofacitinib safety concerns were considered to be clinically managed at an individual patient level. Clinician input also confirmed that tofacitinib is used in UK clinical practice for treating TNFi-naive patients and treatment experienced patients with moderately to severe UC, as it is a fast-acting treatment and reduces the need for corticosteroid use. As such, the ERG subsequently asked the company to provide a revised analysis including tofacitinib as a relevant comparator within both subgroups. The company did not provide this analysis, stating that clinical opinion to the company confirmed that although there may be use of tofacitinib, it is not considered routine practice (refer to B9 of the company's clarification response). In contrast, the ERG noted that a recent multicentre real-world cohort study conducted in the UK by Honap et al. (2020)[22] has found that adverse events requiring curtailment of the treatment were uncommon with no occurrence of thromboembolic events and further concluded that tofacitinib was well-tolerated. The company also stated that in TA633,[20] the committee agreed that the exclusion of tofacitinib was appropriate. Whilst the ERG noted this observation, it should be acknowledged that in TA633,[20] the company had included tofacitinib as a relevant comparator within their model. The ERG were unable to alter the company's model to include a cost effectiveness comparison with tofacitinib (due to inflexibility),

however as an exploratory analysis, a cost comparison was undertaken to determine the comparative difference in drug costs, monitoring costs and adverse event costs between treatments (see 6.1 and 6.1.5 for results).

As a minor point, the ERG acknowledged that in both tofacitinib TA547 and ustekinumab TA633,[20,21] conventional therapy was included as a comparator (in both the biologic naïve and biologic experienced patient populations). However, within this current appraisal for ozanimod, the company did not consider conventional therapy as an active comparator, on the basis that patients were specifically those who have not responded to conventional therapy. The ERG sought clinical expert opinion to comment on the appropriateness of this assumption. Based on input to the ERG, it was considered reasonable to exclude conventional therapy from the analysis on the basis that patients have already failed conventional therapy.

4.2.5. Perspective, time horizon and discounting

The ERG did not identify concerns surrounding discounting. Costs and benefits were discounted at 3.5% which reflects NICE guidance. Furthermore, costs and outcomes were estimated from an NHS and PSS perspective, as appropriate.

The company used a lifetime horizon (58 and 60 years in the TNFi-naïve and TNFi-experienced populations, respectively) in the base case analysis and justified this on the basis that a lifetime horizon has been used in previous UC appraisals including ustekinumab TA633[20] and toficitinib TA547.[21] The ERG noted that using a lifetime horizon is consistent with both TA633[20] and TA547,[21] however in older UC appraisals i.e. vedolizumab TA342[87] and infliximab, adalimumab and golimumab TA329,[88] shorter time horizons (10 years) have been used. Overall, the ERG considered a lifetime horizon to be appropriate as UC is chronic condition characterised by remission and loss of response, thereby affecting patients over the duration of their lifetime.

The ERG noted that a two-week cycle length was used in the model. The company justified this on the basis that it captured the variety of treatment regimens. This is consistent with the cycle length used in ustekinumab TA633,[20] however an 8 week cycle length has been used previously in tofacitinib TA547[21] and vedolizumab TA342.[87] Clinical opinion to the ERG was somewhat mixed regarding the appropriateness of the modelled cycle length, noting that 8 weeks is broadly reasonable for assessing response to treatment, however 2 weeks is also used (particularly with respect to tofacitiib). The ERG did not conduct a scenario analysis using an 8 week cycle length, on the basis that this model parameter was not programmable in the

company's model.Overall, the ERG considered the higher resolution of the 2 week versus 8 week cycle, to allow for greater flexibility within the model, and was therefore reasonable.

Finally, the ERG noted that the model did not incorporate a half-cycle correction. The company justified this on the basis that the model uses a short two-week cycle length, and that a half-cycle correction was not applied in TA547 despite the submitted model having an eight-week cycle length. However, given that TA633[20] included a half-cycle correction (and used a two-week cycle length), the ERG asked the company to include an option in the model that allowed for a half-cycle correction. Based on clarification response B15, including half-cycle correction did not have a meaningful impact on results. Overall, the ERG considered that the company's decision to exclude half-cycle correction did not meaningfully impact results.

4.2.6. Treatment effectiveness and extrapolation

4.2.6.1. Induction period and maintenance period transition probabilities

The proportion of patients achieving 'Remission' and 'Response (No remission)' at the end of the induction phase (for both subgroups) was taken from the NMAs, discussed in Section 3.4.4. The company stated (CS, Document B, p153), that mean absolute probabilities were based on NMA outputs including baseline anchor, response effect, remission effect and standardised mean difference versus baseline for a given treatment in the induction period (Table 39). For the maintenance phase the probability of achieving sustained remission and sustained response were estimated based on the maintenance NMA (See Table 40 below for the mean absolute values).

Table 39: Clinical efficacy at the end of the induction period

Treatment	Induction length (weeks)	Remission	Response (No remission)	No response (Active UC)
TNFi-naïve				
Ozanimod	10	[redacted]	[redacted]	[redacted]
Golimumab	6	[redacted]	[redacted]	[redacted]
Infliximab (biosimilar)	8	[redacted]	[redacted]	[redacted]
Adalimumab (biosimilar)	8	[redacted]	[redacted]	[redacted]
Vedolizumab	6	[redacted]	[redacted]	[redacted]
TNFi-experienced				
Ozanimod	10	[redacted]	[redacted]	[redacted]
Ustekinumab	8	[redacted]	[redacted]	[redacted]
Vedolizumab	6	[redacted]	[redacted]	[redacted]

Treatment	Induction length (weeks)	Remission	Response (No remission)	No response (Active UC)
BSC	10	[redacted]	[redacted]	[redacted]

Abbreviations: BSC, best supportive care; TNFi, tumour necrosis factor inhibitor; UC, ulcerative colitis

Table 40: Clinical efficacy in the maintenance period

Treatment	Sustained remission	Sustained response
TNFi-naïve		
Ozanimod	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Infliximab/biosimilar	[redacted]	[redacted]
Adalimumab/biosimilar	[redacted]	[redacted]
Vedolizumab	[redacted]	[redacted]
Vedolizumab (IV)	[redacted]	[redacted]
Vedolizumab (SC)	[redacted]	[redacted]
TNFi-experienced		
Ozanimod	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]
Vedolizumab	[redacted]	[redacted]
Vedolizumab (IV)	[redacted]	[redacted]
Vedolizumab (SC)	[redacted]	[redacted]
BSC	[redacted]	[redacted]

Abbreviations: BSC, best supportive care; IV, intravenous; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

4.2.6.2. Loss of response

In line with TA633,[20] the model assumed that a proportion of patients lose response over time. This was assumed to be a constant loss of response that extended beyond the trial duration. In TA633,[20] the ERG accepted this assumption given the absence of longer-term follow-up data outlining how absolute or relative loss of response changes over time. Loss of response was estimated using an equation provided in Appendix N, provided by the company.

Table 41: Transition probabilities for loss of response

Treatment	Duration of maintenance period	Loss of response	Loss of response (No remission)
TNFi-naïve			
Ozanimod	42	[redacted]	[redacted]
Golimumab	54	[redacted]	[redacted]
Infliximab/biosimilar	46	[redacted]	[redacted]
Adalimumab/biosimilar	44	[redacted]	[redacted]
Vedolizumab	46	[redacted]	[redacted]
TNFi-experienced			
Ozanimod	42	[redacted]	[redacted]
Ustekinumab	44	[redacted]	[redacted]
Vedolizumab	46	[redacted]	[redacted]
BSC	42	[redacted]	[redacted]

Abbreviations: BSC, best supportive care; TNFi, tumour necrosis factor inhibitor

4.2.6.3. Post active treatment transitions for BSC

In the company model (post active treatment phase), the modelled cohort progress to the 'Active UC' health state where some may continue to receive best supportive care, comprising components of conventional therapy, and still continue to experience 'Remission' or 'Response No Remission' since UC is a relapse-remitting disease. The company submission stated that the transitions among the 'Active UC', 'Remission', and 'Response No Remission' health states for BSC were informed using the 'Loss of Response' and 'Loss of Response No Remission' derived from the pooled placebo arm (from the RCTs included in the NMA) estimates across the subgroups for sustained remission and sustained response.

However, the ERG noted that the loss of overall response (including remission) was used to inform remission transition probabilities for BSC i.e., remission equals overall response (through loss of response) which differed from the approach taken for active treatments (where remission state membership was derived as: overall response (through loss of response) – response no remission (through loss of response no remission). The ERG considered that the remission probabilities for BSC could be calculated through 'loss of remission' (i.e. calculated directly from the sustained remission estimates), as deriving it from overall response slightly overestimates the remission probability. This approach has been incorporated into the ERG base case (see Section 6.3 for results).

Additionally, the ERG noted that the loss of response and loss of response (no remission) estimates were noticeably different between the non-biologic failure and biologic failure subgroups in TA633[20] (Table 43 and Table 44 TA633 committee papers), and that the company had used the data for the TNFi-experienced group in both populations, given patients receiving BSC in the model (regardless of the population selected) do so in the post-active treatment setting, and thus have failed at least one active treatment by definition. The ERG considered that the approach used in TA633 was more appropriate, on the basis that available subgroup data were used to inform loss of response. As a result of using the alternative baseline placebo risk estimates in the ERG's revised base case, loss of response and loss of response (no remission) were based on TNFi-naïve and TNFi-experienced subgroup estimates, as appropriate.

4.2.6.4. Uncertainty surrounding clinical effectiveness estimates

In the base case analysis the company opted to use a FE model in both TNFi-naïve and TNFi-experienced subgroups for the maintenance phase NMAs, as well as the TNFi-experienced subgroup for the induction phase. The ERG acknowledged the company's rationale for using the FE model for the maintenance phase, i.e. that the fit was reasonable and that the RE model did not converge; it also noted the highly uncertain posterior SD in the induction phase NMA for TNFi-experienced participants. However, due to the high degree of heterogeneity amongst the studies included in the NMA, the ERG considered that FE models were inappropriate. As noted in Section 3.5.3.2 the ERG was unable to produce RE models with sufficient convergence (without using an informative prior distribution) and were therefore unable to use a RE model as part of its preferred base case. Furthermore, as noted in Section 3.4.2.4, the ERG identified concerns surrounding the baseline estimation of placebo risk in the NMA. In order to generate estimates of clinical effectiveness for its base case, the ERG used the placebo arm values from individual trials included in the NMA that were more generalisable to the UK context, the results of which are reported in Section 3.5.3.1. The ERG noted that re-running the NMA using the alternative means of estimating baseline placebo risk resulted in fewer total QALYs for all treatments. The ERG used these estimates to inform the ERG's preferred modelling assumptions, as described in Section 6.3.

With respect to treatment discontinuation due to serious adverse events, modelled per cycle probabilities for ozanimod were taken from the TRUENORTH[27,28] study (see Table 47 on p157 of the CS). For vedolizumab, golimumab, infliximab and adalimumab, treatment discontinuation rates were derived from GEMINI 1,[61] VISIBLE 1,[73] PURSUIT M,[66] Kobayshi et al. (2016)[74] and

ULTRA 1[65] and 2[43] respectively. The ERG noted that the discontinuation rate for ozanimod was considerably lower than comparator treatments, particularly in the maintenance phase. The ERG were unclear whether these rates would be reflective of clinical practice and therefore conducted a scenario analysis using an alternative treatment discontinuation rate for ozanimod. See Section 6.1.

4.2.6.5. Validation of model outcomes

In order to assess the validity of the company's base case estimation of QALYs, the ERG reviewed previous economic models (which predominantly considered lifetime horizons), including ERG analyses for previous NICE technology appraisals in UC and other relevant published literature. Though the ERG understands and acknowledges the differences in terms of model structure and methodology across these economic models, the ERG's view is that such a comparison would still serve as a useful means of cross-validating model outcomes (especially the QALYs), irrespective of the differences.

The ERG compared the model generated QALYs of ozanimod (as outlined in the company submission), to that of NICE TA329,[88] TA342[87] and Wu et al.(2018)[86], as summarised in Table 42. It should be noted that the QALY results from tofacitinib (TA547)[21] and ustekinumab (TA633)[20] were not available, as they were commercial in confidence. However, it was noted that in TA633 ERG's validation highlighted that the ustekinumab company model QALY estimates were lower than those from other lifetime models. In general, ERG noted that the QALY estimates from ozanimod company model (though slightly lower) were mostly comparable to the previous appraisals and publications. The difference in the total QALYs between the company's model and previous appraisals could be due to the consideration of additional response health states for the BSC in the post-active treatment phase (which is a change in this model structure compared to the ustekinumab appraisal (TA633)[20]).

Table 42: Comparison of modelled QALYs

Study name (time horizon)	QALYs	
	TNFi-naïve	TNFi-experienced
Ozanimod company model (lifetime)	Oza: [redacted]	Oza: [redacted]
	Ada: [redacted]	Ved: [redacted]
	Inf: [redacted]	Ust: [redacted]
	Ved: [redacted]	
	Gol: [redacted]	
TA342 (lifetime, ERG preferred base case)	Ada:12.39	Ved:11.84
	Ved:12.37	CvT:11.28
	Gol:12.05	Surgery:14.60
	Inf: 12.01	
	Surgery:14.60	
	CvT: 11.73	
TA329 (lifetime, AG model)	**Moderate to severe UC who failed at least 1 prior therapy**	
	Ada:10.82	
	Inf: 10.81	
	Gol: 10.63	
	CvT: 10.47	
Wu et al. (lifetime)	**Moderate to severe UC**	
	CvT:10.49	
	Ved→CvT: 11.48	
	Tof→CvT: 11.51	
	Inf→CvT: 10.87	
	Gol→CvT:10.89	
	Ada→CvT: 10.71	
	Ved→Tof→CvT: 12.37	
	Inf→Tof→CvT:11.81	
	Gol→Tof→CvT:11.83	
	Ada→Tof→CvT:11.67	
	Tof→Ved→CvT:12.37	
	Tof→Inf→CvT:11.84	
	Tof→Gol→CvT:11.86	

Abbreviations: Ada, adalimumab; AG, Assessment Group; CvT, conventional therapy; ERG, Evidence Review Group; Gol, golimumab; Inf, infliximab; Oza, ozanimod; QALY, quality-adjusted life year; TA, technology appraisal; TNFi, tumour necrosis factor inhibitior; Tof, tofacitinib; UC, ulcerative colitis; Ust, ustekinumab; Ved, vedolizumab

4.2.6.6. Probability of surgery and complications

A proportion of patients in the Active UC health state (of the post-active treatment phase) were assumed to require surgery each model cycle. In the company's base case analysis, annual

probabilities were converted into per cycle probabilities (see Table 48 and 49 on p159 of the CS). The company derived the probability of 1st surgery from a published study by Misra et al. (2016),[89] which was a large retrospective analysis of UK Hospital Episode Statistics (HES) for UC, based on 71,966 patients. The ERG noted this study has been used previously in TA633[20] and TA547[21] to estimate the probability of 1st surgery. Based on a review of TA633 and TA547,[20,21] the ERG identified several other alternative published studies which could be used to inform the probability of 1st surgery, including a UK study by Chhaya et al. (2015)[90] However, in both appraisals, the ERG agreed with the company's selection of Misra et al. (2016)[89] in the base case.

The proportion of patients who had complications after having a 1st surgery (1st surgery complications) was derived from a national clinical audit of inpatient care (for adults with UC).[91] This was estimated to be 33.5% (based on a weighted average of 32% elective and 35% non-elective surgeries). The proportion of patients with complications following post-surgery remission (1st surgery remission) was derived from a published UK study by Segal et al. (2018)[92] which assessed long term outcomes of prepouch ileitis in 31 patients. The ERG noted that both of these sources were used previously in TA633.[20] In TA547,[21] the proportion of patients with complications following post-surgery remission (1st surgery remission) was based on a Belgian study by Ferrante et al. (2008).[93] The company and ERG undertook scenario analyses using other published literature sources, however results were not sensitive to this.

In the current appraisal for ozanimod, the company made several simplifying assumptions with respect to modelled surgery rates, including the assumption that the probability of patients requiring a 2nd surgery is the same as the 1st surgery. The ERG noted that this assumption was consistent with TA633[20] and was generally accepted by the ERG. Overall, the ERG considered that the modelled surgery rates were not key drivers of cost effectiveness in this appraisal due to the small proportion of patients who transitioned into the surgical health states.

Table 43: Modelled per cycle probability of surgery

From	To	Per cycle probability
Active UC (post active treatments)	1st Surgery	0.00018
1st Surgery	1st Surgery Complications	0.33500
1st Surgery	1st Surgery Remission	0.66500
1st Surgery Remission	1st Surgery Remission	0.99876
1st Surgery Remission	1st Surgery Complications	0.00124
1st Surgery Complications	2nd Surgery	0.00018

Abbreviation: UC, ulcerative colitis

4.2.6.7. Mortality

The model included all-cause mortality i.e. patients could die in any health state, based on age and gender adjusted background mortality (using UK lifetables). The company assumed that UC is not associated with an additional mortality risk, therefore a standardised mortality ratio of 1 was used for key UC health states (Table 44). Based on clinician input to the ERG, UC was not considered to result in excess mortality. As such the company's assumption appeared reasonable. As per TA633[20] and TA547,[21] the company used a study by Jess et al. (2007)[94] to estimate a 30% mortality risk associated with surgery. This additional mortality risk was also assumed to apply to second surgery. Overall, the ERG considered the company's modelled mortality estimates to be acceptable. Mortality was not considered to be a driver of cost effectiveness, as there is no difference in LY gains between treatments.

Table 44: Standardised mortality ratio by health state

Health state	Standardised Mortality Ratio
Remission	1.0
Response (No remission)	1.0
Active UC	1.0
1st surgery	1.3
1st surgery remission	1.0
1st surgery complications	1.0
2nd surgery	1.3
2nd surgery remission	1.0

Abbreviation: UC, ulcerative colitis

4.2.7. Health-related quality of life

4.2.7.1. Health state utility values

The company conducted a systematic literature review to identify plausible health state utility values for inclusion in the model. On p162 of the CS, the company stated that 27 unique studies which reported HRQoL were identified in patients with moderately to severely active UC and 9 HTA appraisals. The ERG noted that the company did not use a recognised filter for HRQoL studies and restricted the bibliographic database searches to exclude conference abstracts. The company stated that these were separately hand-searched and excluded to avoid double counting. Additional searches in Ovid MEDLINE (Appendix A) conducted by the ERG identified other UK and non UK studies which reported HRQoL data in patients with moderate to severe

UC. However, after screening these studies, the ERG considered the company's search and study identification to be broadly reasonable. In the base case analysis, the company opted to use health state values from ustekinumab TA633.[20] The ERG noted that these values were derived from published literature sources i.e. Woehl et al. (2008)[29] was selected for the remission, response (no remission), active UC and post 1st surgery remission health states. Arseneau et al. (2006)[30] was used for 1st surgery and post 1st surgery complications, as these values were not reported in Woehl et al.[29] The ERG noted that the utility value for 1st surgery was based on a weighted average of the utilities for ileostomy (0.57) and J-Pouch surgery (or ileal pouch-anal anastomosis) (0.68), with weights of 60% and 40% respectively.

The company assumed that the utility associated with the second surgery and post-second surgery remission were equal to the first surgery (due to the lack of published data surrounding second surgery HRQoL values). The ERG found this assumption to be consistent with ustekinumab TA633,[20] where the ERG accepted these values (see discussion below). Based on a review of tofacitinib TA547[21] and vedolizumab TA342,[87] second surgery was not considered, therefore these models did not include second surgery utilities. Health state utilities were adjusted appropriately for age and gender using a published equation by Ara and Brazier (2010)[95] to account for the natural decline in QoL as a result of aging. See Table for the health state utility values used in the company's base case.

The ERG noted Woehl et al. (2008)[29] to be a UK study, which collected HRQoL data on 180 patients with active UC in the UK (using the EQ-5D), whilst Arseneau et al. (2006)[30] collected HRQoL data on 48 US patients from the University of Virginia Health System and Duke University Medical Centre. The ERG considered Woehl et al. (2008)[29] to be generalisable and broadly appropriate (albeit the study is somewhat dated). However, the use of Arseneau et al. (2006)[30] raises some generalisability concerns given the small sample size and participant characteristics i.e. a mean age of 45 years (thereby likely underrepresenting the second disease peak), overwhelmingly Caucasian (96%) and predominantly male (62%), a mean disease duration of 9.8 years (thereby likely underrepresenting the first disease peak) and very little participants who had undergone colectomy (21%). Despite these limitations, Arseneau et al has been considered a reasonable source for use in TA633 (see commentary below).

Table 45: Modelled health state utility values

Health state	Utility	Source
Remission	0.87	Woehl et al. (2008)[29]
Response (no remission)	0.76	Woehl et al. (2008)[29]
Active UC	0.41	Woehl et al. (2008)[29]
1st Surgery	0.61	Arseneau et al. (2006)[30]
Post 1st surgery remission	0.72	Woehl et al. (2008)[29]
Post 1st surgery complications	0.34	Arseneau et al. (2006)[30]
2nd Surgery	0.61	Assumption (as per TA633)[20]
Post 2nd surgery remission	0.72	Assumptions (as per TA633)[20]

Abbreviation: UC, ulcerative colitis

4.2.7.2. Utility value sources used in previous UC appraisals

Ustekinumab TA633 (2020)

The ERG noted that the use of Woehl et al. (2008)[29] and Arseneau et al. (2006)[30] was in line with the recent UC appraisal for ustekinumab TA633,[20] where the ERG considered the values reported in Woehl et al. (2008)[29] and Arseneau et al. (2006)[30] to be 'generally reasonable'. In TA633,[20] the ERG further agreed with the company's decision to not use direct HRQoL data from the pivotal study UNIFI,[52] as they were 'inconsistent with the values used in previous NICE appraisals for UC'. It was not possible to validate this statement as UNIFI utility values were marked as CIC in TA633.[20]

Tofacitinib TA547 (2018)

Based on a review of tofacitinib TA547,[21] Woehl et al. (2008)[29] was also used to derive health state utilities all health states. Trial based utilities from the OCTAVE studies were not considered appropriate for use in the base case due to the re-randomisation design and the lack of intermediate assessment of clinical response and remission between week 8 and week 52. The ERG considered Woehl et al. (2008)[29] to be the most appropriate source for base case utility parameters, and used values reported in Swinburn et al. (2012)[96] as a scenario analysis.

Vedolizumab TA342 (2015)

In the company's base case, health state utility values were derived from the pivotal study GEMINI I,[61] whereby QoL values for the Remission, Mild disease and Moderate to severe disease were estimated based on EQ-5D data. As utility data were not collected for the surgery

health states (post-surgery remission and post-surgery complications), the company used published literature from Punekar and Hawkins (2010)[97] (stated to be an epidemiology and costs study of CD). Although the ERG considered that using trial-based utilities in the base case was appropriate, it was noted that the value for post-surgical remission was lower than the value for moderate to severe UC. This was considered to lack plausibility, as it did not capture any benefit from surgery. The committee agreed that quality of life may be improved after 1st surgery (compared to having moderate to severe UC), although the magnitude of difference was uncertain. Two alternative sources were identified by the ERG and used in scenario analyses i.e. Woehl et al. (2008)[29] and Swinburn et al. (2012).[96] The ERG considered that values from Woehl et al. (2008)[29] (for patients who had surgery) were higher than those reported in Punekar and Hawkins et al.[97] The committee considered that Woehl et al. (2008)[29] and Swinburn et al. (2012)[96] had some important limitations i.e. small patient numbers and uncertainty regarding generalisability to UK practice. However, in the TNFi-experienced population, the committee expressed a preference for using both of these sources.

Infliximab, adalimumab and golimumab TA329 (2015)

The ERG considered Woehl et al. (2008)[29] and Swinburn et al. (2012)[96] to be the most useful sources of utility values in the model as they were UK-based, included a large number of patients (n=180 and n=230 respectively) and reported EQ-5D utility values for most modelled health states. In TA329,[88] the ERG selected utility values by Woehl et al. (2008)[29] to inform their base case analysis and used Swinburn et al. (2012)[96] as a scenario analysis.

4.2.7.3. The availability of direct HRQoL data from TRUENORTH

For ozanimod, quality of life data were available from patients directly in the TRUENORTH[27,28] study (from cohort 1 and cohort 2). Utility values were collected using the EQ-5D-5L at baseline, the end of induction (10 weeks) and the end of maintenance (52 weeks), see Table 46 and Table 47. Values were then cross-walked to EQ-5D-3L values using an appropriate published algorithm by Van Hout et al. (2012)[37] and UK value set from Dolan et al. (1997).[85] The ERG noted that health state values in the induction and maintenance phases were based on the weighted average across placebo arm and ozanimod arms i.e. utility values were health state dependent as opposed to treatment dependent. The company justified this approach on the basis that placebo and ozanimod values were broadly similar across health states.

The ERG noted that despite the availability of direct trial data, the company did not use QoL data from TRUENORTH[27,28] in the base case analysis due to limitations. The company outlined key limitations with the TRUENORTH utility data on p.162 of the CS. These included the following;

- In the induction phase, utility values for the Active UC health state (No response or remission at week 10) may be somewhat overestimated, as Active UC patients in TRUENORTH were receiving ozanimod. However, the modelled Active UC health state assumes that no further treatment would be received. Similarly, in the maintenance phase, patients continued to receive ozanimod in the Active UC health state (No response or remission at week 52). However, the modelled Active UC health state assumes that no further treatment would be received. The QoL of patients in the TRUENORTH Active UC health state was therefore not considered to be reflective modelled patients.
- Maintenance phase utility values were based on small patient numbers and are therefore subject to uncertainty (See Table 53 on p161 of the CS).
- Length of trial data considered too short and may not accurately capture the change in utility over time.
- QoL data for surgical health states were not captured.

In addition to the limitations highlighted by the company, the ERG further noted that in the maintenance phase, the utility value for placebo patients in the Response (no remission) at week 52 health state was higher than re-randomised patients in the same health state who received ozanimod in the induction phase and ozanimod in the maintenance phase ([redacted] versus [redacted] respectively). This result appeared somewhat counterintuitive, as the ERG expected that patients receiving ozanimod in both trial phases would have a higher QoL than those who initially received placebo during induction and then continued to receive placebo in the maintenance phase.

Table 46: TRUENORTH utility data (induction phase)

	Cohort 1		Cohort 2	Weighted average
Health state	**Ozanimod**	**Placebo**	**Ozanimod**	
Baseline (Active UC)	[redacted]	[redacted]	[redacted]	[redacted]
Remission at week 10	[redacted]	[redacted]	[redacted]	[redacted]
Response (No remission) at week 10	[redacted]	[redacted]	[redacted]	[redacted]
No response or remission at week 10 (Active UC)	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: UC, ulcerative colitis

Table 47: TRUENORTH utility data (maintenance phase)

	Re-randomised		Placebo	Weighted average
Health state	**Ozanimod/Placebo**	**Ozanimod/Ozanimod**		
Remission at Week 52	[redacted]	[redacted]	[redacted]	[redacted]
Response (No remission) at Week 52	[redacted]	[redacted]	[redacted]	[redacted]
No response or remission at Week 52 (Active UC)	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: UC, ulcerative colitis

Based on cross-validation, TRUENORTH[27,28] utility values for active UC (No response or remission) and Response (No remission) were considerably higher compared with published literature sources noted in Section 4.2.7.2. As such, using TRUENORTH values in the base case could potentially bias the analysis in favour of treatments with relatively poorer clinical effectiveness estimates, as a high percentage of patients transition to the active UC health state. Overall, the ERG agreed with the company that TRUENORTH utility values were subject to limitation and the use of these values in the base case may have introduced further uncertainty. For completeness, the company provided scenario analyses using alternative sources including TRUENORTH,[27] TA342[87] and TA547[21] (See Section 5.2.3 for results).

4.2.7.4. Disutility associated with adverse events

The base case analysis included disutility associated with serious infection only, which is consistent with previous appraisals including utekinumab TA633[20] and tofacitinib TA547.[21] As per TA633, the company elicited the utility decrement for serious infection from a published systematic review and economic evaluation by Stevenson et al. (2016),[98] which assessed the impact of treatments on rheumatoid arthritis. Modelled disutility associated with a serious adverse event was estimated to be 0.156 and symptoms were assumed to last for 4 weeks (28

days). The ERG noted that the duration of symptoms was considered reasonable and is in line with TA329.[88]

4.2.8. Resources and costs

4.2.8.1. Treatment acquisition costs

Medicine acquisition costs were included in the analysis for active treatments (with the exception of tofacitinib, which was excluded as a relevant comparator by the company) and concomitant treatments. Unit costs were derived from appropriate sources including the British National Formulary (BNF) and the Drugs and Pharmaceutical Electronic Market Information Tool (eMIT). Where more than one formulation of treatment with similar strength was available, the company selected the cheapest for use in the model (see Table 60, p.166 of the CS for the full list of treatments and unit costs). Overall, the ERG considered this approach to be reasonable and likely conservative. However, based on cross validation of the company's medicine acquisition costs with those reported in the BNF and TA633,[20] the ERG noted that the company's cost for adalimumab (£633.60 for 40 mg/0.8 mL) represented the solution for injection pre filled syringes. Another formulation was available for use i.e. solution for injection vials (£316.93 for 40 mg/0.8 mL), which is stated to be for hospital use only. However, based on clinical expert opinion to the ERG, pre-filled syringes are predominently used in practice.

Treatment costs were based on the cost per pack for each treatment and the dosing regimen, as outlined in the SPC for each treatment and/or clinical trials (see Tables 58, 59 and 60 in the CS for the treatment doses used in the model). For each subgroup (TNFi-naïve and TNFi-experienced), the company modelled treatment costs seperately for the induction phase, extended induction phase (scenario analysis only) and the maintenance phase in order to account for variance in dosing and duration. For the induction phase, costs were applied as one off costs at the start of induction. The ERG considered this approach to be reasonable. Clinical opinion to the ERG confirmed that the dosing used in both the induction and maintenance phases were broadly appropriate, albiet there may be some variation in clinical practice with respect to dosing frequency and escalation. The company has provided scenario analysis testing the impact of dose escalation on base case results (see Section 5.2.3). The ERG noted that the company estimated treatment costs on a per model cycle basis which was considered to be appropriate and in line with TA633.[20] However, given that the company's model allowed for the estimation of costs on a per treatment cycle basis, the ERG conducted a scenario analysis using this approach to determine the impact on the ICER. See Section 6.1.6.

Two formulations were available for vedolizumab (SC and IV). In the base case analysis the company assumed that the proportion of patients receiving SC and IV as maintenance therapy was 'evenly distributed' i.e. 50% received SC and 50% received IV. In order to validate the company's base case assumption, the ERG elicited clinical expert opinion. Based on clinical responses, 50% of patients receiving SC vedolizumab for maintenance therapy was considered to be largely reasonable and reflective of current clinical practice. However the proportion is likely to increase over time as one expert noted that not many patients are expected to remain on IV vedolizumab after one year. The company provided scenario analysis which varied the proportion of patients receiving SC vedolizumab. For completeness, the ERG conducted an additional scenario analysis to capture the opinions of clnical experts. See Section 6.1.5.

4.2.8.2. Dose escalation

In the base case analysis, the company assumed that 30% of patients would require dose escalation in the maintenance period i.e. 70% of patients would recive standard dosing (see Table 60 in the CS for). This assumption was applied to all treatments apart from vedolizumab SC and ozanimod. The company stated that dose escalation was not considered for these treatments, as per information contained in their respective SPCs. Based on clinical opinion to the ERG, dose escalation for biologics is common in clinical practice (with between 30%-40% of patients on infliximab receiving an escalated dose). Clinical experts stated that the proportion of patients requiring dose escalation would vary depending on treatment received, however the company's base case assumption of 30% may be somewhat low (with figures more aligned to 40%-50%).

Overall, the ERG considered the company's handling of dose escalation in the base case to be reasonable and in line with TA633,[20] whereby the ERG noted 30% dose escalation was reflective of data within a published study by Lindsay et al. (2017).[99] In order to test uncertainty surrounding dose escalation, the company provided scenario analyses which assumed 0% and 50% of patients required dose escalation. See Section 5.2.

4.2.8.3. Concomitant treatment and conventional therapy costs

Whilst on active treatment, patients received concomitant treatment with conventional therapy. Conventional therapy costs were also applied to patients entering the post active treatment phase of the model i.e. patients in the Active UC health state. As noted in Table below, the per cycle cost (per average patient) was estimated based on the weighted proportion of patients receiving each treatment. The proportion of patients receiving conventional therapy were taken

from previous UC appraisals (ustekinumab TA633[20] and vedolizumab TA342[87]). As stated by the company, patients receiving ozanimod were contraindicated to azathioprine, 6-mercaptopurine and methotrexate, therefore the costs of these treatments were not included in the ozanimod treatment arm. The ERG considered this to be reasonable.

Based on a review of TA547,[21] the ERG noted that alternative conventional therapy proportions were used i.e. these were taken from a national audit of the Royal College of Physicians (RCP) on IBD.[91] The ERG highlighted several concerns surrounding these proportions, namely that it was inappropriate to assume equal usage for the four aminosalicylic (5ASA) drugs, as most patients received mesalazine. As such, the ERG considered the proportions from TA633[20] and TA342[87] to be reasonable.

Table 48: Modelled conventional therapy treatments and proportions

Drug	Dose description	Patient usage (Ozanimod)	Patient usage (other treatments)
Balsalazide	1.5 g twice daily	0.0%	0.0%
Mesalazine	1.2 g/day (divided doses)	13.0%	13.0%
Olsalazine	500 mg twice daily	0.0%	0.0%
Sulfasalazine	500 mg 4 times daily	0.0%	0.0%
Prednisolone	20.0 mg/day for two weeks	36.0%	36.0%
Hydrocortisone	20 mg/day	0.0%	0.0%
Azathioprine	2.5 mg/kg/day	0.0%	39.0%
6-mercaptopurine	1.5 mg/kg/day	0.0%	15.0%
Methotrexate	17.5 mg/week	0.0%	9.0%
Budesonide	3.0 mg/3xday for eight weeks	1.0%	1.0%

4.2.8.4. Administration and monitoring costs

The company's base case analysis included administration costs for all IV treatments only. The cost per IV administration was £186.36 reflecting the average of a consultant and non consultant led face to face attendance. Costs were based on 2019/2020 NHS reference costs, which was considered to be an appropriate source. As per ustekinumab TA633,[20] the company assumed that there to be no cost involved with administering SC treatment, as most patients self administer. Based on clinical opinion to the ERG, most patients would be able to self administer SC treatment, however a small proportion (2%) may require assistance. The ERG noted that the inclusion of administration costs for such a small proportion of patients would not

have an meaningful impact on results and therefore considered the company's base case assumption to be acceptable.

One-off nurse training to teach patients how to self administer was assumed to be incurred by the manufacturer. Based on a review of TA633,[20] the ERG acknowledged that patient education and home delivery is provided by biologic manufacturers. Ozanimod was assumed to incur no administration cost as it is an oral treatment. The ERG considered this assumption to be reasonable.

With respect to monitoring costs, for ozanimod the company inluded the cost of a single ECG during induction which was estimated to be £61.80. This was included to reflect guidance within the SmPC for ozanimod. The cost was derived from 2019/20 NHS reference costs[100] as appropriate. The company assumed that all other monitoring requirements were similar between treatments (as per previous appraisals TA633,[20] TA547[21] and TA342[87]). Based on clinical expert opinion to the ERG, this was cosidered to be a reasonable assumption.

4.2.8.5. Health state costs

The company's analysis included disease management costs and health state specific costs, which applied to all treatments (see Table 49 below for a complete list). Resource use estimates were mostly derived from a published study by Tsai et al. (2008)[101], which estimated annual resource use for each modelled health state based input from a panel of UK gastroenterologists. The ERG noted Tsai et al to be a UK cost effectiveness study which assessed a scheduled maintenance treatment of infliximab in moderate to severe UC. Although the study was somewhat dated, Tsai et al. (2008)[101] has been used and accepted as an appropriate source for resource use estimates in previous UC appraisals including TA633[20]. The ERG noted that Tsai et al. (2008)[101] did not report resource use estimates for surgery health states, as such the company assumed that resource use for 1st surgery and 2nd surgery were the same resource use in the active UC health state. This assumption is in line with TA633.[20]

Table 49: Modelled health state resource use

Resource item	Unit cost	Remission	Remission (no response)	Active UC	1st/2nd Surgery	Post 1st/2nd surgery remission	Post 1st surgery complications
Outpatient							
Consultant visit	£183.43	2	4.5	6.5	6.5	1.5	1.75
Blood test	£1.81	3.25	3.90	6.5	6.5	1.5	3.25
Inpatient							
Emergency endoscopy	£814.46	0	0.25	0.75	0.75	0.50	0.13
Elective endoscopy	£330.51	0.20	0.50	2	2	1.25	0.65
Care without colectomy	£2,301.47	0	0	0.15	0.15	0	3.25
Stoma care (post-colectomy)	£541.75	-	-	-	1	-	-

Abbreviations: UC, ulcerative colitis

Unit costs were based on 2018/2019 NHS reference costs values as appropriate. The cost of stoma care costs (post colectomy), was based on TA547 [21], which appeared reasonable. The model included acute costs associated with 1st and 2nd surgeries. The ERG noted that the costs associated with 1st and 2nd surgeries were estimated to be £14,309.51 and £10,438.22 respectively. These costs were elicited from expert opinion to the company were broadly in line TA633,[20] which reported these to be £15,311 and £10,998 respectively.

Finally, in the economic model, resource use costs were estimated based on a per cycle basis (see Table 50). Overall, the ERG considered the company's handling of health state resource use to be reflective of prior UC appraisals and therefore appropriate.

Table 50: Total per cycle health state costs

Health state	Total cost per cycle
Remission	£16.82
Response (No remission)	£46.05
Active UC	£108.13
1st and 2nd Surgery	£128.90
Post 1st and 2nd surgery (Remission)	£42.09
Post 1st and 2nd Surgery (Complications)	£311.52

Abbreviations: UC, ulcerative colitis

5. COST-EFFECTIVENESS RESULTS

5.1. Company's base case cost-effectiveness results

The company presented both pairwise and fully incremental results for consideration. Pairwise results reported by the company are shown in Table 51 and Table 52 and the fully incremental results are provided in Table 53 and Table 54. As previously highlighted by the ERG, the company has been asked to provide a revised analysis which includes tofacitinib as a comparator in both subgroups. Unfortunately, as this analysis was not provided, the cost effectiveness of ozanimod compared to tofacitinib is unknown. Furthermore, comparator PAS (cPAS) discounts were not included in the company's base case results. These are provided in a confidential appendix.

5.1.1. TNFi-naïve subgroup results (pairwise)

Based on the pair-wise analysis, ozanimod was cost effective compared to adalimumab at a conventional willingness-to-pay threshold of £30,000, resulting in an ICER of £28,686, based on an incremental QALY gain of [redacted] and an incremental cost of [redacted]. Compared to infliximab, golimumab and vedolizumab, ozanimod resulted in southwest (SW) ICERs i.e. ozanimod was less costly and less effective. [redacted].

Table 51: Company (Pairwise) base case results: TNFi-naïve subgroup (Discounted)

	Total costs	LYG	Total QALYs	Incremental costs	Inc. LYG	Inc. QALYs	Cost/ QALY gained
Company deterministic base case							
Ozanimod	[redacted]	[redacted]	[redacted]	-	-	-	-
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£28,686
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£167,024*
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£71,023*
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£52,736*
Company probabilistic base case							
Ozanimod	[redacted]	NR	[redacted]	-	-	-	-
Adalimumab	[redacted]		[redacted]	[redacted]	-	[redacted]	£28,934
Infliximab	[redacted]		[redacted]	[redacted]	-	[redacted]	£155,144*
Golimumab	[redacted]		[redacted]	[redacted]	-	[redacted]	£71,945*

	Total costs	LYG	Total QALYs	Incremental costs	Inc. LYG	Inc. QALYs	Cost/ QALY gained
Vedolizumab	[redacted]	-	[redacted]	[redacted]	-	[redacted]	£63,862*

Abbreviations: LYG, life years gained; NR, not reported; QALY, quality-adjusted life year

Note: * ICER in SW quadrant

5.1.2. TNFi-experienced subgroup (pairwise)

Based on the pair-wise analysis provided by the company, ozanimod was considered less costly and less effective compared to vedolizumab, resulting in a SW ICER of £199,551 [redacted]. Compared to ustekinumab, ozanimod was dominant i.e. less costly and more effective. It should be noted that the probabilistic results presented below are based on the ERG's re-run of the PSA, as the company did not provide these values in the CS.

Table 52: Company (Pairwise) base case results: TNFi-experienced subgroup (Discounted)

	Total costs	LYG	Total QALYs	Incremental costs	Inc. LYG	Inc. QALYs	Cost/ QALY gained
Company deterministic base case							
Ozanimod	[redacted]	[redacted]	[redacted]	-		-	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£199,551*
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Ozanimod dominant
Company probabilistic base case							
Ozanimod	[redacted]	NR	[redacted]	-		-	-
Vedolizumab	[redacted]		[redacted]	[redacted]	-	[redacted]	£1,324,054*
Ustekinumab	[redacted]		[redacted]	[redacted]	-	[redacted]	Ozanimod dominant

Abbreviations: LYG, life years gained; NR, not reported; QALY, quality-adjusted life year

Note: * ICER in SW quadrant

5.1.3. TNFi-naïve subgroup results (fully incremental)

Based on the fully incremental analysis provided by the company, ozanimod was considered the most cost effective treatment compared to adalimumab, resulting in an ICER of £28,686 (based on an incremental QALY gain of [redacted] and an incremental cost of [redacted]. Infliximab was

[redacted] by golimumab, and golimumab was [redacted] by vedolizumab. Vedolizumab resulted in an ICER of £52,736.

Table 53: Company (fully incremental) base case results: TNFi-naïve subgroup (Discounted)

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Cost per QALY gained
Adalimumab	[redacted]	[redacted]	-	-	-
Ozanimod	[redacted]	[redacted]	[redacted]	[redacted]	£28,686
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£52,736

Abbreviations: QALY, quality-adjusted life year

5.1.4. TNFi-experienced subgroup results (fully incremental)

Based on the fully incremental analysis provided by the company ustekinumab was dominated by ozanimod, resulting in an incremental QALY loss of [redacted] and an incremental cost of [redacted]. Compared to ustekinumab, vedolizumab resulted in an ICER of £199,551.

Table 54: Company (fully incremental) base case results: TNFi-experienced subgroup (Discounted)

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Cost per QALY gained
Ozanimod	[redacted]	[redacted]			
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£199,551

Abbreviations: QALY, quality-adjusted life year

5.2. Company's sensitivity analyses

5.2.1. One-way sensitivity analysis

One-way sensitivity analysis (OWSA) was provided and model parameters were varied by +/- 20% (95% CI were used where standard errors of the mean were available). The company presented results based on Net Health Benefit (NHB) and presented results using tornado diagrams (CS, Document B, Section B.3.8.2) as such the impact on ICER was not reported in the CS. Overall, the ERG considered the company's OWSA to be useful in deteriming the sensitivity of model parameters to variation, however the results were be of limited use for

decision making/interpretation as most parameters were varied by an abitrary percentage, and cPAS results were not included for comparator treatments.

5.2.2. Probabilistic sensitivity analysis

The company PSA which varied the model parameters simultaneously to determine the impact on the ICER. The results of the company's PSA (cost-effectiveness plane scatterplots and CEACs) were presented in the CS (Document B, Section B.3.8.1) and results can also be found in Table 51 and Table 52). As there are scatterplots for ozanimod vesus each comparator for both TNF-naïve and TNF-experienced populations they have not presented here again.

The company model used the generated iterations (n=1000) of NMA-derived clinical efficacy parameters related to remission and response. These were hard coded into the Excel model, while the other parameters (namely costs, utilities, discontinuation due to AE, surgery and spontaneous remission related probabilities etc.) used distributions to sample the parameter values probablistically with each PSA run. A table containing the list of parameters varied in the PSA with the respective distributions were presented in the CS (Appendix J.2). The conclusion of PSA results were in line with the base case results; however, in the TNFi-experienced subgroup for the comparison of ozanimod versus vedolizumab the PSA ICER was higher than that of the base case. As per the CS, the company noted that this difference was due to smaller base case incremental QALYs with marginal variations resulting in significant variations in the ICER (though still in the SW quadrant). The ERG noted this difference in incremental QALYs between the base case ([redacted]) and PSA ([redacted]); however, did not find any further issues associated with it. Further, the CS Section B.3.8.1 indicated that AE rates were sampled using a log-normal distribution, utilities were sampled using a beta distribution and the costs using a gamma distribution. However, the ERG noted that a (1-Gamma) distribution was used to sample utilities in the model, although the impact on the results were negligible.

The ERG viewed the approach used to derive the samples for parameters from NMA using Convergence Diagnosis and Output Analysis (CODA) software as appropriate given it takes into account the joint posterior distribution of the parameters included. However, ERG considered that the correlation between the parameters has not been represented adequately as described earlier in Section 3.4.2. The ERG also had reproducibility issues with the PSA as the CODA parameters were hard coded in the model and the settings used for Markov chain Monte Carlo (MCMC) simulations to derive those CODA samples were not fully transparent. Furthermore, the fact that tofacitinib has not been included as a relevant comparator renders the CEAC less

useful for decision making as the probability of ozanimod being cost-effective could change with tofacitinib inclusion.

5.2.3. Scenario analyses

The company conducted scenario analyses to explore uncertainty surrounding key model parameters/assumptions. The company's base case inputs and alternative scenario analysis inputs used are outlined in Table below. The ERG considered the range of scenario analyses conducted by the company to be comprehensive; however it should be noted that results do not include comparator PAS discounts. Furthermore, the company did not conduct a scenario analysis whereby tofacitinib is considered as a relevant comparator in both subgroups. As such, results should be interpreted with caution.

Table 55: Base case and scenario analysis parameters/assumptions used by the company

Model parameter	Base case value	Scenario analysis value(s)
Spontaneous remission	0.5%	0%, 1%
Extended induction	Excluded	Included
Dose escalation	30%	0%, 50%
Treatment waning	Excluded	Included- 25% treatment waning after 2 years
Vial sharing	Excluded	Included
Subsequent treatment	Excluded	Included- applied to TNFi-naïve subgroup only (subsequent treatments were vedolizumab and ustekinumab)
Alternative utility values	Woehl et al.[29] and Arseneau et al. [30]	• TRUENORTH[27,28] • Vedolizumab (TA342)[87] • Tofacitinib (TA547)[21]
CvT/BSC (treatment distribution)		Tofacitinib TA547[21]
Proportion of patients receiving vedolizumab SC	50%	0%, 30%

Abbreviations: BSC, best supportive care; CvT, conventional therapy; NICE, National Institute for Health and Care Excellence; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor; TA, technology appraisal

For the company's full list of scenario analyses results, see Section B.3.8.3 in the CS. The ERG noted that results were mostly sensitive to alternative assumptions with respect to extended induction, dose escalation, utility values and proportion of patients receiving SC vedolizumab.

Incremental results and ICERs for these scenario analyses are presented in Table 56 to Table 59.

Table 56: Extended induction

Scenario	Treatment	Inc. costs	Inc. QALYs	ICERs	NHB (QALY)
TNFi-naïve					
Base case	Adalimumab	[redacted]	[redacted]	£28,686	0.003
	Infliximab	[redacted]	[redacted]	£167,024*	0.175
	Golimumab	[redacted]	[redacted]	£71,023*	0.101
	Vedolizumab	[redacted]	[redacted]	£52,736*	0.205
Extended induction included	Adalimumab	[redacted]	[redacted]	£28,686	0.003
	Infliximab	[redacted]	[redacted]	£95,490*	0.178
	Golimumab	[redacted]	[redacted]	£53,607*	0.116
	Vedolizumab	[redacted]	[redacted]	£49,151*	0.250
TNFi-experienced					
Base case	Vedolizumab	[redacted]	[redacted]	£199,551	0.170
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.156
Extended induction included	Vedolizumab	[redacted]	[redacted]	£81,131	0.234
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.184

Abbreviations: ICER, incremental cost-effectiveness ratio; NHB, net health benefit; QALY, quality-adjusted life years; TNFi, tumour necrosis factor inhibitor

Note: *SW quadrant ICER; costs saved per QALY forgone. Threshold used to calculate NHB was £30,000

Table 57: Alternative dose escalation assumption

Scenario	Treatment	Inc. costs	Inc. QALYs	ICERs	NHB (QALY)
TNFi-naïve					
Base case	Adalimumab	[redacted]	[redacted]	£28,686	0.003
	Infliximab	[redacted]	[redacted]	£167,024*	0.175
	Golimumab	[redacted]	[redacted]	£71,023*	0.101
	Vedolizumab	[redacted]	[redacted]	£52,736*	0.205
0% dose escalation	Adalimumab	[redacted]	[redacted]	£52,734	-0.047
	Infliximab	[redacted]	[redacted]	£105,530*	0.097
	Golimumab	[redacted]	[redacted]	£32,908*	0.007
	Vedolizumab	[redacted]	[redacted]	£41,492*	0.104
50% dose escalation	Adalimumab	[redacted]	[redacted]	£12,655	0.036
	Infliximab	[redacted]	[redacted]	£208,020*	0.228
	Golimumab	[redacted]	[redacted]	£96,434*	0.163
	Vedolizumab	[redacted]	[redacted]	£60,233*	0.272

Scenario	Treatment	Inc. costs	Inc. QALYs	ICERs	NHB (QALY)
TNFi-experienced					
Base case	Vedolizumab	[redacted]	[redacted]	£199,551*	0.170
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.156
0% dose escalation	Vedolizumab	[redacted]	[redacted]	£147,551*	0.118
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.134
50% dose escalation	Vedolizumab	[redacted]	[redacted]	£234,217*	0.205
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.171

Abbreviations: ICER, incremental cost-effectiveness ratio; NHB, net health benefit; QALY, quality-adjusted life years; TNFi, tumour necrosis factor inhibitor

Note: *SW quadrant ICER; costs saved per QALY forgone. Threshold used to calculate NHB was £30,000

Table 58: Alternative utility values

Scenario	Treatment	Inc. costs	Inc. QALYs	ICERs	NHB (QALY)
TNFi-naïve					
Base case	Adalimumab	[redacted]	[redacted]	£28,686	0.003
	Infliximab	[redacted]	[redacted]	£167,024*	0.175
	Golimumab	[redacted]	[redacted]	£71,023*	0.101
	Vedolizumab	[redacted]	[redacted]	£52,736*	0.205
TRUENORTH	Adalimumab	[redacted]	[redacted]	£54,046	-0.026
	Infliximab	[redacted]	[redacted]	£337,782*	0.195
	Golimumab	[redacted]	[redacted]	£143,381*	0.138
	Vedolizumab	[redacted]	[redacted]	£103,454*	0.337
TA342	Adalimumab	[redacted]	[redacted]	£29,933	0.000
	Infliximab	[redacted]	[redacted]	£170,401*	0.176
	Golimumab	[redacted]	[redacted]	£72,272*	0.102
	Vedolizumab	[redacted]	[redacted]	£54,142*	0.212
TA547	Adalimumab	[redacted]	[redacted]	£64,906	-0.032
	Infliximab	[redacted]	[redacted]	£418,880*	0.198
	Golimumab	[redacted]	[redacted]	£175,903*	0.144
	Vedolizumab	[redacted]	[redacted]	£123,157*	0.359
TNFi-experienced					
Base case	Vedolizumab	[redacted]	[redacted]	£199,551*	0.170
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.156
TRUENORTH	Vedolizumab	[redacted]	[redacted]	£440,991*	0.187
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.121
TA342	Vedolizumab	[redacted]	[redacted]	£197,216*	0.170

Scenario	Treatment	Inc. costs	Inc. QALYs	ICERs	NHB (QALY)
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.153
TA547	Vedolizumab	[redacted]	[redacted]	£517,373*	0.189
	Ustekinumab	[redacted]	[redacted]	Ozanimod dominant	0.115

Abbreviations: ICER, incremental cost-effectiveness ratio; NHB, net health benefit; QALY, quality-adjusted life years; TA, technology appraisal; TNFi, tumour necrosis factor inhibitor

Note: *SW quadrant ICER; costs saved per QALY forgone. Threshold used to calculate NHB was £30,000

Table 59: Proportion of patients receiving SC Vedolizumab SC

Scenario	Treatment	Inc. costs	Inc. QALYs	ICERs	NHB (QALY)
TNFi-naïve					
Base case	Vedolizumab	[redacted]	[redacted]	£52,736*	0.205
0% patients receive SC		[redacted]	[redacted]	£68,803*	0.330
30% patients receive SC		[redacted]	[redacted]	£59,039*	0.256
TNFi-experienced					
Base case	Vedolizumab	[redacted]	[redacted]	£199,551*	0.170
0% patients receive SC		[redacted]	[redacted]	£1,982,556*	0.231
30% patients receive SC		[redacted]	[redacted]	£338,194*	0.196

Abbreviations: ICER, incremental cost-effectiveness ratio; NHB, net health benefit; QALY, quality-adjusted life years; SC, subcutaneous; TNFi, tumour necrosis factor inhibitor

Note: *SW quadrant ICER; costs saved per QALY forgone. Threshold used to calculate NHB was £30,000

5.3. Model validation and face validity check

The company described their approach to model validation briefly in the CS Section B.3.10.1, which stated that cell-by-cell model verification was performed by an independent modelling team and clinical opinion was sought to ensure face validity of model structure, inputs and the assumptions. However, the company did not provide a comparison of their model outcomes (QALYs) with that of the previous TAs/publications. Therefore, ERG compared the modelled QALYs from current model with that of the of the previous TAs/publications as discussed in section 4.2.6.5.

Besides a few labelling issues, ERG noted a discrepancy between the CS Document B and the model in the distribution used for utilities in the PSA, as discussed in Section 5.2.2, however it did not have any material impact on the results. Further, during clarification (clarification question B14) the ERG indicated that a fully incremental analysis with the associated CE frontier was missing from the model, after which it was added. Otherwise, no serious errors were found in the company's model that impacted the results.

6. EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

6.1. Exploratory and sensitivity analyses undertaken by the ERG

As noted throughout the report, the ERG conducted a number of scenario analyses to explore uncertainty surrounding certain model parameters and assumptions. The scenario analyses are listed below and the results are presented in Section 6.2.

6.1.1. Cost comparison versus tofacitinib

As noted in 4.2.4 , due to the lack of model flexibility, the ERG were unable to include tofacitinib as a comparator into the analysis. However, in order to explore the uncertainty the ERG conducted a cost comparison which compared ozanimod to tofacitinib in both the TNFi-naïve and TNFi-experienced subgroups. This scenario analysis assumed clinical equivalency between treatments in terms of efficacy and only included differences in drug acquisition costs, monitoring costs and adverse event costs over the modelled time horizon (without considering discontinuation from the active treatment). However, please note that the extended induction and the concomitant medications costs were not considered in this analysis.

Though the clinical equivalency assumption is simplistic, in reality this would likely be a pessimistic assumption for tofacitinib given its clinical response and remission in the maintenance phase were better compared to ozanimod as found in the NMA. The ERG is of the opinion that the committee may find this comparison of costs, though only naïve, to be useful. Further, this analysis could be considered a starting point in addressing the uncertainty associated with the exclusion of tofacitinib as a relevant comparator.

Based on this analysis ozanimod resulted in a cost saving of [redacted] and [redacted] in the TNFi-naïve and TNFi-experienced subgroups respectively, where the PAS price was considered for ozanimod and cPAS was not considered for tofacitinib (see Section 6.2 for results). However, the conclusion changed with the consideration of cPAS for tofacitinib resulting in cost savings compared to ozanimod over lifetime horizon of the model (see cPAS Appendix).

6.1.2. Spontaneous remission

Based on clinical input to the ERG, spontaneous remission is likely to occur for approximately 5% to 30% of flare ups, which would result in a higher per year rate than the company's modelled yearly rate of 12%. This scenario analysis used a higher rate of spontaneous remission reflective of 0.75% per model cycle (18% per year), which also closely corresponds to

the mid-point of clinical expert opinion based estimates (see Section 4.2.2 for further discussion). The ERG noted that this is also in line with the observation mentioned in TA 633 that 1% per model cycle is likely to be an overestimate. Based on this analysis, the total costs were found to decrease across all treatments as the patients from the 'Active UC' state were re-distributed between 'Remission' and 'Response No Remission'. The total QALYs increased as the utility value for the response states were higher. See Section 6.2 for results.

6.1.3. Discontinuation due to AEs

The ERG noted that in the maintenance phase ozanimod was associated with the lowest discontinuation rate compared to all other treatments, resulting in a per cycle discontinuation rate of [redacted]. In order to test uncertainty surrounding modelled discontinuation rate for ozanimod, this scenario analysis assumes a higher discontinuation rate during the maintenance phase i.e. ozanimod treatment discontinuation is assumed to be 5% of the induction discontinuation rate. This rate was chosen as the AE discontinuation rate in the maintenance phase was at least 5% that of the induction for all other treatments based on their respective trials. As expected, the total QALYs and costs decreased marginally with higher discontinuation for ozanimod which resulted in minor impact with regard to cost-effectiveness of ozanimod versus the comparators. See Section 6.2 for results.

6.1.4. Ozanimod AE rate in the maintenance phase

In the company's base case, the per cycle AE rate was based on the rates within the CSR. As noted in 3.2.5.1, the ERG noted there to be some uncertainty surrounding the estimation of ozanimod rates and considered these to be somewhat low when compared to AE rates for comparator treatments (particularly in the maintenance phase). Also, the ERG noted that the rate used in the model was not tested as part of sensitivity analysis (albeit the AE cost per cycle was varied). Therefore, in this scenario, the ERG assumed that the maintenance AE rate for ozanimod was 20% higher, to be in line with the modelled rates for comparator treatments. A very minor increase in the total costs of ozanimod was noted which did not have any impact on its cost-effectiveness versus the comparators. See Section 6.2 for results.

6.1.5. Proportion of vedolizumab SC

As noted in 4.2.8.1, the company assumed that 50% of patients would receive SC vedolizumab and 50% would receive IV vedolizumab. Based on clinical input to the ERG, a 50% split is likely to be a reasonable assumption, however it was noted that patients are being steadily phased

onto SC vedolizumab over time and therefore the majority of patients are likely to receive SC vedolizumab after one year. In order to reflect this opinion, in this scenario, the ERG assumed that 80% of patients receive SC vedolizumab in the maintenance phase (patients typically start treatment on SC vedolizumab after the 6-week induction period). Based on this analysis, ozanimod incremental savings reduced from [redacted] to [redacted], due to reduced administration costs associated with SC vedolizumab and the reduction in the proportion of vedolizumab IV patients modelled to receive dose escalation. See Section 6.2 for results. The results become even more sensitive to the SC proportion when cPAS was considered for vedolizumab (see cPAS Appendix).

6.1.6. Treatment regimen costs applied per treatment cycle

In the company's base case, treatment regimen costs were applied per model cycle in the maintenance phase (in line with TA547).[20] The ERG noted at the clarification stage that the company model included the option of modelling treatment costs per treatment cycle as well, and the results were sensitive to this setting (however it was not tested as a scenario in the CS Section B.3.8.3). Subsequently, the company indicated in the clarification response to question B10 that if the regimen costs were applied per treatment cycle the entire cohort would receive the full cost of the treatment upfront even if they discontinue treatment in subsequent model cycles. As there may be some deviation in the dosing schedule in practice, the company indicated that the application of costs per model cycle was preferred in the base case.

Though the ERG agreed with company's choice of modelling treatment costs per model cycle for the base case, the ERG considered it would still be worth exploring the option of modelling the treatment costs per treatment cycle, given its noticeable impact on the results. Through this scenario the ERG noted the sensitivity of ICER to minor change in treatment costs given the difference in the QALYs between the treatments were lower. For instance, for the comparison of ozanimod versus adalimumab, although the difference in the total drug acquisition costs with per treatment cycle approach was only around [redacted], the ICER increased to >£33k (versus £28k in the base case) as the incremental QALYs were lower ([redacted]). See Section 6.2 for the results.

6.2. Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

The ERG made the changes described in Sections 6.1.2 to Section 6.1.6. Each change has been made individually. Please note that the individual impact of revised modelled efficacy

estimates for BSC was not captured in the ERG base case and hence included here. The results of the ERG's exploratory analyses are provided in Table 60 and Table 61 for the TNFi-naïve and TNFi experienced subgroups respectively. The ERG acknowledged that fully incremental results are considered to be appropriate and suitably robust for decision making by NICE. However, due to the company's exclusion of tofacitinib from the analysis, the ERG have only presented pairwise results on the basis that presentation of fully incremental results (without a relevant active comparator) is likely to be misleading.

Table 60: ERG scenario analysis (TNFi-naïve subgroup)

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
Company base-case					
ozanimod	5.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£28,686	
infliximab		[redacted]	[redacted]	£167,024*	
golimumab		[redacted]	[redacted]	£71,023*	
vedolizumab		[redacted]	[redacted]	£52,736*	
Cost comparison with tofacitinib					
Incremental cost associated with ozanimod	6.1	[redacted]	Not applicable		
Spontaneous remission (0.75% per model cycle)					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£29,830	4%
infliximab		[redacted]	[redacted]	£169,731*	2%
golimumab		[redacted]	[redacted]	£72,123*	2%
vedolizumab		[redacted]	[redacted]	£53,983*	2%
Ozanimod AE discontinuation rate in maintenance phase (5% that of induction)					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£29,790	4%
infliximab		[redacted]	[redacted]	£137,368*	-18%
golimumab		[redacted]	[redacted]	£65,285*	-2%
vedolizumab		[redacted]	[redacted]	£51,677*	-8%
Ozanimod AE rate in the maintenance phase (20% increase)					

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£28,750	0%
infliximab		[redacted]	[redacted]	£166,869*	
golimumab		[redacted]	[redacted]	£70,961*	
vedolizumab		[redacted]	[redacted]	£52,720*	
% patients receiving SC vedolizumab (80% after year 1)					
ozanimod	6.1	-	-	-	-
adalimumab		Not applicable			
infliximab					
golimumab					
vedolizumab		[redacted]	[redacted]	£44,204*	-16%
Treatment regimen costs applied per treatment cycle					
ozanimod	6.1	-	-	-	-
adalimumab		[redacted]	[redacted]	£33,815	18%
infliximab		[redacted]	[redacted]	£188,210*	13%
golimumab		[redacted]	[redacted]	£71,528*	1%
vedolizumab		[redacted]	[redacted]	£53,501*	1%
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	6.3	-	-	-	-
adalimumab		[redacted]	[redacted]	£28,797	0%
infliximab		[redacted]	[redacted]	£167,294*	0%
golimumab		[redacted]	[redacted]	£71,133*	0%
vedolizumab		[redacted]	[redacted]	£52,859*	0%

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year

Table 61: ERG scenario analysis (TNFi-experienced subgroup)

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
Company base-case					
ozanimod	5.1	-	-	-	-
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,725)	
vedolizumab		[redacted]	[redacted]	£199,551*	
Cost comparison with tofacitinib					
Incremental cost associated with ozanimod	6.1	[redacted]	Not applicable		
Spontaneous remission (0.75% per model cycle)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£34,594)	3%
vedolizumab		[redacted]	[redacted]	£198,146*	-1%
Ozanimod AE discontinuation rate in maintenance phase (5% that of induction)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£41,096)	22%
vedolizumab		[redacted]	[redacted]	£160,695*	-19%
Ozanimod AE rate in the maintenance phase (20% increase)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,689)	0%
vedolizumab		[redacted]	[redacted]	£199,367*	
% patients receiving SC vedolizumab (80% after year 1)					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,725)	0%

	Section in ERG report	Incremental costs	Incremental QALYs	ICER £/QALY	% Change from company base case
vedolizumab		[redacted]	[redacted]	£161,152*	-19%
Treatment regimen costs applied per treatment cycle					
ozanimod	6.1	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£47,464)	41%
vedolizumab		[redacted]	[redacted]	£208,721*	5%
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	6.3	-	-	-	
ustekinumab		[redacted]	[redacted]	Dominated by ozanimod (-£33,354)	-1%
vedolizumab		[redacted]	[redacted]	£200,192*	0%

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year

6.3. ERG's preferred assumptions

This section presents the results based on ERG preferred assumptions for the base case. The results below present both the incremental and cumulative impact of ERG preferences.

The ERG's preference would have been to include tofacitinib as a comparator within the economic analysis. However, due to the lack of model flexibility, it was not possible to include tofacitinib in the economic model. As an exploratory analysis, the ERG has conducted a cost comparison versus tofacitinib (see Table 60 and Table 61 for results).

As part of the ERG preferred base case, the ERG considered the following:

- Revised remission and response probability estimates for the treatments and BSC derived from the ERG run of the NMA using the alternative placebo baseline risks (as per 3.4.2.4)
- Revised post-active treatment transition probabilities for BSC which include an alternative means of estimating remission probabilities for BSC based on 'loss of remission' (directly from the sustained remission estimates) as opposed to using the

BSC response rates for the TNFi-experienced population for both populations in the base case.

It is to be noted that due to the use of alternative placebo baseline estimates derived by including only trials which are relevant to decision making, the overall response and remission decreases across all treatments. Due to higher utility associated with the remission, less patients entering that state over the modelled horizon caused reduction in the total QALYs as shown in Table 62. The total costs also decreased owing to a reduction in remission costs which could not be offset by the corresponding increase in active UC state costs.

However, the incremental impact of revised post-active treatment transition probabilities for BSC was different for TNFi-naïve and TNFi-experienced populations. For the TNFi-naïve subgroup, the overall response increased, resulting in marginal total QALY increase while it decreased marginally for the TNFi-experienced subgroup. The increase or decrease in the overall response was driven by the proportional increase or decrease in the 'remission' and 'response no remission' probabilities, which differed between the subgroups. On the other hand, the increase or decrease in total costs was driven by whether the reduction in response health state costs were offset by the corresponding increase in the active UC state costs.

The cumulative effect of these changes in the base case resulted in decreased total costs and QALYs across all treatments for both the subgroups. Please note that the cumulative effect of the ERG base case changes were the same as the incremental impact following revised modelled efficacy estimates for BSC (as shown in Table 62 and Table 63), as there were only two changes as part of the ERG base case.

In the TNFi-naïve subgroup, pairwise deterministic analysis indicated that the ICER for ozanimod compared to adalimumab was £27,794, based on an incremental QALY gain of [redacted] and an incremental cost of [redacted]. Compared to infliximab, golimumab and vedolizumab, ozanimod resulted in SW ICERs i.e., ozanimod was less costly and less effective. Please note that the fully incremental analysis has not been presented here as it would be inaccurate without considering tofacitinib as a relevant comparator.

Probabilistic analysis resulted in similar conclusions with an ICER for ozanimod compared to adalimumab of £27,842. With respect to other comparators, ozanimod was less costly and less effective. Similar to the fully incremental analysis, the CEAC would be inaccurate and not suitable for decision making without considering tofacitinib. Hence, it has not been presented

here. The scatterplots of the cost-effectiveness plane for ozanimod versus each of comparators have been presented in the Appendix B.

Table 62: Summary of ERG's preferred assumptions and ICER (TNFi-naïve subgroup)

Scenario	Total costs	Total QALYs	Incremental cost	Incremental QALYs	ICER: Ozanimod vs. comparators (£/QALY)
Company's base case					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£28,686
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£167,024*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£71,023*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£52,736*
ERG's preferred base case assumptions (applied incrementally over company's base case)					
Re-estimation of baseline placebo risks					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,479
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£169,098*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£82,608*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£56,298*
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,794
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£169,791*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£82,863*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£56,640*
Cumulative impact of ERG preferences (deterministic)					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,794
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£169,791*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£82,863*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£56,640*
Cumulative impact of ERG preferences (probabilistic)					
ozanimod	[redacted]	[redacted]	-	-	-
adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,842

Scenario	Total costs	Total QALYs	Incremental cost	Incremental QALYs	ICER: Ozanimod vs. comparators (£/QALY)
infliximab	[redacted]	[redacted]	[redacted]	[redacted]	£158,721*
golimumab	[redacted]	[redacted]	[redacted]	[redacted]	£87,452*
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£68,470*

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; vs., versus

Note: * ICER in SW quadrant

In the TNFi-experienced subgroup, pairwise deterministic analysis indicated that ozanimod was considered less costly and less effective compared to vedolizumab, resulting in a SW ICER of £436,080 [redacted]. Compared to ustekinumab, ozanimod was dominant i.e. less costly and more effective. Please note that the fully incremental analysis has not been presented here as it would be inaccurate without considering tofacitinib as a relevant comparator.

In the probabilistic analysis, however, ozanimod was found to be dominant compared to both ustekinumab and vedolizumab. As shown in Table 63, for the comparison against vedolizumab the incremental cost savings reduced to [redacted] and the QALY gain increased to [redacted] resulting in the treatment being dominated by ozanimod. However, this should be interpreted with caution as the ICER was found to be highly sensitive to even marginal changes in the incremental costs and QALYs. Furthermore, there is uncertainty around the proportion of patients treated with vedolizumab receiving the treatment as an SC formulation in clinical practice. The ERG noted that it is likely that considering any cPAS for vedolizumab in conjunction with a higher proportion of SC vedolizumab would alter this conclusion, possibly resulting in a SW ICER.

The scatterplots of the cost-effectiveness plane for ozanimod versus each of comparators have been presented in the Appendix B. Like the fully incremental analysis, the CEAC too would be inaccurate and not suitable for decision making without considering tofacitinib. Hence, it has not been presented here.

Table 63: Summary of ERG's preferred assumptions and ICER (TNFi-experienced subgroup)

Scenario	Total costs	Total QALYs	Incremental cost	Incremental QALYs	ICER: Ozanimod vs. comparators (£/QALY)
Company's base case					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£33,725)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£199,551*
ERG's preferred base case (applied incrementally over company's base case)					
Re-estimation of baseline placebo risks					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£71,524)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£427,683*
Revised modelled efficacy estimates for BSC in the post-active treatment phase					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£70,807)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£436,080*
Cumulative impact of ERG preferences (deterministic)					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£70,807)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£436,080*
Cumulative impact of ERG preferences (probabilistic)					
ozanimod	[redacted]	[redacted]	-	-	-
ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£56,635)
vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by ozanimod (-£12,926)

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; vs., versus

Note: * ICER in SW quadrant

6.4. Conclusions of the cost-effectiveness section

For the TNFi-naïve subgroup, based on the ERG's preferred results, ozanimod was cost effective compared to adalimumab at a conventional willingness-to-pay threshold of £30,000, resulting in an ICER of £27,794, based on an incremental QALY gain of [redacted] and an incremental cost of [redacted]. Compared to infliximab, golimumab and vedolizumab, ozanimod resulted in ICERs in the SW quadrant i.e., ozanimod was less costly and less effective. For the TNFi-experienced subgroup, ozanimod was considered less costly and less effective compared to vedolizumab, resulting in a SW ICER of £436,080 [redacted]. Compared to ustekinumab, ozanimod was found to be dominant i.e. less costly and more effective.

The ERG noted that a key strength of the company's submission was the use of precedent to inform the majority of model parameters and assumptions. Furthermore, as discussed throughout the ERG report, the company addressed several key concerns raised previously in prior UC appraisals including TA633 [20] and TA547 [21]. As a result, the ERG's preferred base case assumptions were broadly aligned with the company's (with the exception of baseline placebo risk estimates and revised assumptions with respect to modelled efficacy for BSC). As outlined by the ERG's preferred base case analysis, results were not particularly sensitive to these changes (with the exception of the comparison to vedolizumab in the TNFi-experienced subgroup, see Table 63).

However, there were some key limitations with the company's analysis. In addition to uncertainty surrounding the NMA (and modelled clinical effectiveness estimates), the company did not present results comparing ozanimod to tofacitinib. As noted in Section 4.2.4, the ERG considered tofacitinib to be a potentially relevant comparator. The exclusion of this comparison introduces additional uncertainty and means that the incremental cost effectiveness results (both pairwise and fully incremental) should be interpreted with caution. This concern extends to the interpretation of PSA results as well as the CEAC. Overall, the ERG suggest that NICE deliberate on the appropriateness of tofacitinib as a relevant comparator.

7. END OF LIFE

The ERG considered that ozanimod does not meet NICE end of life criteria as the treatment is not indicated for people with a short life expectancy (normally defined as less than 24 months).

References

1. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. Medical Decision Making. 2013;33(5):657-70.
2. Pasvol TJ, Horsfall L, Bloom S, Segal AW, Sabin C, Field N, et al. Incidence and prevalence of inflammatory bowel disease in UK primary care: a population-based cohort study. BMJ Open. 2020;10(7):e036584.
3. Roberts-Thomson IC, Bryant RV, Costello SP. Uncovering the cause of ulcerative colitis. JGH Open. 2019;3(4):274-6.
4. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG clinical guideline: ulcerative colitis in adults. American Journal of Gastroenterology. 2019;114(3):384-413.
5. NHS. Ulcerative colitis, 2019. Available from: https://www.nhs.uk/conditions/ulcerative-colitis/.
6. Bertani L, Tricò D, Zanzi F, Baiano Svizzero G, Coppini F, de Bortoli N, et al. Oral sucrosomial iron is as effective as intravenous ferric carboxy-maltose in treating anemia in patients with ulcerative colitis. Nutrients. 2021;13(2).
7. Jonefjäll B, Simrén M, Lasson A, Öhman L, Strid H. Psychological distress, iron deficiency, active disease and female gender are independent risk factors for fatigue in patients with ulcerative colitis. United European Gastroenterol Journal. 2018;6(1):148-58.
8. Magro F, Gionchetti P, Eliakim R, Ardizzone S, Armuzzi A, Barreiro-de Acosta M, et al. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. Part 1: Definitions, diagnosis, extra-intestinal manifestations, pregnancy, cancer surveillance, surgery, and ileo-anal pouch disorders. Journal of Crohn's & Colitis. 2017;11(6):649-70.
9. RCGP Learning. Inflammatory bowel disease toolkit, 2019. Available from: https://elearning.rcgp.org.uk/mod/book/view.php?id=12536&chapterid=386.
10. Pola S, Patel D, Ramamoorthy S, McLemore E, Fahmy M, Rivera-Nieves J, et al. Strategies for the care of adults hospitalized for active ulcerative colitis. Clinical Gastroenterology and Hepatology. 2012;10(12):1315-25.e4.
11. Adams SM, Bornemann PH. Ulcerative colitis. American Family Physician. 2013;87(10):699-705.
12. Holvoet T, Lobaton T, Hindryckx P. Optimal management of acute severe ulcerative colitis (ASUC): Challenges and solutions. Clinical and Experimental Gastroenterology. 2021;14:71-81.
13. NHS. Ulcerative colitis. Complications, 2019. Available from: https://www.nhs.uk/conditions/ulcerative-colitis/complications/.
14. Crohn's & Colitis UK. Ulcerative colitis, your guide, 2019. Available from: https://www.crohnsandcolitis.org.uk/about-crohns-and-colitis/publications/ulcerative-colitis.
15. National Institute for Health and Care Excellence (NICE). Ulcerative Colitis: management. NICE guideline [NG130], 2019. Available from: https://www.nice.org.uk/guidance/NG130.
16. Vaizey CJ, Gibson PR, Black CM, Nicholls RJ, Weston AR, Gaya DR, et al. Disease status, patient quality of life and healthcare resource use for ulcerative colitis in the UK: An observational study. Frontline Gastroenterology. 2014;5(3):183-9.
17. NHS. Ulcerative colitis. Diagnosis, 2019. Available from: https://www.nhs.uk/conditions/ulcerative-colitis/diagnosis/.
18. National Institute for Health and Care Excellence (NICE). NICE Pathway: Inducing remission in people with ulcerative colitis, 2021. Available from: https://pathways.nice.org.uk/pathways/ulcerative-colitis/inducing-remission-in-people-with-ulcerative-colitis#content=view-info-category%3Aview-about-menu.

19. NHS Nottingham and Nottinghamshire Clinical Commissioning Group. Nottinghamshire commissioning guidance for the treatment of ulcerative colitis (UC), 2020. Available from: https://www.nottsapc.nhs.uk/media/1636/ulcerative_colitis_algorithm.pdf.
20. National Institute for Health and Care Excellence (NICE). Ustekinumab for treating moderately to severely active ulcerative colitis. Technology appraisal guidance [TA633], 2020. Available from: https://www.nice.org.uk/guidance/TA633.
21. National Institute for Health and Care Excellence (NICE). Tofacitinib for moderately to severely active ulcerative colitis. Technology appraisal guidance [TA547], 2018. Available from: https://www.nice.org.uk/guidance/TA547.
22. Honap S, Chee D, Chapman TP, Patel M, Kent AJ, Ray S, et al. Real-world effectiveness of tofacitinib for moderate to severe ulcerative colitis: A multicentre UK experience. Journal of Crohn's & Colitis. 2020;14(10):1385-93.
23. Celgene Corporation. Zeposia®(ozanimod) capsules for oral use. Package insert, 2021: Bristol Myers Squibb. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/209899s000lbl.pdf.
24. Scott FL, Clemons B, Brooks J, Brahmachary E, Powell R, Dedman H, et al. Ozanimod (RPC1063) is a potent sphingosine-1-phosphate receptor-1 (S1P1) and receptor-5 (S1P5) agonist with autoimmune disease-modifying activity. British Journal of Pharmacology. 2016;173(11):1778-92.
25. Choi D, Stewart AP, Bhat S. Ozanimod: A first-in-class sphingosine 1-phosphate receptor modulator for the treatment of ulcerative colitis. Annals of Pharmacotherapy. 2021:10600280211041907.
26. Cohen JA, Comi G, Selmaj KW, Bar-Or A, Arnold DL, Steinman L, et al. Safety and efficacy of ozanimod versus interferon beta-1a in relapsing multiple sclerosis (RADIANCE): a multicentre, randomised, 24-month, phase 3 trial. Lancet Neurology. 2019;18(11):1021-33.
27. Celgene. Phase 3, multicenter, randomized, double-blind, placebo-controlled trial of oral RPC1063 as induction and maintenance therapy for moderate to severe ulcerative colitis - TRUE NORTH. Clinical Study Report, 2020.
28. Sandborn WJ, Feagan BG, D'Haens G, Wolf DC, Jovanovic I, Hanauer SB, et al. Ozanimod as induction and maintenance therapy for ulcerative colitis. New England Journal of Medicine. 2021;385(14):1280-91.
29. Woehl A, Hawthorne AB, McEwan P. The relation between disease activity, quality of life and health utility in patients with ulcerative colitis. Gut. 2008;57(A1-A172):153-4.
30. Arseneau KO, Sultan S, Provenzale DT, Onken J, Bickston SJ, Foley E, et al. Do patient preferences influence decisions on treatment for patients with steroid-refractory ulcerative colitis? Clinical Gastroenterology and Hepatology. 2006;4(9):1135-42.
31. 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.
32. CRD. Systematic reviews: CRD's guidance for undertaking reviews in healthcare, 2009 York, UK: Centre for Reviews and Dissemination, University of York. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
33. 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.
34. Sandborn WJ, Feagan BG, Wolf DC, D'Haens G, Vermeire S, Hanauer SB, et al. Ozanimod induction and maintenance treatment for ulcerative colitis. New England Journal of Medicine. 2016;374(18):1754-62.
35. 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. Alimentary Pharmacology & Therapeutics. 2020;51(10):922-34.

36. Tran JQ, Zhang P, Ghosh A, Liu L, Syto M, Wang X, et al. Single-dose pharmacokinetics of ozanimod and its major active metabolites alone and in combination with gemfibrozil, itraconazole, or rifampin in healthy subjects: A randomized, parallel-group, open-label study. Advances in Therapy. 2020;37(10):4381-95.
37. Van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health. 2012;15(5):708-15.
38. Gordon JP, McEwan PC, Maguire A, Sugrue DM, Puelles J. Characterizing unmet medical need and the potential role of new biologic treatment options in patients with ulcerative colitis and Crohn's disease: a systematic review and clinician surveys. European Journal of Gastroenterology & Hepatology. 2015;27(7):804.
39. Hoentjen F, Haarhuis BJT, Drenth JPH, de Jong DJ. Elective switching from infliximab to adalimumab in stable Crohn's disease. Inflammatory Bowel Diseases. 2013;19(4):761-6.
40. Patil SA, Rustgi A, Langenberg P, Cross RK. Comparative effectiveness of anti-TNF agents for Crohn's disease in a tertiary referral IBD practice. Digestive Diseases and Sciences. 2013;58(1):209-15.
41. Van Assche G, Vermeire S, Ballet V, Gabriels F, Noman M, D'Haens G, et al. Switch to adalimumab in patients with Crohn's disease controlled by maintenance infliximab: prospective randomised SWITCH trial. Gut. 2012;61(2):229-34.
42. Singh S, Murad MH, Fumery M, Dulai PS, Sandborn WJ. First- and second-line pharmacotherapies for patients with moderate to severely active ulcerative colitis: An updated network meta-analysis. Clinical Gastroenterology and Hepatology. 2020;18(10):2179-91.e6.
43. Sandborn WJ, van Assche G, Reinisch W, Colombel JF, D'Haens G, Wolf DC, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2012;142(2):257-65.e1-3.
44. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. New England Journal of Medicine. 2005;353(23):2462-76.
45. Sands BE, Tremaine WJ, Sandborn WJ, Rutgeerts PJ, Hanauer SB, Mayer L, et al. Infliximab in the treatment of severe, steroid-refractory ulcerative colitis: a pilot study. Inflammatory Bowel Diseases. 2001;7(2):83-8.
46. Panaccione R, Ghosh S, Middleton S, Marquez JR, Scott BB, Flint L, et al. Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. Gastroenterology. 2014;146(2):392-400.e393.
47. Suzuki Y, Motoya S, Hanai H, Matsumoto T, Hibi T, Robinson AM, et al. Efficacy and safety of adalimumab in Japanese patients with moderately to severely active ulcerative colitis. Journal of Gastroenterology. 2014;49(2):283-94.
48. Macaluso FS, Maida M, Ventimiglia M, Renna S, Cottone M, Orlando A. Factors affecting clinical and endoscopic outcomes of placebo arm in trials of biologics and small molecule drugs in ulcerative colitis: A meta-analysis. Inflammatory Bowel Diseases. 2018;25(6):987-97.
49. Olfatifar M, Zali MR, Pourhoseingholi MA, Balaii H, Ghavami SB, Ivanchuk M, et al. The emerging epidemic of inflammatory bowel disease in Asia and Iran by 2035: A modeling study. BMC Gastroenterology. 2021;21(1):204.
50. Prideaux L, Kamm MA, De Cruz PP, Chan FK, Ng SC. Inflammatory bowel disease in Asia: a systematic review. Journal of Gastroenterology and Hepatology. 2012;27(8):1266-80.
51. Motoya S, Watanabe K, Ogata H, Kanai T, Matsui T, Suzuki Y, et al. Vedolizumab in Japanese patients with ulcerative colitis: A phase 3, randomized, double-blind, placebo-controlled study. PLoS ONE. 2019;14(2):e0212989.

52. Sands BE, Sandborn WJ, Panaccione R, O'Brien CD, Zhang H, Johanns J, et al. Ustekinumab as induction and maintenance therapy for ulcerative colitis. New England Journal of Medicine. 2019;381(13):1201-14.
53. Sandborn WJ, Ghosh S, Panes J, Vranic I, Su C, Rousell S, et al. Tofacitinib, an oral Janus kinase inhibitor, in active ulcerative colitis. New England Journal of Medicine. 2012;367(7):616-24.
54. Sandborn WJ, Peyrin-Biroulet L, Zhang J, Chiorean M, Vermeire S, Lee SD, et al. Efficacy and safety of etrasimod in a phase 2 randomized trial of patients with ulcerative colitis. Gastroenterology. 2020;158(3):550-61.
55. Seibold F, Peyrin-Biroulet L, Rubin DT, Feagan B, Oh YS, Tyrrell H, et al. SO2O4O3O1 Efficacy of etrolizumab in patients with moderate to severe UC who had failed TNF antagonist therapy (HICKORY). Swiss Medical Weekly. 2019;149 (Suppl 240):2S.
56. Vermeire S, O'Byrne S, Keir M, Williams M, Lu TT, Mansfield JC, et al. Etrolizumab as induction therapy for ulcerative colitis: a randomised, controlled, phase 2 trial. Lancet. 2014;384(9940):309-18.
57. Rutgeerts P, Feagan BG, Marano CW, Padgett L, Strauss R, Johanns J, et al. Randomised clinical trial: a placebo-controlled study of intravenous golimumab induction therapy for ulcerative colitis. Alimentary Pharmacology & Therapeutics. 2015;42(5):504-14.
58. AbbVie. Study to evaluate the safety and efficacy of two adalimumab dosing regimens in subjects with moderate to severe ulcerative colitis, 2014. Available from: https://clinicaltrials.gov/ct2/show/study/NCT02065622.
59. Probert CS, Hearing SD, Schreiber S, Kuhbacher T, Ghosh S, Arnott ID, et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: A randomised controlled trial. Gut. 2003;52(7):998-1002.
60. Sands BE, Peyrin-Biroulet L, Loftus EV, Jr., Danese S, Colombel JF, Toruner M, et al. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. New England Journal of Medicine. 2019;381(13):1215-26.
61. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. New England Journal of Medicine. 2013;369(8):699-710.
62. Colombel JF, Panes J, D'Haens GR, Schreiber S, Panaccione R, Peyrin-Biroulet L, et al. 945 Higher versus standard adalimumab maintenance regimens in patients with moderately to severely active ulcerative colitis: Results from the SERENE-UC maintenance study. Gastroenterology. 2020;158(6 Suppl 1):S-192.
63. Hibi T, Imai Y, Senoo A, Ohta K, Ukyo Y. Efficacy and safety of golimumab 52-week maintenance therapy in Japanese patients with moderate to severely active ulcerative colitis: a phase 3, double-blind, randomized, placebo-controlled study-(PURSUIT-J study). Journal of Gastroenterology. 2017;52(10):1101-11.
64. Sandborn WJ, Su C, Sands BE, D'Haens GR, Vermeire S, Schreiber S, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. New England Journal of Medicine. 2017;376(18):1723-36.
65. Reinisch W, Sandborn WJ, Hommes DW, D'Haens G, Hanauer S, Schreiber S, et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. Gut. 2011;60(6):780-7.
66. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):96-109.e1.
67. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. International Journal of Epidemiology. 2012;41(3):818-27.

68. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Statistics in Medicine. 2015;34(6):984-98.
69. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. Report by the Decision Support Unit, 2011 (last updated 2016): ScHARR, University of Sheffield. Available from: https://nicedsu.sites.sheffield.ac.uk/tsds/completed-or-in-progress.
70. Dias S, Ades AE, Welton NJ, Jansen J, Sutton AJ. Network meta-analysis for decision making: Wiley; 2018.
71. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. Report by the Decision Support Unit, 2011 (last updated April 2012): ScHARR, University of Sheffield. Available from: https://nicedsu.sites.sheffield.ac.uk/tsds/completed-or-in-progress.
72. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):85-95.
73. Sandborn WJ, Baert F, Danese S, Krznaric Z, Kobayashi T, Yao X, et al. Efficacy and safety of vedolizumab subcutaneous formulation in a randomized trial of patients with ulcerative colitis. Gastroenterology. 2020;158(3):562-72.e12.
74. Kobayashi T, Suzuki Y, Motoya S, Hirai F, Ogata H, Ito H, et al. First trough level of infliximab at week 2 predicts future outcomes of induction therapy in ulcerative colitis-results from a multicenter prospective randomized controlled trial and its post hoc analysis. Journal of Gastroenterology. 2016;51(3):241-51.
75. Jiang XL, Cui HF, Gao J, Fan H. Low-dose infliximab for induction and maintenance treatment in Chinese patients with moderate to severe active ulcerative colitis. Journal of Clinical Gastroenterology. 2015;49(7):582-8.
76. Pugliese D, Variola A, Privitera AC, Allocca M, Bossa F, Cappello M, et al. Real-life study (GORE-UC) evaluating the effectiveness of golimumab for the treatment of ulcerative colitis: An interim analysis from Italian group for the study of inflammatory bowel disease (IG-IBD). United European Gastroenterology Journal. 2017;5(5 Suppl 1):A750-A1.
77. Samaan MA, Cunningham G, Tamilarasan AG, Beltran L, Pavlidis P, Ray S, et al. Therapeutic thresholds for golimumab serum concentrations during induction and maintenance therapy in ulcerative colitis: results from the GO-LEVEL study. Alimentary Pharmacology & Therapeutics. 2020;52(2):292-302.
78. Yu J, Park SJ, Kim HW, Lim YJ, Park J, Cha JM, et al. Effectiveness and safety of golimumab in patients with ulcerative colitis: A multicenter, prospective, postmarketing surveillance study. Gut and Liver. 2021:Online first.
79. Strik A, Van De Vrie W, Bloemsaat-Minekus J, Nurmohamed M, Bossuyt P, Bodelier A, et al. Unchanged infliximab serum concentrations after switching from the originator infliximab to the biosimilar CT-P13 in patients with quiescent inflammatory bowel disease: Results from a prospective study (SECURE). Journal of Crohn's & Colitis. 2018;12(Suppl 1):S477-S8.
80. Park SH, Kim Y-H, Lee JH, Kwon HJ, Lee S-H, Park DI, et al. Post-marketing study of biosimilar infliximab (CT-P13) to evaluate its safety and efficacy in Korea. Expert Review of Gastroenterology & Hepatology. 2015;9 Suppl 1:35-44.
81. Coletta M, Paroni M, Alvisi MF, De Luca M, Rulli E, Mazza S, et al. Immunological variables associated with clinical and endoscopic response to vedolizumab in patients with inflammatory bowel diseases. Journal of Crohn's & Colitis. 2020;14(9):1190-201.
82. Danese S, Subramaniam K, Van Zyl J, Adsul S, Lindner D, Roth J, et al. Vedolizumab treatment persistence and safety in a 2-year data analysis of an extended access programme. Alimentary Pharmacology & Therapeutics. 2021;53(2):265-72.

83. Osterman M, Yarur A, Navaneethan U, James A, Lopez C, Rosario M, et al. Characterization of early clinical and pharmacokinetic response profiles of vedolizumab: An interim analysis of enterpret, a phase 4 clinical study. Gastroenterology. 2020;158(3 Suppl):S15.
84. Vermeire S, Lukas M, Magro F, Adsul S, Lindner D, Rosario M, et al. Vedolizumab efficacy, safety, and pharmacokinetics with reduced frequency of dosing from every 4 weeks to every 8 weeks in patients with Crohn's disease or ulcerative colitis. Journal of Crohn's & Colitis. 2020;14(8):1066-73.
85. Dolan P. Modeling valuations for EuroQol health states. Medical Care. 1997;35(11):1095-108.
86. Wu B, Wang Z, Zhang Q. Cost-effectiveness of different strategies for the treatment of moderate-to-severe ulcerative colitis. Inflammatory Bowel Diseases. 2018;24(11):2291-302.
87. National Institute for Health and Care Excellence (NICE). Vedolizumab for treating moderately to severely active ulcerative colitis. Technology appraisal guidance [TA342], 2015. Available from: https://www.nice.org.uk/guidance/ta342.
88. National Institute for Health and Care Excellence (NICE). Infliximab, adalimumab and golimumanb for treating moderately to severely active ulcerative colitis after the failure of conventional therapy. Technology appraisal guidance [TA329]. 2015. Available from: https://www.nice.org.uk/guidance/ta329.
89. Misra R, Askari A, Faiz O, Arebi N. Colectomy rates for ulcerative colitis differ between ethnic groups: Results from a 15-year nationwide cohort study. Canadian Journal of Gastroenterology and Hepatology. 2016;2016:8723949.
90. Chhaya V, Saxena S, Cecil E, Chatu S, Subramanian V, Curcin V, et al. The impact of timing and duration of thiopurine treatment on colectomy in ulcerative colitis: a national population-based study of incident cases between 1989-2009. Alimentary Pharmacology & Therapeutics. 2015;41(1):87-98.
91. Royal College of Physicians. National clinical audit of inpatient care for adults with ulcerative colitis, UK inflammatory bowel disease (IBD) audit, National Report, 2014. 2014.
92. Segal JP, McLaughlin SD, Faiz OD, Hart AL, Clark SK. Incidence and long-term implications of prepouch ileitis: An observational study. Diseases of the Colon & Rectum. 2018;61(4):472-5.
93. Ferrante M, Declerck S, De Hertogh G, Van Assche G, Geboes K, Rutgeerts P, et al. Outcome after proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis. Inflammatory Bowel Diseases. 2008;14(1):20-8.
94. Jess T, Gamborg M, Munkholm P, Sørensen TI. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. American Journal of Gastroenterology. 2007;102(3):609-17.
95. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health. 2010;13(5):509-18.
96. Swinburn P, Elwick H, Bean K, Curry A, Patel S, Bodger K. The impact of surgery on health related quality of life in ulcerative colitis. Gut. 2012;61:A237.
97. Punekar YS, Hawkins N. Cost-effectiveness of infliximab for the treatment of acute exacerbations of ulcerative colitis. European Journal of Health Economics. 2010;11(1):67-76.
98. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. Health Technology Assessment. 2016;20(35):1-610.
99. Lindsay JO, Armuzzi A, Gisbert JP, Bokemeyer B, Peyrin-Biroulet L, Nguyen GC, et al. Indicators of suboptimal tumor necrosis factor antagonist therapy in inflammatory bowel disease. Digestive and Liver Disease. 2017;49(10):1086-91.

100. NHS England. National schedule of NHS costs, 2019/2020. Available from: https://www.england.nhs.uk/national-cost-collection.
101. Tsai HH, Punekar YS, Morris J, Fortun P. A model of the long-term cost effectiveness of scheduled maintenance treatment with infliximab for moderate-to-severe ulcerative colitis. Alimentary Pharmacology & Therapeutics. 2008;28(10):1230-9.

Appendix A: Additional searches conducted by the ERG

Additional search strategy for phase 4 trials of ozanimod for ulcerative colitis

Ovid MEDLINE (1946 to February 15, 2022)

1 Colitis, Ulcerative/ 37741

2 ((colitis or colorectitis or proctocolitis or procto colitis) adj3 (ulcer* or mucosa* or gravis or idiopathic*)).tw,kf. 46305

3 (((colon or colonic) adj3 ulceration) and chronic*).tw,kf. 48

4 (UC and (ulcer* or colitis*)).tw,kf. 14733

5 1 or 2 or 3 or 4 54678

6 (ozanimod or rpc 1063 or rpc1063 or HSDB 7852 or OZM or Zeposia$2 or UNII-Z80293URPV or Z80293URPV or 1306760-87-1).tw,kf,rn. 153

7 5 and 6 52

8 clinical trial, phase iv/ 2276

9 ("phase 4" or "phase IV").ti,ab. 4739

10 8 or 9 5949

11 7 and 10 0

12 6 and 10 1

Ovid Embase (1974 to February 15, 2022)

1 exp Colitis, Ulcerative/ 81807

2 ((colitis or colorectitis or proctocolitis or procto colitis) adj3 (ulcer* or mucosa* or gravis or idiopathic*)).tw,kf. 73582

3 (((colon or colonic) adj3 ulceration) and chronic*).tw,kf. 94

4 (UC and (ulcer* or colitis*)).tw,kf. 32686

5 1 or 2 or 3 or 4 92875

6 ozanimod/ 504

7 (ozanimod or rpc 1063 or rpc1063 or HSDB 7852 or OZM or Zeposia$2 or UNII-Z80293URPV or Z80293URPV or 1306760-87-1).tw,kf,rn. 574

8 6 or 7 574

9 5 and 8 219

10 phase 4 clinical trial/ 4659
11 ("phase 4" or "phase IV").ti,ab. 7957
12 10 or 11 9728
13 9 and 12 0
14 8 and 12 1

ClinicalTrials.gov (www.clinicaltrials.gov)

Search: Ozanimod (Other terms field). Limited to Phase 4. 1 record – ozanimod for MS
Search: Zeposia (Other terms field). Limited to Phase 4. 1 record – ozanimod for MS
Search: rpc1063 (Other terms field). Limited to Phase 4. 1 record – ozanimod for MS

WHO ICTRP (https://trialsearch.who.int/)

Search: ozanimod (intervention). Recruitment status: ALL. Limited to Phase 4. 0 records
Search: Zeposia (intervention). Recruitment status: ALL. Limited to Phase 4. 0 records
Search: rpc1063 (intervention). Recruitment status: ALL. Limited to Phase 4. 0 records

EU Clinical Trials Register (https://www.clinicaltrialsregister.eu)

Search: ozanimod. Limited to Phase 4. 0 records
Search: zeposia. Limited to Phase 4. 0 records
Search: rpc1063. Limited to Phase 4. 0 records

Additional Ovid MEDLINE and Ovid Embase search strategy for phase 4 trials of comparator treatments for ulcerative colitis

Ovid MEDLINE (1946 to February 16, 2022)

1 Colitis, Ulcerative/ 37738
2 ((colitis or colorectitis or proctocolitis or procto colitis) adj3 (ulcer* or mucosa* or gravis or idiopathic*)).tw,kf. 46303
3 (((colon or colonic) adj3 ulceration) and chronic*).tw,kf. 48
4 (UC and (ulcer* or colitis*)).tw,kf. 14732

5 1 or 2 or 3 or 4 54676

6 clinical trial, phase iv/ 2274

7 ("phase 4" or "phase IV").ti,ab. 4737

8 6 or 7 5947

9 Ustekinumab/ 1437

10 (ustekinumab or "cnto 1275" or cnto1275 or stelara$2 or UNII-FU77B4U5Z0 or FU77B4U5Z0 or 15610-63-0).tw,kf,rn. 2597

11 Infliximab/ 11320

12 (infliximab or CT-P13 or CTP13 or SB2 or "abp 710" or abp710 or avakine$2 or flixabi$2 or "gp 1111" or gp1111 or inflectra$2 or ixifi$2 or "pf 06438179" or "pf 6438179" or pf06438179 or pf6438179 or remicade$2 or remsima$2 or renflexis$2 or revellex$2 or "ta 650" or ta650 or zessly$2).tw,kf,rn. 16945

13 Adalimumab/ 6267

14 (adalimumab or ADA or "abp 501" or abp501 or "abt d2e7" or abtd2e7 or adaly$2 or amgevita$2 or amjevita$2 or "avt 02" or "avt 02" or "bat 1406" or bat1406 or "bax 2923" or bax2923 or "bax 923" or bax923 or "bi 695501" or bi695501 or "chs 1420" or chs1420 or "ct p17" or ctp17 or cyltezo$2 or "da 3113" or da3113 or "dmb 3113" or dmb3113 or exemptia$2 or "fkb 327" or fkb327 or fyzoclad$2 or "gp 2017" or gp2017 or hadlima$2 or halimato$2 or hefiya$2 or "hlx 03" or hlx03 or hulio$2 or humira$2 or hyrimoz$2 or "ibi 303" or ibi303 or imraldi$2 or kromeya$2 or lu 200134 or lu200134 or "m 923" or m923 or mabura$2 or (monoclonal adj3 antibod$ adj3 D2E7) or "msb 11022" or msb11022 or "ons 3010" or ons3010 or "pf 06410293" or "pf 6410293" or pf06410293 or pf6410293 or raheara$2 or "sb 5" or sb5 or solymbic$2 or trudexa$2 or "zrc 3197" or zrc3197 or FYS6T7F842 or 331731-18-1 or 1446410-95-2).tw,kf,rn. 20606

15 (vedolizumab or entyvio$2 or "ldp 02" or ldp02 or "mln 0002" or mln0002 or "mln 02" or mln02 or "ldp 02" or UNII-9RV78Q2002 or 9RV78Q2002 or 943609-66-3).tw,kf,rn. 1430

16 (tofacitinib or "cp 690 550" or "cp 690550" or "cp690 550" or cp690550 or HSDB 8311 or xeljanz$2 or UNII-87LA6FU830 or 87LA6FU830 or 477600-75-2 or 540737-29-9).tw,kf,rn. 2180

17 (golimumab or CNTO-148 or CNTO148 or Simponi$2 or UNII-91X1KLU43E or 91X1KLU43E or 476181-74-5).tw,kf,rn. 1470

18 (filgotinib or "g 146034" or "g 146034 101" or g146034 or "g146034 101" or "glpg 0634" or glpg0634 or "gs 6034" or gs6034 or Jyseleca$2).tw,kf,rn. 195

19 (etrasimod or APD334 or UNII-6WH8495MMH or 6WH8495MMH or 1206123-37-6).tw,kf,rn. 18

20 (etrolizumab or pro 145223 or pro145223 or rhumab beta7 or UNII-I2A72G2V3J or I2A72G2V3J or 1044758-60-2).tw,kf,rn. 87

21 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 37270

22 5 and 8 and 21 13

Ovid Embase (1974 to February 16, 2022)

1 exp Colitis, Ulcerative/ 81818

2 ((colitis or colorectitis or proctocolitis or procto colitis) adj3 (ulcer* or mucosa* or gravis or idiopathic*)).tw,kf. 73598

3 (((colon or colonic) adj3 ulceration) and chronic*).tw,kf. 94

4 (UC and (ulcer* or colitis*)).tw,kf. 32694

5 1 or 2 or 3 or 4 92891

6 phase 4 clinical trial/ 4661

7 ("phase 4" or "phase IV").ti,ab. 7959

8 6 or 7 9731

9 Ustekinumab/ 9542

10 (ustekinumab or "cnto 1275" or cnto1275 or stelara$2 or UNII-FU77B4U5Z0 or FU77B4U5Z0 or 15610-63-0).tw,kf,rn. 9821

11 Infliximab/ 56559

12 (infliximab or CT-P13 or CTP13 or SB2 or "abp 710" or abp710 or avakine$2 or flixabi$2 or "gp 1111" or gp1111 or inflectra$2 or ixifi$2 or "pf 06438179" or "pf 6438179" or pf06438179 or pf6438179 or remicade$2 or remsima$2 or renflexis$2 or revellex$2 or "ta 650" or ta650 or zessly$2).tw,kf,rn. 58355

13 Adalimumab/ 39262

14 (adalimumab or ADA or "abp 501" or abp501 or "abt d2e7" or abtd2e7 or adaly$2 or amgevita$2 or amjevita$2 or "avt 02" or "avt 02" or "bat 1406" or bat1406 or "bax 2923" or bax2923 or "bax 923" or bax923 or "bi 695501" or bi695501 or "chs 1420" or chs1420 or "ct p17" or ctp17 or cyltezo$2 or "da 3113" or da3113 or "dmb 3113" or dmb3113 or exemptia$2 or "fkb 327" or fkb327 or fyzoclad$2 or "gp 2017" or gp2017 or hadlima$2 or halimato$2 or hefiya$2 or "hlx 03" or hlx03 or hulio$2 or humira$2 or hyrimoz$2 or "ibi 303" or ibi303 or imraldi$2 or kromeya$2 or lu 200134 or lu200134 or "m 923" or m923 or mabura$2 or

(monoclonal adj3 antibod$ adj3 D2E7) or "msb 11022" or msb11022 or "ons 3010" or ons3010 or "pf 06410293" or "pf 6410293" or pf06410293 or pf6410293 or raheara$2 or "sb 5" or sb5 or solymbic$2 or trudexa$2 or "zrc 3197" or zrc3197 or FYS6T7F842 or 331731-18-1 or 1446410-95-2).tw,kf,rn. 56135

15 vedolizumab/ 5526

16 (vedolizumab or entyvio$2 or "ldp 02" or ldp02 or "mln 0002" or mln0002 or "mln 02" or mln02 or "ldp 02" or UNII-9RV78Q2002 or 9RV78Q2002 or 943609-66-3).tw,kf,rn. 5799

17 tofacitinib/ 6597

18 (tofacitinib or "cp 690 550" or "cp 690550" or "cp690 550" or cp690550 or HSDB 8311 or xeljanz$2 or UNII-87LA6FU830 or 87LA6FU830 or 477600-75-2 or 540737-29-9).tw,kf,rn. 7079

19 golimumab/ 8467

20 (golimumab or CNTO-148 or CNTO148 or Simponi$2 or UNII-91X1KLU43E or 91X1KLU43E or 476181-74-5).tw,kf,rn. 8667

21 (filgotinib or "g 146034" or "g 146034 101" or g146034 or "g146034 101" or "glpg 0634" or glpg0634 or "gs 6034" or gs6034 or Jyseleca$2).tw,kf,rn. 780

22 (etrasimod or APD334 or UNII-6WH8495MMH or 6WH8495MMH or 1206123-37-6).tw,kf,rn. 103

23 (etrolizumab or pro 145223 or pro145223 or rhumab beta7 or UNII-I2A72G2V3J or I2A72G2V3J or 1044758-60-2).tw,kf,rn. 365

24 filgotinib/ 727

25 etrasimod/ 100

26 etrolizumab/ 347

27 or/9-26 101011

28 5 and 8 and 27 50

Additional search of Ovid MEDLINE to identify HRQoL literature not identified by company searches

Ovid MEDLINE(R) <1946 to February Week 3 2022>

1 Inflammatory Bowel Diseases/ or exp Colitis, Ulcerative/ 61154

2 (inflammatory bowel disease or ulcerative Colitis).ti,ab. 65527

3 1 or 2 78192

4 (hrql or hrqol or patient reported outcome$ or satisfaction or preference or disability adjusted life or daly$ or activities of daily living or adl).ab,ti. 292822

5 ((health adj3 (utility$ or status)) or (utilit$ adj3 (valu$ or measur$ or health or life or estimate$ or elicit$ or disease or score$ or weight)) or (disutility$ and health) or (disutility$ and scor$) or (disutility$ and valu$) or standard gamble or time trade off or time tradeoff or tto or rosser or willingness to pay or visual analog scale or visual analogue scale or discrete choice experiment or qwb or 15d or health utilities index or hui or hui1 or hui2 or hui3).ab,ti. 135634

6 (sf36 or sf 36 or sf6 or sf 6 or short form 6 or sf6d or sf 6d or short form 6d or eq 5d or eq5d or euroqol or euro qol or health status or hye or hyes or rosser index or quality of wellbeing or qwb or CUCQ or (Crohn$ adj1 Ulcerative Colitis Questionnaire) or RFIPC or Rating Form of Inflammatory Bowel Disease Patient Concerns or IBDQ or IBDQ-32 or Inflammatory Bowel Disease Questionnaire or SIBDQ or Short Inflammatory Bowel Disease Questionnaire or (health$ adj year$ adj equivalent$)).ti,ab. 87449

7 3 and (4 or 5 or 6) 1729

8 exp Longitudinal Studies/ or (longitudinal study or retrospective study or prospective study or cohort$ or follow up or cross-sectional study or cross sectional study or followup study or observational study or registry or registries or real world or cross sectional).ti,ab. or exp Retrospective studies/ or exp Prospective studies/ or exp Cohort Studies/ or exp Cross-Sectional Study/ or exp Cohort Studies/ or exp Observational Study/ 3423002

9 7 and 8 867

10 (Ephemera or "Introductory Journal Article" or News or "Newspaper Article" or Editorial or Comment or Overall or Letter or Short Survey or Tombstone or Books).pt. or in vitro Techniques/ or in vitro study/ or (commentary or editorial or comment or letter or mice or rat or mouse or animal or murine).ti. 3583247

11 (case report or case study or case series or woman or man or child or adolescent or female or male or boy or girl or infant).ti. 771190

12 case reports/ or case study/ or case report$.jw. 2094080

13 ((child$ or juvenile or babies or infant$ or adolescent$ or pediatric$ or paediatric$) not ((child$ or juvenile or babies or infant$ or adolescent$ or pediatric$ or paediatric$) and adults)).ti. 1193387

14 review.pt. not (((systematic or meta) and analy$) or ((indirect or mixed) and treatment comparison)).ti,ab. 2583450

15 (exp animal/ or nonhuman/) not exp human/ 4960851

16 or/10-15 12513723

17 9 not 16 774

18 limit 17 to yr="2010-current" 559

19 Quality-Adjusted Life Years/ 14384

20 (quality adjusted or adjusted life year$).ti,ab,kf. 17490

21 (qaly$ or qald$ or qale$ or qtime$).ti,ab,kf. 10972

22 (illness state$1 or health state$1).ti,ab,kf. 6519

23 (hui or hui1 or hui2 or hui3).ti,ab,kf. 1496

24 (multiattribute$ or multi attribute$).ti,ab,kf. 864

25 (utility adj3 (score$1 or valu$ or health$ or cost$ or measur$ or disease$ or mean or gain or gains or index$)).ti,ab,kf. 15133

26 utilities.ti,ab,kf. 6921

27 (eq-5d or eq5d or eq-5 or eq5 or euro qual or euroqual or euro qual5d or euroqual5d or euro qol or euroqol or euro qol5d or euroqol5d or euro quol or euroquol or euro quol5d or euroquol5d or eur qol or eurqol or eur qol5d or eur qol5d or eur?qul or eur?qul5d or euro$ quality of life or european qol).ti,ab,kf. 12212

28 (euro$ adj3 (5 d or 5d or 5 dimension$ or 5dimension$ or 5 domain$ or 5domain$)).ti,ab,kf. 4228

29 (sf36$ or sf 36$ or sf thirtysix or sf thirty six).ti,ab,kf. 21683

30 (time trade off$1 or time tradeoff$1 or tto or timetradeoff$1).ti,ab,kf. 1859

31 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 71426

32 3 and 31 429

33 32 not 18 310

34 limit 33 to yr="2010-current" 188

Appendix B: Scatterplots from probabilistic sensitivity analysis for ERG base case

TNFi-naïve population

For the ERG base case, scatter plots showing the incremental costs and QALYs for ozanimod versus the relevant comparators considered in the TNFi-naïve population across all PSA iterations (n=1000) are presented in Figure 12 to Figure 15.

Figure 12: Cost-effectiveness plane for ozanimod versus adalimumab

Figure 13: Cost-effectiveness plane for ozanimod versus infliximab

Figure 14: Cost-effectiveness plane for ozanimod versus vedolizumab

Figure 15: Cost-effectiveness plane for ozanimod versus golimumab

TNFi-experienced population

For the ERG base case, scatter plots showing the incremental costs and QALYs for ozanimod versus the relevant comparators considered in the TNFi-experienced population across all PSA iterations (n=1000) are presented in Figure 16 and Figure 17.

Figure 16: Cost-effectiveness plane for ozanimod versus vedolizumab

Abbreviations: QALYs, quality-adjusted life years; WTP, willingness-to-pay

Figure 17: Cost-effectiveness plane for ozanimod versus ustekinumab