CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR HTA Programme on behalf of NICE

Tofacitinib for moderately to severely active ulcerative colitis

Produced by Southampton Health Technology Assessments Centre

Authors Neelam Kalita, Research Fellow (Health Economics), SHTAC
Petra Harris, Research Fellow, SHTAC
Joanne Lord, Professorial Fellow, SHTAC
Geoff Frampton, Senior Research Fellow, SHTAC
David A. Scott, Director, Diligent Agile Synthesis Limited
Joanna Picot, Senior Research Fellow, SHTAC

Correspondence to Dr. J. Picot
Southampton Health Technology Assessments Centre (SHTAC)
Wessex Institute
Alpha House
Enterprise Road, University of Southampton Science Park
Southampton SO16 7NS
www.southampton.ac.uk/shtac

Date completed 19/07/2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 17/109/06.

Declared competing interests of the authors

None from the authors. The department Dr Parkes works within has been involved in the Phase III trials of tofacitinib (and its potential future competitor filgotonib) – for which the department receives payment. Dr Parkes does not receive personal payment. Dr Parkes is also on the research advisory board for Crohn's and Colitis UK. Dr Brown currently prescribes all of the comparator drugs in ulcerative colitis and he is currently a principal investigator for a phase IIb/III clinical trial of an investigational product (upadacitinib, which is a JAK1 inhibitor) being developed by Abbvie for treatment of ulcerative colitis.

Acknowledgements

We are very grateful to Dr Miles Parkes, Directory of Gastroenterology and Inflammatory Bowel Disease Lead, Addenbrooke's Hospital Cambridge and Dr. Matthew Brown, Consultant Gastroenterologist, Basingstoke and North Hampshire Hospital who offered clinical advice with Dr. Brown also offering comment on the draft report. We would also like to thank: Karen Welch, Information Scientist, SHTAC, for appraising the literature search strategies in the company's submission, and for running searches where necessary; and Dr Jonathan Shepherd, Principal Research Fellow, SHTAC, for providing a quality assurance review of the draft ERG report.

Rider on responsibility for report

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:

Kalita, N.; Harris, P.; Lord, J.; Frampton, G.; Scott D.A.; Picot, J. Tofacitinib for moderately to severely active ulcerative colitis: A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2018.

Contributions of authors

Neelam Kalita critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Petra Harris critically appraised the clinical effectiveness systematic review and drafted the report; Joanne Lord critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; David Scott critically apprised the network meta-analyses and drafted the report; Geoff Frampton critically appraised the clinical effectiveness systematic review and drafted the report; Joanna Picot critically appraised the clinical effectiveness systematic review, drafted the report, project managed the review and is the project guarantor.

Word count: 67948

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADA	Adalimumab
AE	Adverse event
BID	Twice daily
BSG	British Society of Gastroenterology
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CrI	Credible interval
CSR	Clinical study report
DIC	Deviance information criterion
EMA	European Medicines Agency
ERG	Evidence review group
EQ-5D	5-dimensions EuroQol questionnaire
FAS	Full analysis set
GOL	Golimumab
HDAS	NICE Healthcare Databases Advanced Search
HRQoL	Health-related quality of life
IBDQ	Inflammatory Bowel Disease Questionnaire
ICER	Incremental cost-effectiveness ratio
INF	Infliximab
IQR	Inter quartile range
LLN	Lower limit of normal
MCID	Minimal clinically important difference
MCS	Mental health component summary (of SF-36)
mFAS	Modified full analysis set
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NRI	Non-responder imputation
PAS	Patient access scheme
PBO	Placebo
PCS	Physical component summary (of SF-36)
QALY	Quality-adjusted life year
Q4W	Once every four weeks

Q8W	Once every eight weeks
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SF-36	36-Item Short Form survey
SLR	Systematic literature review
SmPC	Summary of Product Characteristics
STA	Single technology appraisal
SUCRA	Surface under cumulative ranking curve
TA	Technology appraisal
TEAE	Treatment-emergent adverse event
TNF	Tumour necrosis factor
TNFi	Tumour necrosis factor inhibitor
TOF	Tofacitinib
UC	Ulcerative colitis
ULN	Upper limit of normal
VAS	Visual analogue scale
VED	Vedolizumab
WHO	World Health Organisation
WPAI-UC	Work Productivity and Activity Impairment-Ulcerative Colitis

SUMMARY

Scope of the company submission

The company's submission (CS) presents evidence for the clinical effectiveness and cost-effectiveness of tofacitinib (Xeljanz®) for the treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, lost response, or were intolerant to either conventional therapy or a biologic agent. Tofacitinib is an orally administered small-molecule selective inhibitor of the Janus kinase (JAK) family of tyrosine kinases. The inhibition of JAKs by tofacitinib attenuates the signalling of several interleukins and type I and II interferons, which leads to modulation of the immune and inflammatory response in ulcerative colitis. The recommended dose is 10mg twice daily for induction for eight weeks and 5mg given twice daily for maintenance.

Summary of submitted clinical effectiveness evidence

The company's systematic review of clinical effectiveness identified four relevant placebo controlled randomised controlled trials (RCT) of tofacitinib.

- one phase II RCT [treatment arms: TOF 0.5 mg twice a day (BID), 3 mg BID, 10 mg BID, 15 mg BID and placebo]
- two identical phase III induction RCTs (OCTAVE Induction 1 and OCTAVE Induction 2; treatment arms: TOF 10 mg BID and placebo)
- one maintenance RCT (OCTAVE Sustain; treatment arms: 5 mg BID, 10 mg BID and placebo)

The ERG believes the company has identified all the relevant RCTs of tofacitinib. In addition to the RCTs, an open-label uncontrolled long-term extension study of tofacitinib, OCTAVE Open, is ongoing.

The CS focusses on the three large phase III trials OCTAVE Induction 1 and OCTAVE Induction 2 and the OCTAVE Sustain maintenance study. The small phase II trial is included in network meta-analyses (NMAs) but it not reported on in detail in the CS.

The OCTAVE 1 and OCTAVE 2 induction trials followed identical methods and both were multicentre, worldwide RCTs. To be enrolled patients had to have moderately to severely active ulcerative colitis. Eligible patients were randomised on a 4:1 ratio to 10 mg twice a day (BID) of oral tofacitinib or placebo for eight weeks (a third 15 mg BID tofacitinib arm was discontinued prior to full recruitment based on feedback from regulatory authorities). Randomisation was

stratified by previous treatment with TNFi therapies, glucocorticoid use at baseline, and geographic region.

People who had participated in the OCTAVE 1 and 2 induction trials completed 8 weeks of induction therapy and met the criteria for a clinical response were eligible to be re-randomised into the OCTAVE Sustain maintenance study. Eligible patients were randomised in a 1:1:1 ratio to 5 mg BID tofacitinib, 10mg BID tofacitinib or placebo. Randomisation was stratified by induction-trial group assignment and remission status at maintenance-trial entry. The duration of treatment was 52 weeks but any patient who met treatment failure criteria was required to withdraw from the study.

The CS reports the effects of tofacitinib treatment across a range of outcomes relevant to the NICE scope and the company decision problem, which are summarised below.

Remission is the primary outcome of the OCTAVE Induction trials and the OCTAVE Sustain maintenance trial. In both the OCTAVE 1 and OCTAVE 2 Induction trials, a statistically significant difference in remission at week 8 in comparison to placebo was observed in participants who received tofacitinib 10 mg twice daily. The same results were obtained regardless of whether centrally read or locally read endoscopic data were used (albeit the mean differences between the tofacitinib and placebo group were higher in both trials when using locally read endoscopic data). In the OCTAVE Sustain maintenance trial there was a statistically significant difference in remission at week 52 in comparison to placebo for participants who received tofacitinib 10 mg twice daily and those who received tofacitinib 5 mg twice daily. Locally read endoscopic data again produced less conservative results than centrally read endoscopic data. Sustained remission (remission at both week 24 and week 52) results were also in favour of tofacitinib.

Mucosal healing is a key secondary outcome of the OCTAVE trials. A statistically significant difference in the proportion of participants with mucosal healing in favour of tofacitinib was observed both at week 8 in the OCTAVE 1 and OCTAVE 2 induction trials as well as at week 52 in the OCTAVE Sustain trial.

The OCTAVE Sustain maintenance trial reported the outcome of sustained corticosteroid-free remission among those in remission at baseline in this trial. This outcome also favoured the

tofacitinib groups with statistically significant differences between the 5mg and the 10mg tofacitinib arms versus placebo.

Remission, mucosal healing and sustained corticosteroid-free remission did not contribute data to the economic model.

Clinical remission is an outcome with an almost identical definition to the primary outcome of remission. The difference being that the rectal bleeding sub-score of the Mayo score does not have to be zero to achieve clinical remission. The outcomes of clinical remission and clinical response contribute data to the economic model.

Using locally read data (which were used in the base case economic evaluation) in OCTAVE 1, the mean difference between the tofacitinib group and the placebo group was 13.3 percentage points (95% CI 6.5 to 20.2, p=0.0017). The corresponding data for OCTAVE 2 were a mean difference from placebo of 15.6 percentage points (95% CI 9.9 to 21.3, p=0.0002). At week 52 in the OCTAVE Sustain maintenance trial the results for clinical remission also favoured tofacitinib (difference versus placebo 35.1%, 95% CI 26.7 to 43.5, p<0.0001 using locally read data).

Clinical response at both week 8 (OCTAVE Induction trials) and week 52 (OCTAVE Sustain trial) was also statistically significantly higher among participants who received tofacitinib.

Subgroup analyses according to prior TNFi-exposure status were conducted for the main clinical effectiveness outcomes. The results were consistent regardless of prior TNFi-exposure status.

HRQoL was reported using generic (EQ-5D and SF-36) and disease specific (IBDQ and WPAI-UC) instruments. HRQoL was typically improved by tofacitinib treatment however for some HRQoL measures the ERG was uncertain about the impact of missing data. Data from the EQ-5D-3L did not inform the base-case economic model but were included in a scenario analysis.

Safety data for tofacitinib in patients with moderate to severely active ulcerative colitis comes from the Phase II trial, the three Phase III OCTAVE trials and the ongoing OCTAVE Open extension study. Rates of adverse events of any type were broadly similar for the tofacitinib and

placebo arms within each trial with serious adverse events affecting fewer than 10% of patients. Ulcerative colitis was the most frequent serious adverse event and most other serious adverse events were related to ulcerative colitis. Serious infections were uncommon (data on serious infections were included in the economic model). Overall, and in comparison with evidence from the use of tofacitinib in patients with rheumatoid arthritis, no new safety signals were identified.

There are no head-to-head RCTs of tofacitinib versus the comparators defined in the company's decision problem. Therefore the company used NMA to estimate the relative effectiveness and safety of tofacitinib in both the induction and maintenance phases of treatment in comparison to TNF-alpha inhibitors (infliximab, adalimumab and golimumab), vedolizumab and conventional therapies. The company's systematic review identified 21 RCTs that were considered for inclusion in the NMA. Four of these were the tofacitinib RCTs listed above, a further 14 were included in one or more NMA networks and three studies could not be included in any of the NMA networks.

Table 1 NMAs conducted by the company

	TNFi-naïve population subgroup	**TNFi-exposed population subgroup**
Induction phase	Clinical response and clinical remission Mucosal healing	Clinical response and clinical remission Mucosal healing
	Safety outcomes (discontinuation due to AEs, SAEs, serious infections)	
Maintenance phase	Clinical response and clinical remission Mucosal healing	Clinical response and clinical remission Mucosal healing

The ERG judged the NMAs to be generally well conducted but identified nine issues:

- Use of the probit scale to model clinical response/clinical remission is an improvement on a previous approach in NICE guidance TA342 but a multinomial logit model could have been considered.
- Potential inconsistency in a closed loop of the maintenance TNFi-naïve network was not examined

- The ERG would have made different choices regarding model fit, in general for the efficacy outcomes the ERG would have chosen the random effects model as the more conservative approach given the known between study heterogeneity. For the safety outcome of serious infections, the absence of any events in the placebo arms of the tofacitinib trials causes very wide credible intervals.
- The ERG was unable to replicate the same baseline (placebo) credible intervals used in the probit or logit models to estimate absolute probabilities. The company's estimates may be conservative.
- The phase II trial may have had a disproportionate effect on the random effect safety NMA because of the relatively high serious infection rate in the tofacitinib arm of this study.
- No safety NMA was conducted for the maintenance period.
- The company did not attempt to adjust for differences in lengths of induction and maintenance treatment and the ERG is concerned that this could have introduced potential bias against those treatments where studies had shorter induction phase and benefit those treatments with a shorter maintenance phase.
- There are differences between patient populations in the re-randomised design maintenance trials. OCTAVE Sustain re-randomised all responders from the OCTAVE induction trials to either placebo or tofacitinib treatment. In contrast, the other re-randomised studies, only re-randomised patients who had received and responded to active treatment into the maintenance phase of the study.
- Adjustments to treat-through trials were made, and although the ERG does not believe these introduce additional bias, it is nevertheless the case that non-responders at the end of the induction phase are ignored (and these participants potentially could have become responders by the end of the maintenance phase).

For the three outcomes synthesized by NMA which contribute data to the economic model the results were as follows.

The induction phase NMA for the TNFi-naïve population provided strong evidence of benefit for all treatments over placebo with infliximab having the largest treatment effect for both clinical response and clinical remission. In the TNFi-exposed population, tofacitinib had the largest treatment effect on clinical response and clinical remission compared to placebo. Only tofacitinib and vedolizumab showed strong evidence of benefit.

In the maintenance phase NMA for the TNFi-naive population all treatments showed strong evidence of benefit over placebo with tofacitinib 10mg having the largest treatment effect on clinical response and clinical remission. In the TNFi-exposed population, tofacitinib 10mg had the largest treatment effect on clinical response and clinical remission compared to placebo. Tofacitinib 5mg, 10mg and vedolizumab 300mg Q4W and Q8W all showed a strong evidence of benefit over placebo.

[redacted]

Summary of submitted cost effectiveness evidence

The company's submission includes a review of published cost-effectiveness evidence and a new economic model developed for this appraisal. The model compares the cost-effectiveness of Tofacitinib for treating people with moderately to severely active ulcerative colitis who are either intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy (oral corticosteroids and/or immunosuppressants) or a TNF-alpha inhibitor.

The company adheres the NICE scope; but excludes adalimumab as a comparator in TNfi-exposed sub group analysis

Broadly, the company model adheres with the NICE scope. We present a top-line view of the ERG's observations on patient characteristics, sub-groups and comparators included within the company model.

For their base case analyses, patient characteristics (including initial age, weight and gender mix) for the two sub-groups of TNFi-naïve and TNFi-exposed are based on means from the tofacitinib arms in the OCTAVE Induction trials. We view that these baseline characteristics should be assumed similar for people with and without prior exposure to TNFi drugs.

In line with the NICE scope, the company conducts sub-group analysis according to previous treatment with one or more biologics. However, as the NMA results used in the model are

defined by prior exposure to TNF-alpha inhibitors alone, we view it appropriate to label the sub groups based on status of patients' exposure to TNF-alpha inhibitors- i.e. TNFi- naïve and TNFI-exposed. The company presents cost-effectiveness analyses for these two sub-groups. In addition, they also present cost-effectiveness results based on analysis of the whole ITT population. We view that the company's 'ITT' cost-effectiveness scenario is highly uncertain and that it omits relevant comparators (the TNFi drugs), so does not address the specified decision problem. Hence, we focus on separate analyses for the two TNFi exposure subgroups in our discussion and additional analysis.

For patients in TNFi-exposed sub group, the company excludes adalimumab, infliximab and golimumab as comparators. Whilst clinical response and remission rates are not available for infliximab or golimumab in this sub group, but they are available for adalimumab. Hence, we consider adalimumab as a relevant comparator for at least some patients with prior exposure to a TNFi agent, although we understand that further treatment with a TNFi may not be appropriate for all patients in this subgroup.

The structure and assumptions of the submitted model are mostly reasonable, albeit a few issues

The company submitted a Markov cohort model consisting of 9 health states, with a cycle length of 8 weeks and patient lifetime horizon. Costs and QALYs were discounted at 3.5% annually. The model uses 3 sets of input parameters: clinical inputs (governing the rates of response and remission and adverse event rates for comparator treatments, as well as the incidence and complication/mortality rates for surgery), health state utilities; and resource use and costs.

The company assumes treatment effect to be maintained with ongoing treatment and non-responders are given conventional therapy as second-line. We agree with this approach which follows the independent economic analysis in TA329. However, the model does not reflect NICE recommendations for annual assessment of benefit and need for continued treatment in previous appraisals TA329 and TA342. Clinical advice suggests that withdrawal of treatment for patients in remission is unlikely in practice, and the effects of this are difficult to quantify given the model structure and limited evidence over long-term maintenance of remission.

The company model applies a constant risk of relapse across each 8-week cycle of maintenance, with treatment stopping immediately when patients lose response. We consider

this assumption to reflect UK practice. However, we have concerns that the costs of monitoring and follow-up in the company's model do not reflect the full cost of ensuring that treatment can be withdrawn within 8 weeks of a relapse. We address this in our additional analyses.

The company conducted NMA to inform clinical inputs within the model. To populate clinical remission and response, the company used a simple fixed effect approach. The ERG has a general preference for the random effect NMA models, as we believe that the fixed effect models may underestimate uncertainty due to heterogeneity between the studies.

In their base case NMA, the company combined outcomes for subgroups defined as TNFi-failed for vedolizumab with TNFi-exposed subgroups for tofacitinib and adalimumab. We consider that combining results for TNFi-failed and TNFi-exposed subgroups is a potential source of bias in favour of tofacitinib and view that using a more like-for-like comparison between tofacitinib and vedolizumab by using data for the TNFi-failed subgroups from the OCTAVE and GEMINI trials, is reasonable.

The company transformed the results of the clinical response/remission NMAs from the probit scale to the natural scale and converted to absolute probabilities to inform the economic model. For simplicity, they assume a constant ratio of patients in remission and response throughout maintenance phase and beyond in extrapolation. This is inconsistent with the clinical advice to the ERG as experience indicates the risk is greatest in the first 6-12 months; and falls thereafter. However, due to absence of evidence we were unable to adapt the model to reflect clinical evidence. Extrapolation of relapse and discontinuation rates from the maintenance trials is likely to underestimate the average duration of treatment and hence both the costs and QALYs of active treatments. However, it is not possible to estimate the net direction of bias in ICERs between comparators, because trends in long-term risks may vary between TNFi drugs, vedolizumab and tofacitinib.

Adverse events, except serious infections, were excluded from the economic analysis. We agree with the company's approach, but acknowledge that the omission of non-infection SAEs does introduce a risk of bias. However, given the frequency of these events, this is unlikely to influence the cost-effectiveness results. The company estimated risk of serious infections using a binomial logit NMA model in the induction trails and chose random effects model for their base case. However, there was considerable uncertainty around the model estimates. The ERG had

concerns as our verification checks indicated an even higher level of uncertainty around tofacitinib estimates, and we were unable to replicate the company's base case NMA values. We therefore applied a frequentist NMA approach to estimate the risk of serious infection, which we use as a scenario in ERG analysis.

We agree with the company approaches to modelling surgery risks, perioperative- and post-operative complications and mortality.

Health state utilities are estimated from published literature for the base case.

We agree with the company that the utility estimates by a published study by Woehl et al. provide an appropriate source for base case parameters. For scenario analysis, the company also conducted simple and regression-based analyses of EQ-5D data from the OCTAVE trials. However, these estimates are problematic as sources of utility parameters for the economic model due to the re-randomisation design and lack of intermediate assessments of clinical response and remission between week 8 and week 52 which complicate the interpretation of results.

In general, company's approach to costing is appropriate and consistent with related NICE guidance, albeit with a few errors in estimation

Costs and resources associated with drug acquisition, drug administration, monitoring and follow up and treatment of serious infections were included in the company's cost-effectiveness analyses. Overall, the costs inputs and sources used were appropriate although the ERG identified a few inconsistencies:

- We identified an error in the estimation of cost associated with elective surgery with complications which we corrected in the ERG corrected company's base case model.
- The company made an error in estimating weight wastage. Correction of this error had no influence on the base case results as they used 'fitting distribution' approach for wastage calculation.
- We noted a few minor changes in NHS prices for included drugs: sulfasalazine, prednisolone and azathioprine. The price changes lead to a very small decrement in the estimated cost of CT alone, with biologic drugs and with tofacitinib.
- No cost was assumed for administering adalimumab and golimumab which are administered by subcutaneous injection. We address this by assuming an initiation of self-administration of subcutaneous injections by adding the cost of a non-consultant led

clinic attendance to the cost of induction for adalimumab and golimumab in our additional analyses

- Health care usage assumptions were made based on the study by Tsai et al. (2008). Whilst we agree with the company's approach for the base case, we conduct scenarios testing alternative resource use based on expert advice.
- We question company's assumption that maintenance treatment will always stop within 8 weeks of a loss of response which is consistent with the number of outpatient appointments. We test this assumption in our additional analyses
- The company excludes cost of stoma care and the estimated cost of surgery is low compared with previous appraisals. We address these in our additional analysis.

Company's base case results

The company's base case results are presented in

Table 2 and Table 3

Table 2 Cost effectiveness: Company base case, no prior TNFi (with tofacitinib PAS)

Strategy	Total		Incremental analysis			Pairwise ICERs tofacitinib vs. comparator (£/QALY)
	QALYs	Costs (£)	QALYs	Costs (£)	ICER (£/QALY)	
Conventional	[redacted]	[redacted]	-	-	-	£8,554
Adalimumab	[redacted]	[redacted]	-	-	Dominated	Dominated
Golimumab	[redacted]	[redacted]	-	-	Dominated	Dominated
Infliximab	[redacted]	[redacted]	-	-	Dominated	Dominated
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	£8,554	N/A
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£615,057	£615,057

Table 3 Cost effectiveness: Company base case, with prior TNFi (with tofacitinib PAS)

Strategy	Total		Incremental analysis			Pairwise ICERs tofacitinib vs. comparator (£/QALY)
	QALYs	Costs (£)	QALYs	Costs (£)	ICER (£/QALY)	
Conventional	[redacted]	[redacted]	-	-	-	£10,302
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	£10,302	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£7,838,238	£7,838,238

A range of uncertainty analyses were conducted by the company, but they have been selective in the scenarios they present

The company performed a range of deterministic-, probabilitistic- and scenario analyses to assess the methodological as well as parameter uncertainty of their base case analyses. The ERG agrees with their assumptions for DSA and PSA and their results, in general. However, we identified errors in the scenarios relating the use of central read NMA results and tofacitinib maintenance using [redacted] split. The company corrected the error in the latter scenario in their response to clarification question. For the scenario analyses, we view that the company has been selective in the scenarios they present.

Commentary on the robustness of submitted evidence

Strengths

- The model structure is consistent and follows the conventional design for ulcerative colitis appraisals.
- The model generally adheres to the NICE scope for this appraisal.
- The perspective of the analysis aligns with the NICE guide to the methods of Technology Appraisal.
- The model uses a lifetime time horizon to allow estimation of all relevant costs and quantity of life impairment.
- The model uses appropriate sources for costs and resource use and in line with other technology appraisals
- The model allows the flexibility to incorporate treatment sequencing which provides a closer reflection of clinical practice.
- The ERG agrees with the company's approach to modelling surgery and its related risks, source of costs and utilities for the base case and mortality.
- The economic model was of good quality, with very few errors in input parameters, logic or coding.
- In the TNFi-naïve arm, the model results were comparable with the clinical data for the tofacitinib arm.

Weaknesses and Areas of uncertainty

- For their base case analyses, patient characteristics (including initial age, weight and gender mix) for the two sub-groups of TNFi-naïve and TNFI-exposed are based on means from the

tofacitinib arms in the OCTAVE Induction trials. We view that these baseline characteristics should be assumed similar for people with and without prior exposure to TNFi drugs.

- We do not consider the company's cost effectiveness analysis with the ITT population scenario to be reliable due to the high level of uncertainty in the underlying NMA. The scenario also omits relevant comparators (the TNFi drugs), so does not address the specified decision problem.
- The company excludes adalimumab as a comparator for patients with prior exposure to a TNFi, despite available evidence to support this.
- The company assumes equal use of 4 drugs in aminosalicylate class (balsalazide, mesalazine, olsalazine & sulfasalazine). However, clinical advice to ERG suggests most patients receive mesalazine in UK and the doses for active ulcerative colitis are potentially higher than specified in company base case.
- The company assumes treatment effect to be maintained with ongoing treatment and non-responders are given conventional therapy as second-line. However, the economic model does not reflect NICE recommendations for annual assessment of benefit and need for continued treatment in previous appraisals TA329 and TA342.
- The company model applies a constant risk of relapse across each 8-week cycle of maintenance, with treatment stopping immediately when patients lose response. Although this is reflective of UK practice, the costs of monitoring and follow-up in the company's model do not reflect the full cost of ensuring that treatment can be withdrawn within 8 weeks of a relapse.
- The company use fixed effects NMA models to inform the economic model . The ERG has a general preference for the random effect NMA models, as we believe that the fixed effect models may underestimate uncertainty due to heterogeneity between the studies.
- The base case NMAs combine outcomes for subgroups defined as TNFi-failed for vedolizumab with TNFi-exposed subgroups for tofacitinib and adalimumab. Combining results for TNFi-failed and TNFi-exposed subgroups introduces a potential source of bias in favour of tofacitinib.
- The company assume constant ratio of patients in remission and response throughout maintenance phase and beyond in extrapolation. These assumptions might not be realistic as clinical -experience indicates the risk is greatest in the first 6-12 months; and falls thereafter.
- There is considerable uncertainty in the NMA estimates for risks of serious infections. We have reservations about the company's approach to estimating this parameter as our

verification checks indicated an even higher level of uncertainty around tofacitinib estimates, and we were unable to replicate the company's base case NMA values.

- The company's simple and regression-based analyses of EQ-5D data from the OCTAVE trials are problematic as sources of utility parameters for the economic model. They are relevant to the decision problem and clinical evidence, but the re-randomisation design and lack of intermediate assessments of clinical response and remission between week 8 and week 52 complicate the interpretation of results.
- The company did not include any costs associated with an initiation of self-administration of subcutaneous injections for adalimumab and golimumab
- The company excludes cost of stoma care and the estimated cost of surgery is low compared with previous appraisals.

Summary of additional work undertaken by the ERG

The ERG conducted a number of scenario analyses. Our preferred assumptions, alongside the scenarios are presented in Table 4.

Table 4 ERG's preferred assumptions and scenarios

Aspect of the model		ERG preferred	ERG scenarios	
Patients	Age (yrs)	Average of all patients in OCTAVE 1 and 2: 41	Range: 28-52	
	Weight (kgs)	Average for all patients in OCTAVE 1 and 2: 73.5	Range: 70-80 kg	
Comparator	TNFi-exposed	Include adalimumab		
	Treatment sequencing	No change	INF-ADA-CT INF-VED-CT INF-TOF-CT VED-ADA-CT	GOL-ADA-CT GOL-VED-CT GOL-TOF-CT ADA-VED-CT

			TOF-ADA-CT	ADA-TOF-CT
NMA models	Remission and response rates	Use RE except for TNFi-experienced maintenance (RE would not run)	FE for both subgroups, induction and maintenance	
		No change	Use TNFi-failed for both vedolizumab and tofacitinib with TNFi-experienced for adalimumab	
	Serious infections	Frequentist random effects NMA model	Bayesian random effect model	
Utilities	Sources for pre and post-surgery health states	Same as company	• Swinburn et al. • OCTAVE 8 weeks • OCTAVE 52 weeks	
Resource use and costs	Drug stopping rule	Same as company	Additional OP visits to assess response within 8 weeks	
	Conventional drug usage	Same as company	Patient use of mesalazine: 50.3% (CT), 46.2% (concurrent). No other aminoslicylates	
	Health state resource use	Same as company	Reduced admissions, outpatient follow up and endoscopy	
	Drug administration costs	Same as company	Assume 1 OP visit at start of treatment for training on subcutaneous injections	
	Hospitalisation and surgery costs	NHS Reference costs + cost of stoma care post surgery (Buchanan et al. uprated for inflation)	Buchannan et al. estimate of surgery cost (uprated to 2016/17 prices) – includes repeat procedures	
Surgery	Incidence rate	Same as company	Chhaya et al.	
	Complications	Same as company	Tappenden et al.: Probabiity of perioperative complications (elective 0.2386; emergency	

			0.2614), probability of post surgery complications (0.173)

The results of the ERG's preferred assumptions are presented in Table 5. Collectively, our preferred assumptions give very similar results to the company's model. TNF-inhibitors remain dominated (with higher costs and fewer QALYs) than tofacitinib in both the sub-groups. While the pairwise ICER for tofacitinib compared with vedolizumab fall in the south-west quadrant (meaning tofacitinib is less effective but also less costly than vedolizumab) in the TNFi-naïve subgroup; in patients with prior exposure to TNFi, vedolizumab is dominated by tofacitinib under our preferred set of assumptions.

Table 5 Cost effectiveness: ERG preferred assumptions (with Tofacitinib PAS)

TNFi- naïve			
Conventional	[redacted]	[redacted]	£7,815
Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
Golimumab	[redacted]	[redacted]	Tofacitinib dominant
Infliximab	[redacted]	[redacted]	Tofacitinib dominant
Tofacitinib	[redacted]	[redacted]	--
Vedolizumab	[redacted]	[redacted]	£607,571 (SW)
TNFi-exposed			
Conventional	[redacted]	[redacted]	£9,389
Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
Tofacitinib	[redacted]	[redacted]	--
Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

1 Introduction to ERG Report

This report is a critique of the company's submission (CS) to NICE from Pfizer on the clinical effectiveness and cost effectiveness of tofacitinib for moderately to severely active ulcerative colitis. It identifies the strengths and weakness of the CS. Clinical experts were consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the CS was requested from the company by the ERG via NICE on 12 June 2018. A response from the company via NICE was received by the ERG on 27 June 2018 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

2.1 Critique of the company's description of the underlying health problem

The CS provides a generally clear and accurate overview of ulcerative colitis (CS section B.1.3).

Aetiology

Ulcerative colitis is a chronic inflammatory disease that is characterised by relapsing and remitting mucosal inflammation which typically affects the rectum and extends proximally to affect either a variable area of the colon, or its entire mucosal surface.[1,2] Ulcerative colitis is classified as proctitis, left-sided colitis, or extensive colitis, according to its maximal extent seen on colonoscopy. The CS (citing the British Society of Gastroenterology (BSG) guidelines for the management of inflammatory bowel disease in adults[3]) states that about 50% of patients with ulcerative colitis have a relapse in any year (CS section B.1.3.1). The NICE scope concurs that an estimated 30–60% of people with ulcerative colitis will have at least one relapse per year, of which about 80% are mild to moderate and about 20% are severe. However, the BSG guidelines[3] and the NICE scope do not specify the sources of these data. Patients with more extensive disease are at greater risk of developing dysplasia or colorectal cancer, and are generally advised to have surveillance colonoscopy.[4]

The pathogenesis of ulcerative colitis is complex and multifactorial, involving genetic predisposition, defects of the intestinal epithelial barrier, dysfunction of immune responses, and environmental factors.[2] The CS emphasises the importance of understanding the role of the

immune system and inflammatory cascade for understanding the disease and the role of current and future treatment options.

Risk factors

Risk factors for ulcerative colitis are not specified in the CS, but include: a family history of inflammatory bowel disease; Jewish ethnicity; the use of oral contraceptives, hormone replacement therapy, or non-steroidal anti-inflammatory drugs (NSAIDs); and former cigarette smoking. Conversely, in active smokers and some people who have had an appendectomy the risk of developing ulcerative colitis is reduced. Males and females do not differ in their risk of developing ulcerative colitis.[2,4]

Symptoms

Patients typically present with bloody diarrhoea, and some may also have rectal bleeding, urgency, faecal incontinence, nocturnal defecation and fatigue. Greater severity and extent of disease are associated with worsening bloody diarrhoea and the development of systemic signs, but any extent of colitis can be associated with constitutional symptoms, including fatigue and fever.[2,4]

Diagnosis

A gold standard for diagnosing ulcerative colitis is not available. Diagnosis is based on the history of symptoms, endoscopic findings on colonoscopy, histology, and excluding other causes of colonic inflammation (e.g. infection).[1] The key feature of ulcerative colitis on endoscopy is a diffuse continuous mucosal inflammation of the rectum and a variable extent of the colon. Other typical findings include erythema, loss of the normal vascular pattern, bleeding, erosions and ulcerations. The extent of inflammation observed on colonoscopy is related to the risk of disease complications.[1,2] Ulcerative colitis may be diagnosed at any age, but most commonly affects adults aged in their 20s to 40s (CS section B.1.3.1).

Severity of ulcerative colitis is classified as mild, moderate or severe based on a combination of factors which include, among others, the number of bowel movements per day and presence or absence of blood in the stool.[5,6]

Incidence and prevalence

The UK has among the highest incidence and prevalence rates of ulcerative colitis in the world.[7] The incidence of ulcerative colitis in the UK has been estimated at around 13.9 per 100,000 people, with a prevalence around 243 per 100,000 people. The most recent estimate available, for 2011, suggests that there were approximately 146,000 people in the UK who had ulcerative colitis.[3,7] The CS acknowledges that this may be a substantial underestimate, given the broad age of onset and lifelong duration of the condition (CS section B.1.3.1).

2.2 Critique of the company's overview of current service provision

Current treatments for moderately to severely active ulcerative colitis may be pharmacological or surgical, with all patients managed pharmacologically initially, before surgery in some cases. Clinical advice to the ERG is that surgery is reserved for patients who are non-responsive to the available drug treatments. Rarely, surgery may be carried out earlier if absolutely necessary, e.g. if a patient has a high risk of colorectal cancer.

Patients with moderately to severely active ulcerative colitis are typically managed according to a step-up approach based on the patient's history, treatment response and tolerance of individual therapies. Patients who have an inadequate response to conventional therapies (aminosacylates, corticosteroids or thiopurines) may be offered a biological therapy (a tumour necrosis factor (TNF) inhibitor or the anti-integrin agent vedolizumab).[8,9]

The CS briefly describes the clinical pathway of care (CS section B.1.3.3; discussed further below in section 2.3) but does not mention the staff, infrastructure or other resources associated with current service provision for patients with moderately to severely active ulcerative colitis, or whether these would change if tofacitinib is recommended for patients in the NHS. Clinical advice to the ERG is that a nurse-led service is used for intravenous therapies. This would not be applicable for tofacitinib which is administered orally. Tofacitinib can be taken with or without food, and the tablet can be crushed if patients have swallowing difficulties, so we assume that treatment self-administration by patients would be straightforward.

2.3 Critique of the company's definition of the decision problem

Population

The population stated in the NICE scope is "people with moderately to severely active ulcerative colitis who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy (oral corticosteroids and/or immunosuppressants) or a TNF-alpha inhibitor". This is consistent with the indication as specified in the Summary of Product Characteristics (SmPC),[10] as acknowledged by the company in CS Table 2.

In their decision problem table (CS Table 1) the company gives a broader description of the population, as "people with moderately to severely active ulcerative colitis". This description is applied to the NICE scope column within the decision problem table, thereby inaccurately reflecting the NICE scope. The company confirmed that this is a semantic error in CS Table 1 and that the decision problem does reflect the NICE scope (clarification response A1).

We note that the pivotal Phase III trial populations (in OCTAVE 1, OCTAVE 2 and OCTAVE Sustain) are consistent with the population definition as given in the NICE scope. However, according to the trial publication and protocol,[11] patients in the tofacitinib Phase II trial did not have to be intolerant of, or have had an inadequate response or loss of response to conventional therapy. The indication for tofacitinib in the Phase II trial therefore does not appear to be consistent with the NICE scope. In response to a clarification question from the ERG, the company stated that patients were only included in the Phase II trial if they continued to have moderate to severe disease despite previous treatment, and the company provided supporting data on the baseline characteristics of the Phase II trial participants listing the proportions who had failed prior treatments (clarification response A2). Although it is not clear from the CS or trial publication, the Phase II trial therefore does appear to meet the NICE scope.

The population in the pivotal OCTAVE Induction trials had a mean age of around 40-42 years (CS Table 15) and age ranged from 18 to 81 years [Table 13 in each clinical study report (CSR)]. Expert advice from one advisor to the ERG is that patients presenting in NHS clinical practice would typically be younger than the mean age in the trials, with the peak age at presentation being nearer 20 years on average. Although younger patients tend to have more severe ulcerative colitis;[4] the ERG's clinical advisor suggested that the age difference between the trials and clinical practice would be unlikely to affect patients' disease characteristics or their treatment.

Intervention

The intervention is tofacitinib citrate, a 5 mg or 10 mg oral tablet, brand name Xeljanz®. The description of the intervention in CS Table 2 is consistent with the SmPC.[10] The positive CHMP opinion, which is consistent with the NICE scope, was adopted by the EMA on 31st May 2018 for tofacitinib 10 mg to be used in the treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, lost response, or were intolerant to either conventional therapy or a biologic agent.

Tofacitinib is a small-molecule selective inhibitor of the Janus kinase (JAK) family of tyrosine kinases. The inhibition of JAKs by tofacitinib attenuates the signalling of several interleukins and type I and II interferons, which leads to modulation of the immune and inflammatory response in ulcerative colitis.[10] The mode of action of tofacitinib, including its role in inhibition of the JAK-STAT pathway, is summarised in CS section B.1.3.4.

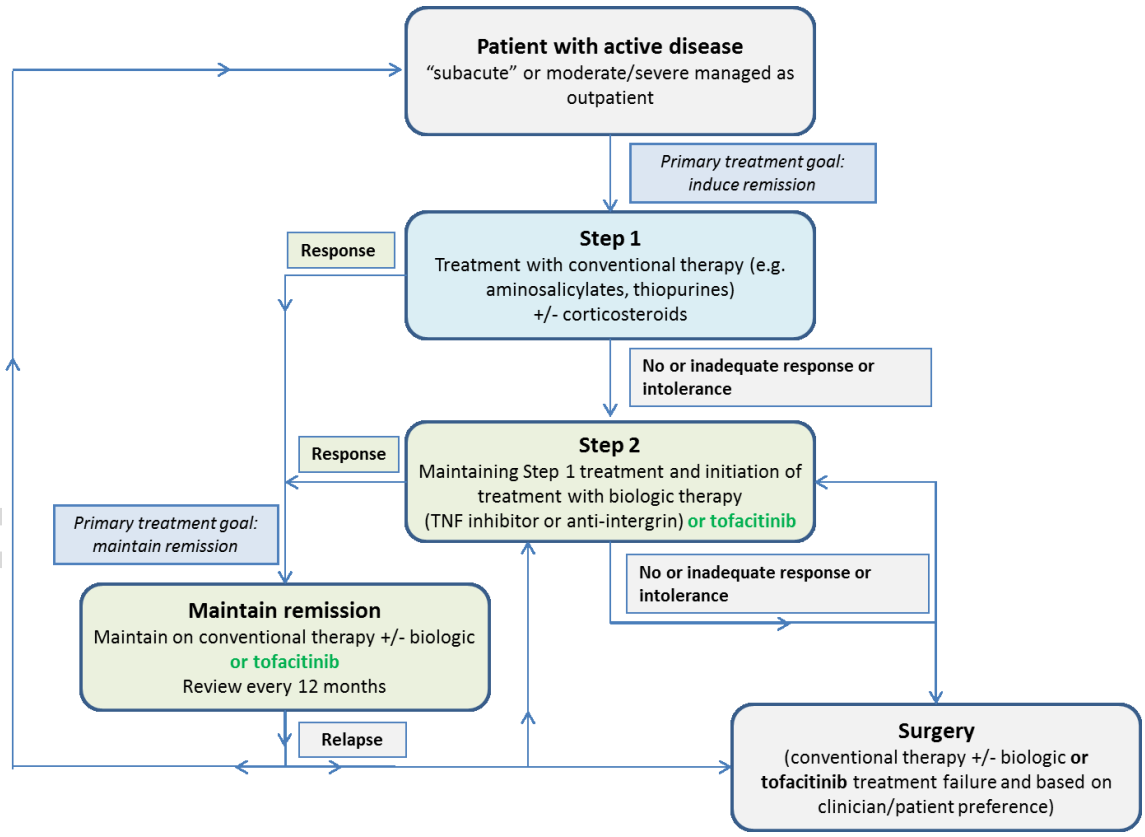
Comparators

The comparators in the company's decision problem (CS Table 1) are as specified in the NICE scope. These are:

- Conventional therapy, which may include a combination of aminosalicylates (sulfasalazine, mesalazine, balsalazide or olsalazine), corticosteroids (beclometasone, budesonide, hydrocortisone or prednisolone), thiopurines (mercaptopurine or azathioprine);
- TNF inhibitors (infliximab, adalimumab and golimumab[8]);
- Vedolizumab[9] (an anti-integrin agent).

These comparators are all used in the NHS (although experts advising the ERG commented that there has been a decline in the usage of thiopurines). CS Figure 1 illustrates the stepwise manner in which these therapies would be used in clinical practice in the NHS, showing that the conventional therapies are the standard first-line approach (step 1) whilst biological therapies (step 2) would not be employed without first trying a conventional therapy. Experts advising the ERG agreed that CS Figure 1 does broadly reflect current NHS practice. In response to a clarification request from the ERG, the company stated that CS Figure 1 is based on current NICE guidelines and clinical practice, but it is "*a simplification of the clinical pathway as the treatment of ulcerative colitis is dependent on multiple factors, the including patient's medical*

history and clinical decision making on the appropriateness of therapies, and therefore may not adequately capture the nuances of clinical practice when comparing to the NICE scope" (clarification response A3). In their clarification response the company provided a simplified version of CS Figure 1 in order to better represent the position of tofacitinib in the treatment pathway in relation to the NICE scope (reproduced in Figure 1).



Source: company's clarification response A3

Figure 1 Proposed position of tofacitinib within the treatment pathway

Outcomes

The outcomes included in the CS are clinically meaningful and are consistent with the NICE scope and EMA guidance on methods for clinical trials in ulcerative colitis.[12] The primary outcome in the phase 3 OCTAVE trials was clinical remission whilst the primary outcome in the phase 2 trial was clinical response. HRQoL was a secondary outcome in all the tofacitinib trials, and mucosal healing was a secondary outcome in the phase 3 trials. Details of the outcome selection are discussed further below in section 3.1.4. In summary, the key issues noted by the ERG are:

- Time to surgical intervention, listed as an outcome in the NICE scope, is not reported in the CS as it was not assessed in the pivotal trials (CS Table 1)
- The CS only provides brief results from the Phase II trial, for the primary outcome only (further results were requested by the ERG)
- The CS does not report all of the patient-reported outcomes that were measured in the pivotal trials (although as noted below in section 3.1.3 this does not appear likely to have resulted in bias)

Other relevant factors

The NICE scope indicates that, if evidence allows, subgroups of people who have been previously treated with one or more biologics and people who have not received prior biologics should be considered. Although the company presents subgroup analyses in their submission (CS Appendix E) their focus is on subgroups of people by TNFi-exposure status. There is no subgroup analysis for subgroups of people by prior biologic therapy (biologic therapy would include not only the TNF inhibitors but also vedolizumab). Nevertheless the ERG is mindful that subgroups by TNFi-exposure status are important, particularly because the existing evidence base for comparator treatments has demonstrated that primary non-response and secondary non-response to TNFi agents are limitations of the existing therapies adalimumab, golimumab and infliximab.

The CS does not identify any inequities that could be associated with the provision (or non-provision) of tofacitinib (CS section B.1.4) and the ERG is not aware of any equality issues with tofacitinib.

3 CLINICAL EFFECTIVENESS

3.1 Critique of company's approach to systematic review

3.1.1 Description of company's search strategy

The company conducted six systematic literature searches (five of which are standard for an STA, plus an additional surgery review):

- Clinical effectiveness: start year unspecified ("no limits") to 16/11/2017
- Non RCT evidence: start year unspecified ("no limits") to 15/11/2017

- Cost effectiveness: start year unspecified (“no limits”) to October 2017
- Health Related Quality of Life: start year unspecified (“no limits”) to 15/11/2017
- Cost and healthcare resource identification: start year unspecified (“no limits”) to 20/10/2017
- Surgery Literature Review - dates not given

The key literature searches were systematic, transparent, well documented and reproducible. A typographical error was found in line 18 of the cost effectiveness searches in Medline and Embase (“mdel*” instead of model*) however correct spelling elsewhere in both of these strategies coupled with accurate spelling in the Cochrane search, should have counteracted this error. The additional surgery review was undertaken to inform the economic analysis on the probability of colectomy and ensuing complications. This search is not fully documented, although a synopsis of the terms used are recorded which is acceptable. Key conferences were adequately searched and ongoing trials were sought via clinicaltrials.gov and the WHO International Clinical Trials Registry Platform.

Overall the searches are deemed fit for purpose. However, all searches in the CS are between six to eight months out of date. Due to time constraints the ERG has prioritised updating the cost effectiveness, HRQoL, and cost & healthcare resource searches, replicating the documented strategies. Two additional cost-effectiveness papers were identified by the ERG’s updated search (see Section 4.2) but no additional relevant references were identified by the updated HRQoL or healthcare resource searches.

To identify any new clinical effectiveness evidence, we conducted a rapid search using HDAS (NICE Healthcare Databases Advanced Search) and Delphis (a broad-scope University of Southampton search engine powered by Ebsco). Four additional full-text publications on tofacitinib clinical effectiveness and/or safety which are not listed in the CS were identified. These reported on: the phase II trial;[13] a NMA comparing tofacitinib against biologic therapies;[14] an analysis of HRQoL in the OCTAVE trials using the Inflammatory Bowel Disease Questionnaire (IBDQ) and SF-36;[15] and a subgroup analysis of effectiveness and safety outcomes from the OCTAVE trials in East Asian participants.[16] Two new conference abstracts reporting results from the OCTAVE trials were also identified.[17,18] These new publications largely duplicate information already present in the CS, or are not directly relevant to the current scope.

3.1.2 Statement of the inclusion/exclusion criteria used in the study selection.

The company provides a clear description of the eligibility criteria (both inclusion and exclusion criteria) for the systematic literature review (SLR) (Appendix D.1.1.3, Table 83). The SLR aimed to identify clinical effectiveness and safety evidence not only for tofacitinib but also for relevant comparators, which could potentially be used in network meta-analysis (NMA).

The population eligibility criteria are specified as 'adult patients with moderately and/or severely active UC' (either treatment-naïve or treatment-experienced). The criteria exclude patients with non-specific inflammatory bowel disease (IBD) and those with acute severe ulcerative colitis or ulcerative colitis exacerbation/flare requiring hospitalisation as well as paediatric patients and animal/in vitro studies. This population is reflective of the decision problem (CS Table 1), although, as noted in section 3.2 above, the population in the company's decision problem is broader than the population specified in the NICE final scope.

The company confirmed (clarification response A1) that the population had been inaccurately described in their decision problem and that their interpretation of the population is in fact consistent with that described in the NICE scope. Whilst the population eligibility criteria as explicitly stated in CS Table 83 are wider than the NICE scope, the ERG is satisfied that the populations in the studies finally included in the company's SLR are consistent with the NICE scope.

The interventions and comparators for the company's SLR generally reflect the NICE scope, the anticipated licensed indication for tofacitinib and current NHS practice. Calcineurin inhibitors and surgical intervention were excluded as comparators in the SLR, whereas placebo was included as a comparator (not specified in the NICE final scope). Clinical expert advice to the ERG suggests that the exclusion of calcineurin inhibitors and surgical intervention is reasonable. It is presumed that the company included placebo as a comparator because it is the comparator in the OCTAVE trials, on which the clinical evidence for tofacitinib is based.

To be included studies had to RCTs (both blinded and open-label RCTs were eligible) and had to report at least one of the following outcome measures: response, remission, mucosal healing, relapse or loss of response/remission, discontinuation, treatment duration, rates of surgical

intervention, time to surgical intervention, Mayo score/Disease activity index, hospitalisation, mortality, adverse events (AEs), serious AEs (SAEs), treatment-related AEs, injection or infusion site reaction and HRQoL (EQ-5D, SF-36, IBDQ).

Setting did not form part of the company's eligibility criteria for the SLR. The company placed no limits on the quality of the included RCTs in their eligibility criteria. The ERG agrees this is appropriate.

The CS provides a flow diagram illustrating the number of records identified in the SLR and reasons for the exclusion of studies at the full text screening stage (CS Appendix D.1.2.1, Figure 39). References linked to the included studies are listed in CS Appendix D.1.2.1, Table 84. It was difficult for the ERG to equate these to the 102 references listed in the CS. In response to a request from the ERG, the company provided the information more clearly, including a correction to a referencing error (clarification response A6 and clarification response Appendix C). A list of the 137 references excluded at the full text stage of the reference screening process is not included in the CS. This was subsequently provided by the company (clarification response A5 and clarification response Appendix B).

The evidence was limited to studies published in the English language, which the ERG considers appropriate for a submission to NICE. However, the company did not discuss any potential bias that may have arisen from the restrictions of the eligibility criteria specified for the SLR. The ERG notes that RCTs are, by design, potentially at a lower risk of bias than other study design and that all the included RCTs were subject to quality assessment using the concise critical appraisal checklists provided by NICE in the STA user guide (CS D.1.2.2.2 Table 86).

3.1.3 Identified studies

The company's SLR included 21 RCTs. In four of these the intervention was tofacitinib:

- one Phase II RCT (treatment arms: tofacitinib 0.5 mg, 3 mg, 10 mg, 15 mg and placebo)[11]
- two identical Phase III induction RCTs (OCTAVE Induction 1 and OCTAVE Induction 2; treatment arms: tofacitinib 10 mg and placebo) (also a 15 mg arm which was discontinued)[19]
- one maintenance RCT (OCTAVE Sustain; treatment arms: 5 mg, 10 mg and placebo)[19]

In all four RCTs the comparator was placebo. All these RCTs were used in support of the company's application for a marketing authorisation and were sponsored by Pfizer, the manufacturer of tofacitinib.

The Phase II trial is not described in detail in the CS but it is included in the company's NMA (CS section B.2.9) and data from this trial are also included in the adverse events section (CS Appendix F Table 166). As the Phase II trial was a small dose-finding study with 194 patients, of whom only 33 received the licensed 10 mg BID dose (company clarification response A16), the CS focuses on the Phase III trials. The ERG agrees that this is reasonable and accordingly the current ERG report also focuses primarily on the Phase III trials.

It was unclear to the ERG from the description of the Phase II trial population reported both in the CS and in the trial publication whether this matched the NICE scope. The company confirmed that it does match the scope, as "*patients were only included if they continued to have moderate to severe disease despite previous treatment*" (clarification response A2). In addition, the company provided a table detailing the failed drug treatments at baseline (clarification response Table 1) and full details of the inclusion and exclusion criteria (clarification response Appendix A).

The number of centres in the studies ranged from 51 (Phase II trial) to 297 (OCTAVE Sustain), but it should be noted that a number of centres in the Phase III trials randomised just one patient (16 centres in OCTAVE 1; 25 centres in OCTAVE 2; and 66 centres in OCTAVE Sustain[19]). While each study included some patients from the UK, this number was low [redacted].

OCTAVE 1 and 2 were double-blind, randomised placebo-controlled tofacitinib induction trials with an 8 week treatment phase, and used identical methods (see Table 6).

In addition to the criteria listed above, patients had to have moderately to severely active disease (6 to 12 on the Mayo score, with a rectal bleeding sub-score of 1 to 3 and an endoscopic sub-score of 2 or 3). Prohibited therapies included TNFi therapies within 8 weeks of baseline; azathioprine, methotrexate, and 6-mercaptopurine within 2 weeks; and ciclosporin and intravenous corticosteroids (CS Tables 9 and 10). Permitted concomitant medications for ulcerative colitis included oral aminosalicylates (stable dose ≥4 weeks prior to baseline and

during study); oral glucocorticoids (maximum dose 25 mg per day of prednisone or a prednisone equivalent; stable dose ≥2 weeks prior to baseline and during study); and antibiotics used for chronic ulcerative colitis (e.g., metronidazole and rifaximin; stable dose ≥2 weeks prior to baseline and during study). Eligible patients were randomised on a 4:1 ratio to 10 mg twice a day (BID) of oral tofacitinib or placebo. The trials initially included a third treatment arm of 15 mg BID oral tofacitinib, but this was discontinued prior to full recruitment based on feedback from regulatory authorities. The company clarified that patients assigned to the tofacitinib 15 mg BID arm continued to receive blinded treatment for the remainder of the induction trial period and, of these, 19 patients were eligible to enter the OCTAVE Sustain trial (clarification response A10).

Patients were eligible to join the OCTAVE Sustain trial if they: met the eligibility criteria of the OCTAVE Induction trials; completed the 8 weeks of induction therapy; and met the clinical response criteria for the induction trials (see Figure 2). This was a randomised, double-blind, placebo-controlled trial lasting 52 weeks. Eligible patients from OCTAVE 1 and 2 were randomised in a 1:1:1 ratio to receive either 5 mg or 10 mg BID oral tofacitinib, or placebo.

The OCTAVE induction and maintenance trials conform to a re-randomisation design. That is, participants are first randomised to tofacitinib or placebo groups of the OCTAVE Induction study. Following 8-weeks of induction therapy, those participants who have met clinical response criteria are re-randomised into one of the three arms of the OCTAVE Sustain maintenance study. An alternative, utilised by some of the other clinical trials that have taken place in this disease area, is a treat-through design. In a treat-through trial participants are randomised to induction therapy and outcomes are measured at the end of the induction phase. Participants then continue in their original randomised group into the maintenance phase and outcomes are measured again at the end of the maintenance phase.

In addition to the four RCTs the OCTAVE study programme also includes the OCTAVE Open extension study which is ongoing.

Table 6 Summary characteristics of tofacitinib RCTs

Phase II trial[11] (efficacy/dose RCT)		**OCTAVE 1[19] (induction RCT)**		**OCTAVE 2[19] (induction RCT)**		**OCTAVE Sustain[19] (maintenance RCT)**		**OCTAVE Open[20] (extension study)**
Tofacitinib 0.5 mg (n=31) 3 mg BID (n=33) 10 mg BID (n=33) 15 mg BID (n=49)	Placebo (n=48)	Tofacitinib 10 mg BID (n=476)[a]	Placebo (n=122)	Tofacitinib 10 mg BID (n=429)[a]	Placebo (n=112)	Tofacitinib 10 mg BID (n=197) 5 mg BID (n=198)	Placebo (n=198)	Tofacitinib[b] 10 mg BID ([redacted]) 5 mg BID ([redacted])
Design: randomised, double- blind, placebo-controlled trial (2:2:2:3:3 ratio tofacitinib 0.5 mg: 3mg: 10 mg: 15 mg: placebo)		*Design:* identical randomised, double-blind, placebo-controlled trials (4:1 ratio tofacitinib: placebo, stratified according to previous treatment with TNFi therapies, glucocorticoid use at baseline, and geographic region)				*Design*: randomised, double-blind, placebo-controlled trial (1:1:1 ratio tofacitinib 5 mg: tofacitinib 10 mg; placebo)		*Design:* open-label extension
Location: 51 sites worldwide (UK = 2, [redacted][d])		*Location:* 144 sites worldwide (UK = 2, [redacted])		*Location:* 169 sites worldwide (UK = 3, [redacted])		*Location:* 297 sites worldwide (UK = 5, [redacted])		*Location:* 215 sites worldwide (UK = 5)
Inclusion: • age ≥18 years • confirmed diagnosis of UC for ≥3 months • score of 6 to 12 on the Mayo scale and		*Inclusion:* • age ≥18 years • confirmed diagnosis of UC for ≥4 months • moderately to severely active disease (6 to 12 on the Mayo score, with a rectal bleeding sub score of 1 to 3 and an endoscopic sub-score of 2 or 3)				*Inclusion:* • entry criteria for the Induction trials • completed 8 weeks induction therapy		*Inclusion:* • completed or demonstrated treatment failure in the OCTAVE Sustain maintenance study or

• moderately or severely active disease (i.e. Mayo-endoscopic findings sub-score of 2 or 3, respectively)	• treatment failure with/to or unacceptable side effects from treatment with ≥1 of: ○ oral or intravenous glucocorticoids ○ azathioprine ○ mercaptopurine ○ infliximab ○ adalimumab	• met clinical response criteria in OCTAVE Induction 1 and 2	• non-responders after completed 8 weeks of treatment in the OCTAVE 1 & 2 induction studies
Background therapy: Oral mesalamine or oral prednisone at a stable dose of ≤ 30 mg per day	*Background therapy:* oral aminosalicylates at a stable dose for ≥4 weeks prior to baseline and during study and oral glucocorticoids (at a maximum dose of 25 mg per day of prednisone or a prednisone equivalent) at a stable dose for ≥2 weeks prior to baseline and during study. Patients on chronic treatment for UC with antibiotics (e.g. metronidazole and rifaximin) were eligible if dose was stable for ≥2 weeks prior to baseline and during study.	*Background therapy:* oral amino-salicylates (stable dose) and chronic treatment for UC with antibiotics (e.g., metronidazole, rifaximin). Oral glucocorticoids at study entry were tapered mandatory starting 1st week at specified rate depending on starting dose (daily dose of prednisone or equivalent was decreased at a rate of 5 mg per week until dose reached 20 mg/day, then 2.5 to 5.0 mg per week until dose reached 10 mg/day, then by 2.5 mg per week until the dose was 0 mg).	*Background therapy:* oral aminosalicylates (stable dose) and chronic treatment for UC with antibiotics (e.g., metronidazole, rifaximin). Oral glucocorticoids at study entry were tapered mandatory as per the OCTAVE Sustain schedule.

Length of follow-up: 8 weeks of treatment and 4 weeks follow-up	*Length of follow-up:* 9 weeks (primary efficacy endpoint at 8 weeks)	*Length of follow-up:* 53 weeks (primary efficacy endpoint at 52 weeks)	*Length of follow-up:* up to 6 years (12-month interim results reported)

Sources Sandborn et al.[11], CS Table 7, 8, 9, 10, 13 and 14, and B.2.6.3.1

BID, twice daily; NR, not reported; UC, Ulcerative colitis.

[a] 15 mg BID tofacitinib treatment was discontinued based on feedback from regulatory authorities (OCTAVE 1: n=38, OCTAVE 2: n=18)

[b] Three subpopulations received tofacitinib 10 mg ([redacted]) in the open label extension study: Induction non-responders tofacitinib 10 mg [redacted]; maintenance completers tofacitinib 10 mg [redacted], maintenance treatment failures 10 mg ([redacted] comprising participants from OCTAVE Induction 1 and 2 who withdrew from OCTAVE Sustain due to treatment failure on tofacitinib (5 mg, [redacted]; 10 mg, [redacted]) or placebo ([redacted]). One subpopulation received tofacitinib 5 mg in the open label extension study: Maintenance: remission tofacitinib 5 mg [redacted]. Note that there appears to be a typographical error in CS Table 8 where the number of patients receiving tofacitinib 10 mg is given as [redacted].

[d] [redacted]

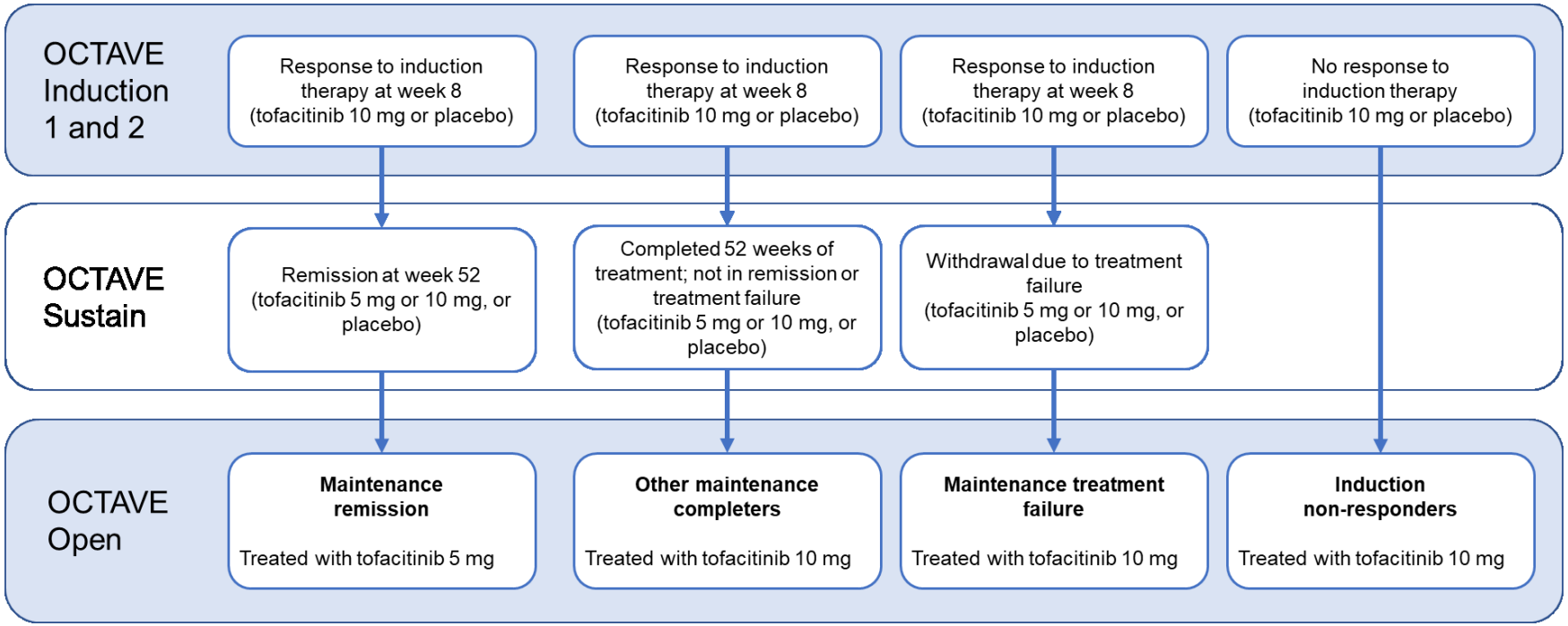


Figure 2 Participant flow in the OCTAVE trials (Source CS Figure 5)

3.1.3.1 OCTAVE RCTs baseline characteristics

The CS states that there were no significant differences in baseline characteristics between groups within each trial, apart from two exceptions. In OCTAVE 2, there was a statistically significant higher proportion of male patients in the tofacitinib group compared with the placebo group (tofacitinib 60.4% versus placebo 49.1%, p = 0.03). However, we note that sex is not considered to be a prognostic factor for ulcerative colitis[2] and a similar difference was not evident in OCTAVE 1. In OCTAVE Sustain, there was a significant difference in smoking status among the tofacitinib and placebo groups (p = 0.03), with a higher proportions of people who had never smoked and lower proportions of current smokers in the two tofacitinib groups than in the placebo group. Smoking is known to be a modifying factor in ulcerative colitis, with former cigarette smoking being a strong risk factor, yet active smokers are less likely to develop ulcerative colitis than former and non-smokers,[2] and active smoking is associated with milder disease.[21] If this imbalance were to affect the results it could disadvantage the tofacitinib groups, since these had a lower proportion of current smokers, yet it might also disadvantage the placebo group since this had a higher proportion of former smokers. We note that the difference between the tofacitinib and placebo arms in the number of patients who were current smokers amounted to only six patients (3.1 percentage points) whilst the difference in the number who were former smokers was only 10 patients for the tofacitinib 10 mg comparison (4.9 percentage points), although it was 24 patients (12.2 percentage points) for the tofacitinib 5 mg comparison. On balance, the risk of selection bias being introduced as a result of these imbalances in smoking status within OCTAVE Sustain appears to be low. Table 7 provides a summary of the trial characteristics of all the tofacitinib trials, including the ongoing extension trial Open.

Generally, patient characteristics appear to be balanced across the different OCTAVE trials, although patients enrolled in OCTAVE Sustain had lower Mayo scores and C-reactive protein levels than in either OCTAVE 1 or 2 (CS Table 15). This may be reflective of patients having had to achieve a response in order to be eligible to join OCTAVE Sustain.

Table 7 Summary of baseline patient characteristics of the OCTAVE 1 and 2 and Sustain

Characteristic	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	TOF 10 mg (N=476)	Placebo (N=122)	TOF 10 mg (N=429)	Placebo (N=112)	TOF 5 mg (N=198)	TOF 10 mg (N=197)	Placebo (N=198)
Male sex, n (%) [a]	277 (58.2)	77 (63.1)	259 (60.4)	55 (49.1)	103 (52.0)	110 (55.8)	116 (58.6)
Age, years [b]	41.3±14.1	41.8±15.3	41.1±13.5	40.4±13.2	41.9±13.7	42.9±14.4	43.4±14.0
Induction trial group assignment, n (%)							
Placebo	—	—	—	—	22 (11.1)	24 (12.2)	24 (12.1)
Tofacitinib, 10 mg BID	—	—	—	—	170 (85.9)	167 (84.8)	167 (84.3)
Tofacitinib, 15 mg BID	—	—	—	—	6 (3.0)	6 (3.0)	7 (3.5)
Remission at maintenance trial entry, n (%)	—	—	—	—	65 (32.8)	55 (27.9)	59 (29.8)
Duration of disease, median yrs [b] (range)	6.5 (0.3–42.5)	6.0 (05–36.2)	6.0 (0.4–39.4)	6.2 (0.4–27.9)	6.5 (0.6–40.3)	6.8 (0.6–35.7)	7.2 (0.6–42.7)
Extent of disease, n/total n (%) [c,d]							
Proctosigmoiditis	65/475 (13.7)	19/122 (15.6)	67/428 (15.7)	16/111 (14.4)	28/196 (14.3)	33/196 (16.8)	21/198 (10.6)
Left-sided colitis	158/475 (33.3)	37/122 (30.3)	149/428 (34.8)	39/111 (35.1)	66/196 (33.7)	60/196 (30.6)	68/198 (34.3)
Extensive colitis or pancolitis	252/475 (53.1)	66/122 (54.1)	211/428 (49.3)	56/111 (50.5)	102/196 (52.0)	103/196 (52.6)	108/198 (54.5)
Total Mayo score [b,e]	9.0±1.4	9.1±1.4	9.0±1.5	8.9±1.5	3.3±1.8	3.4±1.8	3.3±1.8
Partial Mayo score [b,e]	6.3±1.2	6.5±1.2	6.4±1.3	6.4±1.2	1.8±1.3	1.8±1.3	1.8±1.4
C-reactive protein, median mg/litre [b] (range)	4.4 (0.1–208.4)	4.7 (0.1–82.5)	4.6 (0.2–156.0)	5.0 (0.2–205.1)	0.7 (0.1–33.7)	0.9 (0.1–74.3)	1.0 (0.1–45.0)
Oral glucocorticoid use at baseline, n (%) [b]	214 (45.0)	58 (47.5)	198 (46.2)	55 (49.1)	101 (51.0)	87 (44.2)	100 (50.5)
Previous treatment with TNFi, n (%) [c]	254 (53.4)	65 (53.3)	234 (54.5)	65 (58.0)	90 (45.5)	101 (51.3)	92 (46.5)

Characteristic	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	TOF 10 mg (N=476)	Placebo (N=122)	TOF 10 mg (N=429)	Placebo (N=112)	TOF 5 mg (N=198)	TOF 10 mg (N=197)	Placebo (N=198)
Previous treatment failure, n (%) [c,f]							
TNF antagonist	243 (51.1)	64 (52.5)	222 (51.7)	60 (53.6)	83 (41.9)	93 (47.2)	89 (44.9)
Glucocorticoid	350 (73.5)	98 (80.3)	303 (70.6)	83 (74.1)	145 (73.2)	149 (75.6)	151 (76.3)
Immunosuppressant [g]	360 (75.6)	83 (68.0)	301 (70.2)	75 (67.0)	143 (72.2)	141 (71.6)	129 (65.2)
White race, n (%) [h]	395 (84.6)	98 (83.1)	331 (80.3)	88 (83.0)	164 (84.5)	153 (81.8)	155 (80.3)
Weight, kg	72.9 (16.8)	72.7 (16.7)	74.4 (16.8)	73.2 (16.2)	73.4 (17.8)	74.6 (15.1)	76.2 (16.7)
Smoking status, n (%) [c,i]							
Never smoked	301 (63.2)	80 (65.6)	268 (62.5)	81 (72.3)	142 (71.7)	128 (65.0)	113 (57.1)
Current smoker	22 (4.6)	4 (3.3)	25 (5.8)	5 (4.5)	7 (3.5)	6 (3.0)	12 (6.1)
Former smoker	153 (32.1)	38 (31.1)	136 (31.7)	26 (23.2)	49 (24.7)	63 (32.0)	73 (36.9)

Source: CS Table 15

Footnotes: see next page

Footnotes for Table 2

[a] In the OCTAVE Induction 2 trial, there was a significant difference between groups in the proportion of male patients (p = 0.03).

[b] For the OCTAVE Sustain trial, the baseline values were obtained at the time of entry in the OCTAVE Sustain trial.

[c] For the OCTAVE Sustain trial, the baseline values were obtained at the time of entry into one of the induction trials (OCTAVE Induction 1 or 2).

[d] Data on extent of disease are missing for three patients.

[e] The total Mayo score ranges from 0 to 12 and the partial Mayo score (i.e., the total Mayo score excluding the endoscopic subscore) ranges from 0 to 9, with higher scores indicating more severe disease.

[f] Previous treatment failure was determined by the investigator.

[g] Immunosuppressants included agents such as azathioprine and mercaptopurine and did not include biologic agents (e.g., TNF antagonists) or glucocorticoids.

[h] Unspecified race was treated as missing data.

[i] In OCTAVE Sustain, there was a significant difference for smoking status among placebo and tofacitinib groups (*p* = 0.03).

In summary, the CS appears to have identified all relevant RCTs and has provided all relevant study publications electronically, although CSR for the phase II trial had to be requested by the ERG and NICE (Clarification question A4).

3.1.3.2 Non-randomised trials

The company conducted a SLR to identify non-RCT evidence (CS Appendix D.1.4.2.), in order to provide long-term evidence (over 12 weeks for induction and over 52 weeks for maintenance therapy) regarding the efficacy and safety of tofacitinib for the treatment of moderately to severely active ulcerative colitis, hence relevant to the decision problem.

The company included one open-label, ongoing tofacitinib extension trial of up to 6 years duration (OCTAVE Open - NCT01470612).[20] Patients could enter OCTAVE Open from the OCTAVE 1 and 2 Induction trials if they did not have a response or enter from the OCTAVE Sustain trial once they completed 52 weeks of follow-up or if they withdrew due to treatment failure. Consequently, OCTAVE Open has four distinct patient groups (as depicted in Figure 2):

- Induction non-responders: patients from OCTAVE 1 AND 2 who did not have a response to induction therapy and did not enter OCTAVE Sustain (all allocated to 10 mg BID tofacitinib in OCTAVE Open)
- Maintenance remission: patients with a response to induction therapy in OCTAVE 1 and 2 who were in remission at week 52 in OCTAVE Sustain (all allocated to 5 mg BID tofacitinib in OCTAVE Open)
- Maintenance completers: patients who at the end of 52 weeks of maintenance therapy in Sustain were not in remission but did not meet the definition of treatment failure (all allocated to 10 mg BID tofacitinib in OCTAVE Open)
- Maintenance treatment failure: patients with a response in OCTAVE 1 and 2 who withdrew from OCTAVE Sustain due to treatment failure on tofacitinib (all allocated to 10 mg BID tofacitinib in OCTAVE Open)

Patient disposition for OCTAVE Open is presented in a confidential table (CS Appendix D.1.4, Table 119), with demographic and baseline characteristics in Appendix L.1.5 Table 231 and baseline disease characteristics in Appendix L.1.5 Table 232.

Evidence from this trial (which is still ongoing) presented in the CS is predominantly for patients with 12-month data because 24-month data are currently only available for a small number of patients. The CS presents a summary of results in sections B.2.6.3.2 to B.2.6.3.5 with full endpoint results shown in CS Appendix L (Tables 233 to 236). Additionally, a table of 12-month interim data for treatment emergent adverse events (CS Appendix F, Table 167) is presented.

3.1.3.3 Ongoing studies

Apart from the OCTAVE Open trial reported above, which may provide more data within the next 12 months, the CS states that preliminary results from a phase IIIb/IV study of tofacitinib in patients with ulcerative colitis in stable remission (NCT03281304) may also be available within the next 12 months (CS B.2.11). Apart from the ClinicalTrials.gov Identifier, no other information is provided in the CS. Details on the clinical trials website for the trial are shown in Table 8.

Table 8 Ongoing phase IIIb/IV study of tofacitinib

Title:	A Phase 3b/4, Multi-center, Double-blind, Randomized, Parallel Group Study Of Tofacitinib (Cp-690,550) In Subjects With Ulcerative Colitis In Stable Remission
Aim:	To evaluate flexible dosing in patients with ulcerative colitis
Start date:	Nov 2017
Estimated completion date:	Nov 2019
Number randomised:	130
Intervention:	5 mg BID tablet 10 mg BID tablet

3.1.4 Description and critique of the approach to validity assessment

The ERG has assessed the methodological quality of the four tofacitinib RCTs using NICE's recommended criteria (Table 9). The seven questions in Table 9 relate to risks of different types of bias that could arise within the trials. The company has phrased some of their quality assessment questions slightly differently to those recommended by NICE (indicated where appropriate in the table) for the quality assessment based on only the three OCTAVE trials (CS Table 19) and the quality assessment used for all the trials included in the NMA (CS Table 86). For question 5, about imbalances in dropouts, two versions of the question are given in the CS. For clarity we have labelled these as 5a and 5b, since the risk of bias interpretation differs according to how the question is phrased.

The OCTAVE Induction and Sustain RCTs appear to have a low risk of selection bias (questions 1 to 3), as the populations were generally well-balanced across the trial arms.

Participants and investigators in all four tofacitinib RCTs were blinded to the treatment allocations and so the risk of performance or detection bias that could arise through knowledge of treatment allocations appears to be low (question 4). Details of the blinding method of endoscopy readers are provided by the company in clarification response A11b.

The risk of attrition bias as a result of any treatment-related imbalances in dropouts between trial arms appears to be low in the OCTAVE Induction trials (question 5). However, there were some imbalances in dropouts in both the Phase II trial and the OCTAVE Sustain trial which

might have introduced bias. The ERG is uncertain about the direction and magnitude of any bias since there were several different reasons why patients withdrew. The CS acknowledges these imbalances in OCTAVE Sustain but not in the Phase II trial, and does not comment on whether they would have introduced bias.

Not all protocol-specified clinical effectiveness outcomes that were measured in the OCTAVE trials and Phase II trial are reported in the trial publications (question 6). However, the key outcomes are reported and the risk of reporting bias in these trials appears to be low.

The company used the "full analysis set" (FAS) as the primary analysis population but this was defined differently in the Phase II trial and the OCTAVE trials (question 7). The OCTAVE trials conducted an appropriate analysis in which the FAS was consistent with the ITT principle and accounted for missing remission data appropriately. Therefore, the risk of bias in the primary outcome, and all other outcomes analysed according to the FAS, appears to be low in the OCTAVE trials. In contrast, the Phase II trial conducted analyses in which the FAS was defined as being equivalent to a modified ITT population that did not include all randomised patients and not all missing data were included in analyses. As such, there is a risk of attrition bias in the Phase II trial, but with unclear direction and magnitude.

In summary, the OCTAVE Induction trials appear to be generally at low risk of the five types of bias assessed. The OCTAVE Sustain trial and the Phase II trial also appear to be at low risk of selection, performance, detection and reporting biases but could be at risk of attrition bias as a result of unbalanced dropouts between the tofacitinib and placebo arms.

Table 9 Company and ERG assessments of trial quality

Quality assessment question	**Judge-ments**	**Phase II trial**	**OCTAVE 1 & 2**	**OCTAVE Sustain**
1. Was randomisation carried out appropriately? ("yes" indicates low risk of selection bias)	**CS:**	Yes	Yes	Yes
	ERG:	Yes	Yes	Yes
ERG comments: A central randomisation method was employed (CS Tables 9, 13, 86) (not reported in the CS for the phase 2 trial, but stated in the trial publication).				
2. Was the concealment of treatment allocation adequate? ("yes" indicates low risk of selection bias)	**CS:**	Yes	Yes	Yes
	ERG:	Yes	Yes	Yes

ERG comments: Allocation concealment is not explicitly reported in the CS, CSRs or trial publications. However, central randomisation was telephone-based so the ERG assumes that the allocation sequence could not have been known to, foreseen, or influenced by the study investigators prior to them dialling in to receive each patient's random allocation to TOF or PBO.				
3. Were the groups similar at the outset of the study in terms of prognostic factors? ("yes" indicates low risk of selection bias)	**CS:**	Yes	Yes	Yes
	ERG:	Yes	Yes	Yes
ERG comments: Phase II trial (reported in the trial publication): The only statistically significant difference at baseline was in glucocorticoid use (placebo 58%, tofacitinib 10 mg 27%; p=0.03), although due to the small overall sample size this reflects a difference of only six patients. OCTAVE 1 and 2 (CS Table 15): The CS states that the only statistically significant difference between groups was in the proportion of male patients in OCTAVE 2. Where imbalances of >5% between arms occurred in the induction trials these did not systematically affect both trials. OCTAVE Sustain (CS Table 15): The CS states that the only statistically significant difference between groups was in smoking status. The proportion who never smoked differed between all three arms: 71.7% in the TOF 5 mg arm, 65.0% in the TOF 10 mg arm, and 57.1% in the PBO arm, but the difference was relatively small for TOF 10 mg vs PBO.				
4. Were the care providers, participants and outcome assessors blind to treatment allocation? ("yes" indicates low risk of performance and detection bias)	**CS:**	Yes	Yes	Yes
	ERG:	Yes	Yes	Yes
ERG comments: The CS states that the OCTAVE trials were patient-, investigator-, and sponsor-blinded (CS Tables 9 & 13) and the phase 2 trial was double blind (CS Table 86). The ERG assumes that "investigators" and "double blind" cover both the care providers and the outcome assessors, although this is not explicit in the CS. The method of blinding was to use a matching placebo tablet. NB this question is worded slightly differently in CS Table 86 compared to CS Tables 9 and 13, but in both cases a "yes" answer would suggest a low risk of bias.				
5a. Were there any unexpected imbalances in drop-outs between groups? (question as phrased in CS Tables 9 and 13; "no" indicates low risk of attrition bias)	**CS:**	Not reported	No	No
	ERG:	Yes	No	Yes
5b. Were discontinuations similar between groups? (question as phrased in CS Table 86; "yes" indicates low risk of attrition bias)	**CS:**	No	Yes	No
	ERG:	No	Yes	No
ERG comments: Phase II trial: Lower discontinuation rate in the TOF 10 mg group (6%) than the PBO group (27%) (reported in the publication appendix). The TOF discontinuations (n=2) were both due to lack of efficacy. The PBO discontinuations were due to lack of efficacy (n=5), AE (n=3), protocol violation (n=2), consent withdrawn (n=2) and loss to follow up (n=1). OCTAVE 1 & 2: Slight imbalances in discontinuations but these were not consistent in direction across both induction trials (OCTAVE 1: PBO 3.3%, TOF 6.5%; OCTAVE 2: PBO 13.4%, TOF 7.5%) (CS Table 17). OCTAVE Sustain: As noted in CS Table 18, discontinuation rates differed between PBO (73.2%), TOF 5 mg (43.9%) and TOF 10 mg (35.7%). The main reason for discontinuation was lack of clinical response (66.7%, 35.4%, 27.0% respectively); relatively few patients discontinued due to AE (<5% in each arm).				
	CS:	No	No	No

6. Is there any evidence to suggest that the authors measured more outcomes than they reported? (question phrased in CS Table 86 as "unreported outcomes suspected?") ("no" indicates low risk of reporting bias)	**ERG:**	No	Yes, but low bias risk	Yes, but low bias risk
ERG comments: The OCTAVE trial publication (as acknowledged in section 4 of the supplementary appendix)[19] does not report all outcomes that were measured in the OCTAVE Induction and Sustain trials. The publication does not explicitly state reasons why some outcomes were not reported. However, the most important clinical effectiveness outcomes are reported. Where outcomes were measured but not reported in the trial publication (e.g. several patient-reported outcome measures), these appear to favour TOF 10 mg over PBO, according to results in the CSRs. As such, the non-reporting of some outcome measures in the trial publication would appear unlikely to have introduced bias.				
7. Did the analysis (1) include an intention-to-treat (ITT) analysis? (2) If so, was this appropriate and (3) were appropriate methods used to account for missing data? [sub-questions numbered by ERG] ("yes" indicates low risk of attrition bias)	**CS:**	Stated ITT (CS Table 86) but see ERG comment below	Yes	Yes
	ERG:	1. No 2. NA 3. NA	1. Yes 2. Yes 3. Yes	1. Yes 2. Yes 3. Yes
ERG comments: Phase II trial: The trial protocol states that the Full analysis set was the main analysis population, defined as all randomised subjects, who have either withdrawn as a treatment failure or have completed at least one week of dosing and had at least one valid Mayo score during the active double-blind phase of the study (trial protocol section 5). This is a modified ITT rather than a true ITT population. The trial publication describes both a pre-specified analysis and a post-hoc analysis of the primary outcome, neither of which was based on all randomised participants. There is a possible risk of bias but the direction and magnitude are unclear since dropouts from the PBO arm occurred for several different reasons. OCTAVE 1 and 2 and Sustain: Full analysis set was the main analysis population, defined as all subjects as randomly assigned, which is consistent with the ITT principle (CS section B.2.4.1 and section 5 in the trial protocols). Crossovers are not mentioned in the CS, trial publications, protocol, or OCTAVE CSRs and the participant flow in CS Tables 17 and 18 do not mention that any crossovers occurred. Missing values for the primary outcome were analysed by non-responder imputation.				

NA: not applicable

3.1.5 Description and critique of company's outcome selection

The outcomes included in the CS match those in the NICE final scope and appear appropriate. However, time to surgical intervention, although specified in the NICE final scope, was not included, as this was not assessed in the OCTAVE trials.

In clinical trials of therapies for ulcerative colitis the Mayo Score is widely used and was used within the OCTAVE trials (CS Section B1.3.1 and CS Table 3). There are four components to the Mayo score, one of which is 'Endoscopic findings'. In the OCTAVE trials the Mayo endoscopic sub-score was assessed both locally (by the study site investigator) and centrally (from a video recording). Consequently the outcomes in the CS that utilise the endoscopic sub-score were reported separately using the local or the central read of the endoscopic data. The ERG notes that the FDA[22] state that central reading is the preferred approach and the OCTAVE clinical trial programme is the first in ulcerative colitis to use central reads (CS Section B.2.3.1.2.4).

The primary outcome in OCTAVE 1 and 2 was remission at week 8 based on centrally read endoscopic Mayo sub-scores, and at week 52 in OCTAVE Sustain (for definition of remission see Table 10). Higher Mayo scores indicate more severe disease. The company also defined key secondary outcomes: mucosal healing (OCTAVE 1 and 2: week 8; OCTAVE Sustain: week 52), and for OCTAVE Sustain only, sustained corticosteroid-free remission among patients in remission at baseline (week 52). Mucosal healing is associated with lower rates of hospitalisation and surgery,[23] while the use of corticosteroids long-term is not suitable due to side effects so a corticosteroid-free remission is important.[24]

Clinical response and clinical remission based on Mayo sores (for definitions see Table 10) were reported for all three trials (OCTAVE 1 and 2: week 8; OCTAVE Sustain: week 52). As can be observed from Table 10 the difference between the primary outcome of remission and the secondary outcome of clinical remission is that for the former the rectal bleeding sub-score must be zero whereas this is not necessary for the outcome of clinical remission. Clinical response and clinical remission were the only clinical effectiveness outcomes included in the economic model (the primary outcome did not contribute to the economic model), as they were thought to ensure comparability with trials of biological therapies for ulcerative colitis.

The remaining outcomes of disease activity were all based on a Mayo score (for definitions see Table 10):

- Endoscopic remission
- Symptomatic remission
- Deep remission
- Partial Mayo score (range 0-9)
- Total Mayo score (range 0-12)

Health-related quality of life (HRQoL) measures included in the CS were the disease-specific IBDQ, and the Work Productivity and Activity Impairment – Ulcerative Colitis (WPAI-UC) version 2 questionnaire. Generic measures were the 5-dimension EuroQol questionnaire (EQ-5D) and the 36-Item Short Form survey (SF-36). All four HRQoL measures are validated and have been used in other clinical trials in patients with ulcerative colitis.[25,26] However, where different versions of a measure exist (e.g. country specific versions of the IBDQ), the CS did not state which versions were used across the different countries and centres in which the OCTAVE trials took place.

- For the 32-item, disease-specific IBDQ, remission (defined in Table 10) and treatment response were reported for all three OCTAVE RCTs (OCTAVE 1 and 2 at weeks 4 and 8; OCTAVE Sustain at weeks 8, 24 and 52). IBDQ remission scores range from 32 to 224, with higher scores indicating better HRQoL. In HRQoL terms, a total IBDQ score ≥170 points is deemed to constitute clinical remission and a change of ≥16 points has previously been used as a minimal clinically important difference threshold in patients with ulcerative colitis.[27,28]
- For the EQ-5D-3L, both the utility score (based on five dimensions of health status: mobility, self-care, usual activities, pain/discompfort, and anxiety/depression) and visual analogue scale (VAS) outcomes were reported based on EQ-5D-3L version of the instrument, with UK preference weights. This is the only HRQoL measure included in the economic model and it is reported by all three OCTAVE RCTs (OCTAVE 1 and 2 at weeks 2, 8 and change from week 0-8; OCTAVE Sustain at weeks 8, 24, 52 and change from week 0-52).. The CS reports minimal clinically important differences (MCIDs) for UK patients with inflammatory bowel disease of 0.076 for the utility index and 10.9 for the VAS.[29]
    - All three OCTAVE RCTs reported Physical Component Summary (PCS) and Mental Component Summary (MCS) scores from the SF-36 version 2, using the acute form,

which has a recall period of 1 week (in OCTAVE 1 and 2 assessed at baseline and week 8; in OCTAVE Sustain assessed at baseline and weeks 24 and 52). Higher scores indicate better HRQoL. A systematic review[30] of the SF-36 in patients with ulcerative colitis suggests that a group-level clinically important difference threshold of 3 points for both summary scores and responder-level thresholds of 3.1 for PCS and 3.8 for MCS based on the SF-36v2 manual.[31]

- The WPAI-UC score, based on a 6-item questionnaire (version 2) assessing work productivity, is also reported by all three OCTAVE RCTs (OCTAVE 1 and 2 at baseline and week 8; OCTAVE Sustain at baseline and week 52). The questionnaire yields four scores expressed as impairment percentages: absenteeism; presenteeism; work productivity loss; non-work activity impairment. A higher score indicates greater impairment.[32] As part of the response to NICE and the ERG's clarification question A12, the company states that it is not aware of any validated MCID for this outcome in patients with ulcerative colitis. However the company also state that extrapolating from Crohn's Disease suggests a 7% decrease is the MCID for the WPAI.[33,34]

Table 10 Clinical effectiveness outcomes and outcome definitions of the OCTAVE RCTs

Outcome	Definition	When assessed, week		Used in Model
		OCTAVE 1 & 2	OCTAVE Sustain	
Primary: Remission based on centrally-read endoscopic sub-scores	Mayo score ≤2, no individual sub-score >1, rectal bleeding sub-score = 0	8	52	No
Key secondary: Mucosal healing	Mayo endoscopic sub-score ≤1	8	52	No
Key secondary: Sustained corticosteroid-free remission among patients in remission at baseline	Remission (as defined above for the primary outcome) plus no treatment with steroids for ≥4 weeks before the 24-week and 52-week visits	Not assessed	52	No
Clinical response	Mayo score decrease from baseline ≥ 3, and ≥ 30%, with a decrease in rectal	Week 8	52	Yes

	bleeding sub-score of ≥1 or absolute rectal bleeding sub-score of ≤ 1			
Clinical remission	Mayo score ≤2, no individual sub-score >1	8	52	Yes
Endoscopic remissio	Mayo endoscopic sub-score = 0	8	52	No
Symptomatic remission	Mayo score ≤2, no individual sub-score >1, rectal bleeding sub-score and stool frequency sub-score = 0	8	52	No
Deep remission	Mayo score ≤2, no individual sub-score >1, rectal bleeding sub-score and endoscopic sub-score = 0	8	52	No
Partial Mayo score (range 0-9)	Total Mayo score excluding the endoscopic sub-score	2, 4 & 8; change 0-8	Not assessed	No
Total Mayo score (range 0-12)	Sum of 4 sub-scores (stool frequency, rectal bleeding, endoscopic findings, physician's global assessment), each 0-3 with higher scores indicating more severe disease (details in CS Table 3)	Change 0-8	Not assessed	No
HRQoL	**Details/definition**	**OCTAVE 1 & 2**	**OCTAVE Sustain**	**Used in Model**
IBDQ remission	IBDQ score ≥170	4 & 8	8, 24 & 52	No
IBDQ treatment response	IBDQ score increase ≥16 from induction trial baseline	4 & 8	8, 24 & 52	No
EQ-5D score (utility and visual analogue scale versions)	Based on EuroQol-5D 3 level version (no problems, some problems and extreme problems) with UK preference weights	2 & 8; change week 0-8	8, 24 & 52; change 0-52	Yes
SF-36 (PCS and MCS score)	Acute Physical Component Summary & Mental Component Summary scores based on Short-Form 36-item survey (v2)	8; change 0-8	24 & 52; change 0-52	No
WPAI-UC score (assesses work productivity)	6-item Work Productivity and Activity Impairment-Ulcerative Colitis questionnaire (version 2)	8; change 0-8	52; change 0-52	No

3.1.6 Description and critique of the company's approach to trial statistics

The ERG has assessed the approach to trial statistics for the Phase II trial, the OCTAVE 1 and 2 Induction trials and the OCTAVE Sustain trial. The OCTAVE Open study is ongoing and only

summary statistics have been generated (CS Table 16) therefore only the sources of information for this study have been indicated below.

The CS focusses on outcomes from the OCTAVE 1 and 2 Induction trials and the OCTAVE Sustain trial. A brief summary of results from the Phase II trial (which contributes data to the NMAs) is included (CS Figure 18). Interim data from the OCTAVE Open study are summarised in CS Sections B.2.6.3.2 to B.2.6.3.5.

Outcomes and their units of measurement are defined in CS Tables 11 and 12 (OCTAVE 1 and 2), CS B.2.3.1.3.3 (OCTAVE Sustain) and CS B.2.3.1.4.2 (Open study). Outcomes for the Phase II trial are not defined in the CS. Outcomes were defined in the same way in OCTAVE 1, 2 Sustain and Open. The two OCTAVE Induction trials and the OCTAVE Sustain trial are both complete. The only interim data presented in the CS come from the OCTAVE Open study but these do not contribute data to the economic model. The CS has presented appropriate measures of effects (proportions or mean differences with p-values for comparisons between placebo and tofacitinib groups) with uncertainty for continuous outcomes indicated by confidence intervals.

Statistical power

The primary outcome in both the OCTAVE 1 and OCTAVE 2 trials was remission at week 8, based on centrally read Mayo endoscopic subscores. The power calculation is reported in CS Table 16. For each of these trials the company calculated that approximately 545 participants per trial (randomised 4:1, i.e. 436 patients to the 10 mg tofacitinib group and 109 patients to the placebo group) would provide 90% power to detect a difference of 17.5 percentage points between tofacitinib and placebo in the primary and key secondary outcomes. The CS does not justify or explain the rationale for being able to detect a 17.5 percentage point difference between the tofacitinib and placebo groups. One of the ERG's clinical experts thought this was a modest difference but similar to comparator drugs which are used in clinical practice. This power calculation assumed remission rates in the placebo groups of 15% for the primary outcome (remission at week 8) and 35% for the key secondary outcome of mucosal healing. These assumptions are not justified or explained in the CS. The required sample size was achieved for OCTAVE 1 (tofacitinib 10 mg N=476; placebo N=122) and was only narrowly missed for the tofacitinib arm of OCTAVE 2 (tofactinib 10 mg N=429; placebo N=112). Not all of the assumptions made for the power calculation are justified or explained in the CS.

Additionally, the actual sample size was slightly smaller than that calculated and actual rates of remission and mucosal healing were lower than assumed for the power calculation. Nevertheless, the ERG believes that the power calculation was conducted appropriately and the ERG considers that the trials were probably adequately powered.

The primary outcome in the OCTAVE Sustain trial was remission at week 52 based on centrally read Mayo endoscopic subscores. The company calculated that a total of 654 participants (randomised 1:1:1, so 218 in each group) would provide 90% power to detect a 17.5 percentage point difference in remission between the tofacitinib groups (5 mg; 10 mg) and the placebo group, assuming a remission rate in the placebo group of 30% (CS Table 16). The CS does not justify or explain the rationale for being able to detect a 17.5 percentage point difference between the tofacitinib and placebo groups or the assumption of a remission rate in the placebo group of 30%. The required sample size was not achieved, as 593 patients were randomised, which is 61 short of the 654 target (20-22 short per trial arm; CS Table 18). Although the sample size fell short by around 10% per arm, the power calculation was done at a fairly strict level (90% power). On balance the ERG believes that, although there is uncertainty in the statistical power achieved, it is likely to have been adequate.

Statistical power for the Phase II trial and for OCTAVE Open is not reported in the CS.

Analysis populations

The CS defines five main analysis sets (CS B.2.4.1) for the OCTAVE Induction and OCTAVE Sustain trials:

- Full Analysis Set (FAS)
- OCTAVE Induction modified Full Analysis Set (mFAS)
- OCTAVE Sustain mFAS
- Per-Protocol Analysis Set (PPAS)
- Safety Analysis Set (SAS)

FAS – this is the primary analysis population for effectiveness endpoints and is defined as all subjects randomly assigned to either placebo, tofacitinib 10 mg twice daily, or (for OCTAVE Sustain only) tofacitinib 5 mg twice daily. NB this is equivalent to an intention to treat analysis population.

mFAS – this is a subset of the OCTAVE 1 and 2 FAS from which 3 patients were excluded (all from a site in Japan) due to potential unblinding during the study.

OCTAVE Sustain mFAS – this is a subset of the OCTAVE Sustain FAS that included only those patients who had received tofacitinib in the induction trials (i.e. it excluded those patients from the OCTAVE Induction trials who received placebo and met the entry criteria for OCTAVE Sustain).

PPAS – this is a subset of the FAS population who had no major protocol violations that could have potentially had a significant impact on outcomes (this subset was determined by the sponsor prior to database lock).

SAS - included all randomised participants who received at least 1 dose of study medication.

Results from the mFAS and PPAS are not described in full detail in the CS but primary endpoint results are summarised in Appendix L Table 206 to 208.

Analysis populations are not defined in the CS for the Phase II trial or the OCTAVE Open study (NB the trial publication and CSR indicate that FAS in the Phase II trial was defined differently to the Phase III trials and did not include all randomised patients).

Analysis methods

In both the OCTAVE Induction trials, binary outcomes were analysed using a Cochran-Mantel-Haenszel (CMH) Chi-square test, stratified by prior treatment with TNFi therapy, corticosteroid use at baseline, and geographic region. This analysis was applied to the primary outcome (proportion of participants with remission), the key secondary outcome (proportion with mucosal healing), and other binary secondary outcomes (proportions with outcomes derived from the Mayo score, proportion with IBDQ remission, and proportion with IBDQ treatment response) (CS Table 16; OCTAVE Induction trials CSRs sections 9.7.4.2 to 9.7.4.4).

Binary outcomes in the OCTAVE Sustain trial were also analysed using a CMH Chi-square test, but stratification was by treatment received in the induction trials and remission status at baseline (OCTAVE Sustain CSR 9.7.5.2).

In both the OCTAVE Induction trials, continuous outcomes measured only at baseline and week 8 (e.g. the secondary endpoint of change from baseline to week 8 in the total Mayo score) were analysed with an analysis of covariance (ANCOVA) model with observed-cases data. Factors in the ANCOVA were prior treatment with TNFi therapy, corticosteroid use at baseline, and geographic region, whilst baseline score was a covariate. For continuous outcomes measured repeatedly over time (e.g. partial Mayo score at baseline and weeks 2, 4 and 8) data were analysed using a linear mixed-effects model with baseline, treatment group, prior treatment with TNFi therapy, corticosteroid use at baseline, geographic region, visit, and treatment group by visit interaction as fixed effects and subject as a random effect (CS Table 16 and OCTAVE 1 and 2 CSRs section 9.7.4.4)

Continuous outcomes in the OCTAVE Sustain trial (e.g. Mayo scores at baseline, weeks 24 and 52) were analysed using a linear mixed-effects model with induction study treatment assignment included as a baseline stratification factor. (CS Table 16 and OCTAVE Sustain CSR 9.7.5.4).

In OCTAVE 1 and 2 the type 1 error rate was controlled at 0.05 by a fixed-sequence testing procedure for the primary outcome and the key secondary outcome. In OCTAVE Sustain the type 1 error rate was controlled at the 0.05 level for the primary outcome and both of the key secondary outcomes by using a sequentially rejective Bonferroni-based iterative multiple test procedure.

Analysis methods are not reported in the CS for the Phase II trial. The CS states that summary statistics for the OCTAVE Open study have been produced for the interim analysis of the available data (CS Table 16).

In summary, the ERG is satisfied that the analysis methods for the OCTAVE trials were pre-specified and appear appropriate for binary and continuous outcome data. However, the type 1 error rate was controlled only for the primary and key secondary outcomes of the OCTAVE trials, with no adjustments made for multiple comparisons among the other secondary outcomes and therefore caution is needed in interpreting these analyses.

Missing data

Missing data for binary outcomes derived from the total or partial Mayo score were managed in the same way across OCTAVE 1, 2, OCTAVE Sustain and OCTAVE Open (CS Table 16). Patients with missing data for these outcomes were considered as not having had a response (i.e. a non-responder imputation was applied). The ERG agrees that for the binary outcomes based on the Mayo score this is a conservative approach.

For continuous secondary effectiveness outcomes only measured at two timepoints (e.g. at baseline and week 8, or at baseline and week 52) and for continuous effectiveness outcomes (e.g. partial Mayo score) measured repeatedly over time, missing values were not imputed. In the case of continuous effectiveness outcomes measured repeatedly, a linear mixed-effects model was used for the analyses where the missing data were assumed to be missing at random. No justification for the choice of methods to manage data missing from continuous outcomes is provided in the CS.

The CSRs for the OCTAVE 1, 2 and Sustain trials indicate that sensitivity analyses with different approaches for handling missing data (last observation carried forward and observed-cases analyses) were performed for the primary and the key secondary outcomes (OCTAVE 1 and 2 CSRs section 9.7.4.2, OCTAVE Sustain CSR section 9.7.3.1.1) but this is not commented on in the CS. According to the CSRs the results of the sensitivity analyses were consistent with the primary analyses using non-responder imputation.

Missing data for the patient reported outcomes were initially handled using the rules suggested by the developers of the questionnaires (OCTAVE 1 and 2 CSRs section 9.7.2.1; SUSTAIN CSR section 9.7.3.1), but the CS does not state what these rules were. For IBDQ binary outcomes if missing data could not be imputed using the tool developers' rules then they were treated as non-responders. The CS does not state how many of the missing data were accounted for using the developers' rules and how many were imputed as non-responders.

Missing data for the other patient reported outcome measures in the OCTAVE trials were handled differently between the outcomes and between the OCTAVE Induction and OCTAVE Sustain trials (CS Table 16):

- OCTAVE Induction trials: Missing data for EQ-5D continuous outcomes were assumed to be missing at random whilst missing values for SF-36 and WPAI were not imputed.

- OCTAVE Sustain trial: The missing at random assumption was applied to both EQ-5D and SF-36 continuous outcomes, whilst missing WPAI values were not imputed.

The company does not explain these methodological differences and no alternative methods to account for missing data are reported in the CS. The ERG notes that the proportions of data missing from the OCTAVE 1 and 2 trials (as calculated by the ERG from data presented in CS Appendix L Tables 95, 218 and 219) vary among the different patient-reported outcomes. These were lowest for the EQ-5D (0.8% to 8.0% missing data per arm at week 8) and highest for the SF-36 (4.9% to 12.5% missing data per arm at week 8). Furthermore there appear to be imbalances in missing data between trial arms but the company does not comment on this.

Methods for handling missing data are not described for the Phase II trial.

In summary, the ERG would have preferred the company to have provided a justification of the different approaches to handling missing data. For the primary outcomes and key secondary outcomes the company conducted appropriate sensitivity analyses which gave results consistent with the primary analysis. Different methods for accounting for missing data were not explored for patient-reported outcomes. The ERG would therefore interpret the patient reported outcome measures more cautiously than the primary outcome and key secondary outcomes where the exploration of the impact of missing data has been more thorough.

Subgroups

Subgroup analyses are reported in CS section B.2.7. Subgroups based on prior biologic therapy (people previously treated with one or more biologics and people who have not received prior biologic therapy) are listed in the NICE scope under 'Other considerations'. The CS does not report on subgroups based on prior biologic therapy but instead focuses on results according to the subgroups of patients who are TNFi-naïve and those who are TNFi-exposed (i.e. not a wider group of people who have received prior biologic therapy which could include vedolizumab which had been received by some participants in the OCTAVE Induction trials).

The CS highlights (CS Table 4) that the limitations to existing therapy with TNFi agents include that some patients will fail to respond to induction therapy (primary non-response to TNFi-agents) and up to 50% of initial responders will lose response over time (secondary non-

response). Consequently prior TNFi-therapy is an important factor in decisions regarding treatment options.

Pre-planned subgroup analyses were conducted for outcomes according to four factors in OCTAVE 1 and 2 (CS Table 9): prior TNFi exposure (yes vs no); prior TNFi failure (yes vs no), baseline corticosteroid use (yes vs no), and geographic region. However, results of subgroup analyses are not presented in the CS for geographic region. Two of the four factors were pre-specified subgroup analysis factors in the OCTAVE Sustain trial [prior TNFi exposure (yes vs no); prior TNFi failure (yes vs no)]. The OCTAVE Sustain trial included additional pre-planned subgroups, six of which are listed alongside the two noted above in CS Table 13: duration of disease (<6 years vs ≥ 6 years); prior corticosteroid failure at induction study baseline (yes vs no); induction study treatment assignment (tofacitinib 10 mg vs tofacitinib 10 mg or 15 mg vs placebo); remission at maintenance study baseline (yes vs no); mucosal healing at maintenance study baseline (yes vs no); corticosteroid use at maintenance study baseline (yes vs no). Results of these subgroup analyses in OCTAVE Sustain are presented in CS Appendix E.

In OCTAVE 1 and 2 two of the factors assessed by subgroup analyses, prior TNFi exposure and corticosteroid use, were stratification factors at randomisation. Similarly two of the factors assessed by subgroup analyses in OCTAVE Sustain, induction-trial group assignment and remission status at maintenance-trial entry, were stratification factors at randomisation. This would help to ensure that the patient characteristics in these subgroups were well-balanced between the trial arms (confirmed for OCTAVE 1 and 2 by the baseline characteristics of the TNFi exposure subgroups provided by the company in clarification response A7).

The ERG presumes that type 1 error (a false positive, identifying an effect that isn't real) was not controlled for in the subgroup analyses as no statement relating to this has been identified in the CS.

The CS points out (CS section B.2.7.2) that the OCTAVE trials were not powered to test the statistical significance of subgroup analyses due to the limited patient numbers in the subgroups. To increase statistical power, subgroup analyses were also conducted for the pooled OCTAVE 1 and 2 trial population, although the CS does not comment on the statistical power that would have been achieved in these analyses.

In summary, the company has pre-specified the factors for which subgroup analyses were conducted, which is good practice. The CS focuses on the analyses by prior TNFi exposure which, being a randomisation stratification factor in the OCTAVE Induction trials, should improve the balance in patient characteristics between the tofacitinib and placebo arms (i.e. reduce the risk of selection bias) for these subgroups. The ERG agrees that pooling subgroups for the OCTAVE 1 and 2 trials was appropriate for maximising the available statistical power for the TNFi exposure subgroups to increase confidence in the subgroup analyses of OCTAVE 1 and 2. However, the subgroup analyses for OCTAVE Sustain were not powered to test the statistical significance of effects and thus should be interpreted cautiously.

3.1.7 Description and critique of the company's approach to the evidence synthesis

The ERG describes and critiques the company's approach to evidence synthesis by NMA. The ERG identified a number of issues which are discussed in section 3.1.7.1 to section 3.1.7.9.

In an absence of direct head-to-head comparisons between active treatments, the company conducted a network meta-analysis (NMA). NMA is an extension of pairwise network meta-analysis which combines direct and indirect evidence through a connected network of comparators. The NMA compared the relative effects of tofacitinib (5 mg and 10 mg) with adalimumab (40/80/160mg), golimumab (200/100mg and 100mg), infliximab (5 mg/kg), vedolizumab (300mg Q4W and Q8W), and placebo. EMA-licensed doses were included and treated as separate treatments in the NMA. All studies in a moderate to severely active ulcerative colitis population who had failed to tolerate conventional therapy were included.

Effectiveness outcomes included in the NMA consisted of clinical response, clinical remission, and mucosal healing. Safety outcomes included discontinuations due to adverse events, serious adverse events, and serious infections. We have focused our critique on those outcomes included in the economic model: clinical response, clinical remission, and serious infections.

Baseline characteristics of included studies are presented in CS Table 87. The company noted heterogeneity between studies in terms of certain patient characteristics (including prior TNFi exposure, disease duration, and studies in Asian patients) and study design (treat-through or re-randomisation for the maintenance period).

To reduce heterogeneity the company undertook separate NMAs for the TNFi-naïve and TNFi-experienced/failure populations. This choice was informed by subgroup analysis from the OCTAVE programme, a similar assumption in NICE TA342, and a "single integrated induction phase NMA" conducted by the company which showed a statistically significant effect for the interaction between treatment and prior TNFi exposure.

Separate analyses were conducted for the induction (6 to 8 weeks) and maintenance periods (up to one year). Evidence networks and included studies are shown in Figure 3 to Figure 5 below. Most treatments were compared to placebo apart from the Mshimesh 2017 trial[35] which compared adalumimab to infliximab (Induction TNFi-naïve and safety networks), and the UC-SUCCESS study[36] which compared azathioprine to infliximab (safety network only).

Safety outcomes were analysed independently of TNFi exposure status to maximise statistical power for rare events and assumed that prior TNFi-exposure has no effect on safety outcomes. The company stated no NMA was conducted for the safety outcomes in the maintenance period due to the differences in study design.

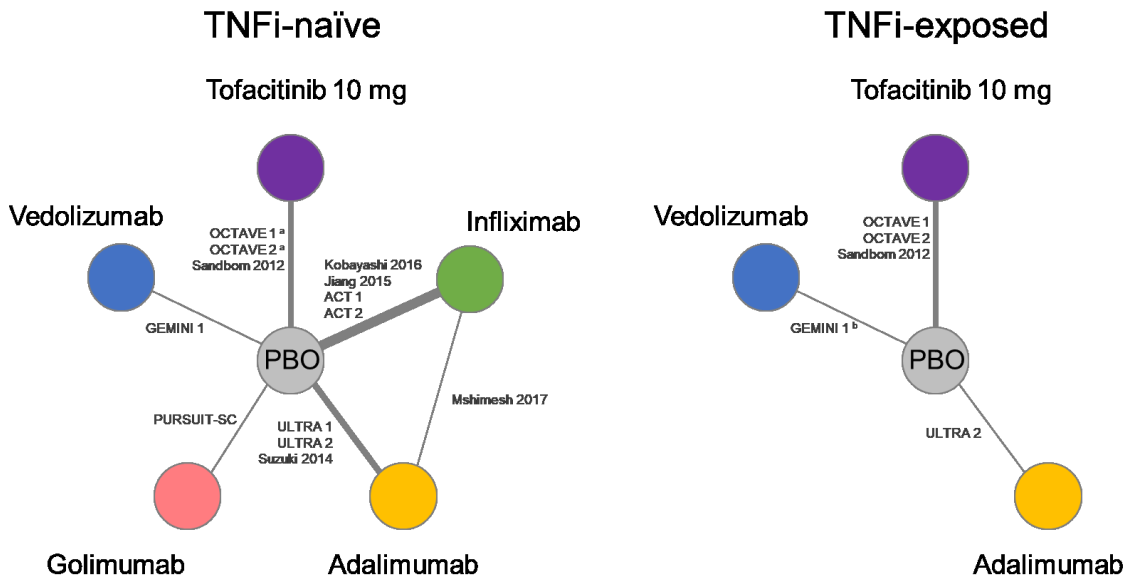


Figure 3 Base-case network of evidence for induction phase clinical response and clinical remission by TNFi-exposure subgroup (taken from CS Figure 28)

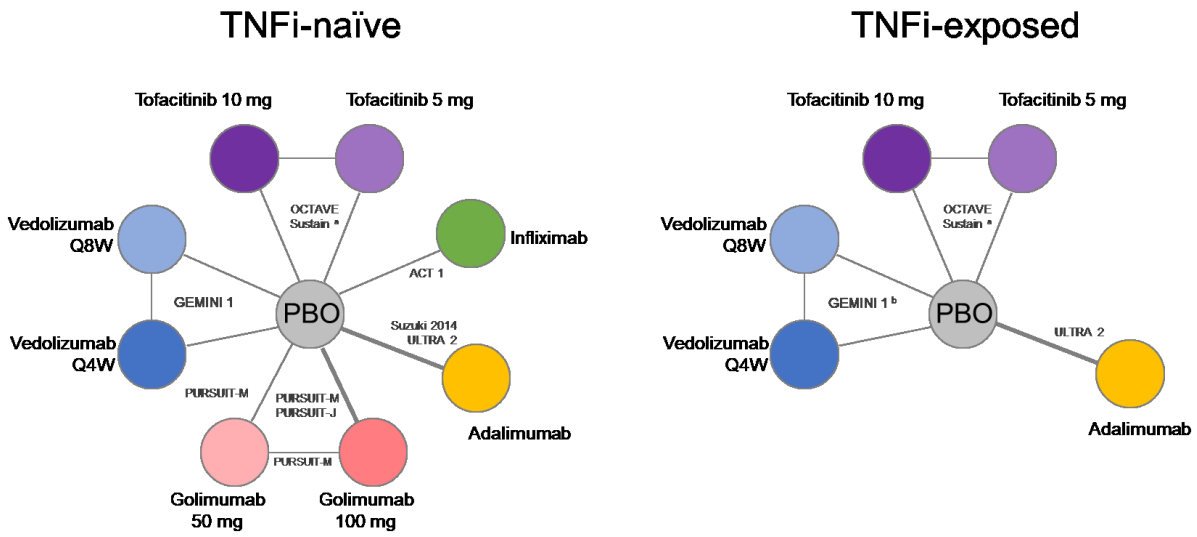


Figure 4 Base-case network of evidence for maintenance phase clinical response and clinical remission by TNFi-exposure subgroup (taken from CS Figure 29)

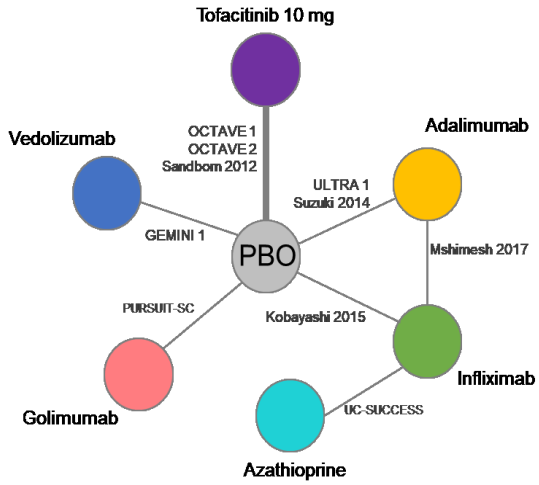


Figure 5 Base-case network of evidence for induction phase safety outcomes (discontinuation due to AEs, serious AEs and serious infections) (taken from CS Figure 30)

Fixed and random effects models were conducted. Where there was a difference in the deviance information criterion (DIC) of less than 3, the company favoured the fixed effects model.

Table 11 summarises the outcomes and comparators included in the analyses undertaken by the company.

Table 11 Outcomes and comparators included in the NMA analyses reported in the CS

Treatment	Clinical response/clinical remission, mucosal healing				Safety (discontinuations due to AEs, serious AEs, serious infections)
	Induction phase, TNFi-naive	Induction phase, TNFi-exposed	Maintenance phase, TNFi-naive	Maintenance phase, TNFi-exposed	
Tofacitinib	X	X	X	X	X
Adalimumab	X	X	X	X	X
Golimumab	X		X		X
Infliximab	X		X		X
Placebo	X	X	X	X	X
Vedolizumab	X	X	X	X	X
Azathioprine[a]					

[a] azathioprine was included in the safety evidence network but not in the NMA results

The company used a multinomial probit model for clinical response and clinical remission. Essentially this modelled clinical response and clinical remission jointly, treating them as ordered categorical data, thus maintaining the correlation between outcomes. This also assumed a common relative treatment effect across response categories. A binomial logit model was used for the safety endpoints.

As noted above, two alternative study designs were used in the maintenance phase. The tofacitinib, golimumab, and vedolizumab studies used a "re-randomised" design, whilst the adalimumab and infliximab studies used a "treat-through" design. Whilst the "treat-through" studies followed a traditional parallel design, randomising patients at baseline, "re-randomised" studies only included induction phase responders in the maintenance phase and re-randomised them to the active treatment or placebo.

The company adjusted for the differences in maintenance study design by adjusting the treat-through study results (ULTRA 2, Suzuki 2014[37] [Adalimumab]; ACT 1[38] [infliximab]) to match those of the re-randomised studies (OCTAVE Sustain[19] [tofacitinib]; PURSUIT-M[39] and PURSUIT-J[40] [Golimumab]; GEMINI 1[41] [vedolizumab]) using similar methods to Takeda in TA342.[9]

Response and remission results are presented on the probit scale (where a negative coefficient indicates treatment is more effective than placebo), and as odds ratios and absolute probabilities. Safety outcomes are presented as log odds, odds ratios, and absolute probabilities.

The company conducted three sets of sensitivity analyses for the effectiveness outcomes: centrally read (as opposed to locally read) endoscopic sub-scores; excluding Asian studies;[35,37,40,42,43] and using prior TNFi-failure as opposed to prior TNFi-exposure data. A further sensitivity analysis in the response to clarification questions (question A16) excluded the Phase II tofacitinib study.

One sensitivity analysis was conducted on safety outcomes: excluding Asian studies[35,37,42,43] and the tofacitinib Phase II (non-Asian) study (Sandborn 2012[11]).

The company's approach to data synthesis by NMA was generally well conducted. A summary of the ERG's appraisal of the company's approach is presented in Table 12.

However, a number of issues were identified which are discussed in sections 3.1.7.1 to 3.1.7.9 which follow Table 12.

Table 12 ERG appraisal of the NMA approach

Checklist		Response
Does the MS present an NMA?		Yes
Are the NMA results used to support the evidence for the clinical effectiveness of the intervention		Yes
Are the NMA results used to support the evidence for the cost-effectiveness of the intervention		Yes, selected endpoints
Homogeneity		
	1. Is homogeneity considered?	Yes
	2. Are the studies homogenous in terms of patient characteristics and study design?	No, difference in TNFi exposure status and study design
	3. Is the method used to determine the presence of statistical heterogeneity adequate? (e.g. Chi-squared test, I-squared statistic)	Yes, meta-regression (interaction between treatment/TNFi exposure status)
	4. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. subgroup analysis, sensitivity analysis, meta-regression)	Yes, separate analyses by TNFi-exposure status. Adjustments made for differences in study design.
Similarity		
	1. Is the assumption of similarity stated?	No
	2. Have they justified their assumption?	Yes, see above
Consistency		
	1. Does the analysis explicitly assess consistency?	Yes, partially

	2. Does the method described include a description of the analyses/ models/ handling of potential bias/ inconsistency/ analysis framework?	Yes
	3. Are patient or trial characteristics compared between direct and indirect evidence trials?	No
	4. If Q3 is yes, and inconsistency is reported, is this accounted for by not combining the direct and indirect evidence?	n/a no inconsistency reported (p<0.05)

3.1.7.1 Use of the probit scale to model clinical response/clinical remission

The company used the multinomial ordered probit scale for clinical response and clinical remission. By modelling clinical response and clinical remission jointly, the company avoided a situation where "*it would be possible to end up with a model that makes impossible predictions, for example that more patients experience clinical remission than experience clinical response*" (CS section B.1.1.1.1). The ERG agrees with this assessment.

In the previous NICE TA342,[9] Takeda used separate binomial logit models for clinical response and clinical remission which was criticised by the ERG:
The results for clinical response and remission should be interpreted with further caution because these were estimated without considering the dependence/correlation between response and remission (TA342, ERG report, p65).

We concur with the company on this point. Hence, the use of a multinomial probit model is an improvement as it takes account of this correlation between outcomes, which is fundamental for the economic model. It is also consistent with the Mayo score, which is essentially a continuous score divided into ordered categories. However, interpreting coefficients on the probit scale is difficult and non-intuitive. We suggest an alternative, the logit model, could have been considered which would have the advantage that the coefficients would be more interpretable.

We queried the company's use of the probit model in the clarification questions (question A18). We agree that separate binomial logit models for response and remission could have introduced inconsistent results across categories of response. However, a multinomial logit analysis could have been considered. The multinomial logit has been previously used in psoriatic arthritis for ordered categorical data.[44] We do not expect such a model would have

resulted in different results but would have aided the interpretability and readability of the company's submission.

Whilst the main analyses tables report the odds ratios and probabilities along with the probit or log odds (e.g. CS Tables 25 and 26), other tables in the sensitivity analysis report results on just the probit scale (e.g. CS Tables 27 and 28). Furthermore, some tables headings are labelled as median treatment effect without acknowledging the scale (e.g. CS Tables 43 and 45, CS section B.3.3.1.1, should read "probit scale", CS Table 48, B.3.3.3, should read "log odds"). This lack of clarity added to the difficulty interpreting the probit scale impedes the readability of the company's submission.

3.1.7.2 Assessment of inconsistency

We noted the presence of closed loops in some of the networks. The company provided details of inconsistency checking and results in their response to the clarification questions (A19). They found no statistically significant inconsistency in the TNFI-naïve subgroup induction network (CS Figure 28) nor safety network (CS Figure 30). However, inconsistency in the maintenance TNFi-naïve network between the two-arm and three-arm trial was not examined.

3.1.7.3 Validation of company results and assessment of model fit.

The ERG replicated selected results to validate the analysis. No errors were found in the company's code. Our validation prioritised the following outputs which contributed to the economic model but we also looked at serious adverse events given the rarity of the serious infections endpoint.

1. CS Table 26 – response/remission fixed effects model in TNFi-naive subgroup (Maintenance phase), using input data from CS Table 93. Probit.
2. CS Table 25 – response/remission fixed effects model in TNFi-exposed subgroup (Induction phase), using input data from CS Table 43. Probit.
3. CS Table 34 - Serious infections random effects model (Induction phase), using input data from CS Table 96. Log odds.
4. CS Table 33 - Serious adverse events fixed effects model (Induction phase), using input data from CS Table 96. Log odds.
5. CS Table 28 – response/remission fixed effects model in TNFi-failure subgroup (maintenance phase), using data from Table 99. Probit.

Furthermore, the ERG conducted a number of additional analyses based around best model fit.

The company states in Appendix D (CS section D.1.3.3) that "*where the difference in DIC suggested indifference [i.e. a difference of less than 3 points], the simpler fixed effects model was preferred*". The ERG would have chosen the random effects model as the more conservative approach in such circumstances to account for between-study heterogeneity.

The ERG believes there is potential heterogeneity between studies which would favour using a random effects model in the base case, model fit being equal. Selected baseline characteristics of studies included in the NMA are presented in CS Table 87. A visual inspection of this table shows disease duration varies from 4.3 to 10.9 years, and IBDQ score varies from 114 to 167. One of our experts identified disease extent (extensive/pan-colitis vs left sided disease) which is not well reported but varies between studies (CS Table 87) as well as albumin, haemoglobin, and baseline C-reactive protein as other potential effect modifiers. These are potentially unobserved sources of heterogeneity. Furthermore, baseline characteristics are not compared by TNFi-exposure status which further precludes an effective qualitative assessment of heterogeneity. In addition to the differences in maintenance design, there is also a difference in the inclusion criteria of re-randomised trials. GEMINI 1,[41] PURSUIT-M,[39] and PURSUIT-J[40] allowed only active treatment responders to enter the maintenance period, whereas OCTAVE Sustain allowed all responders, whether on active treatment or placebo, to enter the maintenance period. The ERG in NICE TA342 also noted that due to the presence of heterogeneity, the fixed effects model would underestimate uncertainty.

Model fit statistics are presented in CS Table 23 (response/remission - Induction phase; CS section B.2.9.2.1.1), Table 24 (response/remission - maintenance phase; CS section B.2.9.2.1.1), and Table 31 (safety outcomes– Induction period; CS section B.2.10.8.1). The first column of CS Table 24 is mislabelled as Induction whereas it is for the Maintenance phase. We have summarised the choice of company base-case model and the ERG preferred model for each of the analyses in Table 13.

Table 13 Company choice of base-case and ERG preference

	Company base-case model	ERG favoured model
Clinical response/clinical remission, Induction TNFi-naive	Random effects	Random effects
Clinical response/clinical remission, Induction TNFi-exposed	Fixed effects	Random effects
Clinical response/clinical remission, Maintenance TNFi-naive	Fixed effects	Random effects
Clinical response/clinical remission, Maintenance TNFi-exposed	Fixed effects	Fixed effects
Serious infections, Induction	Random effects	Fixed effects

In the induction phase TNFi-exposed subgroup, the fixed effects model was preferred despite similar DIC and similar total residual deviance. The ERG would have selected the random effects model as the more conservative analysis. Whilst the base case models are presented in the main NMA results (CS Table 25) the alternative model is not reported. We would prefer to have seen this explored as a sensitivity analysis.

Similarly, the company preferred the fixed effects model in the maintenance phase TNFi-naïve population for clinical response/remission. The ERG would have chosen the random effects model for both the lower DIC and total residual deviance. The ERG would prefer to have seen this explored as a sensitivity analysis.

Finally, the company chose the random effects model for serious infections. In response to a clarification request the company provided the random effect standard deviation (1.82, 95%CrI 0.15, 4.59) (clarification question A22). This wide CrI indicates weak support for the random effects model which has a similar DIC, thus we might have favoured the fixed effects model. The ERG would prefer to have seen the fixed effects model included in a sensitivity analysis.

Table 14 and Table 15 show the results of the ERG validation and exploratory analysis for the response and remission analyses. The ERG ran the same number of chains, burn-in and

simulations reported by the company (section D.1.3.3). Models converged and our results concur to two decimal places.

The alternative choice random effects models show wider credible intervals and some variation in the median estimates for adalimumab and golimumab in the maintenance analysis for the TNFi-naïve population as smaller studies are given more weight under the random effects than the fixed effects model.

Table 14 ERG replication and additional analysis on model choice - clinical response and clinical remission for TNFi-naïve subgroup

Comparator	Treatment effect vs placebo, median (95% CrI), probit scale[a]		
	Company base-case (fixed effects)	ERG replication of base-case (fixed effects)	ERG alternative model selection (random effects)
Maintenance phase			
Tofacitinib 5 mg	[redacted]	[redacted]	[redacted]
Tofacitinib 10 mg	[redacted]	[redacted]	[redacted]
Infliximab 5 mg/kg	[redacted]	[redacted]	[redacted]
Adalumimab 40 mg Q2W	[redacted]	[redacted]	[redacted]
Golimumab 50 mg	[redacted]	[redacted]	[redacted]

Source of company base-case (fixed effects) is CS Table 26
[a] On the probit scale, negative coefficients indicate improvement over placebo. Where the upper and lower CrI are both negative, treatments show strong evidence of benefit versus placebo.

Table 15 ERG replication and additional analysis on model choice - clinical response and clinical remission for TNFi-exposed subgroup

Comparator	Treatment effect vs placebo, median (95% CrI), probit scale[a]		
	Company base-case (fixed effects)	ERG replication of base-case (fixed effects)	ERG alternative model selection (random effects)
Induction phase			
Tofacitinib 10 mg	[redacted]	[redacted]	[redacted]
Adalumimab 160/80/40 mg	[redacted]	[redacted]	[redacted]
Vedolizumab 300 mg	[redacted]	[redacted]	[redacted]

Source of company base-case (fixed effects) is CS Table 25
[a] On the probit scale, negative coefficients indicate improvement over placebo. Where the upper and lower CrI are both negative, treatments show strong evidence of benefit versus placebo.

However, when we attempted to replicate the serious infections results there was a higher level of uncertainty around the coefficients particularly for tofacitinib (Table 16). The wider credible intervals persisted under the fixed effects model conducted by the ERG.

Table 16 ERG replication and additional analysis on model choice - serious infections

Comparator	Treatment effect vs placebo, median (95% CrI), logit scale		
	Company base-case (random effects)	ERG replication of base-case (random effects)	ERG alternative model selection (fixed effects)
Tofacitinib 10 mg	[redacted]	41.42 (4.66, 125.3)	38.72 (3.52, 96.9)
Infliximab 10 mg/kg	[redacted]	-0.56 (-6.82, 5.61)	-0.51 (-2.8, 1.52)
Adalumimab 160/80/40 mg	[redacted]	-0.21 (-5.86, 5.44)	-0.1 (-1.74, 1.49)
Golimumab 200/100 mg	[redacted]	-2.28 (-10.07, 5.28)	-2.12 (-5.50, -0.17)
Vedlizumab 300 mg	[redacted]	-1.90 (-9.71, 5.79)	-1.78 (-5.23, 0.47)
Azathioprine	[redacted]	-0.59 (-10.74, 9.6)	-0.55 (-4.8, 3.63)

Source of company base-case (fixed effects) is CS Table 34

The very wide credible intervals for tofacitinib are caused by the lack of any serious infections across placebo arms in the three tofacitinib studies, hence the difficulty to estimate a relative treatment effect compared to placebo (Table 17). There was also considerable autocorrelation in the tofacitinib coefficient despite thinning and running an extended number of simulations.

The reasons for the difference in our results are unclear, particularly how the company arrived at their estimate for tofacitinib.

Table 17 Tofacitinib induction phase serious infections used in NMA (data from CS Table 96)

Study name	Treatment arm	Serious Infections, n/N (%)
OCTAVE Induction 1	Placebo	0/122 (0%)
	Tofacitinib 10 mg	6/476 (1%)

Study name	Treatment arm	Serious Infections, n/N (%)
OCTAVE Induction 2	Placebo	0/112 (0%)
	Tofacitinib 10 mg	1/429 (0%)
Phase II trial	Placebo	0/48 (0%)
	Tofacitinib 10 mg	2/33 (6%)

As an alternative, we ran the induction phase serious infections analysis in a frequentist framework using the NMA web app developed by Owen and colleagues at the Complex Reviews Support Unit (CRSU) [https://crsu.shinyapps.io/metainsightc/]. The engine underneath this app is Netmeta, which being frequentist, adds 0.5 to zero cells, which results in better convergence and a smaller variance for tofacitinib (Figure 6).

We acknowledge the controversy over adding an arbitrary 0.5 to cells.[45] Nevertheless, we would argue this is a reasonable approximation under the circumstances. If we assume the placebo arms across studies are homogeneous then it seems unjust to encumber tofacitinib with a huge variance for not having a serious infection in any of their placebo arms (the OCTAVE Sustain placebo arm had two serious infections, akin to active treatment which had two in the 5 mg dose, and one in the 10 mg dose but no safety NMA was conducted for the maintenance phase). Of the other five studies with a placebo arm included in the safety analysis, only one had zero events (Suzuki 2014[37]), but similar treatment comparisons in other studies enabled relative treatment effects to be calculated. Random effects results (Figure 6) were generally consistent with the CS albeit all credible intervals were much smaller as was the mean effect for tofacitinib.

Although the UC-SUCCESS study[36] comparing azathioprine to infliximab was included in the safety network it is unclear why azathioprine was not included in the NMA results. We have retained azathioprine in our additional analysis as it appears to meet the inclusion criteria.

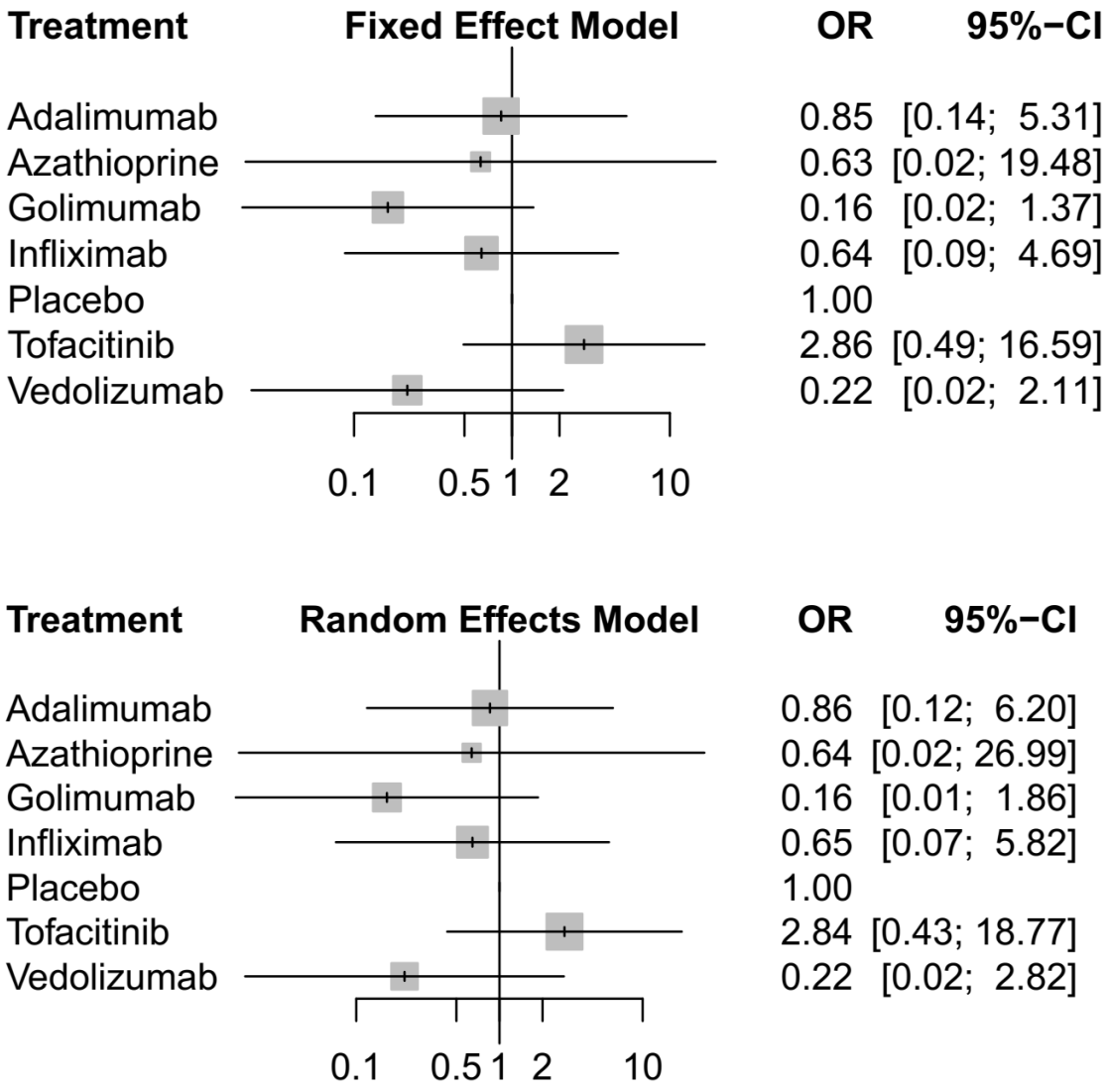


Figure 6 ERG additional analysis - frequentist models for serious infections, courtesy of CRSU web app

Finally, in the NMA base-case analysis for the TNFi-exposed subgroup in the maintenance phase, only TNFi-failed data were available from GEMINI 1 (vedolizumab) (CS Table 22). This may have introduced bias against vedolizumab.

The company conducted a sensitivity analysis using TNFi-failure data from both OCTAVE Sustain and GEMINI 1. However, the maintenance phase analysis "*could not be run because there were too few data points to estimate the multinomial probit model parameters*" (CS section B.2.9.3.2, CS Table 28), essentially because ULTRA 2 was dropped from this analysis. The ERG conducted the analysis using the TNFi-exposed data from ULTRA 2 (adalimumab). In our opinion, this introduced no more bias than the base

case which combined TNFi-exposed data for tofacitinib and adalimumab with TNFi-failure data for vedolizumab. Our scenario analysis at least included comparable data for tofacitinib and vedolizumab.

In the event, as Table 18 shows, use of TNFi-failure data makes little difference to the response/remission results for tofacitinib.

Table 18 ERG scenario analysis using TNFi-failure data from both OCTAVE Sustain and GEMINI 1

Comparator	Treatment effect vs placebo, median (95% CrI), probit scale[a]		
	Company base-case (fixed effects)	ERG replication of base-case (fixed effects)	ERG exploratory scenario analysis (fixed effects)
Maintenance phase			
Tofacitinib 5 mg	[redacted]	[redacted]	[redacted]
Tofacitinib 10 mg	[redacted]	[redacted]	[redacted]
Adalumimab 40 mg Q2W	[redacted]	[redacted]	[redacted]
Vedolizumab 300 mg Q8W	[redacted]	[redacted]	[redacted]
Vedolizumab 300 mg Q4W	[redacted]	[redacted]	[redacted]

Source of company base-case (fixed effects) is CS Table 28
[a] on the probit scale, negative coefficients indicate improvement over placebo. Where the upper and lower CrI are both negative, treatments show strong evidence of benefit versus placebo.

3.1.7.4 Baseline response models – uncertainty around absolute probabilities

To estimate absolute probabilities of each event, treatment effects from the NMA were combined with an estimate of the placebo (baseline) response from the placebo arms of included studies. In response to clarification request A17 the company provided the data, priors and output (meanA, precA) in WinBUGs code format for the probit baseline models. We were able to replicate selected median estimates for the baseline calculations. However, despite running the CS code [validated against NICE DSU Technical Support Document (TSD) 2[46]] and data we were unable to replicate the baseline credible intervals used in the

probit or logit models. The company models tended to lead to wider credible intervals compared to our calculations, thus would lead to conservative results. A summary of the differences in our findings is provided in Table 19 below.

Table 19 ERG replication of baseline (placebo) response results

Comparator	Treatment effect vs placebo, median (95% CrI)	
	Company baseline	ERG replication of company baseline
Induction TNFi-exposed, probit scale		
Response/remission	[redacted]	[redacted]
Maintenance TNFi-naïve, probit scale		
Response/remission	[redacted]	[redacted]
Induction, logit scale		
Serious Infections	[redacted]	[redacted]
Serious adverse events	[redacted]	[redacted]

3.1.7.5 Inclusion of the tofacitinib phase II trial

The Sandborn 2012 Phase II (induction) tofacitinib trial[11] is less well described in the CS despite being included in the NMAs. Furthermore, the company state:
All studies, except for one [Sandborn 2012], were conducted in patients with moderately to severely active ulcerative colitis who had an inadequate response to or had failed to tolerate one or more of the following conventional therapies: oral or intravenous corticosteroids, azathioprine, and/or 6-mercaptopurine (CS section B.2.9.1.1).

The ERG thus questioned the eligibility of this trial. The company confirmed that the Phase II trial met the inclusion criteria for the NMA and they also provided selected NMA results obtained with the Phase II trial excluded from the NMA (Table 7 in clarification response A16). These results for response and remission for the TNFi-naïve and TNFi-exposed populations in the induction period were similar to the base case (CS Table 25).

Base case results without the Phase II trial were not provided for the safety outcomes. However, given the relatively high serious infection rate in the tofacitinib arms of the Phase II trial compared to the OCTAVE trials (6% [2/33] patients had an event compared to 1% [6/476] in OCTAVE Induction 1 and none in OCTAVE Induction 2), the Phase II trial may

have had a disproportionate effect on the random effects NMA results, and we consider this to be a conservative analysis.

3.1.7.6 No safety NMA in the trials' maintenance period

The company said they were unable to perform an NMA for safety in the maintenance phase due to the aforementioned differences in study design, and carryover effects of active treatment on the placebo responders (OCTAVE trials only). These are the same reasons given for the need to adjust the treat-through trials for the response/remission outcomes. Of course, the latter bias could have been averted by using the mFAS population (i.e. excluding the placebo responders from OCTAVE Sustain thereby matching the GEMINI 1 population) of OCTAVE Sustain. Clinical experts advising the ERG suggested that adverse events are likely to increase with drug exposure over time, but it is unclear whether this would have introduced bias and, if so, in which direction.

3.1.7.7 Adjustment for differing lengths of the induction and maintenance periods across trials

The length of the induction phase ranged from six weeks for golimumab and vedolizumab to eight weeks for the other treatments. The maintenance phase ranged from 44 weeks to 54 weeks. Adalimumab had the shortest maintenance phase and golimumab the longest. These are summarised in Table 44 of the company's submission (CS section B.3.3.1.2) which is summarised here (Table 20).

Table 20 Duration of induction and maintenance phases of trials (CS Table 44)

	Induction phase (weeks)	Maintenance phase (weeks)	Total duration (weeks)	Maintenance design
Tofacitinib	8	52	60	Re-randomised
Adalimumab	8	44	52	Treat-through
Golimumab	6	54	60	Re-randomised
Infliximab	8	46	54	Treat-through
Vedolizumab	6	46	52	Re-randomised

In response to our request for clarification the company confirmed that they did not attempt to adjust for different lengths of the induction and maintenance phases across studies. The company noted:
it would have been impossible to properly estimate what difference was due to the treatment effect and what difference was the effect of an earlier measure (company's clarification response A20).

The meaning of this is unclear. However, one of our clinical experts suggested that a shorter induction phase may influence response and that it was entirely possible to see a higher response rate at week 8 than week 6. Our expert referred to the GEMINI 3 (vedolizumab) study in Crohn's disease where it became clear that the 6-week induction phase had failed to capture a majority of responders.

In our opinion, this could have introduced potential bias against studies with shorter induction phases, namely golimumab and vedolizumab. The company in TA342 performed a complementary log-log model (TSD2) to adjust for differences in follow-up in the induction phase (TA342 Company's submission, section 6.7.5). This assumes a Poisson process for each trial arm and a constant event rate and can be applied to binomial and multinomial models (TSD2[46]).

Furthermore, in the induction phase, CS Table 96 (Appendix D, p233) suggests 12-week induction data for the Phase II trial[11] and 14-week data for Kobayashi 2015[43] were used in the safety analysis. This appears to contradict CS Table 44.

With respect to the maintenance phase, the company referred to previous NICE appraisals, in particular that the ERG and appraisal committee for NICE TA342[9] did not believe differences in the length of the maintenance phase would impact results. However, this seems to refer to a difference of between 52 and 54 weeks in the maintenance period. (6.7.3, p125 TA342 company's submission) which are smaller than the differences in Table 1 above. In any case, we concur that the company's base case is likely to be a conservative assumption. Studies with a shorter maintenance phase would experience fewer responders losing response, given the assumption that response wanes slowly over time. Hence this could benefit those treatments with a shorter maintenance phase (i.e. golimumab and vedolizumab) but would be conservative for tofacitinib.

3.1.7.8 Differences between patient populations in the re-randomised maintenance trials.

As noted in section 3.1.7.3 above, unlike the other re-randomisation trials OCTAVE Sustain allowed all responders, whether on active treatment or placebo, to enter the maintenance period. This is a source of heterogeneity and might also be a potential source of bias if placebo responders in OCTAVE Sustain were less able to sustain their response or potentially more susceptible to active treatment, although the direction of any bias is unclear.

However, the company conducted an analysis using a modified Full Analysis Set (mFAS) population which explicitly excluded placebo responders (CS section B.2.4.1). This mFAS population is consistent with the GEMINI 1, PURSUIT-M, and PURSUIT-J maintenance populations and would also have ensured comparability across the placebo groups of the re-randomised trials. Selected results from the mFAS population for OCTAVE Sustain are presented in Appendix L, but only include centrally-read clinical remission. Consequently there are insufficient data to conduct this analysis for the NMA or economic model. As a proxy for the direction of effect of any bias, we compared centrally read remission at 52 weeks in the FAS (CS Figure 10, section B.2.6.2.1.1) and mFAS (CS Table 207, Appendix L.1.2) populations. Remission at 52 weeks was slightly lower in both tofacitinib arms using the mFAS population, suggesting that the base case NMA results may be slightly biased in favour of tofacitinib.

Hence, the ERG believes the mFAS population could have been made the base case or at least explored in a sensitivity analysis.

3.1.7.9 Adjustments to treat-through trials

The company considered that heterogeneity in the study design in the maintenance phase would have introduced bias had they used the reported clinical response and clinical remission data. Furthermore, some placebo patients in the maintenance phase had also received active treatment in the induction phase (OCTAVE Sustain only).

The company considered two methods to adjust for these differences in design. The first was to adjust the re-randomised trials to better match the treat-through design (following an approach used by Thorlund 2015a[47]) and the second was to adjust the threat-through studies to match the re-randomised (the approach used by Takeda[48] in NICE TA342).

The company favoured the latter approach similar to Takeda in NICE TA342 because it required "less data manipulation" and was more aligned with clinical practice and the economic model. The ERG concurs with this choice, which is also acknowledged by Thorlund 2015b[49] for whom "Published data did not allow us to adjust [treat-through] results to fit a re-randomised design" but recognised that the "Re-randomisation design … may mimic a more realistic clinical application of biologic therapy, wherein patients are given a trial of therapy for induction and those who respond are subsequently considered for maintenance dosing."

In NICE TA342, Takeda assumed that patients who responded at 12 months must also have responded at the end of induction, and they used inflation factors to adjust the event rates in both the active treatment and placebo arms for the treat-through trials. However, the exact calculations utilised are unknown since details are unavailable on the NICE website. Hence we cannot tell if the same methods were used in the CS.

The ERG in NICE TA342 criticised Takeda's approach since it "ignores the fact that non-responders at the end of induction could have become responders at the end of the maintenance phase" and
The ERG believes that the adjustment applied to the trials without re-randomisation at the end of the induction phase by the company did not adjust the bias sufficiently, rather, it is possible that their adjustment method actually introduced more bias into the analysis
(TA342, ERG report, p64)

In the CS, the company made the same assumption that the "*number of responders at end of induction period is a proxy for the total number of patients entering maintenance*" (CS Appendix D.1.3.2.1). This could potentially introduce bias against comparators in those studies which had a shorter induction phase as noted above.

The company made the following adjustments to the data in the treat-through trials to better match the re-randomised trials:

- the proportion of patients achieving "sustained clinical response" was used as the clinical response for the treat-through trials "*as this mitigates the risk of counting maintenance phase responders who were induction phase non-responders*" (CS Appendix D.1.3.2.1).
- For the Suzuki 2014 trial,[37] sustained clinical response was not reported for the placebo arm; instead, the ratio of sustained clinical responders to clinical responders

was estimated by the company from the ULTRA 2 (adalimumab) trial[50] and applied to Suzuki 2014.

- The average proportion of clinical remitters among clinical responders in the re-randomised placebo arms was applied to the placebo arms of the treat-through trials.
- For active treatments, the company used numbers of clinical remitters who were induction phase responders.

The ERG is unclear whether these calculations would have introduced any further bias beyond the original criticism in NICE TA342 that non-responders at the end of the induction phase are ignored. Whilst the re-randomisation design ignores non-responders, bias could have been introduced if relative treatment effects on non-responders differ or if the induction phase were of a different length. We believe the use of sustained clinical response has the potential to introduce additional bias against the "treat-through" studies albeit we are unclear whether it has done so.

Summary of the ERG's critique of the NMA approach

The company's NMAs were generally well conducted and made a number of efforts to minimise bias. Nevertheless, the ERG believes a number of potential biases remain.

- Our choice of random effects models for the induction TNFi-exposed and maintenance TNFi-naïve subgroups may have mitigated some concerns over heterogeneity
- Our choice of a frequentist model for serious infections may have mitigated bias from high uncertainty around rare events
- The differences in the uncertainty around our baseline response calculations compared to the company's may lead to conservative results
- There may be undetected inconsistency in the maintenance TNFi-naïve network which could lead to bias in the golimumab estimates. The direction of effect is unclear.
- The lack of safety analysis in the maintenance period may have introduced bias from longer drug exposure but the direction of effect is unclear.
- In the adjustment to the treat-through maintenance trials to match the re-randomised trials, the use of sustained clinical response has the potential to introduce bias albeit we are unclear whether it has done so.
- Potential bias remains with respect to the differences between the re-randomised populations (inclusion of placebo responders in OCTAVE Sustain). The direction of bias is uncertain but may favour tofacitinib.

- Bias may remain with respect to the different lengths of the induction and maintenance phases. The direction of bias may be in favour those studies with a shorter maintenance phase analysis and against those studies with a shorter induction phase.

3.2 Summary statement of company's approach

The ERG's assessment of the company's approach to the evidence synthesis is summarised in Table 21.

Table 21 Quality assessment (CRD criteria) of CS review

CRD Quality Item with ERG comments	
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	1. Yes. Eligibility criteria are tabulated (CS Appendix D.1.1.3 Table 83) and generally appropriate. An exception is that the stated population eligibility criteria are broader than the NICE scope; however, the populations of the studies that were finally included in the company's SLR do match the NICE scope. Outcome measures did not form part of the eligibility criteria.
2. Is there evidence of a substantial effort to search for all relevant research? i.e. all studies identified	2. Yes. All literature searches were systematic and transparent, and are well-documented and reproducible, although over 6 months out of date. An adequate range of bibliographic databases was searched. Supplementary sources and key conferences were also searched. The ERG conducted a rapid update search, which identified four additional relevant full-text publications not listed in the CS and two new conference abstracts reporting results from the OCTAVE trials (see Section 3.1.1). However, these publications either duplicated information already present in the CS or are not directly relevant to the current scope.
3. Is the validity of included studies adequately assessed?	3. Yes. The company assessed the risk of bias in the OCTAVE RCTs (CS Table 19) and the RCTs included in the CS that form part of the NMA (CS Table 86), using the critical appraisal checklist provided by NICE in the Single Technology Appraisal (STA) user guide.[51] (for further details see Section 3.1.4).
4. Is sufficient detail of the individual studies presented?	4. Yes. The CS presents sufficient detail of OCTAVE 1, 2 and OCTAVE Sustain, including general methods (CS

	Tables 9 and 13), eligibility criteria (CS Table 10), participant baseline characteristics (CS Table 15), statistical methods (CS Table 16), outcomes (CS Tables 9, 11, 12, 13), subgroups (CS Tables 9 & 13) and participant flow (CS Tables 17 and 18; Figures 46 and 47). The CS reports limited information on the Phase II trial, but the company and ERG considered this trial to be of less importance than the Phase III trials and the ERG considers the brevity of reporting acceptable.
5. Are the primary studies summarised appropriately?	5. Yes. Clinical effectiveness results from the OCTAVE 1, 2 and OCTAVE Sustain trials are clearly summarised (CS sections B.2.6 and B.2.7, with results from the NMA trials summarised in CS section B.2.9 and Appendix D.1.2). Adverse events are summarised in CS section B.2.10 and CS Appendix F.

The company's evidence synthesis is generally well structured and uses standard methodology. The company's search for clinical effectiveness studies is over 6 months out of date. However, an ERG search update did not identify any missing tofacitinib trials.

The population eligibility criteria for the company's SLR as stated in the CS are broader than the NICE final scope, but the populations of the studies finally included in the SLR are consistent with the NICE scope. With the exception of time to surgical intervention (not reported in the OCTAVE trials), outcome measures reported in the CS match the outcome categories listed in the NICE final scope.

The CS does not include all of the patient-reported outcomes (PROs) that were measured in the OCTAVE Induction and OCTAVE trials, although the PROs that are reported appear adequate (EQ-5D informs the economic analysis) and the risk of reporting bias appears to be low (see Table 9).

Overall, there appears to be a low risk of systematic error in the systematic review of the CS based on the methods employed.

3.3 Summary of submitted evidence

A noted earlier (section 3.1.5) the Mayo endoscopic sub-score was assessed both locally and centrally in the OCTAVE trials. Consequently outcomes that utilise the endoscopic sub-

score were reported separately using the local or the central read of the endoscopic data in the CS. Locally read data were used in the base-case NMAs for the outcomes that contribute data to the economic model.

3.3.1 Summary of results for Remission (Primary endpoint in OCTAVE 1 and 2 and Sustain)

OCTAVE Induction trials 1 and 2:

In both the OCTAVE 1 and OCTAVE 2 trials, a statistically significant difference in remission at week 8 in comparison to placebo was observed in participants who received tofacitinib 10 mg twice daily (Table 22). When endoscopic sub-scores were centrally read in OCTAVE 1, 18.5% of those in receipt of tofacitinib 10 mg were in remission at week 8 in comparison to 8.2% of the placebo group (mean difference from placebo 10.3 percentage points, 95% CI 4.3 to 16.3, p-value 0.007). The corresponding data for OCTAVE 2 are 16.6% in the tofacitinib group in remission at week 8 versus 3.6% of the placebo group, mean difference from placebo 13.0 percentage points, 95% CI 8.1 to 17.9, p-value <0.001).

The locally read data produced mean differences that were 2-3 percentage points higher than those of the centrally read data, but were still statistically significant (Table 22),

This pattern was also observed for the pooled induction population (central read difference from placebo 11.6 versus 14.3 for the local read, p-values in both cases <0.0001).

As previously stated, these remission data are not used for economic modelling.

Table 22 Remission at week 8 in induction trials (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 10 mg	PBO	Difference vs PBO, mean (95% CI)
OCTAVE 1	**N=476**	**N=122**	
Central read	88 (18.5)	10 (8.2)	10.3 (4.3–16.3); p=0.007
Local read	118 (24.8)	14 (11.5)	13.3 (6.5–20.2); p=0.0017
OCTAVE 2	**N=429**	**N=112**	
Central read	71 (16.6)	4 (3.6)	13.0 (8.1–17.9); p<0.001
Local read	89 (20.7)	6 (5.4)	15.4 (9.7–21.1); p=0.0002
OCTAVE 1 & 2 pooled data	**N=905**	**N=234**	
Central read	159 (17.6)	14 (6.0)	11.6 (7.7–15.5); p<0.0001
Local read	207 (22.9)	20 (8.5)	14.3 (9.8–18.8); p<0.0001

Source CS Figure 6 and Appendix L.1.4 Table 213

OCTAVE Sustain

In the OCTAVE Sustain maintenance trial, a statistically significant difference in remission at week 52 in comparison to placebo was observed both for participants who received tofacitinib 10 mg twice daily and those who received tofacitinib 5 mg twice daily (Table 23). When endoscopic sub-scores were centrally read, 34.3% of those in receipt of tofacitinib 5 mg were in remission at week 52 in comparison to 11.1% of the placebo group (mean difference from placebo 23.2 percentage points, 95% CI 15.3 to 31.2, p-value <0.001). A greater proportion of the tofacitinib 10 mg group were in remission at week 52 (40.6%), so consequently the difference in comparison to placebo was also greater (29.5 percentage points, 95% CI 21.4 to 37.6, p<0.001).

The local read data again produced less conservative results than the central read data, with the percentage difference between tofacitinib 5 mg and placebo approximately 3 percentage points higher at 26.3 (95% CI 19.0 to 34.5, p<0.0001) and that between tofacitinib 10 mg and placebo approximately 5 percentage points higher at 34.6 (95% CI 26.2 to 43.0, p<0.0001).

In addition to remission at week 52, data were also reported for participants with sustained remission (i.e. remission at both week 24 and week 52). For both the 5 mg and 10 mg tofacitinib doses and regardless of whether the central or local read data were used, the results were statistically significantly in favour of tofacitinib, with the greater percentage difference in comparison to placebo being obtained with the 10 mg dose (Table 23).

Lastly, remission and sustained remission were also reported for the subset of patients who entered the OCTAVE Sustain maintenance trial in remission. Among these patients, less than 12% of those in the placebo group maintained their remission, whereas in the 5 mg tofacitinib group there mean percentage difference in comparison to placebo was over 30 percentage points and was over 42 percentage points in the 10 mg tofacitinib group. Differences against placebo (at either tofacifinib dose and using both central and local read data) were all statistically significant (Table 23).

None of the remission data are used for economic modelling.

Table 23 Remission outcomes in maintenance trial (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 5 mg N=198	PBO N=198	% difference vs PBO (95% CI)	TOF 10 mg N=197	% difference vs PBO (95% CI)
Remission at week 52					
Central read	68 (34.3)	22 (11.1)	23.2 (15.3–31.2); p<0.001	80 (40.6)	29.5 (21.4–37.6); p<0.001
Local read	78 (39.4)	26 (13.1)	26.3 (18.0–34.5); p<0.0001	94 (47.7)	34.6 (26.2–43.0); p<0.0001
Sustained remission at weeks 24 and 52					
Central read	44 (22.2)	10 (5.1)	17.2 (10.6–23.7); p<0.001	50 (25.4)	20.3 (13.5–27.1); p<0.001
Local read	62 (31.3)	19 (9.6)	21.7 (14.1–29.4) p<0.0001	73 (37.1)	27.5 (19.6–35.4); p<0.0001
Remission at week 52 among patients in remission at baseline, n/total n (%)					
Central read	30/65 (46.2)	6/59 (10.2)	36.0 (21.6–50.3); p<0.001	31/55 (56.4)	46.2 (31.0–61.4); p<0.001
Local read	32/65 (49.2)	7/59 (11.9)	37.4 (22.7–52.1) p<0.0001	32/55 (58.2)	46.3 (30.9–61.7); p<0.0001
Sustained remission at wks 24 & 52 among patients in remission at baseline, n/total n (%)					
Central read	24/65 (36.9)	3/59 (5.1)	31.8 (18.8–44.8); p<0.001	26/55 (47.3)	42.2 (27.9–56.5); p<0.001
Local read	32/65 (49.2)	7/59 (11.9)	37.4 (22.7–52.1) p<0.0001	32/55 (58.2)	46.3 (30.9–61.7); p<0.0001

Source CS Figure 10 and Appendix L.1.4 Table 221

3.3.2 Summary of results for mucosal healing (Key secondary endpoint in OCTAVE 1 and 2 and OCTAVE Sustain)

OCTAVE Induction trials 1 and 2

The proportion of participants with mucosal healing at week 8 was statistically significantly greater in the tofacitinib 10 mg group in both the OCTAVE 1 and OCTAVE 2 trials in comparison to the placebo group (Table 24). For centrally read data in OCTAVE 1, 31.3% of those in the tofacitinib group had mucosal healing at week 8 in comparison to 15.6% of the placebo group (mean difference from placebo 15.7 percentage points, 95% CI 8.1 to 23.4, p-value 0.001). The corresponding data for OCTAVE 2 are 28.4% in the tofacitinib group in

remission at week 8 versus 11.6% of the placebo group, mean difference from placebo 16.8 percentage points, 95% CI 9.5 to 24.1, p-value <0.001).

Greater differences between the two arms of the trials in favour of tofacitinib 10 mg twice daily were observed when using the local read data. For OCTAVE 1 the local read difference from placebo was almost four percentage points higher than the central read difference at 19.5 (95% CI 10.8 to 28.2) and the local read difference was over four percentage points higher than the central read difference for OCTAVE 2 (21.2, 95% CI 13.1 to 29.2). In both trials the local read difference from placebo was statistically significant (p<0.0001).

In the pooled induction population the central read difference from placebo was 16.3 versus 20.3 for the local read, with p-values in both cases <0.0001.

These mucosal healing data are not used for economic modelling, but an NMA was conducted for this outcome (Section 3.3.9.2 below).

Table 24 Mucosal healing week 8 in induction trials (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 10 mg	PBO	Difference vs PBO, mean (95% CI)
OCTAVE 1	**N=476**	**N=122**	
Central read	149 (31.3)	19 (15.6)	15.7 (8.1–23.4); p<0.001
Local read	202 (42.4)	28 (23.0)	19.5 (10.8–28.2); p<0.0001
OCTAVE 2	**N=429**	**N=112**	
Central read	122 (28.4)	13 (11.6)	16.8 (9.5–24.1); p<0.001
Local read	156 (36.4)	17 (15.2)	21.2 (13.1–29.2); p<0.0001
OCTAVE 1 & 2 pooled data	**N=905**	**N=234**	
Central read	271 (29.9)	32 (13.7)	16.3 (11.0–21.6); p<0.0001
Local read	358 (39.6)	45 (19.2)	20.3 (14.4–26.3); p<0.0001

Source CS Figure 7 and Appendix L.1.4 Table 214

OCTAVE Sustain maintenance trial

At week 52 in the OCTAVE Sustain maintenance trial, the proportion of participants with mucosal healing was statistically significant better in comparison to placebo for participants who received tofacitinib 10 mg twice daily, and those who received tofacitinib 5 mg twice daily (Table 25). When endoscopic sub-scores were centrally read, 37.4% of those in

receipt of tofacitinib 5 mg were in remission at week 52 in comparison to 13.1% of the placebo group (mean difference from placebo 24.2 percentage points, 95% CI 16.0 to 32.5, p-value <0.001). In the tofacitinib 10 mg group, a greater proportion were in remission at week 52 (45.7%, percentage difference vs placebo 32.6 percentage points, 95% CI 24.2 to 41.0, p<0.001).

The local read data again produced less conservative results than the central read data (Table 25).

In addition to mucosal healing at week 52, data were also reported for participants with sustained mucosal healing (i.e. mucosal healing at both week 24 and week 52). For both the
5 mg and 10 mg tofacitinib doses and regardless of whether the central or local read data were used, the results were statistically significantly in favour of tofacitinib (Table 25).

Lastly, among the subset of patients who entered the OCTAVE Sustain maintenance trial with mucosal healing, mucosal healing at week 52 and sustained mucosal healing at weeks 24 and 52 were reported. Differences against placebo (at either tofacitinib dose and using both central and local read data) were statistically significant (Table 25).

These mucosal healing data are not used for economic modelling, but an NMA was conducted for this outcome (Section 3.3.9.2 below).

Table 25 Mucosal healing outcomes in OCTAVE Sustain (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 5 mg N=198	PBO N=198	% difference vs PBO (95% CI)	TOF 10 mg N=197	% difference vs PBO (95% CI)
Mucosal healing week 52					
Central read	74 (37.4)	26 (13.1)	24.2 (16.0–32.5); p<0.001	90 (45.7)	32.6 (24.2–41.0); p<0.001
Local read	89 (44.9)	31 (15.7)	29.3 (20.7–37.9) p<0.0001	106 (53.8)	38.2 (29.5–46.8); p<0.0001
Sustained mucosal healing at weeks 24 and 52					
Central read	55 (27.8)	13 (6.6)	21.2 (14.1–28.3); p< 0.001	65 (33.0)	26.4 (19.0–33.8); p< 0.001

Local read	82 (41.4)	25 (12.6)	28.8 (20.5–37.1) p< 0.0001	98 (49.7)	37.1 (28.7–45.5); p< 0.0001
Mucosal healing at week 52 among patients with mucosal healing at baseline, n/total n (%)					
Central read	44/105 (41.9)	12/101 (11.9)	30.0 (18.7–41.4); p< 0.001	49/89 (55.1)	43.2 (31.1–55.3); p< 0.001
Local read	48/105 (45.7)	14/101 (13.9)	31.9 (20.2–43.5) p< 0.0001	56/89 (62.9)	49.1 (37.0–61.1); p< 0.0001
Sustained mucosal healing at weeks 24 and 52 among patients with mucosal healing at baseline, n/total n (%)					
Central read	35/105 (33.3)	9/101 (8.9)	24.4 (13.8–35.0); p< 0.001	44/89 (49.4)	40.5 (28.7–52.3); p< 0.001
Local read	48/105 (45.7)	13/101 (12.9)	32.8 (21.3–44.4) p< 0.0001	53/89 (59.6)	46.7 (34.6–58.8); p< 0.0001

Source CS Figure 11 and Appendix L.1.4 Table 222

3.3.3 Summary of results for sustained corticosteroid-free remission among those in remission at baseline (Key secondary endpoint in OCTAVE Sustain)

OCTAVE Sustain

Among the 593 participants who had a response in either the OCTAVE 1 or 2 induction trials and were randomised into the OCTAVE Sustain trial, 179 were in remission at OCTAVE Sustain baseline. Of these participants, 35.4% in the tofacitinib 5 mg arm and 47.3% of the tofacitinib 10 mg arm were in a sustained corticosteroid-free remission at weeks 24 and 52 in comparison to 5.1% of the placebo group (based on central read data) (Table 26). The differences between the tofacitinib arms and the placebo arm were statistically significant (p<0.001 for both doses of tofacitinib vs placebo). Results based on local read data gave slightly higher percentage differences between tofacitinib and placebo.

These results did not contribute to economic modelling.

Table 26 Sustained corticosteroid-free remission at weeks 24 and 52 among those in remission at baseline (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 5 mg N=65	PBO N=59	% difference vs PBO (95% CI)	TOF 10 mg N=55	% difference vs PBO (95% CI)
Central read	23 (35.4)	3 (5.1)	30.3 (17.4–43.2); p<0.001	26 (47.3)	42.2 (27.9–56.5); p<0.001
Local read	31 (47.7)	7 (11.9)	35.8 (21.1–50.5); p<0.0001	32 (58.2)	46.3 (30.9–61.7); p<0.0001

Source: CS Figure 12 and Appendix L.1.4 Table 225

3.3.4 Summary of results for clinical remission

Note that the definition of clinical remission is almost identical to the definition of the primary outcome remission, except that the rectal bleeding sub-score does not have to be zero (Section 3.1.5).

OCTAVE 1 and 2

Due to the similarity of the definitions for clinical remission and remission, the proportion of participants achieving clinical remission in OCTAVE 1 (Table 27) were identical to those achieving remission reported above (Table 22). In OCTAVE 2 a single patient in the tofacitinib group, who met the criteria for clinical remission but who had not met the criteria for remission, accounted for the difference between the remission and clinical remission outcomes. Consequently, the results were statistically significantly in favour of the tofacitinib 10 mg group for this outcome.

As observed with other outcomes, use of the local read data led to a greater percentage difference between tofacitinib and placebo than with the central read data. The local read data were used in the NMA (Section 3.3.9.1 below), which then contributed to the economic model.

Table 27 Clinical remission week 8 in induction trials (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 10 mg	PBO	Difference vs PBO, mean (95% CI)
OCTAVE 1	**N=476**	**N=122**	
Central read	88 (18.5)	10 (8.2)	10.3 (4.3–16.3); p=0.007
Local read	118 (24.8)	14 (11.5)	13.3 (6.5–20.2); p=0.0017

OCTAVE 2	N=429	N=112	
Central read	72 (16.8)	4 (3.6)	13.2 (8.3–18.1); p<0.001
Local read	90 (21.0)	6 (5.4)	15.6 (9.9–21.3); p=0.0002
OCTAVE 1 & 2 pooled data	**N=905**	**N=234**	
Central read	160 (17.7)	14 (6.0)	11.7 (7.8–15.6); p<0.0001
Local read	208 (23.0)	20 (8.5)	14.4 (9.9–18.9); p<0.0001

Source: CS Figure 8 and Appendix L.1.4 Table 215

OCTAVE Sustain

Clinical remission outcomes were very similar to remission outcomes and favoured the tofacitinib groups. For the central read data, one participant in the tofacitinib 10 mg group attained clinical remission who had not met the criteria for remission, but outcomes in the tofacitinib 5 mg and placebo groups were identical to those for the primary outcome (Table 28). When locally read data were used, two patients (one in the tofacitinib 5 mg and one in the 10 mg tofacitinib group) met the criteria for clinical remission.

The local read data were used in the NMA (Section 3.3.9.1 below), which then contributed to the economic model.

Table 28 Clinical remission week 52 in maintenance study (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 5 mg N=198	PBO N=198	% difference vs PBO (95% CI)	TOF 10 mg N=197	% difference vs PBO (95% CI)
Central read	68 (34.3)	22 (11.1)	23.2 (15.3–31.2); p<0.001	81 (41.1)	30.0 (21.9–38.2); p<0.001
Local read	79 (39.9)	26 (13.1)	26.8 (18.5–35.1); p<0.0001	95 (48.2)	35.1 (26.7–43.5); p<0.0001

Source: CS Figure 13 and Appendix L.1.4 Table 226

3.3.5 Summary of results for clinical response

OCTAVE 1 and 2

Over half of the participants in OCTAVE 1 and 2 achieved a clinical response by week 8 of treatment with tofacitinib 10 mg twice daily. In contrast, just under a third of participants in the placebo group had a clinical response (Table 29). The percentage difference between the tofacitinib group and the placebo group was statistically significant in both trials and for

both the central and locally read data. The 593 participants with a clinical response (central read) were eligible to enter the OCTAVE Sustain maintenance study.

The local read data were used in an NMA (Section 3.3.9.1) which contributed data to the economic model.

Table 29 Clinical response week 8 induction trials (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 10 mg	PBO	Difference vs PBO, mean (95% CI)
OCTAVE 1	**N=476**	**N=122**	
Central read	285 (59.9)	40 (32.8)	27.1 (17.7–36.5); p<0.001
Local read	289 (60.7)	42 (34.4)	26.3 (16.8–35.8); p<0.0001
OCTAVE 2	**N=429**	**N=112**	
Central read	236 (55.0)	32 (28.6)	26.4 (16.8–36.0); p<0.001
Local read	249 (58.0)	33 (29.5)	28.6 (18.9–38.2); p<0.0001
OCTAVE 1 & 2 pooled data	**N=905**	**N=234**	
Central read	521 (57.6)	72 (30.8)	26.8 (20.1–33.5); p<0.0001
Local read	538 (59.4)	75 (32.1)	27.4 (20.6–34.2); p<0.0001

Source: CS Figure 9 and Appendix L.1.4 Table 216

OCTAVE *Sustain*

Just over 60% of participants who had achieved a response to induction therapy and were re-randomised into the OCTAVE sustain study tofacitinib 10 mg twice daily group had a clinical response at week 52 (Table 30). In the tofacitinib 5 mg twice daily arm just over 50% of participants had clinical response at week 52. For both the tofacitinib 10 mg and 5 mg groups the percentage difference in comparison to the placebo group (in which approximately 20% had a clinical response) was statistically significant (central reads TOF 5 mg vs placebo a difference of 31.3 percentage points, 95% CI 22.4 to 40.2, p<0.001; TOF 10 mg vs placebo 41.7, 95% CI 32.9 to 50.5, p<0.001).

The local read data, which were very similar to the central read data, were used in an NMA (see Section 3.3.9.1 below) which contributed data to the economic model.

Table 30 Clinical response week 52 in maintenance study (FAS, NRI, central and local reads)

Parameter, n (%)	TOF 5 mg N=198	PBO N=198	% difference vs PBO (95% CI)	TOF 10 mg N=197	% difference vs PBO (95% CI)
Central read	102 (51.5)	40 (20.2)	31.3 (22.4–40.2); p<0.001	122 (61.9)	41.7 (32.9–50.5); p<0.001
Local read	101 (51.0)	41 (20.7)	30.3 (21.3–39.3); p<0.0001	121 (61.4)	40.7 (31.9–49.5); p<0.0001

Source: CS Figure 14 and Appendix L.1.4 Table 227

3.3.6 Summary of other clinical effectiveness endpoints

Other clinical effectiveness outcomes are not reported in detail in the CS and do not contribute data to the economic model. The majority of outcomes were statistically significantly in favour of tofacitinib, but they are based on Mayo scores and no adjustment has been made for multiple testing, therefore the statistical significance of these should be interpreted cautiously.

3.3.7 Summary of health related quality of life

3.3.7.1 EQ-5D

EQ-5D outcomes (utility index score and VAS score) were obtained at week 2 and week 8 of the OCTAVE induction trials and results are reported for each trial arm as adjusted mean change from baseline to week 2 and week 8 (Table 31). Missing values were assumed to be missing at random in the linear mixed-effects model used to analyse these data. The mean difference in the change from baseline between the tofacitinib and placebo arm favoured the tofacitinib arm for both the EQ-5D based outcomes and at both time points and was statistically significant except for the EQ-5D utility score difference from placebo at 8 weeks in OCTAVE 2. The CS notes in section B.2.6.1.2 that the benefits observed with tofacitinib exceeded the estimated M for patients with inflammatory bowel disease (utility index 0.076; VAS 10.9).

In the OCTAVE Sustain maintenance trial EQ-5D outcomes were obtained at weeks 4, 8, 16, 24, 32, 40 and 52 (Table 32). Missing values were assumed to be missing at random in the linear mixed-effect model used to analyse these data. In comparison to baseline values the EQ-5-D Utility Index values rose slightly over the 52-week analysis period in both the

tofacitinib 5 mg and 10 mg trial arms whereas in the placebo group values fell, indicating worsening quality of life. Differences between the tofacitinib arms and placebo favoured tofacitinib and statistically significant differences were obtained both for the 5 mg dose and the 10 mg dose from week 8 onwards. A similar pattern was observed for the EQ-5D VAS score and statistically significant differences between the tofacitinib arms and placebo were obtained at week 4.

Caution is advised in interpreting these results as the proportions of missing observations differed between arms, the reasons for the data being missing are not explained, and the appropriateness of the missing at random assumption is not discussed. EQ-5D data do not contribute to the company's economic base-case model but are included in a scenario analysis. Estimates of health state utilities obtained from OCTAVE EQ-5D data are discussed in this report in section 4.3.5.

Table 31 Change from baseline to week 8 in EQ-5D utility index and VAS scores. Summary for OCTAVE induction trials (FAS, without imputation)

Outcome	Time point	Adjusted mean ± SE		Difference vs PBO, mean (95% CI)
		TOF 10 mg	PBO	
OCTAVE 1		**N=476**	**N=122**	
EQ-5D utility index score	Week 2	0.13 ± 0.01 n=466	0.08 ± 0.02 n =122	0.04 ± 0.02 (0.00–0.08), p=0.0264
	Week 8	0.15 ± 0.01 n=452	0.08 ± 0.02 n=121	0.08 ± 0.02 (0.04–0.12), p<0.0001
EQ-5D VAS score	Week 2	13.11 ± 0.83 n=466	9.09 ± 1.52 n=122	4.02 ± 1.67 (0.75–7.29); p=0.0162
	Week 8	17.67 ± 0.84 n=451	9.49 ± 1.52 n=121	8.19 ± 1.67 (4.90–11.48); p<0.0001
OCTAVE 2		**N=429**	**N=112**	
EQ-5D utility index score	Week 2	0.12 ± 0.01 n=420	0.04 ± 0.02 n=109	0.08 ± 0.02 (0.04–0.12); p=0.0001
	Week 8	0.14 ± 0.01 n=414	0.11 ± 0.02 n=103	0.03 ± 0.02 (−0.02, 0.07); p=0.2201
EQ-5D VAS score	Week 2	13.32 ± 0.91 n=421	5.31 ± 1.67 n=110	8.01 ± 1.84 (4.39–11.62); p<0.0001
	Week 8	16.52 ± 0.91 n=414	8.29 ± 1.70 n=104	8.23 ± 1.87 (4.55–11.91); p<0.0001

Source: CS Appendix L, Table 218

Table 32 Change from baseline to week 8 in EQ-5D utility index and VAS scores. Summary for Octave Sustain (FAS, without imputation)

Time point	Change, adjusted mean ± SE			Change, adjusted mean ± SE	
	TOF 5 mg N=198	PBO N=198	Difference vs PBO (95% CI)	TOF 10 mg N=197	Difference vs PBO (95% CI)
EQ-5D Utility Index					
Wk 4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 16	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D VAS Score					
Wk 4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 16	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Wk 52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: based on Table 46 in OCTAVE Sustain CSR

3.3.7.2 IBDQ

A statistically significant difference in the proportion of participants achieving IBDQ remission had emerged in favour of tofacitinib at the week 4 time point in both OCTAVE 1 and 2 (Table 33). The proportion of participants with IBDQ remission increased in all trial arms at week 8 with the difference between tofacitinib and placebo also increasing.

In the OCTAVE Sustain maintenance trial statistically significant differences in the proportions of participants with IBDQ remission and IBDQ response emerged by week 8 in favour of both tofacitinib 5 mg and 10 mg in comparison to placebo (Table 34 and CS Figure 15).

[redacted]

The ERG were uncertain about how much of the missing data were accounted for by IBDQ developers' rules and how much were treated as non-responders and consequently these data should be interpreted cautiously. IBDQ data are not included in the economic model.

Table 33 IBDQ results summary for OCTAVE induction trials (FAS, NRI)

Outcome, n (%)	Time point	TOF 10 mg	PBO	Difference vs PBO, mean (95% CI)
OCTAVE 1		**N=476**	**N=122**	
IBDQ remission (IBDQ score of ≥ 170)	Week 4	167 (35.1)	27 (22.1)	13.0 (4.4–21.5); p=0.008
	Week 8	206 (43.3)	32 (26.2)	17.0 (8.1–26.0); p<0.001
IBDQ treatment response (increase in IBDQ score of ≥ 16 points from induction trials baseline)	Week 4	299 (62.8)	55 (45.1)	17.7 (7.9–27.6); p<0.001
	Week 8	307 (64.5)	56 (45.9)	18.6 (8.8–28.4); p<0.001
OCTAVE 2		**N=429**	**N=112**	
IBDQ remission (IBDQ score of ≥ 170)	Week 4	124 (28.9)	9 (8.0)	20.9 (14.3–27.5); p<0.001
	Week 8	173 (40.3)	20 (17.9)	22.5 (14.0–30.9); p<0.001

Outcome, n (%)	Time point	TOF 10 mg	PBO	Difference vs PBO, mean (95% CI)
IBDQ treatment response (increase in IBDQ score of ≥ 16 points from induction trials baseline)	Week 4	266 (62.0)	44 (39.3)	22.7 (12.6–32.9); p<0.001
	Week 8	288 (67.1)	54 (48.2)	18.9 (8.7–29.2); p<0.001

Source: CS Appendix L, Table 217

Table 34 IBDQ results summary for OCTAVE Sustain trial (FAS, NRI)

Time point	TOF 5 mg N=198, n (%)	PBO N=198	% Difference vs PBO (95% CI)	TOF 10 mg N=197, n (%)	% Difference vs PBO (95% CI)
IBDQ Remission (IBDQ score of ≥ 170)					
Baseline	[redacted]	[redacted]		[redacted]	
Week 8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IBDQ Response (increase in IBDQ score of ≥ 16 points from induction trials baseline)					
Baseline	[redacted]	[redacted]		[redacted]	
Week 8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Week 40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Table 42 in OCTAVE Sustain CSR

3.3.7.3 SF-36

In the OCTAVE 1 and 2 trials the PCS and MCS scores for SF-36 increased (i.e. improved) from baseline to week 8 in both the tofacitinib and placebo arms, but the improvement was statistically significantly greater in the tofacitinib arms (Table 35).

In OCTAVE Sustain the PCS and MCS outcomes were analysed as changes from baseline at week 24 and at week 52. At week 52 both the PCS and MCS scores had decreased (i.e. deteriorated) in the placebo and tofacitinib arms, with the decrease being largest for the placebo group, whilst the scores in the tofacitinib 10 mg arm had increased. The difference in change from baseline versus placebo was statistically significant for both the tofacitinib 5 mg and 10 mg arms at both time points (Table 36).

The company do not discuss the clinical significance of the SF-36 results and in the analysis of SF-36 PCS and MCS scores, missing data were not imputed. The ERG observes that the proportion of missing data in the OCTAVE Sustain trial was greater in the placebo arm than in either of the two tofacitinib arms (28% missing from tofacitinib 10 mg, 35% from tofacitinib 5 mg and 64% from placebo arms at 52 weeks). We assume that the patients who had not contributed data are most likely to be those who had failed treatment, although the CS does not state this or provide any further explanation for the missing data. The robustness of these SF-36 results is therefore unclear and they should be interpreted with caution.

Table 35 Change from baseline to week 8 in SF-36 component summary scores for OCTAVE induction trials (FAS, without imputation)

Outcome	Adjusted means ± SE		Difference vs PBO, mean (95% CI)
	TOF 10 mg	PBO	
OCTAVE 1	**N=476**	**N=122**	
PCS score change from baseline	6.8 ± 0.3 n=443	2.5 ± 0.6 n=116	4.2 ± 0.7 (2.9–5.5); p<0.0001

MCS score change from baseline	6.8 ± 0.5 n=443	3.5 ± 0.9 n=116	3.4 ± 1.0 (1.5–5.3); p =0.0005
OCTAVE 2	**N=429**	**N=112**	
PCS score change from baseline	6.8 ± 0.4 n=397	4.6 ± 0.7 n=98	2.2 ± 0.7 (0.7–3.6); p=0.0035
MCS score change from baseline	7.6 ± 0.5 n=397	4.4 ± 1.0 n=98	3.2 ± 1.1 (1.1–5.4); p=0.0037

Source: CS Appendix L, Table 219

Table 36 Change from baseline to week 52 in SF-36 component summary scores in OCTAVE Sustain (FAS, without imputation)

Outcome	Adjusted mean ± SE		% Difference vs PBO (95% CI)	Adjusted mean ± SE	% Difference vs PBO (95% CI)
	TOF 5 mg N=198	PBO N=198		TOF 10 mg N=197	
Change from baseline in PCS at wk 24	−0.3 ± 0.7 n=189	−5.0 ± 0.7 n=180	4.8 ± 0.8 (3.2–6.4); p<0.0001	0.4 ± 0.7 n=187	5.4 ± 0.8 (3.8–7.0); p<0.0001
Change from baseline in PCS at wk 52	−0.0 ± 0.8 n=129	−5.2 ± 0.9 n=71	5.1 ± 1.0 (3.1–7.2); p<0.0001	0.3 ± 0.7 n=141	5.5 ± 1.0 (3.4–7.5); p<0.0001
Change from baseline in MCS at wk 24	−1.1 ± 0.9 n=189	−7.3 ± 0.9 n=180	6.3 ± 1.0 (4.2–8.3); p<0.0001	−0.4 ± 0.9 n=187	6.9 ± 1.0 (4.8–9.0); p<0.0001
Change from baseline in MCS at wk 52	−1.0 ± 1.0 n=129	−6.7 ± 1.2 n=71	5.8 ± 1.3 (3.1–8.4); p<0.0001	0.1 ± 1.0 n=141	6.8 ± 1.3 (4.2–9.4); p<0.0001

Source: CS Appendix L, Table 228

3.3.7.4 WPAI-UC

For the analysis of the WPAI-UC missing data were not imputed. We assume that the high proportion of missing data for some elements of the WPAI-UC is likely to be because participants were not in employment, but the CS does not provide an explanation. For the 'non-work activity impairment' item, which is answered by all people whether or not in employment, the proportion of missing data is low.

After 8 weeks WPAI-UC scores in the OCTAVE Induction trials for all four elements had decreased (which indicates an improvement) but the effect was greater in the tofacitinib group (Table 37). Differences from placebo were in favour of tofacitinib and statistically significant for three of the four measures (presenteeism, work productivity loss and non-work activity impairment).

In the OCTAVE Sustain trial at week 52 WPAI-UC scores had increased (i.e. worsened) in the placebo group for all four elements but had decreased in the tofacitinib 5 mg and 10 mg trial arms. The difference versus placebo was statistically significant for presenteeism and non-work activity impairment (Table 38).

The company has not discussed the clinical significance of these observed changes in WPAI-UC scores, instead relying only on statistical significance for their interpretation.

The WPAI-UC scores are not included in the economic model.

Table 37 Summary of change from baseline in WPAI-UC scores in the OCTAVE induction trials (FAS, without imputation)

Outcome	Adjusted mean ± SE		Difference vs PBO, mean (95% CI)
	TOF 10 mg	PBO	
OCTAVE 1	**N=476**	**N=122**	
Absenteeism	−11.2 ± 1.3 n=270	−7.1 ± 2.7 n=55	−4.2 ± 2.9 (−9.9, 1.6); p=0.1565
Presenteeism	−22.1 ± 1.6 n=273	−9.2 ± 3.3 n=60	−12.9 ± 3.5 (−19.8, −6.0); p=0.0003
Work Productivity Loss	−19.1 ± 2.0 n=180	−8.5 ± 3.9 n=43	−10.6 ± 4.3 (−19.1, −2.1); p=0.0143
Non-Work Activity Impairment	−25.4 ± 1.3 n=442	−11.5 ± 2.3 n=119	−14.0 ± 2.6 (−19.0, −8.9); p<0.0001
OCTAVE 2	**N=429**	**N=112**	
Absenteeism	−7.3 ± 1.6 n=223	−9.3 ± 3.0 n=52	2.1 ± 3.3 (−4.4, 8.5); p=0.5295
Presenteeism	−18.6 ± 1.7 n=235	−13.7± 3.3 n=56	−4.9 ± 3.6 (−12.0, 2.2); p=0.1767
Work Productivity Loss	−14.7 ± 2.2 n=168	−11.2 ± 3.8 n=45	−3.5 ± 4.3 (−11.9, 4.9); p=0.4123

Non-Work Activity Impairment	−24.0 ± 1.3 n=398	−12.2 ± 2.5 n=98	−11.8 ± 2.7 (−17.2, −6.4); p<0.0001

Source: CS Appendix L, Table 212

Table 38 Summary of change from baseline to week 52 in WPAI-UC scores in the OCTAVE Sustain trial (FAS, without imputation).

Outcome	Adjusted mean ± SE		% Diff vs PBO (95% CI)	Adjusted mean ± SE	% Diff vs PBO (95% CI)
	TOF 5 mg N=198	PBO N=198		TOF 10 mg N=197	
Absenteeism	−4.5 ± 2.2 n=63	1.1± 2.8 n=33	−5.6 ± 3.4 (−12.2, 1.0); p=0.0953	−3.1 ± 2.2 n=68	−4.2 ± 3.3 (−10.7, 2.4); p=0.2131
Presenteeism	−3.6 ± 2.8 n=67	7.2 ± 3.5 n=34	−10.9 ± 4.1 (−18.9, −2.8); p=0.0081	−4.3 ± 2.9 n=70	−11.5 ± 4.1 (−19.5, −3.4); p=0.0052
Work Productivity Loss	−3.4 ± 4.9 n=22	1.0 ± 5.4 n=17	−4.4 ± 6.8 (−17.8, 9.0); p=0.5198	−6.6 ± 4.8 n=26	−7.6 ± 6.6 (−20.6, −5.4); p=0.2528
Non-Work Activity Impairment	−2.8 ± 2.2 n=112	11.3 ± 2.8 n=54	−14.1 ± 3.3 (−20.6, −7.5); p<0.0001	−3.1 ± 2.2 n=125	−14.4 ± 3.3 (−20.8, −7.9); p<0.0001

Source: CS Appendix L, Table 229

3.3.8 Sub-group analyses results

3.3.8.1 Prior TNFi exposure status

The CS focuses on subgroup results according to prior TNFi exposure (yes vs no) (CS sections B.2.7.3 to B.2.7.5). This is a more restricted subgroup than prior biologic therapy listed in the NICE scope; prior biologic therapy would include other biologics such as vedolizumab, in addition to the TNF inhibitors.

Detailed sub-group analyses by TNFi exposure status are presented here only for the outcomes that contribute data (from NMAs) to the economic model. Subgroup analyses by TNFi-exposure status for outcomes that do not contribute to data to the economic model (which include the primary outcome of the OCTAVE trials, remission) are presented in

Appendix 1. As stated in section 3.1.6 of this report the OCTAVE trials were not powered to test the statistical significance of subgroup analyses and although pooling the OCTAVE 1 and 2 trials maximises the available statistical power for the TNFi-exposure subgroups the results should nevertheless be interpreted cautiously.

Clinical remission

[redacted] (Table 39). The p-value from the Breslow-Day test did not indicate any significant heterogeneity in effect between the subgroups. When the data from OCTAVE 1 and 2 were pooled the differences between tofacitinib and placebo were [redacted] whilst the differences with locally-read outcomes were [redacted] for the TNFi-naive subgroup and [redacted] for TNFi-exposed subgroup.

At week 52 in OCTAVE Sustain the proportion of participants with clinical remission was higher in those who had received tofacitinib (either the 5 mg or 10 mg maintenance dose) than those who had received placebo in both the prior TNFi-exposed and TNFi-naïve subgroups, both for centrally-read and locally-read outcomes (Table 40). However, the tofacitinib versus placebo difference was greater in the TNFi-naive subgroup than the TNFi-exposed subgroup and this is particularly apparent for the tofacitinib 5 mg versus placebo comparison (e.g. for centrally-read outcomes the differences between tofacitinib 5 mg and placebo were [redacted] in the prior TNFi-naive subgroup and [redacted] in the prior TNFi-exposed subgroup). No test for heterogeneity of effects among the subgroups is reported for OCTAVE Sustain.

Table 39 Proportion of patients in clinical remission in OCTAVE Induction 1 and 2 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup: prior-TNFi treatment	TOF 10 mg n/N (%)	PBO n/N (%)	Difference (95% CI); p-value	p-value for hetero-geneity
OCTAVE 1, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	
TNFi-naïve	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]

Local read				
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]; p=[redacted]	
OCTAVE 2, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]; p=[redacted]	
OCTAVE 1 & 2 pooled data, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	Not reported
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]; p[redacted]	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]; p=[redacted]	Not reported
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]; p=[redacted]	

Source: CS Appendix E Table 123

Table 40 Proportion of patients in clinical remission in OCTAVE Sustain at week 52 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup: prior-TNFi treatment	TOF 5 mg n/N (%)	PBO n/N (%)	Difference vs PBO (95% CI)	TOF 10 mg n/N (%)	Difference vs PBO (95% CI)
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]; p[redacted]	[redacted]	[redacted] p[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]	[redacted] p[redacted]
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]; p[redacted]	[redacted]	[redacted] p[redacted]

TNFi-exposed Local read	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]	[redacted] p[redacted]

Source: CS Appendix E Table 127

Clinical response

Results of the subgroup analyses of clinical response by prior TNFi treatment differed between OCTAVE 1 and OCTAVE 2 (Table 41). In OCTAVE 1 for both centrally-read and locally-read data the difference in clinical response at 8 weeks favouring tofacitinib was greater among TNFi-exposed participants than TNFi-naïve participants. The p-value from the Breslow-Day test suggests that there was significant heterogeneity in treatment effect between the subgroups. This was not the case for OCTAVE 2, in which clinical response results for the TNFi-exposed and TNFi-naïve subgroups were more similar (heterogeneity test not significant), and the difference favouring tofacitinib over placebo was slightly larger in the treatment-naïve subgroup for both centrally- and locally-read data. When the data from OCTAVE 1 and 2 were pooled the central read differences between tofacitinib and placebo were [redacted] in the TNFi-naïve subgroup and [redacted] in the TNFi-exposed subgroup.

At week 52 in OCTAVE Sustain the proportion of participants with a clinical response was higher among those who had received tofacitinib (either the 5 mg or 10 mg maintenance dose) than those who had received placebo in both the TNFi-exposed and TNFi-naïve subgroups, both for centrally-read and locally-read outcomes (Table 42). The proportions of participants with a clinical response were consistently higher in all three trial arms in the prior TNFi-naïve subgroup than in the TNFI-experienced subgroup; however, the relative treatment effect (difference versus placebo) was almost identical in the TNFi-naïve and TNFi-exposed subgroups for both the 5 mg versus placebo and the 10 mg versus placebo comparisons, for both the centrally-read and locally-read data.

Table 41 Proportion of patients with a clinical response in OCTAVE Induction 1 and 2 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup: prior-TNFi treatment	**TOF 10 mg n/N (%)**	**PBO n/N (%)**	**Difference (95% CI); p-value**	**p-value for hetero-geneity**
OCTAVE 1, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]; p=[redacted]	[redacted]

TNFi-exposed Central read	[redacted]	[redacted]	[redacted] p[redacted]	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted] p=[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted] p[redacted]	
OCTAVE 2, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted] p=[redacted]	[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted] p=[redacted]	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted] p=[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted] p=[redacted]	
OCTAVE 1 & 2 pooled data, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted] p[redacted]	Not reported
TNFi-exposed Central read	[redacted]	[redacted]	[redacted] p[redacted]	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted] p[redacted]	Not reported
TNFi-exposed Local read	[redacted]	[redacted]	[redacted] p[redacted]	

Source: CS Appendix E Table 124

Table 42 Proportion of patients with a clinical response in OCTAVE Sustain according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup: prior TNFi treatment	TOF 5 mg n/N (%)	PBO n/N (%)	Difference vs PBO (95% CI)	TOF 10 mg n/N (%)	Difference vs PBO (95% CI)
TNFi-naïve Central read	[redacted]	[redacted]	[redacted] p[redacted]	[redacted]	[redacted] p[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted] p[redacted]	[redacted]	[redacted] p[redacted]

TNFi-naïve Local read	[redacted]	[redacted]	[redacted] p[redacted]	[redacted]	[redacted] p[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted] p[redacted]	[redacted]	[redacted] p[redacted]

Source: CS Appendix E Table 128

In addition to the subgroup analyses by TNFi-exposure status reported above for clinical remission and clinical response, subgroup analyses by TNFi-exposure status were also reported for remission (the primary outcome of both OCTAVE 1, 2 and Sustain) and for sustained corticosteroid-free remission among patients who were in remission at baseline (OCTAVE Sustain) (CS sections B.2.7.4 and B.2.7.5). Neither of these outcomes contribute to data to the economic model. These subgroup analyses are summarised in Appendix 1.

3.3.8.2 Other subgroups

Pre-planned subgroup analyses for the outcomes of remission, mucosal healing, clinical response and clinical remission are reported for OCTAVE 1 and OCTAVE 2 at 8 weeks and for OCTAVE Sustain at 52 weeks according to prior TNFi failure (yes vs no) and corticosteroid use at baseline (yes vs no) (CS Appendix E Tables 130 to 141). As noted above, the OCTAVE trials were not specifically powered statistically for subgroup analyses and adjustments to account for multiple testing were not performed, so these results should be interpreted with caution.

Prior TNFi failure

For all four of these outcomes tofacitinib was [redacted] in both the TNFi failure and no-failure subgroups, for both the centrally-read and locally-read outcomes. Overall, the data suggest that for these four outcomes the treatment effect was [redacted] in the no-failure subgroup than in the prior TNFi-failure subgroup, or there was [redacted] between the subgroups. However, this apparent [redacted] between the subgroups was generally [redacted] (OCTAVE 1 and 2) or the significance was not reported (OCTAVE Sustain).

Corticosteroid use at baseline

For all four outcomes tofacitinib was [redacted] in both the subgroup who had corticosteroid use at baseline and those without corticosteroids, for both the centrally-read and locally-read outcomes. The treatment effects for remission, clinical response and clinical

remission in OCTAVE 1 and OCTAVE 2 generally [redacted] between the two subgroups. The treatment effect for mucosal healing in OCTAVE 1 and OCTAVE 2 tended to be [redacted] in the no-corticosteroid subgroup, but only for centrally-read data (subgroup heterogeneity was [redacted] in OCTAVE 2 only for the centrally-read data). In OCTAVE Sustain the treatment effect for all four outcomes was [redacted] in the no-corticosteroid subgroup, although a test of subgroup statistical heterogeneity is not reported.

Other subgroup analyses in OCTAVE Sustain

For OCTAVE Sustain further pre-planned subgroup analyses are reported for the outcomes of remission, mucosal healing, clinical response, clinical remission, sustained corticosteroid-free remission at weeks 24 and 52 and sustained clinical response at weeks 24 and 52 (CS Appendix E, Tables 142 to 153). The majority of the subgroup analyses were conducted according to the following factors: treatment assignment during the induction study, in remission at maintenance study baseline (yes vs no), mucosal healing at maintenance study baseline (yes vs no), prior corticosteroid failure (yes vs no) and disease duration (<6 years vs ≥6 years). The ERG notes that in Appendix E Tables 149, 150, and 151 an additional subgroup of 'Gender' is listed for some comparisons in place of disease duration (e.g. Table 149 local read data for tofacitinib 10 mg and placebo) but the ERG believes this may be an error and that these data are likely to be disease duration data. Overall, across the different subgroups investigated, a higher proportion of participants in the tofacitinib groups (5 mg and 10 mg) consistently achieved the desired outcome than in the placebo group.

Results from one further potentially relevant subgroup, geographic region, are not reported in the CS.

3.3.9 Network meta-analysis results

In this section we present a summary of the base-case NMA results, with clinical remission and clinical response presented together because these were modelled jointly using the multinomial probit model described earlier in section 3.1.7. Results are presented on the probit scale (for clinical response and clinical remission) or the logit scale (for mucosal healing), as odds ratios and absolute probabilities. On the probit scale a negative coefficient indicates treatment is more effective than placebo whereas an odds ratio greater than one indicates that the comparator treatment had a greater treatment effect than placebo (for columns headed 'Comparator vs PBO) or that tofacitinib had a greater treatment effect than the comparator (for columns headed 'TOF vs comparator). The 95% credible interval indicates the lower and upper extremes in which the odds ratio is expected to lie with a

probability of 95%). The surface under cumulative ranking curve (SUCRA) value is used to rank treatments based on their probability of ranking first through to last among the treatment options. If the SUCRA probability is 0% the treatment always ranks last and if it is 100% the treatment always ranks first.

3.3.9.1 Summary of NMA results for clinical response and clinical remission

In the induction phase for the TNFi-naïve population analysis all treatments were included. Infliximab had the largest treatment effect on the secondary outcomes of clinical remission and clinical response compared to placebo, whilst adalimumab had the smallest effect. All treatments showed strong evidence of benefit over placebo. In the TNFi-exposed population, tofacitinib, adalimumab, and vedolizumab were included. Tofacitinib had the greatest treatment effect on clinical remission and clinical response compared to placebo. Both tofacitinib 10 mg and vedolizumab 300 mg showed strong evidence of benefit over placebo (Table 43).

Table 43 Induction Phase base-case NMA results – comparative effects and probabilities of achieving clinical response and clinical remission

Comparator	Comparator vs PBO			TOF vs comparator		Absolute probability		SUCRA [a]
	Treatment effect, median (95% CrI)	Odds ratio, median (95%CrI)		Odds ratio, median (95%CrI)				
	Probit scale	Clinical response	Clinical remission	Clinical response	Clinical remission	Clinical response	Clinical remission	
TNFi-naïve subgroup								
PBO				[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]	[redacted]			[redacted]	[redacted]	[redacted]
INF 10 mg/kg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA 160/80/40 mg [b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 200/100 mg [c]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg [d]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed subgroup								
PBO				[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]	[redacted]			[redacted]	[redacted]	[redacted]
ADA 160/80/40 mg [b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg [d]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 25

[a] based on treatment effect on probit scale. [b] 160 mg at week 0, 80 mg at week 2, 40 mg at weeks 4 and 6. [c] 200 mg at week 0, 100 mg at week 2. [d] At weeks 0 and 2.

The maintenance phase NMA results for clinical response and clinical remission showed a similar pattern to those for the induction phase analyses (Table 44). In the TNFi-naive population, tofacitinib 10 mg had the largest treatment effect on clinical response and clinical remission compared to placebo, with all treatments showing strong evidence of benefit over placebo.

In the TNFi-exposed population, tofacitinib 10 mg also had the largest treatment effect on clinical response and clinical remission compared to placebo

The NMA results for clinical remission in both the induction and maintenance phases of treatment are included in the economic model, with the exception that adalimumab is not presented as a comparator for the TNFi-exposed subgroup.

Table 44 Maintenance phase base-case NMA results – comparative effects and probabilities of achieving clinical remission

Comparator	Comparator vs PBO			TOF 5 mg vs comparator		Absolute probability		SUCRA [a]
	Treatment effect, median (95% CrI)	Odds ratio, median (95%CrI)		Odds ratio, median (95%CrI)				
	Probit scale	Clinical response	Clinical remission	Clinical response	Clinical remission	Clinical response	Clinical remission	
TNFi-naïve subgroup								
PBO								
TOF 5 mg								
TOF 10 mg								
INF 5 mg/kg								
ADA 40 mg Q2W								
GOL 50 mg								
GOL 100 mg								
VED 300 mg Q8W								
VED 300 mg Q4W								

Comparator	Comparator vs PBO			TOF 5 mg vs comparator		Absolute probability		SUCRA [a]
	Treatment effect, median (95% CrI)	Odds ratio, median (95%CrI)		Odds ratio, median (95%CrI)				
	Probit scale	Clinical response	Clinical remission	Clinical response	Clinical remission	Clinical response	Clinical remission	
TNFi-exposed subgroup								
PBO				[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF 5 mg	[redacted]	[redacted]	[redacted]			[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA 40 mg Q2W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg Q8W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg Q4W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 26

[a] based on treatment effect on probit scale

3.3.9.2 Summary of NMA results for Mucosal healing

Mucosal healing was a key secondary outcome for the OCTAVE Induction 1 and 2 trials and the OCTAVE Sustain trial. This outcome is not included in the economic model.

In the induction phase for the TNFi-naïve subgroup all treatments showed strong evidence of benefit over placebo at achieving mucosal healing. Infliximab had the largest treatment effect on mucosal healing compared to placebo and adalimumab had the lowest. In the TNFi-exposed subgroup, tofacitinib 10 mg had the largest treatment effect compared to placebo (Table 45).

Table 45 Induction phase base-case NMA results – comparative effects and probabilities of achieving mucosal healing

Comparator	Comparator vs PBO		TOF vs Comparator	Absolute probability, median (95% CrI)	SUCRA
	Treatment effect, median (95% CrI) Logit scale	Odds ratio, median (95% CrI)	Odds ratio, median (95% CrI)		
TNFi-naïve subgroup					
PBO			[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]		[redacted]	[redacted]
INF 10 mg/kg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA 160/80/40 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 200/100 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed subgroup					
PBO			[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]		[redacted]	[redacted]
ADA 160/80/40 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Appendix D.1.3.5.2.2. Table 109 (doses have been added by the ERG based on CS Table 25)

In the maintenance phase for the TNFi-naïve subgroup tofacitinib 10 mg had the greatest effect on mucosal healing in comparison to placebo. Infliximab and adalimumab had the smallest effects on mucosal healing in comparison to placebo. The remaining treatments, golimumab and vedolizumab, showed strong evidence of benefit over placebo in mucosal healing. In the TNFi-exposed subgroup vedolizumab had the greatest effect on mucosal

healing, with tofacitinib (5 mg and 10 mg) also providing greater benefit than placebo (Table 46).

Table 46 Maintenance phase base-case NMA results – comparative effects and probabilities of achieving mucosal healing

Comparator	Comparator vs PBO		TOF vs comparator	Absolute probability, median (95% CrI)	SUCRA
	Treatment effect, median (95% CrI) Logit scale	Odds ratio, median (95% CrI)	Odds ratio, median (95% CrI)		
TNFi-naïve subgroup					
PBO			[redacted]	[redacted]	[redacted]
TOF 5 mg	[redacted]	[redacted]		[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
INF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 50 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 100 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED Q8W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED Q4W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed subgroup					
PBO			[redacted]	[redacted]	[redacted]
TOF 5 mg	[redacted]	[redacted]		[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED Q8W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED Q4W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Appendix D.1.3.5.2.2. Table 110

3.3.9.3 NMA sensitivity analyses

The company conducted sensitivity analyses to test the impact of the following factors on NMA outcomes:

- Studies in which the majority of participants were Asian were excluded. These studies were Suzuki 2014, Mshimesh 2017, Jiang 2015, Kobayashi 2015 and Pursuit J. The CS does not provide an explanation for excluding Asian studies, but states

that this "*sensitivity analysis is aligned with the base-case assumptions made in the NMA supporting TA329*".

- Centrally read endoscopic subscores (instead of locally-read subscores) were analysed for the clinical response, clinical remission and mucosal healing outcomes.
- TNFi-failure subgroup: this sensitivity analysis limited the data from the OCTAVE trials and the ULTRA 2 trial to patients who had prior TNFi failure (i.e. a subset of the base case data which included all patients with prior TNFi-exposure)
- Overall ITT analysis: data were not divided into two subgroups by TNFi-exposure status but instead an overall analysis was conducted regardless of prior TNFi-exposure status.

Condensed versions of results tables for these sensitivity analyses are presented in Appendix 2 of this report and are available in full in CS Appendix D.1.3.5.

On the whole the NMA results were relatively robust to the changes made in the sensitivity analyses described above.

[redacted]

The ERG notes that for the sensitivity analyses using data from the TNFi-failure population, [redacted] (CS Table 28).

3.3.10 Summary of adverse events

3.3.10.1 Adverse events in the OCTAVE research programme

The CS presents safety data in patients with moderate to severe ulcerative colitis from the Phase II trial, the two OCTAVE Induction trials, the OCTAVE Sustain trial and the OCTAVE Open extension study. In total, tofacitinib has been evaluated in 1157 patients with ulcerative

colitis, equivalent to 1986 patient-years of tofacitinib exposure with a maximum of 4.4 years of treatment (CS section B.2.10).

The CS classifies adverse events as: common adverse events; serious adverse events; adverse events leading to discontinuation; and adverse events of special interest (CS section B.2.10 and CS Table 29). In CS Appendix F, adverse events are also classified as being treatment-emergent, although the data presented for the overall frequencies of serious adverse events and treatment-emergent serious adverse events in the OCTAVE Induction and OCTAVE Sustain trials are identical (see Table 47). The CS lists adverse events of special interest as being infections (in general), herpes zoster infections, malignancies, gastrointestinal perforations and cardiovascular events, but does not give an explicit rationale. The company presents data on a wide range of adverse events (CS Appendix F), but the only adverse events that inform the economic analysis are serious infections (discussed further below). The CS presents less detailed information on adverse events for the Phase II trial and the OCTAVE Open extension study than for the OCTAVE Induction and Sustain trials. Where data are available, we have summarised the frequency of the main classes of adverse events for the Phase II and Phase III trials in Table 47 and for the OCTAVE Open study in Table 48.

[redacted]

Overall incidence of adverse events

The proportion of patients with of adverse events of any type ranged from 42% to 80% across the Phase II and Phase III trials, being highest in the OCTAVE Sustain trial 10 mg tofacitinib arm. Rates of any adverse event were broadly similar for the tofacitinib and placebo arms within each trial (Table 47 and Table 48). The most frequent specific adverse events were worsening ulcerative colitis, nasopharyngitis, arthralgia, and headache.

Serious adverse events

Serious adverse events affected fewer than 10% of patients in the tofacitinib trials, ranging from 3% in the 10 mg tofacitinib arm of OCTAVE 1to 8% in the placebo arm of OCTAVE 2 (Table 47), and [redacted] of the OCTAVE Open study (Table 48). The most frequent serious adverse event was ulcerative colitis and most serious adverse events were related to ulcerative colitis (CS section B.2.10.3).

Infections

The frequency of any infections ranged from 15% to 40% across the Phase II and Phase III trials, and was highest (24% to 40%) in the OCTAVE Sustain trial (Table 47). In addition to

nasopharyngitis, a range of other types of infection occurred but most of these each affected ≤2% of patients (CS Tables 156 to 158). The only type of infection (besides nasopharyngitis) that occurred in ≥5% of patients was Herpes zoster, which affected 5.1% of patients in the tofacitinib 10 mg arm of the Sustain trial (<5% in all other trial arms).

Most infections were mild or moderate in severity (CS section B.2.10.5). Serious infections were uncommon, affecting a maximum of only 2 patients in any trial arm (≤2%). Serious infections occurred only in the tofacitinib arm within each trial, with the exception of OCTAVE Sustain where 2 patients in the placebo arm had serious infections. The CS lists the specific serious infections that occurred in the OCTAVE Induction and Sustain trials but does not specify those which occurred in the Phase II trial (n=2) or the OCTAVE Open study (n not reported). The patients who had serious infections in OCTAVE 1 (n=6), OCTAVE 2 (n=1) and OCTAVE Sustain (n=5) are notable in that they each had a different type of infection, i.e. no individual type of serious infection occurred in more than one patient (CS Table 162).

CS Table 168 summarises the incidence of serious adverse events that have occurred in tofacitinib-treated patients across the company's clinical research programme on ulcerative colitis. The data show that [redacted].

Discontinuation due to adverse events

The frequency of adverse events leading to discontinuation ranged from 2% to 8% in the Phase II trial and OCTAVE Induction trials, but was higher in the OCTAVE Sustain trial (9% to 19%) (Table 47). The most common reason for discontinuation was worsening ulcerative colitis (CS section B.2.10.4).

Adverse events of special interest

The CS is slightly inconsistent in the reporting of adverse events of special interest, since infections and Herpes zoster are not listed under adverse events of special interest in CS Table 29, although they are reported elsewhere in the table. Where reported, adverse events of special interest affected a maximum of 3 patients in any trial arm.

Abnormal laboratory test results

The CS tabulates, but does not comment on, selected abnormal laboratory test results relating to cholesterol and triglyceride metabolism, and also reports the frequency of abnormal creatine kinase results. Monitoring cholesterol and other lipid parameters is recommended in the SmPC due to known short-term effects of tofacitinib on these.[10] The CS does not report whether any other laboratory test results (e.g. relating to liver or renal function) were abnormal, although the SmPC lists abnormal liver function tests as being a possible uncommon adverse event.[10] Overall, 5% to 27% of patients in the Phase II and Phase III trials had elevated total cholesterol (>1.3 x the upper limit of normal [ULN]) and 8% to 31% of patients had elevated low-density lipoprotein (>1.2 x ULN), with the rates being consistently higher in the tofactinib than placebo arms (Table 47). A smaller proportion of patients had abnormalities in high-density lipoprotein (1% to 9%) and triglycerides (0% to 8%) without a consistent within-trial difference between arms. The proportion of patients with elevated creatine kinase ranged from 2% to 28% and was higher in the tofacitinib than placebo arms in OCTAVE 1 and OCTAVE Sustain, but not in OCTAVE 2 (not reported for the Phase II trial).

Malignancies

The CS reports the frequency of malignancies across the company's tofacitinib ulcerative colitis research programme (total 1157 patients), divided into non-melanoma skin cancer and all other malignancies (CS Table 30). In total, 15 patients (1.3%) had non-melanoma skin cancer and 13 patients (1.2%) had a malignancy other than non-melanoma skin cancer. The company comments that a potential elevated risk of non-melanoma skin cancer was identified during the ulcerative colitis clinical trial programme compared to the company's rheumatoid arthritis trial programme, which likely reflects an increased malignancy risk in patients who have inflammatory bowel disease. However, the CS also states that the draft SmPC includes effective routine risk minimisation measures (CS section B.2.13.1). We note that the licensed indication for tofacitinib in rheumatoid arthritis is different to that in ulcerative colitis,[10] [e.g, tofacitinib is often administered with methotrexate, and at a different daily dose], so comparisons between the ulcerative colitis and rheumatoid arthritis trials programmes should be made with caution.

Mortality

The CS reports that there were 5 deaths across the OCTAVE programme (CS section B.2.10.6). These were: 1 death in the tofacitinib arm of OCTAVE 1, caused by dissecting aortic aneurysm, assessed as unrelated to the study drug; and 4 deaths in OCTAVE Open, all in the 10 mg tofacitinib group. Three of the deaths in the OCTAVE Open study occurred

>28 days after the last dose of tofacitinib and were due to malignancies. The remaining patient had died of hepatic angiosarcoma, in which tofacitinib was considered to have played a contributory role.

Table 47 Summary of adverse events in the tofacitinib Phase II and Phase III trials

Adverse event (AE)	Phase II trial		OCTAVE Induction 1		OCTAVE Induction	
	TOF 10 mg (N=33)	PBO (N=48)	TOF 10 mg (N=476)	PBO (N=122)	TOF 10 mg (N=429)	PBO (N=112
Any AE, n (%)	14 (42)	23 (48)	269 (57)	73 (60)	232 (54)	59 (
Serious AE, n (%)	2 (6)	4 (8)	16 (3) [a]	5 (4) [a]	18 (4) [a]	9 (
Most frequent AE, n (%) [b]						
Worsening ulcerative colitis	2 (6)	9 (19)	11 (2)	5 (4)	13 (3)	6
Nasopharyngitis	1 (3)	1 (2)	34 (7)	9 (7)	21 (5)	4
Arthralgia	2 (6)	0	14 (3)	6 (5)	11 (3)	6
Headache	3 (9)	2 (4)	37 (8)	8 (7)	33 (8)	9
Infections, n (%)						
Any infection [c]	9 (27)	7 (15)	111 (23)	19 (16)	78 (18)	17 (
Serious infection	2 (6)	0	6 (1)	0	1 (0.2)	
Herpes zoster	1 (3)	0	3 (1)	1 (1)	2 (1)	
AE leading to discontinuation, n (%) [d]	1 (3)	4 (8)	18 (4)	2 (2)	17 (4)	8
AE of special interest, n						
Intestinal perforation	Not reported		1	0	0	
Cancer other than non-melanoma skin cancer	Not reported		0	0	0	
Non-melanoma skin cancer	Not reported		1	0	1	
Cardiovascular events	Not reported		2	0	2	
Abnormal laboratory test results, n (%) [f]						
N for laboratory data	33	48	471	122	424	
Total cholesterol >1.3× ULN	8 (24)	5 (10)	80 (17)	11 (9)	73 (17)	
Low-density lipoprotein >1.2× ULN	9 (27)	4 (8)	91 (19)	11 (9)	92 (22)	12
High-density lipoprotein <0.8× LLN	3 (9)	2 (4)	6 (1)	2 (2)	7 (2)	
Triglycerides >1.3× ULN	2 (6)	0	15 (3)	1 (1)	12 (3)	
Creatine kinase >2× ULN	Not reported		45 (10) (n=474)	2 (2)	40 (9) (n=425)	10
Addition or increase in dose of lipid lowering agent, n (%)	Not reported		4 (1)	0	2 (1)	1

Source: CS Tables 29 ad 166 and the Phase II trial publication.

[a] The CS reports that these data are serious AE (CS Table 29; not marked as confidential) and also reports that these same data are treatment-emergent serious AE (CS Tables 159-161; marked as academic in confidence).

[b] The four most frequent adverse events in OCTAVE Sustain.

[c] Reported as “adverse effects from infection” (CS Table 166).

[d] Including patients who discontinued treatment because of worsening ulcerative colitis.

[e] Invasive ductal breast carcinoma.

[f] Laboratory data were missing for some patients.

Table 48 Summary of treatment-emergent adverse events in OCTAVE Open

Adverse event (AE)	Tofacitinib 5 mg	Tofacitnib 10 mg	Total
Number of AE			
Patients with AE, n (%)			
Patients with serious AE, n (%)			
Patients with severe AE, n (%)			
Patients discontinued due to AE, n (%)			
Patients with dose reduced or temporary discontinuation due to AE, n (%)			

Source: CS Table 167

Except for the number of AE, subjects were counted only once per treatment in each row.

3.3.10.2 Adverse events NMA

Three safety outcomes were analysed by NMA: discontinuations due to adverse events; serious adverse events; and serious infections. Only data from the serious infections NMA contribute to the economic model. As stated in section 3.1.7, safety outcomes were analysed only for the induction phase of the included studies and with data for TNFi-exposed and TNFi-naïve subgroups combined in one analysis.

Discontinuation due to adverse events does not contribute to the economic model.

(Table 49).

Table 49 Induction phase base-case NMA results – comparative effects and probabilities of discontinuing due to AEs

Comparator	Odds ratio, median (95%CrI)		Absolute probability median (95% CrI)	SUCRA[a]
	Comparator vs PBO	TOF vs comparator		
PBO		[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]		[redacted]	[redacted]
INF 10 mg/kg	[redacted]	[redacted]	[redacted]	[redacted]
ADA 160/80/40 mg [b]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 200/100 mg [c]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg [d]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 32

[a] Based on treatment effect on probit scale. [b] 160 mg at week 0, 80 mg at week 2, 40 mg at weeks 4 and 6. [c] 200 mg at week 0, 100 mg at week 2. [d] At weeks 0 and 2.

Serious adverse events do not contribute data to the economic model.

[redacted]

Table 50 Induction phase base-case NMA results – comparative effects and probabilities of serious AEs

Comparator	Odds ratio, median (95%CrI)		Absolute probability median (95% CrI)	SUCRA[a]
	Comparator vs PBO	TOF vs comparator		
PBO		[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]		[redacted]	[redacted]
INF 10 mg/kg	[redacted]	[redacted]	[redacted]	[redacted]
ADA 160/80/40 mg [b]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 200/100 mg [c]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg [d]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 33

[a] Based on treatment effect on probit scale. [b] 160 mg at week 0, 80 mg at week 2, 40 mg at weeks 4 and 6. [c] 200 mg at week 0, 100 mg at week 2. [d] At weeks 0 and 2.

Serious infections are included in the economic model.

[redacted] (Table 51).

Table 51 Induction phase base-case NMA results – comparative effects and probabilities of serious infections

Comparator	Odds ratio, median (95%CrI)		Absolute probability median (95% CrI)	SUCRA[a]
	Comparator vs PBO	TOF vs comparator		
PBO		[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]		[redacted]	[redacted]
INF 10 mg/kg	[redacted]	[redacted]	[redacted]	[redacted]
ADA 160/80/40 mg [b]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 200/100 mg [c]	[redacted]	[redacted]	[redacted]	[redacted]
VED 300 mg [d]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 34

[a] Based on treatment effect on probit scale. [b] 160 mg at week 0, 80 mg at week 2, 40 mg at weeks 4 and 6. [c] 200 mg at week 0, 100 mg at week 2. [d] At weeks 0 and 2.

3.4 Overall summary of clinical effectiveness evidence

The systematic review of clinical effectiveness evidence in the CS identified four RCTs of tofacitinib as a treatment for people with moderately to severely active ulcerative colitis. One is a Phase II dose finding study (in which only one arm received the licensed 10 mg BID dose), two were identical Phase III RCTs of tofacitinib as an induction therapy (OCTAVE Induction 1 and OCTAVE Induction 2), and the fourth was an RCT of tofacitinib as a maintenance therapy (OCTAVE Sustain). In all four trials the comparator was placebo. The OCTAVE clinical trial programme followed a re-randomisation design in which participants who had been randomised into one of the two induction phase trials (OCTAVE 1 or OCTAVE 2) and who had achieved a response to 8-weeks of induction therapy were eligible to be re-randomised into the 52-week OCTAVE Sustain trial.

In addition to the four tofacitinib RCTs there is also an ongoing open label extension study (OCTAVE Open) which all participants in the OCTAVE trial programme are eligible to enter.

The focus in the CS is on the two Phase III induction therapy RCTs (OCTAVE 1 and OCTAVE 2) and the maintenance therapy RCT (OCTAVE Sustain) which provide the bulk of the evidence presented in the CS.

The ERG judged the two OCTAVE Induction trials to be at a low risk of the five types of bias assessed. The OCTAVE Sustain trial and the Phase II trial could be at risk of attrition bias as a result of unbalanced dropouts between the tofacitinib and placebo arms but appear to be at a low risk of bias for the other four types of bias assessed. Overall the studies appear to have been well conducted. The main clinical effectiveness outcomes reported in the CS are remission (primary outcome), mucosal healing, sustained corticosteroid-free remission (OCTAVE Sustain only), clinical remission, and clinical response. Health related quality of life outcomes (both generic and disease specific) and adverse events were also reported.

The company's systematic review had broad inclusion criteria to enable the identification of evidence not only for tofacitinib but also for relevant comparators. It identified 21 RCTs in total, the four tofacitinib RCTs plus 17 RCTs, most of which compared an active treatment to placebo. In the absence of direct head-to-head comparisons between active treatments, these studies could potentially be used in NMA.

Key clinical effectiveness outcomes within the OCTAVE trials were based on components of the Mayo Score. One of the four components to the Mayo score, 'Endoscopic findings', was assessed both locally (by the study site investigator) and centrally (from a video recording). Results including this component of the Mayo score were reported separately using the local or the central read of the endoscopic data.

Remission, as opposed to clinical remission, was the primary outcome of the OCTAVE induction and maintenance trials but this outcome did not contribute to economic modelling. At week 8 the mean differences between tofacitib and placebo in OCTAVE 1 and 2 were in favour of tofacitinib and statistically significant regardless of whether central read or local read data were used. Results using local read data were less conservative regarding the effectiveness of tofacitinib than those using central read data. A similar effect of tofacitinib was observed in the OCTAVE Sustain trial for both the 5 mg and 10 mg tofacitinib

maintenance doses, with the percentage difference in comparison to placebo being greater for the 10 mg tofacitinib dose at week 52.

Mucosal healing was designated a key secondary outcome for the OCTAVE trials. A greater proportion of participants in the tofacitinib group achieved mucosal healing at week 8 in comparison to the placebo group in both OCTAVE 1 and OCTAVE 2 and the differences versus placebo were statistically significant for both central and local read data. In the OCTAVE Sustain maintenance trial, statistically significant differences in both mucosal healing at week 52 and sustained mucosal healing at weeks 24 and 52 were reported for the 5 mg and 10 mg tofacitinib maintenance doses in comparison to the placebo arm of the trial.

Sustained corticosteroid-free remission among those in remission at baseline (a further key secondary outcome) in the OCTAVE Sustain trial, was statistically significantly greater in the tofacitinib 5 mg and 10 mg arms than in the placebo arm.

Clinical remission, which has a very similar definition to the primary outcome of remission, contributed data to the economic model via the NMA. Due to the similarity of outcome definition the results from the OCTAVE trials were almost identical to those reported above for remission, favouring tofacitinib.

The outcome of clinical response also contributes data to the economic analysis via NMA. The percentage difference between the tofacitinib group and the placebo group in favour of tofacitinib was statistically significant in both OCTAVE induction trials and the OCTAVE Sustain maintenance trial and for both the central and locally read data.

HRQoL was reported using both generic (EQ-5D and SF-36) and disease specific (IBDQ and WPAI-UC) instruments. Results showed HRQoL was typically improved by tofacitinib treatment; however, for some HRQoL measures we are uncertain about the impact of the missing data. Data from the EQ-5D-3L do not inform the base-case economic model but were included in a scenario analysis.

Subgroup analyses focused on results according to prior TNFi-exposure. Note that this is a more restricted subgroup than that of prior biologic therapy (which would also include other biological therapies such as vedolizumab) which is listed in the NICE scope. The OCTAVE trials were not powered to test the statistical significance of subgroup analyses so the results should be interpreted cautiously. Overall, the results were consistent regardless of prior TNFi-exposure status.

Safety data for tofacitinib in patients with moderate to severely active ulcerative colitis comes from the Phase II tofacitinib trial, the two OCTAVE Induction trials, the OCTAVE Sustain trial and the ongoing OCTAVE Open extension study. In total tofacitinib has been evaluated in 1157 patients with ulcerative colitis with a maximum exposure to tofacitinib of 4.4 years.

Rates of adverse events of any type were broadly similar for the tofacitinib and placebo arms within each trial. Serious adverse events affected fewer than 10% of patients in the tofacitinib trials. The most frequent serious adverse event was ulcerative colitis, and most serious adverse events were related to ulcerative colitis.

Serious infections were uncommon, affecting a maximum of only 2 patients (≤2%) in any trial arm and occurred only in the tofacitinib arm within each trial, with the exception of OCTAVE Sustain where two patients in the placebo arm had serious infections. As previously noted, data on serious infections was included in the economic model.

There were five deaths across the OCTAVE trials programme and tofacitinib was considered to have played a role in one of these (the death of a patient with hepatic angiosarcoma).

Overall, and in comparison with evidence from the use of tofacitinib in patients with rheumatoid arthritis, no new safety signals were identified.

NMA was used to compare tofacitinib to other potential treatment options where there was available evidence. Analyses were conducted for the outcomes of clinical response, clinical remission, mucosal healing and safety (discontinuations due to AEs, serious AEs, serious infections). Of these outcomes, clinical response, clinical remission and serious infections contributed data to the economic model and results for these outcomes are summarised below.

Heterogeneity was present among the studies available to include in NMA. There were differences in study design (re-randomised design as for the OCTAVE trial programme versus treat-through design in which participants entering the maintenance phase of a study remain in the arm they were allocated to during the induction phase of the study) and some patient characteristics (TNFi-exposure status, disease duration, studies in predominantly Asian patients).

Separate analyses by NMA were undertaken for the induction and maintenance phases of treatment. In addition, to reduce heterogeneity, NMAs were conducted separately for the TNFi-naïve and TNFi-exposed populations (Table 11). An adjustment was also made to the outcomes from studies with a treat-through design to try to better align these outcomes with those from studies with a re-randomised design.

Because clinical response and clinical remission are correlated outcomes (both are based on Mayo score data) the company used a multinomial probit model which maintained the correlation between these outcomes. A binomial logit model was used for the safety outcomes. Fixed and random effects models were conducted and sensitivity analyses were also undertaken (using centrally-read rather than locally-read endoscopic subscores; exclusion of studies with predominantly Asian patients; using TNFi-failure data instead of TNFi-exposed data; and (in response to a clarification question), excluding the Phase II tofacitinib trial).

The ERG judged the NMA to be generally well conducted but identified the following issues:

- Use of the probit scale to model clinical response/clinical remission. Whilst an improvement on a previous approach in NICE TA342, the use of a probit model did not aid interpretability and readability of the CS. A multinomial logit model could have been considered.
- Assessment of inconsistency did not examine any potential inconsistency in the maintenance TNFi-naïve network between the two-arm and three-arm trial.
- Exploration of best model fit. The ERG conducted additional analyses and would have made different choices regarding model fit. In general, for the effectiveness outcomes, the ERG would have chosen random effects models as the more conservative approach given the known between-study heterogeneity. In contrast, for the safety outcome of serious infections, the absence of any events in the placebo arms of the tofacitinib trials caused very wide credible intervals even when the ERG investigated a fixed effect model. The ERG therefore also ran an analysis using a frequentist framework for this outcome which allows a value of 0.5 to be added to cells when a zero value is present in the input data.
- Uncertainty around absolute probabilities from baseline models. To estimate absolute probabilities of each event, treatment effects from the NMA were combined with an estimate of the placebo (baseline) response from the placebo arms of included studies. The ERG was unable to replicate the placebo credible intervals used in the probit or logit models. The company models tended to lead to wider

credible intervals compared to our calculations, thus would lead to conservative results.

- Inclusion of the tofacitinib Phase II trial. Results from NMAs for response and remission for TNFi-naïve and TNFi-exposed subgroups are similar to the base case when the tofacitinib Phase II trial is excluded. However, results without the Phase II trial were not provided for safety outcomes and in this NMA the Phase II trial may have had a disproportionate effect on the random effect NMA because of the relatively high serious infection rate in the tofacitinib arm of this study.
- No safety analysis for the maintenance period. The company stated they were unable to conduct a NMA for safety outcomes in the maintenance phase. However, the ERG believe this could have been achieved by using the mFAS population of OCTAVE sustain. However, the issue of combining studies with treat-through and re-randomised designs would still remain
- Lack of adjustment for differing lengths of induction and maintenance periods across trials. The company did not attempt to adjust for differences in lengths of induction and maintenance treatment and the ERG is concerned that this could have introduced potential bias against those treatments where studies had shorter induction phase and benefit those treatments with a shorter maintenance phase.
- Differences between patient populations in the re-randomised design maintenance trials. OCTAVE sustain re-randomised all responders from the OCTAVE induction trials to either placebo or tofacitinib treatment. In contrast, in the three other studies with a re-randomised design, only patients who had received and responded to active treatment were eligible to be re-randomised into the maintenance phase of the study. In the ERG's view the base-case may be biased in favour of tofacitinib and it would have been useful to have explored the mFAS population for OCTAVE Sustain in a sensitivity analysis.
- Adjustments to treat-through trials. Although the ERG does not believe the adjustments made by the company introduce additional bias, it is nevertheless the case that non-responders at the end of the induction phase are ignored (and these participants potentially could have become responders by the end of the maintenance phase).

In the induction phase for the TNFi-naïve population all treatments showed strong evidence of benefit over placebo with infliximab having the largest treatment effect for both clinical response and clinical remission. In the TNFi-exposed population, tofacitinib had the largest treatment effect on clinical response and clinical remission compared to placebo. Only tofacitinib and vedolizumab showed strong evidence of benefit.

In the maintenance phase for TNFi-naive population all treatments showed strong evidence of benefit over placebo, with tofacitinib 10 mg having the largest treatment effect on clinical response and clinical remission. In the TNFi-exposed population, tofacitinib 10 mg had the largest treatment effect on clinical response and clinical remission compared to placebo. Tofacitinib 5 mg, 10 mg and vedolizumab 300mg Q4W and Q8W all showed a strong evidence of benefit over placebo.

In the safety analysis, which was only conducted for the induction phase [redacted].

In final summary, the ERG has identified the following key limitations of the evidence presented in the CS:

- All the direct evidence on the effectiveness of tofacitinib is from trials of tofacitinib versus placebo. The majority of evidence for other active treatments also comes from placebo controlled trials. In the absence of direct head-to-head comparisons of the available active treatments NMAs were undertaken.
- Heterogeneity was present among the studies included in the NMAs. Although the company took steps to try and reduce heterogeneity the ERG would have preferred random effects models for the effectiveness outcomes.
- The NMA for serious infections in the induction phase was potentially affected disproportionately by the Phase II tofacitinib trial. In addition, for this outcome the placebo arms of the tofacitinib trials experienced zero events. The ERG would therefore have preferred a fixed effects model as a sensitivity analysis for this outcome. However, very wide credible intervals persisted even with a fixed effects model and therefore the ERG has run an alternative frequentist analysis to investigate the impact of adding a value to cells in analyses where there are no events in the tofacitinib or placebo arms.
- No NMA for safety outcomes was conducted for the maintenance phase.
- Biases may exist due to differing lengths of induction and maintenance periods across trials (with may bias against treatments with shorter induction phases and

benefit treatment with shorter maintenance phases), and differences between studies with a re-randomisation design (the base-case NMA may be biased in favour of tofacitinib).

4 COST EFFECTIVENESS

4.1 Overview

The company submission includes:

- A systematic review of published economic evaluations of tofacitinib and other therapies for people with moderately to severely active ulcerative colitis (CS B.3.1 pages 118 to 120 and Appendix G);
- A description of the methods and results of their model developed to assess the cost-effectiveness of tofacitinib in relation to the comparators and population specified in the NICE scope for this appraisal (CS B.3.2 to B.3.11 pages 120 to 172 and Appendices H, I, J and M).

We summarise and critique these elements of the CS in sections 4.2 and 4.3 below and present additional work conducted by the ERG in section 4.4, including model validation, corrections to the company's analyses and additional analysis.

All of the results in this chapter include a confidential patient access scheme (PAS) price discount that has been agreed for tofacitinib but not an existing confidential PAS discount for the comparator vedolizumab. Results including both PAS discounts are presented in a confidential addendum to the ERG report.

4.2 Company's review of published economic evaluations

The company conducted a systematic review of the literature to identify economic evaluations of tofacitinib or any other therapy for moderately or severely active ulcerative colitis. The methods and results of this review are described in section B.3.1 and Appendix G of the CS. The ERG considers that the company's search strategy and inclusion/exclusion criteria were appropriate. However, as the search was conducted in October 2017, we updated it to identify any more recent relevant publications.

The company included 53 publications, described in Table 175 (CS Appendix G.1.2.2). The main submission focusses on 10 UK studies reported in six full papers[25,52-56] and seven abstracts[57-61] (see Table 35 CS page 120). Three of the full papers reported analyses conducted by the Evidence Review Groups for previous NICE technology appraisals: Archer et al. (2015)[25] and Tappenden et al. (2016)[54] relate to TA329 of infliximab, adalimumab and golimumab;[8] and Essat et al. (2016)[53] relates to TA342 of vedolizumab. A paper by Wilson

et al. (2017)[62] reported on the cost-effectiveness analysis of vedolizumab compared with TNF-alpha inhibitors from the Takeda submission for TA342. Tsai et al. (2008)[55] reported a cost-effectiveness analysis of maintenance treatment for infliximab compared with standard care based on the ACT I and ACT II RCTs. This analysis was used to inform resource utilisation and cost estimates in TA329, TA342 and in this current appraisal – see section 4.3.6.3 below (page 163). The final paper identified in the company's search - Buckland et al. (2008)[52] - compared high and low dose mesalazine, so is not relevant to the current appraisal.

The ERG update search identified two additional publications: a full paper by Wilson et al. (2018),[62] reporting cost-effectiveness of vedolizumab compared with conventional therapy from the Takeda TA342 analysis; and a paper by Wu et al. (2018)[63] reporting a cost-utility analysis comparing sequenced strategies including conventional therapies, tofacitinib, adalimumab, vedolizumab, golimumab and infliximab from a UK and Chinese perspective.

The analysis by Wu et al. indicated that one of the treatment sequences shown in Table 52 would be optimal in the UK context, depending on the incremental cost-effectiveness ration (ICER) threshold. Other sequences gave fewer QALYs for a higher cost than one or more alternatives (simple or extended dominance). At a cost-effectiveness threshold of £30,000 per QALY gained, the optimal treatment sequence would be adalimumab at first line, tofacitinib at second line and then conventional therapy.

Table 52 Non-dominated treatment sequences, UK perspective (Wu et al. 2018)[63]

Sequence	Cost (£)	QALYs	ICER (£ per QALY)	Comparator
CT	132,769	10.49	-	-
ADA-CT	134,598	10.71	8,438	CT
ADA-TOF-CT	153,333	11.67	19,407	ADA-CT
TOF-ADA-CT	154,216	11.70	30,989	ADA-TOF-CT
TOF-VED-CT	182,728	12.37	42,511	TOF-ADA-CT

CT: conventional therapy; ADA adalimumab; TOF tofacitinib; VED vedalimumab

We consider the cost-effectiveness of sequential treatment strategies in exploratory ERG analysis, see section 4.3.2.2 below.

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

In most regards the company's economic evaluation follows the NICE reference case and the NICE scope for this appraisal (see Table 53). The exception is the company's exclusion of adalimumab, infliximab and golimumab as comparators for patients with prior exposure to a TNFi. For this subgroup, clinical response and remission rates are not available for infliximab or golimumab, but they are available for adalimumab. Therefore, the company could have included adalimumab as a comparator for the TNFi-exposed subgroup. We discuss the appropriateness of this comparison in section 4.3.2.2 below.

Table 53 NICE reference case

Criteria	Included?	Comment
Decision problem as in scope	Yes	
Comparators as listed in scope	No	Adalimumab, infliximab and golimumab not included for people with prior TNFi exposure
Perspective on costs: NHS and PSS	Yes	
Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	Yes	
Cost utility analysis with fully incremental analysis	Yes	
Synthesis of evidence on outcomes based on a systematic review	Yes	
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	
Health effect expressed in QALYs. EQ-5D is preferred measure of health related quality of life	Yes	
Health related quality of life reported directly by patients and/or carers.	Yes	
Preference data from representative sample the UK population	Yes	
An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	Yes	
Discount rate: 3.5% pa for costs & health effects	Yes	

4.3.2 Modelled decision problem

4.3.2.1 Population

The population in the company model aligns with the NICE scope - people with moderately to severely active ulcerative colitis who are either intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy or a TNF-alpha inhibitor.

Subgroup analysis by TNFi exposure

The scope requests subgroup analysis according to previous treatment with one or more biologic drugs. The company base case is presented for two separate subgroups, labelled as biologic-naive and biologic-exposed (CS B.3.2.1). They argue that this division is appropriate as clinical evidence indicates that prior exposure to biologics is an important treatment effect modifier and that patients' treatment history is a deciding factor in the treatment pathway in clinical practice. *We agree with this approach but note that labelling the subgroups according to 'biologic' exposure is misleading, as the NMA results used in the model are defined by prior exposure to TNF-alpha inhibitors alone (not vedolizumab).*

Analysis for the whole population (ITT NMA)

The company also presents a scenario analysis using results from an NMA for all patients in the induction trials and the re-randomisation responder trials of maintenance therapy (CS B.3.7.2.1 and D.1.3.5.1.2). The CS notes that this 'ITT' scenario analysis is susceptible to heterogeneity in the proportion of patients with prior TNFi exposure in the trials. In particular, the TNFi trials only included TNFi-naïve patients, whereas the vedolizumab and tofacitinib trials included a mixture of patients with and without prior TNFi exposure. The company argues that the comparison between tofacitinib and vedolizumab represents the 'least confounded' results from the ITT scenario and they exclude the TNFi drugs from the table of cost-effectiveness analysis (Table 63 CS page 156).

We note that there is a high degree of uncertainty over the results of the ITT NMA. In particular, the odds ratios for vedolizumab compared with tofacitinib are very close to 1 with wide credible intervals: for example, for maintenance therapy with 8-weekly vedolizumab compared with daily 5 mg tofacitinib, the estimated odds ratios are [redacted] for clinical response and [redacted] for clinical remission (CS Table 106 D.1.3.5.1.2). *The ERG does not consider the company's 'ITT' cost-effectiveness scenario to be reliable because of the high level of uncertainty in the underlying NMA. The scenario also omits relevant comparators (the TNFi drugs), so does not address the specified decision problem. We therefore focus on separate analyses for the two TNFi exposure subgroups in our*

discussion and additional analysis. This approach is consistent with committee considerations in the NICE appraisal of vedolizumab (TA342).[9]

Baseline characteristics

In the company model, utility values and mortality rates depend on the age and gender mix of the cohort. Assumptions about the distribution of body weight are used to estimate dose and hence costs for some medications (infliximab and azathioprine). The company base case assumes the following baseline characteristics for the two subgroups:

- *TNFi-naïve*: age 41.1 years, 59.7% male and body weight 74.8 kg
- *TNFi-exposed*: age 41.3 years, 58.8% male and body weight 72.6 kg

These characteristics are based on means from the tofacitinib arms in the OCTAVE Induction trials, see Table 54.

Table 54 Patient baseline characteristics (OCTAVE Induction trials)

Subgroup	Treatment	N	Male, n (%)	Age, mean (95% CI)	Weight, kg mean (95% CI)
TNFi-naive	Tofacitinib	417	249 (59.7)	41.1 (39.8, 42.4)	74.8 (73.2, 76.4)
	Placebo	104	64 (61.5)	43.2 (40.5, 45.9)	73.7 (70.8, 76.6)
	Total	**521**	**313 (60.1)**	**41.5 (39.9, 43.2)**	**74.6 (72.6, 76.5)**
TNFi-exposed	Tofacitinib	488	287 (58.8)	41.3 (40.0, 42.6)	72.6 (71.1, 74.1)
	Placebo	130	68 (52.3)	39.4 (36.9, 41.9)	72.3 (69.3, 75.3)
	Total	**618**	**355 (57.4)**	**40.9 (39.3, 42.5)**	**72.5 (70.6, 74.4)**
All patients	Tofacitinib	905	536 (58.8)	41.2 (39.9, 42.5)	73.6 (72.1, 75.2)
	Placebo	234	132 (52.3)	41.1 (38.5, 43.7)	72.9 (70.0, 75.9)
	Total	**1139**	**668 (58.6)**	**41.2 (39.6, 42.8)**	**73.5 (71.5, 75.4)**

Source: CS Table 36, page 121. Subgroup and treatment totals estimated by ERG.

We consider it more appropriate to characterise the modelled population using all patients randomised in the OCTAVE induction trials, including patients in tofacitinib and placebo arms. Furthermore, we note that the small differences between the subgroups may well be due to chance – a suggestion that is supported by the observation that the mean age of randomised patients in the TNFi-exposed subgroup (40.9 years) is less than that for those in the TNFi-naïve subgroup (41.5). This appears counter-intuitive, although clinical advice to the ERG is that most exacerbations requiring drug change occur in the first year. Thus, the average of patients in the two subgroups may well be similar.

For comparison, the median age at diagnosis of ulcerative colitis in the 2016 RCP audit was 32 years (interquartile range (IQR) 24 to 45) and the median age at initiation of biologic treatment was 39 years (IQR 28 to 52).[64] The gender distribution in the audit was 59% (529/903), similar to that in the OCTAVE trials.

We consider that the gender mix, initial age and weight of the model cohort should be assumed similar for people with and without prior exposure to TNFi drugs. In ERG analysis, we assume 59% males, initial age 41 years and weight 73.5 kg, based on means for both arms in the OCTAVE Induction trials. We conduct scenario analysis to assess the impact of age (28 to 58) and body weight (range 70 kg to 80 kg) on the results.

4.3.2.2 Comparators

The model assumes that patients start treatment with tofacitinib or the biologic comparators with an induction phase of treatment. Patients who respond during induction continue to receive maintenance treatment with the same drug (with concomitant use of conventional drugs) until loss of response or an acute exacerbation requiring surgery. Patients who do not respond to induction treatment and those who relapse during maintenance continue to receive conventional treatment alone, until planned or emergency surgery, or death.

Inclusion of comparators in economic analysis

Tables 40 and 41 in the CS (page 130) outline the comparators used in the company's economic analysis:

- **TNFi-naïve subgroup**, all comparators specified in the scope (infliximab, adalimumab, golimumab, vedolizumab, tofacitinib and conventional therapy (CT));
- **TNFi-exposed subgroup**, only vedolizumab, tofacitinib and CT are included. Cost-effectiveness is not reported for infliximab, adalimumab or golimumab.

For patients with prior exposure to TNFi drugs, infliximab and golimumab could not be included in the company's NMA due to a lack of trial evidence (CS section B.2.9.2.1). However, the TNFi-exposed NMA does include adalimumab, so the company could have included adalimumab in the cost-effectiveness analysis for this subgroup. The CS does not give a clear rationale for omitting adalimumab for the TNFi-exposed subgroup.

Clinical experts have advised the ERG that treatment with a TNFi would sometimes be considered for a patient with prior exposure to another TNFi. There is a group of patients who lose response to first TNFi (usually infliximab) for a variety of reasons, such as

pharmacokinetics and anti-drug antibody formation. If they have initially responded and then lost response (secondary loss of response) it would be current practice to switch to a second line TNFi (in-class switch). Those who do not respond to a first line TNFi (primary non-responders), and those who lose response with therapeutic serum trough TNFi levels and without anti-drug antibody formation, are usually switched out-of-class (e.g. to vedolizumab or tofacitinib).

The occurrence of in-class switching is also supported by evidence from the UK IBD Audit: 21% of patients starting adalimumab (17/83) had previously not responded or been intolerant to a TNFi (RCP 2015, page 49).[65]

The ERG considers that adalimumab is a relevant comparator for at least some patients with prior exposure to a TNFi agent. We therefore include adalimumab in ERG analysis for this subgroup. However, we understand that further treatment with a TNFi may not be appropriate for all patients in this subgroup.

Drug use and dosage

SmPC dose regimens and recommendations about when to stop treatment with tofacitinib and biologic comparators are set out in Table 38 (CS page 128). This table also summarises dose assumptions used for costing in the model, see Appendix M (CS M.1.1) for further explanation. Table 38 incorrectly specifies the doses of adalimumab in the model. Based on the licensed dose, patients would receive 160 mg + 80 mg + 2 x 40 mg = 320 mg during the 8-week induction period and 40 mg x 4 = 160 mg per 8 weeks of maintenance. We confirm that the correct doses for adalimumab have been coded in the model.

Dosing and use of conventional drugs are detailed in Table 39 (CS page 129), with further explanation in Appendix M (CS M.1.1). CT is assumed to comprise a combination of aminosalicylates (balsalazide, mesalazine, olsalazine and sulfalazine), corticosteroids (hydrocortisone rectal foam and oral prednisolone) and the immunomodulator azathioprine. Clinical advice to the ERG suggests that the company's assumption of equal usage for the four aminosalicylic acid (5ASA) drugs does not reflect UK practice, as mesalazine is much more commonly prescribed for this patient group. See section 4.3.6.1 (page 159) below for discussion of drug utilisation and costing assumptions.

Stopping rules for drug treatment

- *Discontinuation due to lack of response to induction therapy*
  CS Table 38 summarises SmPC recommendations about when to stop tofacitinib and biologic drug treatment. These recommendations relate to early assessment of response following induction treatment (from 8 to 16 weeks after initiation). In contrast, the clinical trials provide evidence of response at 6 weeks for golimumab and vedolizumab and at 8 weeks for other comparators, and the model assumes a fixed 8-week induction period followed by immediate cessation of treatment for patients whose disease does not show a response in this time. *If in practice clinicians assess response to induction later than 8 weeks, the average cost of induction therapy will be higher than that estimated by the company model. However, effectiveness may also be higher if some patients have a late response to induction. The direction and magnitude of the bias from assuming a fixed 8-week period of induction for all comparators is unclear.*

- *Discontinuation due to loss of response during maintenance*
  Guidance for the TNF-alpha inhibitors (TA329) and vedolizumab (TA342) recommend annual assessment of response, with treatment continuing only if there is clear evidence of ongoing benefit. Clinical advice to the ERG is that the benefit of biologic treatment is usually considered annually, in line with NICE guidance. However, treatment would usually be withdrawn earlier for patients who lose response, as the patient will seek an appointment when symptoms recur or get worse and this will trigger consideration of changing or stopping treatment.

  The company model applies a constant risk of relapse across each 8-week cycle of maintenance, with treatment stopping immediately when patients lose response. Thus, it assumes that maintenance treatment is stopped within 8 weeks of a loss of response. To achieve this, all patients on maintenance treatment must have fast access to clinical assessment on relapse or be seen routinely every 8 weeks. The company model assumes an average of 2 outpatient visits for patients in remission on maintenance treatment and 4.5 visits per year for patients with a response but no remission.

  The ERG considers that the assumption that treatment will be withdrawn following relapse reflects UK practice. However, we have concerns that the costs of monitoring and follow-up in the company's model do not reflect the full cost of ensuring that treatment can be withdrawn within 8 weeks of a relapse. We consider a scenario with

additional costs for outpatient visits to enable treatment cessation within 8 weeks of a relapse - see section 4.3.6.3 below.

- *Trial of withdrawal for patients in stable remission on maintenance treatment*
  TA329 and TA342 also recommend a trial of withdrawal for patients with stable remission after 12 months of treatment, with the option to restart treatment following relapse. The company model does not reflect these recommendations, as maintenance treatment is assumed to continue for as long as patients have a response. We have been advised that in practice, patients in sustained clinical remission are more likely to continue maintenance treatment, as clinicians and patients are reluctant to stop a drug that appears to be working.

- *Other causes of treatment discontinuation*
  The model assumes that all drug treatment, including conventional therapy, stops after emergency or elective surgery.

  The only AEs included in the model were serious infections (SI) (see section 4.3.4.2 below) but the model assumes that treatment continues following SI. The company's NMA of safety outcomes from the induction trials includes discontinuation due to AEs (CS B.2.10.8.2 pages 110 to 112).

  [redacted]

  However, these results were not used in the model, as the company argue that this would lead to double counting because definitions of adverse events in OCTAVE and other trials included worsening of ulcerative colitis, which is already accounted for. The company state that risks of discontinuation due to AE or other causes are low and likely to be outweighed by discontinuation due to lack of efficacy (CS B.3.2.5).

  Clinical advice suggests that tofacitinib or biological treatment would be temporarily withheld following serious infection. If the drug had been clinically effective prior to the infection, withholding the drug until the infection has cleared, and then re-starting the drug again would be an option: e.g. for infections such as tonsillitis, pneumonia and urinary infections. If the infection was opportunistic or severe, such as disseminated herpes virus or meningitis, it is likely that the drug would be permanently stopped. Other SAEs likely to result in treatment cessation include malignancy (e.g. lymphoma), a major cardiovascular event, severe infusion reactions, drug-induced lupus reactions,

hypersensitivity reactions, neurological events (such as demyelination, neuropathy, focal neurology) and joint pains. Some rashes also warrant cessation, especially psoriasis-like eruptions. Leucopenia would always require dose reduction or temporary cessation.

The ERG considers it likely that including discontinuation due to AE from the NMA in the model would cause some degree of double-counting. However, the assumption of no discontinuation due to serious infections or other AEs is also unrealistic and likely to introduce bias.

Surgical treatment

Unlike previous NICE TAs for ulcerative colitis - TA329 and TA342 - surgery is not specified as a comparator in the scope for this current appraisal. This reflects the TA329 and TA342 committee conclusions that patients and clinicians would rather avoid or delay surgery because of adverse effects on wellbeing, potential for complications and the irreversible nature of the intervention that were not captured in the economic evaluations. The company model treats elective surgery as an option for patients with moderately to severely active ulcerative colitis treated with conventional treatment alone. The model also includes a risk of acute exacerbation requiring emergency surgery for patients not in remission (active disease or response without remission).

Drug sequencing

The CS presents results for one line of treatment with tofacitinib or biological comparator, followed by CT or surgical treatment. However, the model includes the facility to compare scenarios with two lines of active tofacitinib/biological treatment before CT/surgical treatment, as in the analysis by Wu et al. (2018) described above (4.2).[63] Our clinical advisors have indicated that after an initial trial of CT alone, patients with moderately to severely active ulcerative colitis would start treatment with a TNFi agent (usually infliximab). Patients without a response in the induction phase and those who lose response on maintenance treatment would then either switch within-class to another TNFi (adalimumab or golimumab) or outside-class (currently vedolizumab).

We conduct scenario analysis to compare the cost-effectiveness of sequenced treatment with biologic/tofacitinib for people without prior TNFi exposure, including in-class switching (e.g. infliximab-adalimumab), step up (e.g. infliximab-vedolizumab, infliximab-tofacitinib) and step-down (e.g. vedolizumab-infliximab, tofacitinib-infliximab) strategies: see 4.4.3.

4.3.3 Model structure

The company describes the structure and key features of their model in CS Section B.3.2.2. The model structure is similar to that in previous ulcerative colitis appraisals TA329 and TA342. It is a Markov cohort model, with a cycle length of 8 weeks and patient lifetime horizon. The half-cycle correction is not incorporated. Costs and QALYs are discounted at an annual rate of 3.5%. The company's illustration of the model is reproduced in Figure 7.

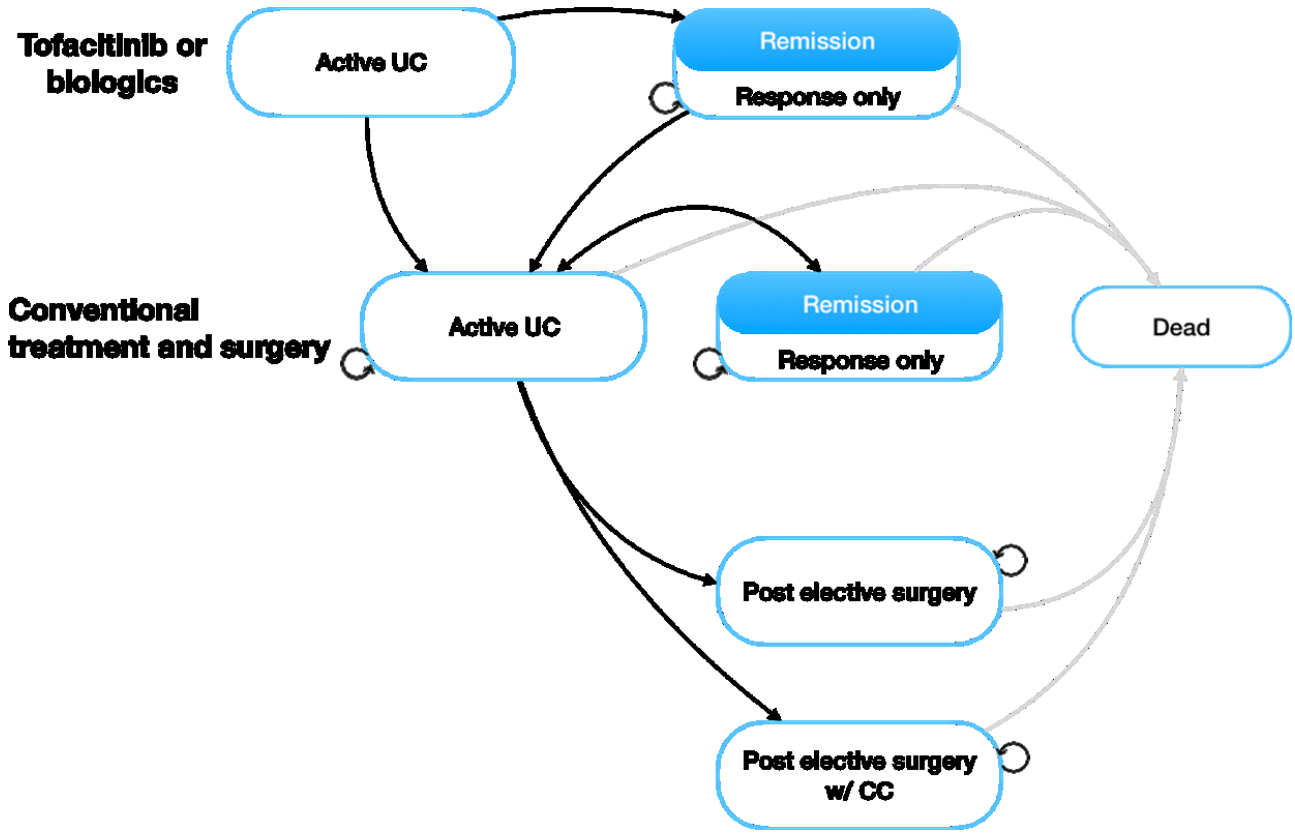


Figure 7. Company's model structure (Figure 31, CS B.3.2.2)

w/ CC = with colectomy complications; UC = Ulcerative Colitis

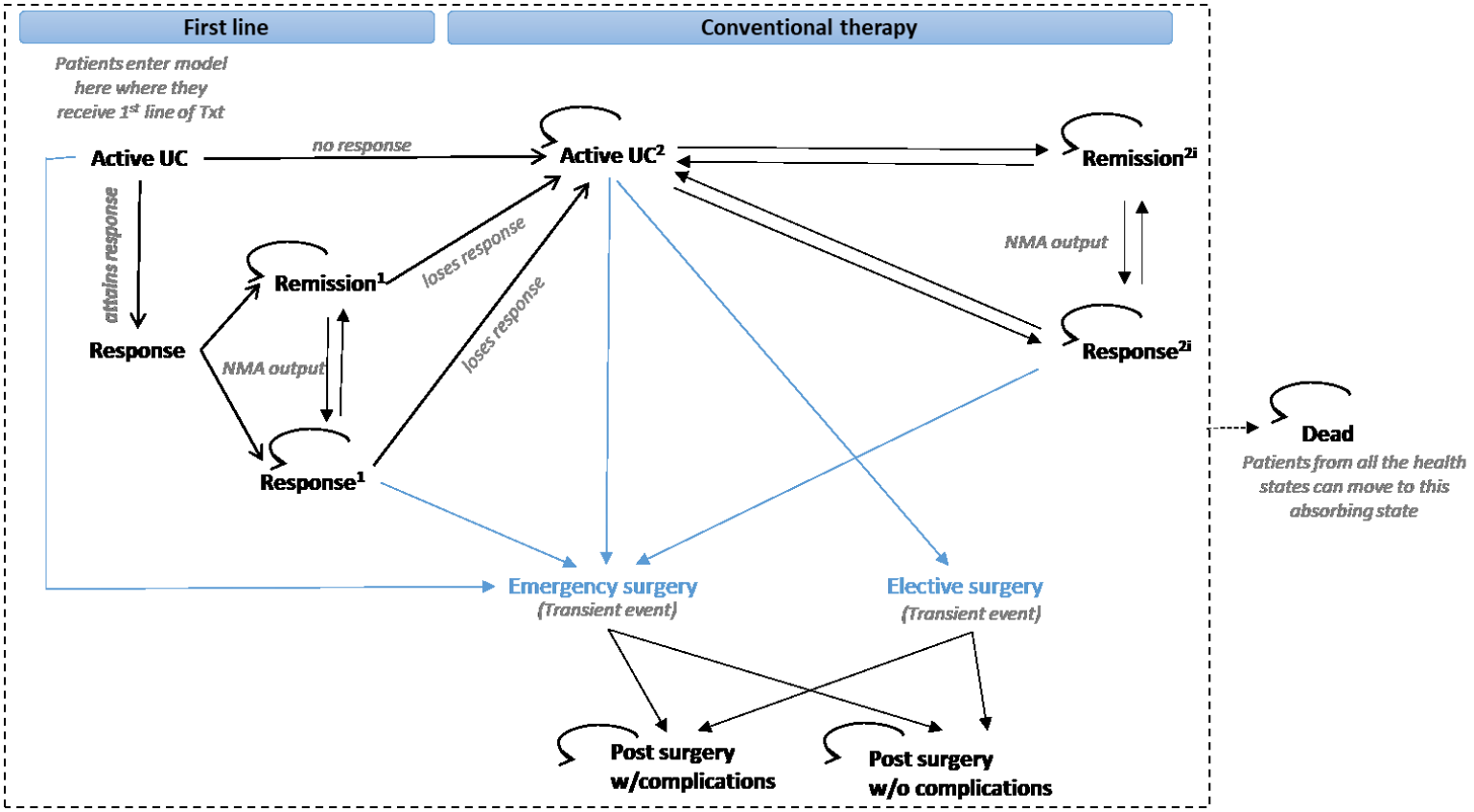
The model consists of nine health states, defined by stage of treatment (first line treatment with tofacitinib or biologic; conventional treatment; or post-surgery) and level of disease control (active UC; clinical response without remission; or remission), which we describe in Table 55. The transitions between the health states are further illustrated in Figure 8.

The company summarise key model assumptions and compare against previous ulcerative colitis appraisals in Tables 37 and 60 of the CS (pages 126 and 153 respectively). We critique of the model features and base case assumptions in section 4.3.7 below.

Table 55: Description of the model health states

	Health states	Description
Tofacitinib / biologic	1. Active UC	Patients enter the model with moderately to severely active ulcerative colitis following intolerance, inadequate response or loss of response to conventional therapy or a TNFi-alpha inhibitor. They commence treatment with an 8-week induction phase of treatment with tofacitinib or a biologic comparator.
	2.Remission	Of those who respond to induction treatment, a proportion attain remission (using clinical definitions of remission and response). Patients continue to receive maintenance treatment so long as they remain in response. For each 8-week maintenance cycle, the proportions of patients with a response and the proportion of responders in remission are estimated from the NMA.
	3.Response only	
Conventional treatment	4.Active UC	Patients transition to the Active UC state on conventional treatment following: • Non-response to tofacitinib/biologic induction • Loss of response in tofacitinib/biologic maintenance For the CT comparator arm, patients start in this state.
	5.Remisson	Patients may attain response with or without remission while on conventional therapy. Transitions between active UC, remission and response only health states continue to occur while patients receive ongoing conventional treatment.
	6.Response only	
Surgery	Emergency surgery *	In each model cycle, a proportion of patients who are not in remission (health states 3, 4 and 6) require emergency surgery due to an acute exacerbation.
	Elective surgery *	A proportion of patients in the Active UC health state are assumed to undergo elective surgery in each cycle.
	7.Post surgery without complications	Surgery is associated with perioperative risks of complications and mortality. Patients who survive surgery transition to one of two health states: with- or without long-term complications.
	8.Post surgery with complications	
	9.Dead	Absorbing state; the model accounts for: • Death from UC only occurs from surgery • Death from other causes (background mortality) occurs from all the health states

*The model treats surgery as a transient event: it is *NOT* a health state



[1] Patients who maintain response here continue to receive the 1st line of treatment in the maintenance phase

[2] Patients who fail to respond receive conventional therapy as 2nd line of treatment

[2i] Patients who maintain response here continue to receive the 2nd line of treatment which is CT in the maintenance phase

Figure 8. ERG illustration of patient transition in the model

The model uses three sets of input parameters:

- **Clinical inputs** that govern rates of response and remission and adverse event rates for comparator treatments, as well as the incidence and complication/mortality rates for surgery;
- **Utilities** for health states and disutilities for adverse events;
- **Resource use and costs** for drug acquisition and administration; monitoring and follow up, treatment of serious infections and surgery.

Values and sources of these parameters are summarised in Table 59 of the CS (page 149). We discuss and critique the parameter sources in sections 4.3.4, 4.3.5 and 4.3.6 below.

4.3.4 Clinical parameters

4.3.4.1 Response and remission

Choice of NMA models for economic analysis

The model uses NMA results to estimate the proportions of patients achieving clinical response and clinical remission in the induction and maintenance phases of treatment. The NMA results used in the company base case are reported in Tables 25 and 26 of the CS (pages 95 and 96). These correspond to the economic model inputs shown in Tables 43 and 45, respectively (CS pages 131 and 134).

See section 3.1.7 for the ERG summary and critique of the NMAs. We highlight key issues related to the company's choice of NMA models to use in their economic analysis.

- **Definitions of response and remission**
  The model uses locally read clinical response and clinical remission outcomes from OCTAVE and other trials (see Table 10 above for outcome definitions). The primary outcome for the OCTAVE trials - remission based on centrally-read endoscopic sub-scores – was not available from other studies in the networks. The company argue that local reading is "closer to real-world data", because clinicians make their own assessment of endoscopy results to inform treatment decisions (CS B.2.3.1.2.4). The NMA sensitivity analysis of centrally-read outcomes are gives similar results to the locally-read analysis. *The ERG agrees that locally-read clinical response/remission results are most relevant for the economic analysis.*

- **Choice of fixed effects versus random effects**

  The company state that their choice of NMA models was based on DIC measures of model fit, but that they preferred the simpler fixed effect approach when DIC statistics were similar (CS B.2.9.2.1.1). Table 56 below summarises the NMA models chosen for the company base case analysis.

Table 56 Selection of response/remission NMA models

	Patient subgroup	Induction	Maintenance
Company base case	TNFi-naive	Random effects	Fixed effects
	TNFi-exposed	Fixed effects	Fixed effects
ERG preference	TNFi-naive	Random effects	Random effects
	TNFi-exposed	Random effects	Fixed effects *

* Random effects model would not run for the maintenance NMA

The ERG has a general preference for the random effect NMA models, as we believe that the fixed effect models may underestimate uncertainty due to heterogeneity between the studies. We test the impact of different NMA models on cost-effectiveness results in section 4.4.3 below.

- **Combination of TNFi-failed and TNFi-exposed subgroups**

  The base case NMAs combine outcomes for subgroups defined as TNFi-failed for vedolizumab with TNFi-exposed subgroups for tofacitinib and adalimumab (CS Table 22). The company conducted a sensitivity analysis for the TNFi-failure subgroup, which reduced the probit score for tofacitinib by -0.13 in the induction phase, bringing it closer to vedolizumab. (CS Table 28). They reported that results were not available for adalimumab and that the analysis could not be run for the maintenance phase. Therefore, the TNFi-failure NMA sensitivity analysis does not provide the required input parameters and was not used in the economic model.

The ERG considers that combining results for TNFi-failed and TNFi-exposed subgroups is a potential source of bias in favour of tofacitinib. We conduct a scenario analysis using a more like-for-like comparison between tofacitinib and vedolizumab, using data for the TNFi-failed subgroups from the OCTAVE and GEMINI trials (see Table 18 *in section 3.1.7*

Transformation of NMA results to transition probabilities

Results of the clinical response/remission NMAs were transformed from the probit scale to the natural scale and converted to absolute probabilities for use in the model. The probability of response is calculated using the P = 1- ϕ(θ) formula, where ϕ is the inverse of the cumulative normal distribution and θ is the sum of the probit scores for placebo and active treatment. The probability of remission is calculated using P = 1- ϕ(θ) formula, where ϕ is the inverse of the cumulative normal distribution and θ is the sum of the probit scores for placebo, active treatment, and remission.

For the induction phase, the proportions of the cohort with active disease, response but no remission and remission at the end of the first 8-week model cycle are shown in Table 57 below, by treatment and TNFi exposure subgroups.

Table 57 Distribution of cohort by health state at end of induction

	TNFi naïve subgroup			**TNFi exposed subgroup**		
	Active UC	Response only	Remission	Active UC	Response only	Remission
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Conventional	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

NA, results not available from network meta-analysis

Some further assumptions are needed to calculate 8-week transition probabilities from the 52-week NMA response/remission rates. The company describes the approaches taken in previous NICE technology appraisals in section B.3.3.1.2 (page 132) of the CS.

- In the TA329 MTA (adalimumab, infliximab and golimumab), the assessment group had access to mid-point response and remission data for the maintenance period.[25] They used these data to estimate transition probabilities for two phases of maintenance - week 8 to 32 and week 32 to 52. The results are generally more favourable for the TNFi drugs in the second period than in the first.
- In the TA342 STA (vedolizumab), the company used a calibration approach to fit transition probabilities to the 52 week NMA results. This involved applying certain constraints, such as that no more than 20% of people with mild disease would enter

remission. This approach was criticised by the TA342 ERG for using arbitrary constraints and assumptions.

In the present appraisal, the company note that they considered both of these approaches, but without success: due to a lack of mid maintenance period results for some comparators; and a failure to accurately predict the target data with calibration.

Instead, the company take a simpler approach by assuming constant risks within and beyond the one-year trial data. The probability of loss of response is calculated from the probability of no response over 52 weeks from the NMA (1 minus the probability of response), adjusted to the 8-week model cycle. Members of the cohort who maintain a response in each cycle are then split between remission and response only health states using a fixed proportion (the ratio of 52-week probabilities of response with and without remission). The resulting estimates of the 8-week probabilities of loss of response and the proportions of patients in response with and without remission are shown in Table 58.

Table 58 Parameters used to model change of health state during maintenance

	TNFi naïve subgroup		TNFi exposed subgroup	
	Probability of losing response (per 8 weeks)	Percentage of responders in remission	Probability of losing response (per 8 weeks)	Percentage of responders in remission
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]
Golimumab 50mg	[redacted]	[redacted]	[redacted]	[redacted]
Golimumab 100mg	[redacted]	[redacted]	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]	[redacted]	[redacted]
Tofacitinib 10mg	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab Q8W	[redacted]	[redacted]	[redacted]	[redacted]
Vedolizumab Q4W	[redacted]	[redacted]	[redacted]	[redacted]
Conventional	[redacted]	[redacted]	[redacted]	[redacted]

Adapted from CS Table 45 page 134.

These calculations are mathematically correct, but we emphasise that they rely on assumptions of a constant risk of loss of response and constant ratio of patients in remission and response throughout maintenance treatment. Clinical advice to the ERG is that these assumptions might not be realistic. Experience with TNFi agents suggests that most serious exacerbations requiring drug change occur in the first year of treatment. Loss of response continues after a year of therapy but tails off in the second and subsequent years. Further,

the proportion of patients with a response and in remission is likely to increase over time, because responders without remission are more likely to stop or switch therapy (or have surgery) whereas those in remission will continue. Thus, the only-responders will tend to drop out faster than those in remission.

Similar concerns were raised by the NICE committee for TA329, which noted a discrepancy between modelled estimates of treatment duration and expert advice that of patients who start a TNFi, one third to one half are expected to continue therapy in the long term (paragraph 4.71).[8]

Results from the OCTAVE Open study are suggestive of similar trends in long-term maintenance of response and remission with tofacitinib (CS B.2.6.3.1 and Appendix L Table 233).

[redacted]

We conclude that the model assumption of constant risk of loss of response for patients on maintenance treatment does not reflect clinical experience. Extrapolation of relapse and discontinuation rates from the maintenance trials is likely to underestimate the average duration of treatment and hence both the costs and QALYs of active treatments. However, it is not possible to estimate the net direction of bias in ICERs between comparators, because trends in long-term risks may vary between TNFi drugs, vedolizumab and tofacitinib.

4.3.4.2 Adverse events: serious infection rates

The company conducted three NMAs on safety, based on data from induction phase RCTs, as described in sections 3.1.7 above (CS B.2.10.8.1). These include discontinuations due to adverse events and incidence of serious adverse events (SAE), but the company model only uses results from the serious infection (SI) NMA (CS B.3.3.3).

Exclusion of other serious adverse events

The company explain that they excluded adverse events other than serious infections because the most common SAEs reported in the trials were GI events, events related to ulcerative colitis, or “worsening of disease”, which may already be accounted for in the model through loss of response and remission, as described above.

Advice from our clinical expert suggests that there are other SAEs, such as malignancy and cardiac events, which though small in number are significant for patients and incur considerable cost to the NHS. This observation is in line with the approach taken in NICE TA 342 which included TB, malignancy (due to lymphoma), acute hypersensitivity reactions and skin site reactions, in addition to SIs.

We agree that there would have been a risk of double-counting the costs and effects of ulcerative colitis exacerbations had all SAEs had been included in the model. The omission of non-infection SAEs does introduce a risk of bias but given the frequency of these events this omission is unlikely to change the cost-effectiveness results.

NMA method for serious infections

The company applied a binomial logit NMA model to estimate the risk of serious infections in the induction trials (CS Table 34 page 111). They chose the random effects model for their base case because the DIC statistic was lower than for the fixed effects model.

The company acknowledges substantial uncertainty in the precision of estimates from the SI NMA, which gave a very high upper limit to the credible interval for all comparators and for tofacitinib in particular because there were no cases of serious infection in the placebo arms of the tofacitinib induction trials. The company note that if the credible interval limits for the SI risks are used in deterministic sensitivity analysis, this parameter has the greatest impact on the ICERs. They argue that this would be misleading and instead apply arbitrary limits around the SI risk for tofacitinib of 0% to 50% increase from placebo.

Whilst the ERG agrees that there is considerable uncertainty associated with the risk of serious infections, we have reservations about the company's approach to estimating this parameter (discussed in detail in section 3.3.10.2). Our verification checks indicated an even higher level of uncertainty around tofacitinib estimates, and we were unable to replicate the company's base case NMA values. We therefore applied a frequentist NMA approach to estimate the risk of serious infection, which we use as a scenario in ERG analysis (details in section 4.4.3).

Transformation of NMA results to SI probabilities

Table 59 shows the probabilities of serious infections used in the company base case, with ranges for sensitivity analysis. The probabilities are estimated from incidence during the induction phase (assumed to be 8 weeks), which is assumed to apply to each subsequent 8-week cycle of maintenance treatment. Except for the tofacitinib arm, the central estimates

from the company's NMA are similar to the ERG frequentist estimates, but the latter approach give more plausible ranges of uncertainty.

Table 59 Probabilities of serious infections used in model (per 8-week cycle)

Treatment	Company (Bayesian RE)			ERG (Frequentist RE)		
	Base case	Lower limit	Upper limit	Base case	Lower limit	Upper limit
Placebo	[redacted]	[redacted]	[redacted]	0.67%		
Adalimumab	[redacted]	[redacted]	[redacted]	0.58%	0.08%	4.15%
Golimumab	[redacted]	[redacted]	[redacted]	0.11%	0.01%	1.25%
Infliximab	[redacted]	[redacted]	[redacted]	0.44%	0.05%	3.90%
Tofacitinib *	[redacted]	[redacted]	[redacted]	1.90%	0.29%	12.57%
Vedolizumab	[redacted]	[redacted]	[redacted]	0.15%	0.01%	1.89%

* By assumption, the company limits range for tofacitinib sensitivity analysis

The company made a number of assumptions in relation to serious infections. First, the risk of serious infection is assumed to be same regardless of patients' prior experience of treatment with TNFi-agents. The duration of serious infections is also assumed to be the same for all comparators: the model applies a disutility for the duration of the 8-week cycle in which the infection occurs. These are simplifying assumptions that appear reasonable.

A rather stronger assumption is that the risk of serious infection is constant throughout treatment (i.e. probability of SI is same in the induction and maintenance phases and regardless of the length of maintenance). The company test this assumption with a scenario in which serious infections are only assumed to occur only in the induction phase.

4.3.4.3 Incidence of emergency and elective surgery

The company conducted a focused search to identify estimates for the probability of colectomy and related complications (see CS section B.3.3.2 and CS Appendix M, section M.3). Misra et al. (2016)[66] is chosen to inform estimates of the cumulative risks of emergency and elective surgery in the base case – see Table 60.The company argues that this study is the most appropriate source as it: was based on a retrospective analysis of UK Hospital Episode Statistics (HES) for ulcerative colitis with a follow-up of 15 years since diagnosis; consisted of a larger and more contemporary cohort; excluded surgery due to colorectal cancer and provided a split for elective and emergency surgery rates.

Table 60 Colectomy rates used in the base case model

	Cumulative risks	Risk per cycle	Value used in the model
Colectomy	6.9%	0.073%	--
Elective colectomy	5.5%	0.058%	0.058%
Emergency colectomy	2.0%	0.021%	0.021%

Source: Misra et al. 2016 (UK HES Data)[66]

The CS reports on 3 other studies: Chhaya et al. 2015; Solberg et al. 2009 (used in TA329 AG model) and Frolkis et al. 2013 (used in TA342).[67-69] The Company use estimates from Frolkis et al. to inform their sensitivity analysis.

The ERG agrees with the company's selection of the Misra et al. study for the base case estimate of surgery risks. For completeness, we test rates from Chhaya et al. in scenario analysis, although we consider it unlikely to influence the results.

4.3.4.4 Colectomy complications and mortality

Perioperative complications

In their base case, based on UK Inflammatory Bowel Disease (IBD) 2014 audit, the company assumed that 32% of patients who underwent elective surgery and 35% of patients who underwent emergency surgery had perioperative complications.[64] Although the rates were doubled in the sensitivity analysis, they did not influence the base case results.

Post-operative complications

The company also included an ongoing risk of pouchitis after elective or emergency surgery. The base case risk was 1.46% per 8-week cycle, based on a Belgian study by Ferrante et al. (2007).[26] The risk was varied in company sensitivity analysis based on a Japanese study by Suzuki et al (2014) [37]. *To explore the sensitivity of results to pouchitis risk, we conducted a scenario analysis similar to that by Tappenden et al. (2016)[54] using a Japanese study by Arai et al. (2010)[70] which reported overall incidence of early and late complications (see section 4.4). It is worth noting that we do not anticipate change in this parameter to have a substantial impact on the base case results.*

Perioperative mortality

The company assumed the same perioperative mortality rate for patients undergoing elective and emergency surgery. In the base case, the mortality risk per operation was estimated to be 2.8% based on the reduction in overall mortality by 19% between round 3

and round 4 of the IBD audit.[64] Our clinical advisor has noted that although the overall surgical mortality may be around 2.8%, emergency surgery will carry a higher risk. The ECCO guidelines quote a mortality rate of 5-8% for emergency surgery and <1% for timely elective surgery in "specialised centres". We view the company's approach of taking an average rate across elective and emergency surgery as a reasonable simplification.

4.3.4.5 All-cause mortality

The model assumes that ulcerative colitis and treatment does not have any influence on mortality, with the exception of perioperative deaths. All-cause mortality risks, adjusted for age and gender-mix, for the general population from the UK Life tables are applied to patients in pre- and post-surgery states. The same approach was used in the assessment group model for TA329, although in TA342 the company applied state-specific relative risks to include an excess risk of death due to ulcerative colitis: 1.9 for moderately to severely active ulcerative colitis and 1.3 for post-surgery ulcerative colitis states .[9] *We consider that the approach in the current appraisal is acceptable. Although there are additional mortality risks not reflected in the model – e.g. for colorectal cancer – the relative risk estimates are likely to include perioperative deaths already accounted for.*

4.3.5 Health related quality of life

The model includes 7 utility parameters:

- A baseline utility for people without ulcerative colitis, adjusted for age and gender;
- 4 multipliers to reflect reduced utility (compared with no ulcerative colitis) for the health states:
    - Active ulcerative colitis;
    - Clinical response without clinical remission
    - Clinical remission;
    - Post-surgery.
- A utility multiplier for the effect of surgical complications;
- And a utility multiplier for the adverse effect of serious infections.

Parameter estimates were obtained from a systematic review of the literature on utility in ulcerative colitis (CS B.3.4.3 and Appendix H) and analysis of EQ-5D utility data from the OCTAVE trials (CS B.3.4.1 and Appendix M).

Utility estimates from published literature

The company conducted a systematic search for utility estimates (CS B.3.4.3 and Appendix H). We consider that the search strategy was satisfactory. As the search was conducted

over six months ago, we updated it, but did not identify any additional relevant studies. The company included 115 studies in their review, 44 of which reported EQ-5D utilities (Table 185, CS Appendix H). In the main submission, the company focus on 11 published studies reporting EQ-5D utility estimates for more than one relevant health state, in addition to economic analyses conducted for NICE TA329 (Archer et al. 2016)[25] and TA342 (Takeda 2014)[48] (see CS Table 50 B. 3.4.3 page 141). Utility parameters from published sources used in the company analysis are shown in Table 61. The company use estimates from Woehl et al. (2008)[71] in their base case and estimates from Swinburn et al. (2012),[72] in order to align with previous NICE technology appraisals for ulcerative colitis (TA329 and TA342).

Table 61 Utility parameters from the literature used in model

Source	Health state	Utility	ERG comments
Ara & Brazier[73]	No disease	Initial values TNFi-naive 0.8968 Prior TNFi 0.8960 Declines over time	Depends on age and gender of cohort. Formula derived from Health Survey for England 2003 and 2006 EQ-5D-3L (n=25,080). Regression coefficients not included in PSA.
Woehl et al.[71]	Active UC	0.4713	Utility multipliers calculated with respect to remission state. Used to adjust 'no disease' in company base case.
	Response	0.8736	
	Remission	1.0000	
	Post-surgery	0.8161	
Swinburn et al.[72]	Active UC	0.6317	Utility multipliers with respect to remission state. Active UC mean of 'severe' and 'moderate' utilities. Used in company scenario analysis.
	Response	0.8944	
	Remission	1.0000	
	Post-surgery	0.6596	
Diamantopoulos[74]	Serious infections	0.9858	Utility multiplier with respect to remission state
Kosmas (2015)[75]	Post-surgery complication	0.7889	Utility multiplier with respect to post-surgery state

Utility estimates from the OCTAVE trials

EQ-5D outcomes from the OCTAVE 1 and 2 induction trials and the OCTAVE sustain maintenance trial are outlined in CS B.2.6.1.2 and B.2.6.2.2, with further information in Table 218 (CS L.1.4) and Figures 54 to 61 (CS M.4). We discuss EQ-5D results from the OCTAVE induction and maintenance trials in section 3.3.7 above. To summarise, patients randomised in OCTAVE 1 and 2 were given an EQ-5D-3L questionnaire at baseline, 2 and 8 weeks, and patients in OCTAVE Sustain were given the questionnaire at baseline, 4, 8, 16, 24, 32, 40 and 52 weeks. Utility scores were calculated using UK preference weights, so are consistent with the NICE Reference Case.[76]

The CS reports analysis of EQ-5D data from the OCTAVE Induction trials to assess change in utility over time based on final health state at week 8. It is stated that the analysis was conducted separately for the TNFi naïve and exposed subgroups, using the full analysis dataset and a 'non-responder imputation method' (CS M.4). The company concluded that this analysis showed 'homogeneity' in mean EQ-5D index by final health state, although no statistical analysis was presented to support these claims. The company then used simple methods to estimate utility parameters from the OCTAVE data, which they used in scenario analysis their original submission (see Table 62 below).

Table 62 Simple estimates of health state utilities from OCTAVE EQ-5D data

Health state	N	Assumed utility	Assumed range (Min-Max)	Comments / assumption
Active UC	[redacted]	[redacted]	[redacted]	Mean of EQ-5D scores at baseline for participants in OCTAVE Induction 1 and 2 trials
Response no remission	[redacted]	[redacted]	[redacted]	Mean area under EQ-5D curves over one year for OCTAVE Sustain participants in remission or response-no-remission states at end of trial (see CS M.4 Table 238)
Remission	[redacted]	[redacted]	[redacted]	

Adapted from CS B.3.4.1 Table 49

In response to clarification questions, the company conducted further analysis of OCTAVE trial data. Linear mixed effect models were applied, grouping patients by health state (clinical remission, clinical response but not remission, active UC) at the trial endpoints (week 8 for OCTAVE 1 and 2, and week 52 for OCTAVE Sustain). Covariates tested included baseline EQ-5D, treatment, prior TNFi exposure, corticosteroid use at baseline, geographic region. We reproduce the results from the company response to clarification question B2 in Table 63 below.

The order of health state mean utilities are logical: for each trial dataset, estimates are highest for patients in remission and lowest for patients without a response. The company note that the mean utility estimates for each health state are higher in the maintenance trial than in the induction trials (although the confidence intervals overlap). This might support the view that primary non-responders (participants in the induction trials who had not had a response by week 8 and were excluded from the maintenance trial) are different to secondary non-responders (participants who started maintenance therapy with a response but lost this over the year of follow up). The company use these results to conduct two scenario analyses around their base case analysis, see Table 64. However, the company emphasise that both these regression-based estimates and their earlier simple estimates of

health state utility values do not sufficiently address the difficulties relating to the re-randomisation design of the OCTAVE Sustain study.

Table 63 Linear mixed model estimates of utility by health state from OCTAVE EQ-5D data (reproduced from Table 11 clarification response question B2)

Efficacy endpoint [a]	OCTAVE Induction 1 and 2 Values at week 8		OCTAVE Sustain Values at week 52	
	N [b]	Adjusted mean (95% CI) [c]	N [d]	Adjusted mean (95% CI) [e]
Non-clinical response	[redacted]	[redacted]	[redacted]	[redacted]
Clinical response (but not clinical remission)	[redacted]	[redacted]	[redacted]	[redacted]
Clinical remission	[redacted]	[redacted]	[redacted]	[redacted]

[a]Efficacy endpoints are based on NRI and Local Read of Endoscopy.
[b] N = number of subjects with non-missing EQ-5D data at week 8
[c] Adjusted mean derived from the linear mixed effects model: Score = Treatment + Prior treatment with TNFi therapy + Corticosteroid use at baseline + Geographic region + Week + Treatment*Week + Baseline EQ-5D with subjects as random effect
[d] N = number of subjects with non-missing EQ-5D data at week 52
[e] Adjusted mean derived from the linear mixed effects model: Score = Treatment + Induction Treatment + Baseline Remission Status + Week + Treatment*Week + Baseline EQ-5D with subjects as random effect.
Abbreviations: CI, confidence interval

Table 64 Additional company scenarios for OCTAVE utility estimates (Clarification Response question B2)

Scenario	Health state	Induction (first cycle)	Maintenance
1	Active UC	Week 8: [redacted]	Week 8: [redacted]
	Response only	Week 8: [redacted]	Week 52: [redacted]
	Remission	Week 8: [redacted]	Week 52: [redacted]
2	Active UC	Week 8: [redacted]	Week 52: [redacted]
	Response only	Week 8: [redacted]	Week 52: [redacted]
	Remission	Week 8: [redacted]	Week 52: [redacted]

The company's simple and regression-based analyses of EQ-5D data from the OCTAVE trials are problematic as sources of utility parameters for the economic model. They are relevant to the decision problem and clinical evidence, but the re-randomisation design and lack of intermediate assessments of clinical response and remission between week 8 and week 52 complicate the interpretation of results. We therefore agree with the company that the utility estimates by Woehl et al. [71] provide a more appropriate source for base case parameters that are consistent with previous NICE appraisals for ulcerative colitis. We use these estimates in ERG preferred analyses, but also test scenarios based on the company's OCTAVE analyses and published sources (Swinburn et al.) [72].

4.3.6 Resource use and costs

4.3.6.1 Drug acquisition

The assumptions underlying drug cost calculations are outlined in section B.3.5.1 (CS pages 143 to 145). Further detail is given in Appendix M (CS M.5).

Tofacitinib and biologic comparators

Table 52 (CS page 144) lists total costs per 8-week cycle for induction and maintenance treatment for tofacitinib and the biologic drugs. However, we note that several of the per-cycle costs in this table do not match the figures in the company's model - see Table 65 below for the drug acquisition costs from the model.

Table 65 Drug acquisition cost for tofacitinib and biologics

Drug	Induction (per 8 weeks)		Maintenance (per 8 weeks)	
	Dose	Cost	Dose	Cost
Tofacitinib [a]	10 mg twice daily	[redacted]	5 mg twice daily [d]	[redacted]
			10 mg twice daily	[redacted]
Adalimumab	160 mg week 0, 80 mg week 2 & 40 mg week 4 & 6	£2,817	40 mg every other week [d]	£1,409
			40 mg every week	£2,817
			27% every week [e]	£1,789
Golimumab [b]	200 mg week 0, 100 mg week 2 & 50 mg week 6	£3,052	50 mg every 4 weeks [d]	£1,526
			100 mg every 4 weeks	£1,526
Infliximab [c] (biosimilar)	5 mg/kg week 0, 2 & 6	£5,269 (£4,742)	5 mg/kg every 8 weeks	£1,756 (£1,581)
Vedolizumab	300 mg week 0, 2 & 6	£6,150	300 mg every 8 weeks [d]	£2,050
			300 mg every 4 weeks	£4,100

a Includes confidential PAS discount for tofacitinib. [redacted]

b Costs for golimumab assume provision of 100 mg dose at same cost as 50 mg dose as agreed in patient Access Scheme (TA329)

c Base case analysis assumes use of infliximab biosimilar (Remsima or Inflectra). Costs allow for wastage (no vial sharing) estimated by simulated distribution of body weight based on means and standard deviations for patients at baseline in the OCTAVE Induction trials.

d Base case analyses in bold. Alternative doses used in scenario analysis.

e Following assumption by ERG in TA329: in maintenance, 73% of patients have 40 mg of adalimumab every other week and 27% of patients have 40 mg of every week

The model includes a confidential Patient Access Scheme (PAS) discount for tofacitinib and the golimumab PAS agreement to supply 100 mg tablets at the same price as 50 mg tablets (TA329). All other drugs are at list price. We note that there is a PAS discount in place for

vedolizumab that is not factored into these costs. We present results including the vedolizumab PAS discount in a confidential addendum to this report.

In addition to the standard dose used in the base case calculations (in bold in the above table), the company presents scenarios for higher maintenance doses for tofacitinib (20 mg per day), adalimumab (27% of patients have 40 mg every week), golimumab (100 mg every 4 weeks) and vedolizumab (300 mg every 4 weeks). In the base case, the company assumed use of a biosimilar for infliximab (Remsima or inflectra). We have been advised that a significant minority of patients on infliximab will be on 6-weekly dosing (around 25-30%, compared with more than 50% on 8-weekly dosing). However, as the cost-effectiveness results are not sensitive to an increase in the cost of infliximab, we do not explore this further.

Cost calculations in the model are correct based on the stated assumptions about dosage and current NHS list prices (MIMS June 2018). Estimates are similar to those in the company model for TA342 (vedolizumab), with the exception of the induction cost for golimumab (in TA342 the company assumed 6, 50 mg doses). We consider the assumption of 3 100mg and 1 50mg dose (as in the current company's submission) to be more reasonable.

Conventional treatment

The costs of conventional drug treatment as a comparator and concomitant with biologic or tofacitinib are summarised in Table 53 of the CS (page 145). These costs match those used in the company base case model, with the exception of azathiopine which is costed in the model allowing for wastage. We summarise the costs used in the company base case analysis in Table 66 below.

Estimated usage is based on reported concomitant medication in the 2016 RCP audit of biological treatment for IBD.[64] The company assumes that for patients on conventional therapy alone, the proportions of patients prescribed the three main classes of drugs (aminosalicylates, corticosteroids and immunomodulators) are similar to reported use at initiation of biological therapy in the audit (50.3%, 47.9% and 46.4% respectively). Concomitant usage rates were based on reported use after three months of biological treatment (46.4%, 20.1% and 37.3% respectively). Azathioprine is excluded from the estimated cost of conventional therapy concomitant with tofacitinib, as this combination is not recommended. Further assumptions were made about usage within the drug classes and dosage – see Appendix M (CS M.5).

Table 66 Drug acquisition cost for conventional treatment

Drug	Per 8 weeks		Usage (% of patients) [c]		
	Dose	Cost	CT alone	With biologic	With tofacitinib [b]
Aminosalicylates					
Balsalazide	1.5 g twice daily	£52.42	12.6%	11.6%	11.6%
Mesalazine	1.2 g daily	£54.90	12.6%	11.6%	11.6%
Olsalazine	500 mg twice daily	£300.53	12.6%	11.6%	11.6%
Sulfasalazine	0.5 g twice daily	£6.87	12.6%	11.6%	11.6%
Corticosteroids					
Prednisolone	20 mg daily	£6.79	44.1%	19.9%	19.9%
Hydrocortisone [d]	Once every other day	£18.66	3.8%	0.6%	0.6%
Immunomodulators					
Azathioprine	2 mg/kg daily [a]	£7.48	46.4%	37.2%	0.0%
		Total cost	**£59.30**	**£52.18**	**£49.40**

a Costs for azathioprine allow for wastage estimated by simulated distribution of body weight based on means and standard deviations for patients at baseline in the OCTAVE Induction trials.
b Azathioprine not recommended for concomitant use with tofacitinib
c Usage estimated from RCP national IBD audit (2016): at initiation of biologic treatment for CT alone; after 3 months of biologic treatment for concomitant treatment.
d Rectal foam

The company has assumed equal usage of the four aminosalicylic acid (5ASA) drugs. However, we have been advised that almost all 5ASA use in the UK is mesalazine. Sulphasalazine is restricted to those with joint disease, and Olsalazine and Balsalazide are very rarely prescribed. Given the high cost of olsalzine, this suggests that the cost of 5ASA drugs is over-estimated. However, doses of 5ASA in patients with active disease, such as those starting tofacitinib or biological therapies, are likely to be maximised; e.g. mesalazine 4.8 g per day.

The above estimates of the cost of conventional therapy are lower than those used in the previous NICE appraisal of vedolizumab (TA342): £204.80 for CT alone and £102.40 concomitant with biologic therapy. However, the TA342 estimates were based on expert opinion, with the assumption that CT costs would be halved it taken with a biologic drug. We consider that the estimates in Table 66 are likely to be more reflective of NHS practice, since they are based on national audit data.

Overall, we consider the drug acquisition costs used in the company model to be realistic. We note that there have been some small changes in NHS prices for included drugs; sulfasalazine (£7.83), prednisolone (£0.47) and azathioprine (£2.20) (MIMS June 2018).

These changes result in a very small reduction in the estimated cost of CT alone (£58.02), with biologic drugs (£51.68) and with tofacitinib (£48.86).

Drug wastage calculations

The dose of infliximab and azathioprine are based on body weight. The company apply assumptions about wastage in their cost calculations, assuming no vial sharing. The wastage calculation methods are described in CS Appendix M (section M.5). In the base case, the company uses the method recommended by Hatswell et al. (2016)[77], with the distribution of body weight simulated from means and standard deviations for men and women in the OCTAVE Induction trials.

The ERG agrees with the company's approach to costing wastage for IV drugs. The model includes an option to use mean body weight from the trials, assuming vial sharing, but we do not consider this further as it is not realistic for NHS practice.

4.3.6.2 Drug administration

Vedolizumab and infliximab are administered by IV infusion and require an outpatient appointment with a healthcare professional. The company assumed 3 appointments for induction and 1 for maintenance per 8-week model cycle. The cost per visit was estimated at £137.37, based on the weighted mean for consultant led and non-consultant led, face-to-face, non-admitted, follow-up gastroenterology clinic appointments (NHS Reference Costs 2016-17) – CS Table 54, page 145. This estimate is similar to that used in the NICE appraisal of vedolizumab (TA342).

Adalimumab and golimumab are administered by subcutaneous injection. The company assume that patients can self-administer these treatments at zero cost to the NHS, due to the available of support from the drug manufacturers.

The company conduct sensitivity analysis around the cost of IV administration for vedolizumab and infliximab (varying the cost per dose from £70.20 to £161.72. We consider this range appropriate. We conduct additional scenario analysis to assess the impact of assuming an initiation of self-administration of subcutaneous injections: adding the cost of a non-consultant led clinic attendance (£107) to the cost of induction for adalimumab and golimumab.

4.3.6.3 Monitoring and follow up

Assumptions about the use and cost of monitoring and follow-up are summarised in section B.3.5.2 Tables 55 and 56 (CS pages 146 to 147) – see Table 67 below.

Table 67 Health state resource use and costs

Resource	Unit cost [a]	Resource use per year by health state [b]				
		Active UC	Response only	Remission	Post-surgery (no compl.)	Post-surgery (complications)
Outpatient visits [c]	£137	6.50	4.50	2.00	1.50	1.75
Blood tests [d]	£3.06	6.50	3.90	3.25	1.50	3.25
Endoscopy [e]	£277	2.00	0.50	0.20	1.25	0.65
Hospital episodes [f]	£2,985	1.50	1.20	0.30	0	3.25
Total per year		**£5,944**	**£4,350**	**£1,236**	**£557**	**£10,131**
Cost of surgery [g]		-	-	-	**£6,091**	**£7,295**

a Unit costs from NHS Reference Costs 2016-17
b Resource use from expert opinion in Tsai et al. (2008)[55], except hospital episodes for response only and remission health states from expert advice to company.
c Weighted average for consultant led and non consultant led (WF01A)
d Directly accessed haematology service (DAPS05)
e Diagnostic colonoscopy, 19 years and over (FE32Z)
f Non-elective inpatient (codes not specified)
g Elective proximal and distal colon procedures, 19 years and over with/without complications (FF32 and 33 (see CS Table 58 page 148).

Resource use assumptions were based on opinion from a panel of UK gastroenterologists, reported by Tsai et al. (2008).[55] The company state that they chose this source because the definition of the health states aligns with those used in the model: with Mayo scores similar to those in the OCTAVE trials. The Tsai et al. estimates of resource use have also been used in other NICE appraisals for ulcerative colitis (TQ329 and TA342).

Tsai et al. reported the same rate of 0.30 hospital admissions per year under standard care, for active ulcerative colitis, response only and remission states. The company changed this to assume more hospital episodes per year for the active UC and response only health states based on clinical expert opinion. Clinical advice to the ERG is that this is unrealistic, and that hospital admission is only undertaken for acute severe colitis (which is already included in the model), moderately severe ulcerative colitis not responding to oral prednisolone (which would not be treated with tofacitinib) and post-surgery with complications (admitted about once a year). Some other usage assumptions are also high in a current NHS context, including outpatient visits for patients in remission and post-surgery without complications, and endoscopy for uncomplicated post-surgery.

Health care usage assumptions from Tsai et al. (2008) are consistent with health state definitions in the model and with previous NICE appraisals for ulcerative colitis (TA329 and TA342). However, we have been advised that some estimates of the number of outpatient visits and endoscopies are high, and that the company's additional assumptions about hospital episodes are unrealistically high, particularly as admission for acute exacerbation requiring emergency surgery is already included in the model. We therefore test an alternative resource use scenario, suggested by our clinical expert (Table 68).

Table 68 ERG scenario for resource use by health state

Resource	Unit cost [a]	Resource use per year by health state [b]				
		Active UC	Response only	Remission	Post-surgery (no compl.)	Post-surgery (complications)
Outpatient visits [c]	£137	6.50	4.00	1.00	1.00	2.00
Blood tests [d]	£3.06	6.50	4.0	4.0	1.00	3.25
Endoscopy [e]	£277	2.00	0.50	0.20	0.2	0.65
Hospital episodes [f]	£2,985	0	0	0	0	1.0

We also question whether the assumption that maintenance treatment will always stop within 8 weeks of a loss of response is consistent with the number of outpatient appointments. We test two scenarios to align the costs of assessing patients on maintenance treatment with the model assumption that treatment will always be discontinued within 8 weeks of a relapse:

- *Add one additional outpatient appointment consultation when patients have a relapse while on maintenance treatment. In this case, all patients are assumed to seek and obtain an additional appointment when they experience symptoms.*
- *Assume 6.5 outpatient visits per year for all patients on maintenance treatment. This would be necessary if patients do not seek or cannot obtain an earlier appointment when they experience symptoms of moderately or severely active ulcerative colitis, so routine appointments would be needed to assess patients every 8 weeks.*

The company model omits ongoing costs of stoma care for the post-colectomy health states. This issue was addressed in the NICE vedolizumab appraisal TA342, and the committee concluded that these costs should be included but that the ERG estimate of £315 for a 6-month period was low. We revisited stoma cost estimates by Buchanan et al. (2011)[78] and uprated them for nurse costs (PSSRU 2017) and HCHS inflation for consumables: estimating an annual cost of £1,065.90 per person with a stoma, or £426.36 per person in the post-surgery health states (assuming 40% have a stoma). We include these costs in ERG preferred analysis.

The unit costs for health resources are also reasonable, although we note that the source for the mean cost per hospital episode is unclear (the CS and model does not specify which NHS Reference Cost codes are included). However, in comparison with estimates in TA329 and TA342, some of the unit costs are low. In particular, the estimated costs of surgery are lower than estimates from previous appraisals, which were based on the analysis by Buchanan et al. (2011): £13,176 for Europe or £11,620 in the UK. The model also omits ongoing stoma are costs for stoma care: estimated at £466 by Buchanan et al.[78]

We conduct additional scenario analysis to test the sensitivity of the results to higher estimates of the cost of surgery and the inclusion of stoma care costs in the post-surgery health states.

4.3.6.4 Treatment of serious infections

Finally, company estimates of the costs of treating serious infections are listed in CS Table 57 (page 147). The cost of £2,539 was estimated as a weighted average of inpatient care for six types of infection, with unit costs and incidence based on NHS Reference Cost data (2016-17). The company explored a wide range around this estimate (£722 to £11,471) in sensitivity analysis, which is appropriate given additional uncertainty due to the omission of other types of adverse events.

4.3.7 ERG critique of model assumptions and inputs

We summarise the key model assumptions alongside ERG's critique in Table 69. Broadly, we agree with company's approach albeit a few concerns, as highlighted in the table.

Table 69 Other model features and base case assumptions

Factor	Company justification		ERG comments
Model framework	Markov model	Allows the modelling of recurrent risks, such as response to treatment after induction and maintenance	We agree with the company general approach (Markov cohort structure) and representation of health states and transitions. The model structure and assumptions are similar to TA329.
Time horizon	Patient lifetime	UC is a chronic condition, so a patient lifetime horizon allows calculation of all relevant costs and quality of life impairment	Agree
Cycle length	8 weeks	Based on maintenance phase assessment intervals in the clinical trials of tofacitinib and other comparators. A fixed cycle length was used to allow the flexibility to adding a continuous sequence of treatments.	Agree.
Half cycle correction	Not applied	Relatively short cycle length	Agree

Factor	Company justification		ERG comments
Source of clinical effectiveness estimates	NMA for clinical response and remission (locally read) for subgroups with/without prior exposure to TNFi drugs	Locally read clinical response/remission reflects real-life practise. Choice of NMA models based on DIC statistics, with preference for fixed effects if no difference	Agree with use of locally-read clinical definitions of response and remission in economic model. We prefer random effects models to better reflect uncertainty related to heterogeneity. Combining TNFi-exposed subgroup for tofacitinib with TNFi-failed subgroup for vedolizumab is likely to have biased results for this comparison. We test alternative NMA model in ERG additional analysis, in section 4.4.3.
Calculation of transition probabilities	Outputs from NMA for response and remission transformed to 8-week probabilities	Simple approach; assumes constant risk through maintenance phase and beyond in extrapolation, as well as a constant ratio of response to remission. Company attempted calibration to fit 8-week transitions but this did not work.	We view it as unrealistic to assume constant risk of loss of response. Clinical experience indicates the risk is greatest in the first 6-12 months; and falls thereafter. The proportion of patients with response and remission is likely to increase over time as per our clinical advice. This is because responders (without remission) are more likely to stop or switch therapy whereas those in remission would continue with treatment. However, in the absence of evidence it is difficult to adapt the model

Factor	Company justification		ERG comments
Treatment waning of effects and discontinuation	Treatment effect was assumed to be maintained with ongoing treatment. Non-responders are given conventional therapy as second-line	Follows the approach taken in the independent economic analysis in NICE TA329	Agree with discontinuation for failure to respond in induction or loss of response in maintenance. We note this assumes that in practice patients who experience exacerbations of symptoms can be assessment and, if appropriate, treatment stopped within 8 weeks. The model does not reflect NICE recommendations for annual assessment of benefit and need for continued treatment in previous appraisals TA329 and TA342. Clinical advice suggests that withdrawal of treatment for patients in remission is unlikely in practice, and the effects of this are difficult to quantify given the model structure and limited evidence over long-term maintenance of remission.
Continuation of conventional therapy	Patients on CT and/or those who previously achieved but lost response to biologics were assumed to continue on CT irrespective of disease state	Simplifying assumption consistent with previous TAs and published literature [25]	Agree
Surgery	A proportion of non-responders and those who discontinue CT undergo elective colectomy. Patients from all health states (except remission) may undergo emergency surgery.	Consistent with clinical practice	Agree

Factor	Company justification		ERG comments
Risk of surgery	Assumed to be time-independent	Consistent with prior TAs; the base case model combined existing evidence with population in the model	Agree. Note surgery is treated as a transient event rather than a health state.
Source of utilities	Background utility ('no disease') based on EQ-5D by age and gender in general population. Health state utilities (EQ-5D) for pre and post-surgical states from Woehl *et al.* 2008. Sensitivity analyses using OCTAVE trial EQ-5D estimates and Swinburn et al.	Woehl *et al.* used as base case in previous TAs, with Swinburn in scenario analysis. Use of age/gender dependent background utility consistent with scenarios in previous TAs. Results consistent in scenarios with simple and regression-based utility estimates from trial EQ-5D data	Agree with the company's approach for the background utility estimates. We also agree with the use of Woehl *et al.* estimates of health state utilities, for consistency with other TA. Improved analysis of trial EQ-5D and scenario analysis in response to clarification questions. But we agree that the re-randomisation design of the maintenance trial complicates interpretation of within-trial utility estiamtes. We conduct additional scenario analysis in section 4.4.3
Source of unit costs	NHS reference costs, eMIT and MIMS for drug costs	Consistent with the NICE reference case	Agree
Biologic treatments	Golimumab formulation	It was assumed that the 100 mg vials of golimumab were used in induction (2x100 mg vial at week 0 and 1x100 mg vial at week 2) and the 50 mg vials were used for the maintenance dose (1x50 mg vial Q4W)	Agree

Factor	Company justification		ERG comments
Conventional therapy	The RCP audit data about use of conventional drugs by drug class at biologic initiation assumed to reflect CT alone for active UC	Assumption in absence of evidence on the CT mix	Agree
	Assumed equal use of 4 drugs in aminosalicylate class (balsalazide, mesalazine olsalazine & sulfasalazine)	Assumption in absence of evidence	Advice to ERG is that most patients receive mesalazine in UK. Doses for active UC higher than specified in company base case. Net effect on costs in base case likely to be neutral.
	Hydrocortisone was considered as a topical treatment (rectal foam); prednisolone was assumed to represent the oral corticosteroid treatment group and beclomethasone is used as add-on treatment to 5-ASA. Azathioprine assumed to represent the immunomodulator group	Simplifying assumptions	Agree
Concomitant medication	Use of conventional drugs concomitant with biologics/ tofacitinib based on 3-months follow-up in RCP audit. Azathioprine was excluded from concomitant use with tofacitinib	The evidence at 3-months follow-up were assumed to be reflective of continuous concomitant use.	Agree

Factor	Company justification		ERG comments
Administration cost for injections	No administration cost for self-administered sub-cutaneous injections assumed	Consistent with clinical practice	Agree. We conducted additional scenario analysis to assess the impact of assuming one outpatient consultation to support initiation of self-administered injections to the cost of induction for adalimumab and golimumab in section 4.4.3.
Health state resource use	Mostly based on Tsai *et al.* 2008, except increased frequency of hospitalisation was assumed for more severe disease	Consistent with structure of economic model and previous Tas. Gradient of hospitalisation with disease severity is realistic	Agree with use of Tsai et al. as base case. But clinical advice to ERG suggests frequency of outpatient visits and endoscopy exceed current UK practice and additional assumptions about hospital episodes are unrealistic. We test alternative resource use scenario in section 4.4.3 We also conduct scenario analysis to assume additional outpatient consultations to achieve 8-weekly assessment of response and cessation of treatment if indicated (see section 4.4.3) The company excludes cost of stoma care and the estimated cost of surgery is low compared with previous appraisals. The test the inclusion of stoma care costs and higher surgery costs in additional analysis, section 4.4.3

Factor	Company justification		ERG comments
Cost of serious infection	The cost of a serious infection was considered to be a weighted average of six types of infection: sepsis, tuberculosis, pneumonia, skin and soft tissue infection, bone and joint infection and urinary tract infection	Simplifying assumption in the absence of other evidence	Agree
Adverse events	Costs and utility loss associated with serious infection risk included	Evidence on the incidence of serious infections was available for all drugs. SIs are often associated with immunosuppressants. In the base case, the range of SIs with tofacitinib was assumed to increase between 0-50% from the base case value	There is a lot of uncertainty associated with SIs due to the rarity of events. Tofacitinib had the highest number of serious infections whilst golimumab had the lowest, We detail our concerns in section 3.1.7 and 4.3.4.2 and conduct additional analysis using an alternative frequentist NMA in section 4.4.3.
Mortality	Death from surgery and other cause mortality (as general population)	Consistent assumption on death from surgery as in TA329. Evidence on death from other cause in UC is sparse.	Agree

Source: CS Table 37 and Table 60

4.3.8 Model validation

The company describes their approach to model validations in CS section B.3.10. They state that they engaged UK clinical experts, statisticians and health economists to validate model inputs and assumptions in a UK advisory board meeting. Further details on the key aspects of validation are summarised in CS Table 78.

The CS stated that clinical experts validated model methods pertaining to the patient population; subgroup analysis by prior TNFi-exposure; time on treatment and discontinuation rates; costs (including monitoring cost for tofacitinib, health state costs and resource use, including rate of hospitalisation); emergency surgery; quality of life and maintenance dose of tofacitinib. The experts are reported to agree with the company's assumptions in most of these aspects, except for:

- **Patient population:** Although the baseline characteristics of the patient population in OCTAVE reflect UK practice, the duration of disease in OCTAVE trials (which was 6-7 years) is longer than that in clinical practice (which is ~2-4 years).
- **Health state unit costs and resource use, including rate of hospitalisation:** Tsai et al. was confirmed to reflect an accurate representation of unit costs and resource use as per clinical practice. However, the experts suggested that the model base-case assumptions relating to annual medical resource use (CS Table 55) underestimated the resource use per patient per year.
- **Tofacitinib maintenance dose:** Experts observed that the company assumption relating to [redacted] of patients benefitting from maintenance dose of 10mg twice daily may not be limited to patients in the TNFi-exposed group only.

The economic model was quality checked by health economists. For face validity, the company compared the proportion of patients in response and remission predicted by the model against the estimated values from the NMA, shown below in Figure 9.

Further, the model results were compared with previous TA329; however, the CS did not report any comparison of the results in TA329 with those in the current appraisal. We discuss this in detail in section 4.4.1. For internal validity, the CS stated that a second modeller reviewed the model; conducted extreme value tests alongside inspecting model code, formulae and references. An independent health economist was reported to have reviewed the model structure, parameter inputs and core model assumptions.

Figure 9 NMA results and model predictions of patient allocation and treatment survival

Source: CS Appendix M.2

4.3.9 Company cost effectiveness results

4.3.9.1 Base case deterministic results

The company present their base case results in CS section B.3.7, page 155. These incorporate the confidential PAS discount for tofacitinib but not the PAS discount for vedolizumab. The base case assume use of biosimilar drugs for infliximab. We report results including all available PAS discounts in a confidential addendum to this report.

People without prior exposure to TNF-alpha inhibitors

Results for the subgroup with no prior TNFi exposure are shown in Table 70.

- Adalimumab, golimumab and infliximab are dominated by tofacitinib – they are estimated to cost more and produce fewer QALYs;
- Tofacitinib gives a mean QALY gain of [redacted] QALYs for a mean additional cost of [redacted] compared with conventional therapy: giving an incremental cost-effectiveness ratio (ICER) of £8,554 per QALY gained;
- Compared with tofacitinib, vedolizumab gives an additional QALY gain of [redacted] QALYs for an additional cost of [redacted]: an ICER of £615,057 per QALY gained.

Table 70 Cost effectiveness: Company base case, no prior TNFi (with tofacitinib PAS)

Strategy	Total		Incremental analysis			Pairwise ICERs tofacitinib vs. comparator (£/QALY)
	QALYs	Costs (£)	QALYs	Costs (£)	ICER (£/QALY)	
Conventional	[redacted]	[redacted]	-	-	-	£8,554
Adalimumab	[redacted]	[redacted]	-	-	Dominated	Dominated
Golimumab	[redacted]	[redacted]	-	-	Dominated	Dominated
Infliximab	[redacted]	[redacted]	-	-	Dominated	Dominated
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	£8,554	N/A
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£615,057	£615,057

Reproduced from CS B.3.7.1 Table 61 page 155

People with prior TNF-alpha inhibitor exposure

Company base case results for the subgroup of people with prior TNFi exposure are shown in Table 71. The company omits adalimumab as a comparator in this subgroup. Clinical response/remission rates are not available for this subgroup for infliximab or golimumab.

- Compared with conventional therapy, tofacitinib gives a mean gain of [redacted] QALYs for an additional cost of [redacted]: resulting in an ICER of £10,302 per QALY gained;
- Compared with tofacitinib, vedolizumab gives an additional QALY gain of [redacted] QALYs for an additional cost of [redacted]: giving an ICER of over £7.8m per QALY gained.

Table 71 Cost effectiveness: Company base case, prior TNFi exposure (tofacitinib PAS)

Strategy	Total		Incremental analysis			Pairwise ICERs tofacitinib vs. comparator (£/QALY)
	QALYs	Costs (£)	QALYs	Costs (£)	ICER (£/QALY)	
Conventional	[redacted]	[redacted]	-	-	-	£10,302
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	£10,302	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£7,838,238	£7,838,238

Reproduced from CS B.3.7.1 Table 62 page 155

Disaggregated model results

The company report QALY and cost results from the model disaggregated by stages of treatment and health state in Appendix J to the CS (pages 382 to 388). We show key results for assessment of the face validity of the model in Table 72 and Table 73 below.

Table 72 shows the break down for patients with no prior exposure to TNF-alpha inhibitors. Predicted survival is very similar for the alternative treatments, at around [redacted] years from model entry (up to age [redacted] years). For all comparators, a large proportion of the estimated lifetime is spent with active ulcerative colitis, under management with conventional drug treatments. After discounting, life expectation is about [redacted] years, with very little difference between the comparators. QALY differences between treatments are slightly larger (from [redacted] to [redacted] discounted QALYs), due to estimated effects on rates of response and remission for the TNF-alpha inhibitors, tofacitinib and vedolizumab. Cost differences between the comparators are largely driven by the cost of the initial drug treatment, which are offset to some degree by savings in the cost of monitoring and managing the condition for the more effective drugs. Other cost differences are small.

Disaggregated results for patients with prior TNFi exposure are shown in Table 73. Modelled health outcomes are less favourable for the TNFi-exposed subgroup than for the TNFi naive subgroup, reflecting the lower response and remission rates from the NMA

Table 72 Disaggregated model results: company base case, no prior TNFi exposure

	CT	Adalimumab	Golimumab	Infliximab	Tofacitinib	Vedolizumab
Years of treatment (undiscounted)						
Initial treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Conventional	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Years by health state (undiscounted)						
Active UC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response nr	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery wc	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Life years (discounted)						
Active UC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery wc	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QALYs (discounted)						
Active UC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post surgery wc	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Costs (discounted)						
Initial treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Conventional	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adverse events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Surgery	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Health state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

UC, ulcerative colitis; nr, no remission; wc, with complications

Table 73 Disaggregated model results: company base case, prior TNFi exposure

	Conventional	Tofacitinib	Vedolizumab
Years of treatment (undiscounted)			
Initial treatment	[redacted]	[redacted]	[redacted]
Conventional	[redacted]	[redacted]	[redacted]
Post surgery	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]
Years by health state (undiscounted)			
Active UC	[redacted]	[redacted]	[redacted]
Response nr	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]
Post surgery	[redacted]	[redacted]	[redacted]
Post surgery wc	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]
Life years (discounted)			
Active UC	[redacted]	[redacted]	[redacted]
Response nr	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]
Post surgery nc	[redacted]	[redacted]	[redacted]
Post surgery c	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]
QALYs (discounted)			
Active UC	[redacted]	[redacted]	[redacted]
Response nr	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]
Post surgery nc	[redacted]	[redacted]	[redacted]
Post surgery c	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]
Costs (discounted)			
Initial treatment	[redacted]	[redacted]	[redacted]
Conventional	[redacted]	[redacted]	[redacted]
Adverse events	[redacted]	[redacted]	[redacted]
Surgery	[redacted]	[redacted]	[redacted]
Health state	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]

UC, ulcerative colitis; nr, no remission; wc, withouth complications

4.3.9.2 Deterministic sensitivity analyses

The CS presents the parameters and ranges included in their Deterministic Sensitivity Analysis (DSA) in CS Table 59. Parameters for safety, efficacy and utilities were varied using confidence intervals and published literature. For certain parameters such as risk of serious infections, the company conducted exploratory scenarios based on assumptions (see section 4.3.4.2 above). Results of the DSA are tabulated in CS Table 69 and CS Table 70 and presented as tornado plots in CS Figure 37 and CS Figure 38. The tornado plots for both TNFi- naïve and TNFi-exposed subgroups compare tofacitinib against conventional therapy alone. These show that the costs of serious infections, costs of conventional treatment and response estimates for the maintenance phase are key drivers of model results. Other parameters such as risk of colectomy, health state related resource use, response estimates in induction also influence the base case results, but to a lesser extent. The company has not presented tornado plots comparing tofacitinib with other comparators. In particular, the comparison with vedolizumab is important as the effectiveness of the two drugs are comparable. This makes it difficult to draw any robust conclusions from the DSA results. We address this issue in ERG additional analyses in section 4.4.

4.3.9.3 Probabilistic Sensitivity Analysis

The company conducted a probabilistic sensitivity analysis (PSA) on their base-case model to assess parameter uncertainty. Assumptions used to characterise uncertainty are described in CS Section B.3.6.1. Briefly, the company uses CODA samples for safety and efficacy parameters obtained from the NMA. We view this approach as appropriate as this preserves the joint posterior distribution and any correlation of treatment effects in the simulated outputs. Beta distributions are used for colectomy rates, perioperative complications and mortality, post-surgery complications, mortality and utility estimates. Parameters for costs and resource use are assigned gamma distribution. We consider that the parameters are assigned appropriate distributions and the PSA is correctly implemented. The results of the PSA are presented in CS Table 67 and CS Table 68; scatter plots are presented in CS Figure 33 and CS Figure 34; and cost effectiveness acceptability curves (CEACs) are in CS Figure 35 and CS Figure 36. The overall conclusion of the PSA results are similar to the base case results; however, in both the sub groups, total QALYs and costs are higher in the PSA results compared to the base case results. The company attributes this difference in PSA and base case results to the CODA samples used in the PSA. The CS states that at a willingness-to-pay threshold of £20,000 per QALY, tofacitinib had the highest probability of being cost-effective amongst the comparators at 80.5% in the TNFi-naïve group and 56.3% in the TNFi-exposed group, respectively.

4.3.9.4 Scenario Analysis

The company conducted a range of scenario analyses to assess the impact of key variables on the model outcomes. We were unable to replicate the following scenarios as the CS did not provide sufficient explanation: NMA results for the ITT population, maintenance dose mix of tofacitinib and centrally read NMA results. The company provided further information in their response to clarification question B6. They also acknowledged an error in incremental QALYs and incremental costs for the scenario relating to mix maintenance dose of tofacitinib in TNFi-naïve subgroup (CS Table 65) which they corrected in their response. Despite incorporating the changes suggested by the company, we were unable to replicate the company's cost-effectiveness results pertaining to scenario using central read NMA results. We present our results for this scenario in section 4.4.2. A summary of the company's scenarios, alongside their justifications and results obtained are presented in Table 74. The company concluded that the cost effectiveness results in both the sub-groups- TNFi-naïve and TNFi-exposed were predominantly influenced by change in utility estimates.

The ERG considers that the company has been selective in the scenarios that they present to explore the robustness of their base case cost-effectiveness results. In particular, they do not explore the impact of key assumptions such as inclusion of costs associated with stoma care, cost-effectiveness results from alternative NMA models. We extend the range of scenario analyses in ERG additional analyses below.

Table 74 Company scenario analyses

Company scenarios	Brief rationale/assumption	ICERs for Tofacitinib vs CT (£/QALY) TNFi-naïve	TNFi-exposed
Base case		**£8,554**	**£10,302**
Overall ITT population			£7,805
Tofacitinib maintenance dose mix	[redacted] of patients receiving 5mg; [redacted] of patients receiving 10mg	£12,628	£13,947
Fixed baseline utility instead of age-adjusted	Assumption that patient quality of life stays constant over time.	£8,760	£10,589
OCTAVE trial utilities	EQ-5D data were collected in Tofacitinib Phase III clinical trials	£15,508	£18,276
Swinburn utilities	To compare with previous analyses	£11,932	£14,487
Emergency surgery from any state	Due to the uncertainty on the likely protection from acute events based on the level of response/remission, patients are assumed to undergo emergency surgery regardless of state membership	£8,194	£9,962
Emergency surgery only from active UC	As above but assuming response to treatment offers the same level of protection from acute events, as remission	£8,652	£10,475
No emergency surgery	As above, but assuming no emergency surgery in the model	£8,710	£10,593
Central read NMA results	Central read was the primary endpoint in OCTAVE trials.	£9,469	£10,793
Discounting every cycle	It tested the sensitivity of the model when the discounting of outcomes is applied every 8 weeks.	£8,606	£10,398
Adalimumab maintenance 73% 40 mg Q2W and 27% 40 mg QW	Dose escalation of adalimumab was assumed in Archer *et al*.	£8,554	--
Golimumab 100 mg every 4 weeks in maintenance	A 100 mg Q4W maintenance dose was assessed as part of the clinical trials and is recommended for consideration in some patients, such as those who have experienced a decrease in their response	£8,554	--
Vedolizumab 300 mg every 4 weeks in maintenance	A 300 mg Q4W maintenance dose was assessed as part of the clinical trials and is recommended for consideration in some patients who have a body weight ≥ 80 kg	£8,554	Dominated

Source: CS Table 63 to 66; 71 to 77

4.4 Additional work undertaken by the ERG

4.4.1 ERG model validation

4.4.1.1 Model verification procedures

We checked the economic model for transparency and validity. The visual basic code used within the model was accessible. The NMA code in WinBUGs was provided in Appendix D.1.3.4.

We conducted a range of 'white box' tests to verify model inputs, calculations and outputs:

- Cross-checking of all parameter inputs against values in the CS and cited sources;
- Checking the individual equations within the model;
- A range of extreme value and logic tests to check the plausibility of changes in results when parameters are changed
- Checking the VBA code for treatment sequencing
- Checking all model outputs against results cited in the CS, including the base case, PSA, DSA and manually ran all the scenarios
- Running the NMA code in WinBUGs to replicate selected results (see section 3.1.7).

In addition, we checked the model calculations of patient transitions through the health states, costs and QALYs by re-coding the model independently based on the inputs from the company's submitted model.

Overall, we found the economic model to be of a good quality, with very few errors in input parameters, logic or coding. We identified a few small errors that we correct in section 4.4.2 below, which did not make any substantive difference to the results. We were also successful in replicating outputs from most of the company's NMA models, with the exception of the serious infection NMA (section 3.1.7).

4.4.1.2 External validity

We have tabulated the model predictions against the observed clinical data for the maintenance phase, in Table 75 below. While the model results appear comparable with the clinical data for the tofacitinib arm in the TNFi-naïve group, there are large differences in the estimates for TNFi- exposed subgroup for this arm, along with the placebo arms for both induction and maintenance phases.

Table 75 Comparison of the predicted model results of Tofacitinib and Placebo (CT) against the observed clinical data – INDUCTION Phase

Study	Treatment	TNFi-naive		TNFi-exposed	
		Clinical response	Clinical remission	Clinical response	Clinical remission
OCTAVE Induction 1	Placebo	[redacted]	[redacted]	[redacted]	[redacted]
	Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]
OCTAVE Induction 2	Placebo	[redacted]	[redacted]	[redacted]	[redacted]
	Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]
Model	Placebo	[redacted]	[redacted]	[redacted]	[redacted]
	Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix J.1.2. Table 199

4.4.1.3 Cross validation

In section 4.2 above (page 134), we state that the CS reported previous economic models, including published literature and analyses conducted by ERGs for previous NICE TAs, for patients in ulcerative colitis. Whilst we acknowledge that there are methodological differences between the economic models across these studies, nonetheless we view that they provide sources for cross-validation of results from the company base-case analysis. Of the reported studies, we cross-validate the modelled findings of the current appraisal with 2 previous NICE TAs (TA342 and TA329) and 1 published study as summarised in Table 76. The most relevant analysis for the current appraisal is the final version from the NICE TA of vedolizumab (TA342). This appraisal relates to same patient population as the current appraisal and comparators overlap, except Tofacitinib and surgery.

Table 76 Comparison of modelled outcomes

Study name (time horizon)	QALYs		Life years	
	TNFi- naive	***TNFi-exposed***	***TNFi- naive***	***TNFi-exposed***
Current appraisal (lifetime)	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
TA342 (10 years)	Ada: 5.76	Ved: 5.46	Not reported	Not reported
	Gol: 5.79	CT: 5.37		
	Inf: 5.82	Surgery: 4.28		
	Surgery: 4.28			
	Ved: 5.90			
	CT: 4.28			
TA329 (Lifetime, AG model)	***Moderate to severe UC who failed at least 1 prior therapy***			
	Ada: 10.82		Not reported	
	Inf:10.81			
	Gol: 10.63			
	CT: 10.47			
Wu et al. (lifetime)	***Moderate to severe UC***			
	CT:10.49		Not reported	
	Ved→CT: 11.48			
	Tof→CT: 11.51			
	Inf→CT: 10.87			
	Gol→CT:10.89			
	Ada→CT: 10.71			
	Ved→Tof→CT: 12.37			
	Inf→Tof→CT:11.81			
	Gol→Tof→CT:11.83			
	Ada→Tof→CT:11.67			
	Tof→Ved→CT:12.37			
	Tof→Inf→CT:11.84			
	Tof→Gol→CT:11.86			
	Tof→Ada→CT:11.70			

4.4.2 ERG corrections to company model

We identified a few errors in the company's model, as shown in Table 77 below. The company corrected issue 2(ii) and provided further information to address issue 2(i) as response to the clarification questions. However, the ERG was unable to replicate the company's results for scenario in issue 2(i), although the differences in ICERs, obtained by the company and ERG, were minimal. The ERG implemented the corrections in Issues 1 and 3. These are discussed in the following sub-sections.

Table 77 ERG corrections to company model

<table>
<tr><th>Aspect of model</th><th>Problem</th><th>ERG Correction</th></tr>
<tr><td rowspan="2">1.Cost calculations</td><td>i. Cost of elective surgery with complications: the company used the cost of surgery without complications</td><td>Recoded column FA in 'Engine2L' sheet</td></tr>
<tr><td>ii. Cost of CT : We noted a few small changes in prices for sulfasalazine (£6.87), prednisolone (£0.91) and azathioprine (£2.17)</td><td>Values used by the ERG:
Sulfasalazine: £7.83;
Prednisolone: £0.47;
Azathioprine: £2.20

(MIMS June 2018)</td></tr>
<tr><td rowspan="2">2. Scenario analysis</td><td>i. Centrally read NMA results: ERG was unable to replicate the cost-effectiveness results presented by the company in CS Table 72 (scenario 7) and CS Table 76 (scenario 7)</td><td>We were unable to replicate the ICERs for tofacitinib vs CT (£/QALYs) reported by the company for this scenario (shown below)
<table>
<tr><td></td><td>Company</td><td>ERG</td></tr>
<tr><td>TNFi-naïve</td><td>£9,469</td><td>£9,524</td></tr>
<tr><td>TNFI-exp</td><td>£10,793</td><td>£10,789</td></tr>
</table></td></tr>
<tr><td>ii. CS Table 65: Error in incremental costs and incremental QALYs</td><td>Company corrected this as response to clarification question B6 (b). The corrections did not change the ICER.</td></tr>
<tr><td>3.Weight - wastage</td><td>i. Error in estimation of weight – wastage in cell N17:N18 and cell Q17:Q18 in sheet!Cost_Drug</td><td>Recoded the cells in sheet!Cost_Drug. The corrections do not have any impact on the base case CE results as these use 'fitting distribution' approach for wastage calculation.</td></tr>
</table>

4.4.2.1 Results for TNFi-naive subgroup

Making the corrections in Table 77 to the company's base case model resulted in a small increase in the ICERs for people without prior exposure to TNFi (Table 78**)**. The results were robust to deterministic and probabilistic sensitivity analysis and scenario analyses.

Table 78 Deterministic company base case (ERG corrected) -TNFi-naive (tofacitinib PAS)

Strategy	Total		Incremental analysis			Pairwise ICERs TOFA vs. comparator (£/QALY)
	QALYs	Costs (£)	QALYs	Costs (£)	ICER (£/QALY)	
Conventional	[redacted]	[redacted]	[redacted]	[redacted]	-	£8,564
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Tofa. dominant
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Tofa. dominant
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Tofa. dominant
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	£8,564	N/A
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£615,077	£615,077

Table 79 DSA results company base case (ERG corrected) - TNFi-naïve (tofacitinib PAS)

	ICER TOFA vs. CT (£/QALY)	
Base case	**£8,564**	
Parameter	**Low limit**	**High limit**
Serious infection costs	£7,622	£13,191
Conventional treatment costs (min-max)	£9,559	£4,137
Response/remission treatment effect - maintenance	£6,292	£11,920
Colectomy risk (No risk - Frolkis 10y)	£7,388	£11,109
Health-state related resource use per patient per year	£8,334	£10,994
Response/remission treatment effect - induction	£7,609	£10,180
Serious infection risk	£7,259	£9,382
Hospitalisation cost	£9,850	£7,604
Pre-surgery health state utilities	£8,105	£9,493
OP visit + blood test costs	£9,140	£8,353
Endoscopy cost	£9,067	£8,082
Remission (z) - maintenance	£8,315	£8,838
Post-surgery health state utilities	£8,511	£8,617
Periorative mortality risk (0 - 3%)	£8,587	£8,559
Remission (z) - induction	£8,545	£8,581
Post-operative pouchitis (0.7 - 2%)	£8,576	£8,552
Colectomy cost	£8,573	£8,553
Serious infection utility reduction (0% - 3%)	£8,555	£8,572
Periorative complications (No risk - double the risk)	£8,566	£8,561
Post-surgery complication utility weight reduction (0% - 40%)	£8,566	£8,561
OP administration cost (£70 - £161)	£8,564	£8,564

Table 80 Probability of being cost-effective - TNFi-naïve subgroup

Treatments	£20k per QALY WTP	£30k per QALY WTP
Conventional	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Infliximumab	[redacted]	[redacted]
Tofacitinib	[redacted]	[redacted]
Vedolizumab	[redacted]	[redacted]

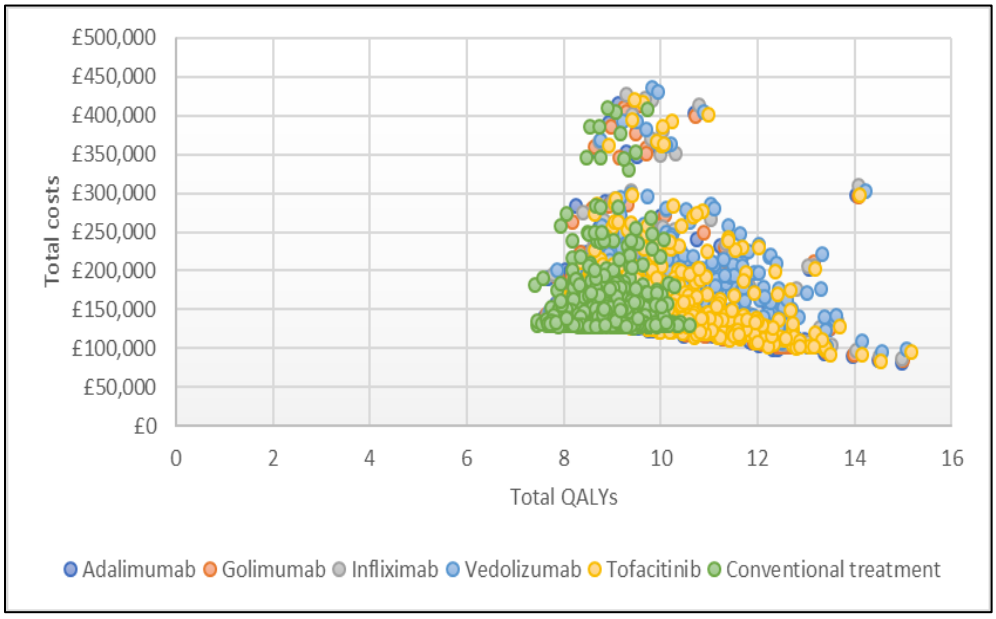


Figure 10 PSA scatter plot - TNFi-naïve subgroup

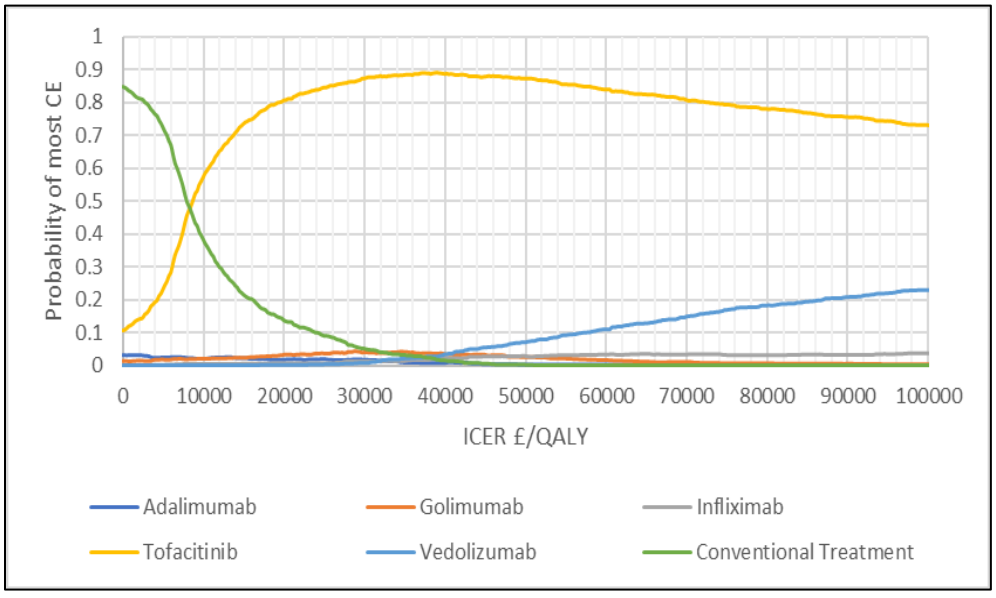


Figure 11 Cost effectiveness acceptability plane - TNFi-naïve subgroup

Table 81 Scenario analyses, company base case (ERG corrected) – TNFi-naive subgroup

Scenarios	Assumption	ICER for tofacitinib vs.	
		CT	Vedolizumab
Base case		**£8,564**	**£615,077**
Tofacitinib maintenance dose mix	[redacted] of patients receiving 5mg; [redacted] of patients receiving 10mg	£12,637	Tofacitinib dominant
Fixed baseline utility instead of age-adjusted	Assumption that patient quality of life stays constant over time.	£8,770	£634,346
OCTAVE trials utilities	EQ-5D data were collected in Tofacitinib Phase III clinical trials	£15,525	£1,079,814
Swinburn utilities	To compare with previous analyses	£11,945	£853,228
Emergency surgery from any state	Due to the uncertainty on the likely protection from acute events based on the level of response/remission, patients are assumed to undergo emergency surgery regardless of state membership	£8,204	£606,872
Emergency surgery from active UC only	As above but assuming response to treatment offers the same level of protection from acute events, as remission	£8,661	£618,151
No emergency surgery	As above, but assuming no emergency surgery in the model	£8,719	£618,068
Central read NMA	Central read was the primary endpoint in OCTAVE trials.	£9,534	£187,809
Discounting every cycle	It tested the sensitivity of the model when the discounting of outcomes is applied every 8 weeks.	£8,616	£617,451
Vedolizumab dose 300 mg Q4W	A 300 mg Q4W maintenance dose was assessed as part of the clinical trials and is recommended for consideration in some patients who have a body weight ≥ 80 kg	£8,564	Tofacitinib dominant

4.4.2.2 Results for TNFi-exposed subgroup

Table 82 Deterministic company base case (ERG corrected), TNFi-exposed (TOF PAS)

Strategy	Total		Incremental analysis			Pairwise ICERs TOF vs. comparator (£/QALY)
	QALYs	Costs (£)	QALYs	Costs (£)	ICER (£/QALY)	
Conventional	[redacted]	[redacted]	[redacted]	[redacted]	-	£10,311
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Tofa. dominant
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	£10,311	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	£7,838,381	£7,838,381

Table 83: DSA results for TNFi-exposed subgroup (compared to CT)

	ICER (£/QALY)	
Base case	**£10,311**	
Parameter	**Low limit**	**High limit**
Serious infection costs	£9,376	£14,909
Conventional treatment costs (min-max)	£11,302	£5,903
Response/remission treatment effect - maintenance	£7,825	£13,342
Health-state related resource use per patient per year	£9,531	£12,383
Colectomy risk (No risk - Frolkis 10y)	£9,108	£11,909
Serious infection risk	£9,013	£11,126
Response/remission treatment effect - induction	£9,461	£11,501
Hospitalisation cost	£11,481	£9,439
Pre-surgery health state utilities	£9,751	£11,374
Remission (z) - maintenance	£9,758	£10,946
OP visit + blood test costs	£10,857	£10,112
Endoscopy cost	£10,818	£9,827
Post-surgery health state utilities	£10,250	£10,373
Remission (z) - induction	£10,250	£10,371
Periorative mortality risk (0 - 3%)	£10,339	£10,305
Post-operative pouchitis (0.7 - 2%)	£10,323	£10,299
Colectomy cost	£10,321	£10,301
Serious infection utility reduction (0% - 3%)	£10,301	£10,321
Post-surgery complication utility weight reduction (0% - 40%)	£10,314	£10,308
Perioperative complications (No risk - double the risk)	£10,314	£10,309
OP administration cost (£70 - £161)	£10,311	£10,311

Table 84: Probability of being cost-effective - TNFi-exposed subgroup

Treatments	£20k per QALY WTP	£30k per QALY WTP
Tofacitinib	[redacted]	[redacted]
Vedolizumab	[redacted]	[redacted]
CT	[redacted]	[redacted]

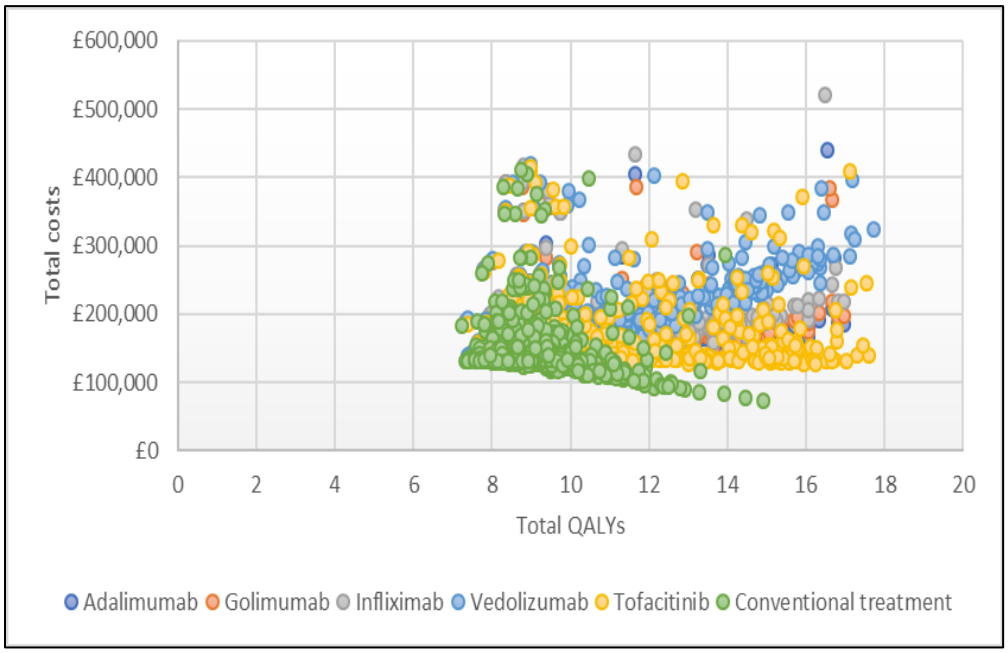


Figure 12 PSA scatter plot for TNFi-exposed subgroup

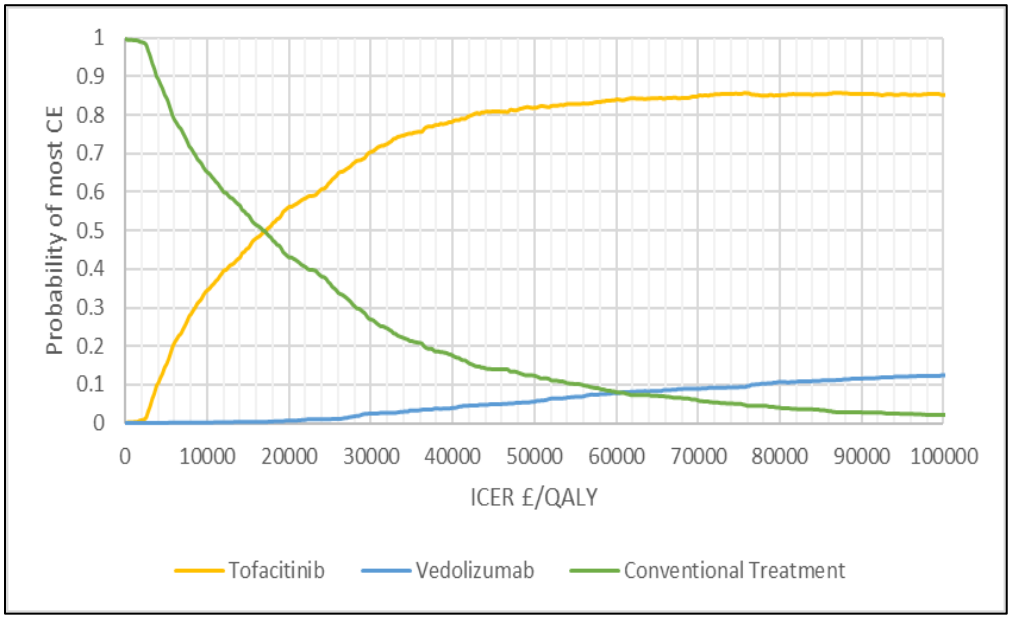


Figure 13 Cost effectiveness acceptability plane for TNFi-exposed subgroup

Table 85 Scenario analyses, company base case (ERG corrected) – TNFi-exposed

Scenarios	Assumption	ICER for Tofacitinib vs.	
		CT	**Vedolizumab**
Base case		**£10,311**	**£7,838,381**
Tofacitinib maintenance dose mix	[redacted] of patients receiving 5mg; [redacted] of patients receiving 10mg	£13,956	Tofacitinib dominant
Fixed baseline utility instead of age-adjusted	Assumption that patient quality of life stays constant over time.	£10,599	£6,502,288
OCTAVE trials utilities	EQ-5D data were collected in Tofacitinib Phase III clinical trials	£18,292	Tofacitinib dominant
Swinburn utilities	To compare with previous analyses	£14,501	£7,087,359
Emergency surgery from any state	Due to the uncertainty on the likely protection from acute events based on the level of response/remission, patients are assumed to undergo emergency surgery regardless of state membership	£9,971	£7,612,076
Emergency surgery from active UC only	As above but assuming response to treatment offers the same level of protection from acute events, as remission	£10,485	£6,780,235
No emergency surgery	As above, but assuming no emergency surgery in the model	£10,603	£6,781,118
Central read NMA	Central read was the primary endpoint in OCTAVE trials.	£10,798	Tofacitinib dominant
Discounting every cycle	It tested the sensitivity of the model when the discounting of outcomes is applied every 8 weeks.	£10,408	£8,260,662
Vedolizumab dose 300 mg Q4W	A 300 mg Q4W maintenance dose was assessed as part of the clinical trials and is recommended for consideration in some patients who have a body weight ≥ 80 kg	£10,311	Tofacitinib dominant

4.4.3 ERG additional analysis

Table 86 below summarises ERG assumptions and scenario analyses that we ran to further explore uncertainties over the model results.

Table 87 shows the cumulative impact of applying ERG preferred assumptions to the company base case model. The change that has the biggest impact is the use of alternative NMA models to populate the input parameters for serious infections. This is consistent with the company's observation based on their sensitivity analyses. Varying the age of the patients; using different NMA models for clinical response and remission and adding the costs of stoma care have little impact on the results. Collectively our preferred assumptions give very similar results to the company's model. TNF-inhibitors remain dominated (with higher costs and fewer QALYs) than tofacitinib in both the sub-groups across the range of assumptions tested. The pairwise ICERs for tofacitinib compared with vedolizumab mostly fall in the south-west quadrant (meaning tofacitinib is less effective but also less costly than vedolizumab), under our preferred set of assumptions vedolizumab is dominated by tofacitinib. However, we note again that these results do not take account of the PAS discount for vedolizumab. Final results including all PAS discounts are provided in the confidential addendum to this report.

We performed a range of additional scenario analyses on the ERG preferred base case, as specified in Table 86. Results are summarised in Table 88 and Table 89 below, with full cost-effectiveness results in Appendix 9.3. In the TNFi-naïve subgroup, Tofacitinib dominated the TNFi-agents (adalimumab, infliximumab and golimumab) across all the scenarios. The ICERs for Tofacitinib vs CT were most sensitive to sources for utilities and assumptions about health service use but remained below £20,000 per QALY for all scenarios. For Tofacitinib vs Vedolizumab, the ICERs moved between the south–east (indicating, tofacitinib was cheaper and more effective than vedolizumab) and south west quadrants (indicating, tofacitinib was cheaper and less effective compared with vedolizumab).

Similarly the TNFi-experienced subgroup, tofacitinib dominated TNFi- agents in all scenarios. The ICER for tofacitinib vs CT remained low, reaching a maximum of £21,376 per QALY with OCTAVE EQ-5D utility estimtes Tofacitinib dominated vedolizumab across all the scenarios, except in the the company's preferred NMA models for response and remission (favouring fixed effect models).

Table 86 ERG preferred assumptions and scenarios

Aspect of the model		Company base case	ERG preferred	ERG scenarios		Reason for analysis
Patients	Age (yrs)	TNFi-naïve: 41.5 TNFi-exposed: 40.9	Average of all patients in OCTAVE 1 and 2: 41	Range: 28-52		To explore the impact of patient characteristics on the cost-effectiveness results
	Weight (kgs)	TNFi-naïve: 74.6 TNFi-exposed: 72.5	Average for all patients in OCTAVE 1 and 2: 73.5	Range: 70-80 kg		
Comparator	TNFi-exposed	Excludes adalimumab	Include adalimumab			NMA results available for adaliumuab in TNFi-exposed group. Clinical advice that some patients would switch to another TNFi
	Treatment sequencing	The base case includes only 1st line and 2nd line treatments	No change	INF-ADA-CT INF-VED-CT INF-TOF-CT VED-ADA-CT TOF-ADA-CT	GOL-ADA-CT GOL-VED-CT GOL-TOF-CT ADA-VED-CT ADA-TOF-CT	To test effect of switching within or between classes and compare 'step-up' and 'step-down' strategies
NMA models	Remission and response rates	Use FE models except for TNFi-naive induction (FE better fit)	Use RE except for TNFi-experienced maintenance (RE would not run)	FE for both subgroups, induction and maintenance		ERG prefers RE models, given study heterogeneity
		Combined TNFi-failed for vedolizumab with TNFi-exposed for tofacitinib and adalimumab	No change	Use TNFi-failed for both vedolizumab and tofacitinib with TNFi-experienced for adalimumab		To provide a more like-for-like comparison between tofacitinib and vedolizumab - main competitors.
	Serious infections	Bayesian random effect model	Frequentist random effects NMA model	Bayesian random effect model		Due to rarity and null events credible intervals for Bayesian NMA are implausibly wide.
Utilities	Sources for pre and post-surgery health states	Background age/gender specific general population EQ-5D for remission. Utility multipliers for other heallth states from Woehl et al. 2008	Same as company	• Swinburn et al. • OCTAVE 8 weeks • OCTAVE 52 weeks		Woehl et al. used in previous TAs. For scenario analysis, we use results analysis of EQ-5D data from OCTAVE provided in company clarification response

Resource use and costs	Drug stopping rule	8 weekly loss of response or surgery	Same as company	Additional OP visits to assess response within 8 weeks	Include costs required to allow rapid assessment and change of therapy following exacerbations
	Conventional drug usage	Estimated from RCP IBD audit 2016	Same as company	Patient use of mesalazine: 50.3% (CT), 46.2% (concurrent). No other aminoslicylates	Clinical expert advice
	Health state resource use	Based on Tsai et al. plus additional admissions	Same as company	Reduced admissions, outpatient follow up and endoscopy	To reflect advice on current NHS clinical practice
	Drug administration costs	OP visit for IV infusion (infliximab, vedolizumab) No administration cost for self-administered subcutaneous injections (golimumab, adalimumab)	Same as company	Assume 1 OP visit at start of treatment for training on subcutaneous injections	Company states that support for self-administration of injections is provided by manufacturers. But this may not always be available in NHS.
	Hospitalisation and surgery costs	NHS Reference costs 2016-17 for colectomy procedures	NHS Reference costs + cost of stoma care post-surgery (Buchanan et al. uprated for inflation)	Buchannan et al. estimate of surgery cost (uprated to 2016/17 prices) – includes repeat procedures	Stoma costs To align with previous TA 342
Surgery	Incidence rate	Misra et al. (UK HES Data)	Same as company	Chhaya et al.	Exploratory analyses
	Complications	IBD audit	Same as company	Tappenden et al.: Probabiity of perioperative complications (elective 0.2386; emergency 0.2614), probability of post-surgery complications (0.173)	To align with previous TA 342

Table 87 Cumulative effect of ERG preferred assumptions

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Company base case (ERG corrected)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£8,564
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	£615,077 (SW)
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£10,311
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	£7,838,381 (SW)
Average age: 41 years	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£8,562
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	£614,916 (SW)
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£10,304
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	£7,798,892 (SW)
+ ERG preferred NMAs for remission and response	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£8,584
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	£590,046 (SW)
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£10,148
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
+ Frequentist NMA for serious infections	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,886
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£607,642 (SW)
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,458
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
+ Cost of stoma-care = ERG preferred analysis	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,815
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	£607,571 (SW)
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,389
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 88 Scenario analyses, ERG base case (Tofacitinib PAS) – TNFi-naive subgroup

Scenarios	ICER for tofacitinib versus	
	CT	Vedolizumab
ERG preferred base case	**£7.815**	**£607,571**
Age: 28 years	£7,644	£589,024
Age: 52 years	£8,019	£628,794
Weight: 70 kg	£7,827	£607,395
Weight: 80 kg	£7,819	£607,504
NMA: FE for response and remission	£7,793	£633,458
NMA: TNFi-failed (Ved) + TNFi-exp (tof and ada)	Not relevant	Not relevant
NMA: FE for Serious Infections	£8,513	£589,976
Utility: Swinburn et al.	£10,898	£845,865
Utility: OCTAVE 8 weeks	£17,764	£1,360,239
Utility: OCTAVE 52 weeks	£18,256	£1,373,067
Drug stopping: 6.5 OP visits for all patients in maintenance	£9,090	£608,793
Reduced health state resource use (clinical scenario)	£13,938	£613,289
CT drug usage	£7,827	£607,576
Drug admin cost for subcutaneous injection	£7,815	£607,571
Stoma care costs (£81.66 per cycle based on TA342)	£7,804	£607,561
Surgery costs (based on Buchannan et al.)	£7,764	£607,522
Surgery incidence rate (based on Chhaya et al.)	£7,980	£611,440
Surgery complications (based on Tappenden et al.)	£7,556	£605,226
Treatment sequencing	£13,951	£614,361

Table 89 Scenario analyses, ERG base case (Tofacitinib PAS) – prior TNFi experience

Scenarios	ICER for tofacitinib versus	
	CT	Vedolizumab
ERG preferred base case	£9,389	Tofacitinib dominant
Age: 28 years	£9,170	Tofacitinib dominant
Age: 52 years	£9,648	Tofacitinib dominant
Weight: 70 kg	£9,401	Tofacitinib dominant
Weight: 80 kg	£9,394	Tofacitinib dominant
NMA: FE for response and remission	£9,541	£8,801,245
NMA: TNFi-failed (Ved) + TNFi-exp (tof and ada)	£9,669	£2,521,513
NMA: FE for Serious Infections	£10,080	Tofacitinib dominant
Utility: Swinburn et al.	£13,198	Tofacitinib dominant
Utility: OCTAVE 8 weeks	£21,376	Tofacitinib dominant
Utility: OCTAVE 52 weeks	£21,283	Tofacitinib dominant
Drug stopping: 6.5 OP visits for all patients in maintenance	£10,597	Tofacitinib dominant
Reduced health state resource use (clinical scenario)	£14,950	Tofacitinib dominant
CT drug usage	£9,402	Tofacitinib dominant
Drug admin cost for subcutaneous injection	£9,389	Tofacitinib dominant
Stoma care costs (£81.66 per cycle based on TA342)	£9,379	Tofacitinib dominant
Surgery costs (based on Buchannan et al.)	£9,341	Tofacitinib dominant
Surgery incidence rate (based on Chhaya et al.)	£9,558	Tofacitinib dominant
Surgery complications (based on Tappenden et al.)	£9,134	Tofacitinib dominant
Treatment sequencing	£9,389	Tof-Ada-CT dominant

5 End of life

The NICE end of life treatment criteria are not applicable and are not included in the CS.

6 Innovation

The CS highlights six aspects of tofacitinib therapy for moderately to severely active ulcerative colitis in making the case for innovation (CS B.2.12). These six aspects are:

- Tofacitinib is the first in a new class of treatments and has a novel mechanism of action (inhibitor of JAKs).
- Tofacitinib is an oral therapy in contrast to the available biologic therapies for people with moderately to severely active ulcerative colitis which are administered either as infusion or by subcutaneous injection.
- Tofacitinib is a small synthetic molecule which means the formation of anti-drug antibodies (which reduce the efficacy of large protein biologics such as the TNF-alpha inhibitors) is not likely to occur, the risk of immunogenicity is reduced, and therapeutic drug monitoring is not required.
- Tofacitinib is a monotherapy, which would be expected to have a more favourable safety profile than combination therapies of a biologic therapy plus immunomodulatory agent. (NB the ERG notes that in the company's safety NMA, tofacitinib had the second-highest probability of serious adverse events after placebo (section 3.3.10.2).
- Tofacitinib treatment may be interrupted without the expectation of a reduced response
- Tofacitinib has a rapid onset of action.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

Choice of model fit for clinical response and clinical remission NMAs

Heterogeneity was present among the studies included in the NMA. Heterogeneity was due to differences in the designs of the included studies and differences between the baseline characteristics of the patients recruited to the trials included in the NMAs. In some cases the fit of the fixed-effect and random-effects models were comparable in terms of fit and the company chose the fixed-effects model in these circumstances. The ERG would have chosen the

random effects model to account for between study heterogeneity and provide a more conservative analysis.

Choice of model fit for serious infection NMA

Serious infections NMA under both random-effects (company choice) and fixed-effect (ERG alternative) resulted in very wide credible intervals. There are two issues with the available tofacitinib data on serious infections. Firstly the number of serious infections that occurred in the Phase II tofacitinib trial is higher than for the other OCTAVE trials. Secondly, in the Phase II trial and both of the OCTAVE Induction trials there were no serious infection events in the placebo arms. The ERG therefore ran an alternative NMA using a frequentist framework that allows for a value of 0.5 to be added to zero cells. Whilst adding a value to a zero cell is controversial this analysis does not adversely impact the confidence intervals for tofacitinib on account of the absence of serious infections among any of the placebo arms in the OCTAVE trials programme.

Absence of maintenance phase safety NMAs

No NMA for safety outcomes was conducted for the maintenance phase. The ERG believe this could have been achieved by using the mFAS population of OCTAVE Sustain. Whilst the use of mFAS would have aligned the re-randomised studies these would still have to be combined with data from the studies with a treat-through design and hence would only have been a partial solution.

No exploration of correction for different durations of induction and maintenance phases or differences between studies with a re-randomisation design.

Not all studies included in the NMAs had the same induction and maintenance phase durations as the OCTAVE tofacitinib studies. In particular the studies of golimumab and vedolizumab had a shorter induction phase (6 weeks versus OCTAVE studies 8 weeks) and the maintenance phases of the adalimumab (44 weeks), infliximab (46 weeks) and vedolizumab (46 weeks) studies were shorter than those of the tofacitinib (52 weeks) and golimumab studies (54 weeks). It is possible that there could be a bias against studies with a shorter induction period (if a higher response could be possible if measured at 8 weeks instead of 6 weeks). If this were the case this would bias against golimumab and vedolizumab in the induction phase. Similarly it is possible that there could be a bias in favour of studies with a shorter maintenance period (if

fewer responders lose response in the shorter time frame). If this were the case the bias would work against tofacitinib which has one of the longer maintenance phases.

7.2 Summary of cost effectiveness issues

Baseline characteristics of patient population included in the economic model

For their base case analyses, patient characteristics (including initial age, weight and gender mix) for the two sub-groups of TNFi-naïve and TNFI-exposed are based on means from the tofacitinib arms in the OCTAVE Induction trials. We view that these baseline characteristics should be assumed similar for people with and without prior exposure to TNFi drugs. We explore this in our additional analyses.

Analysis for the whole population: ITT NMA

The company has conducted an ITT NMA for the whole population and performed a cost-effectiveness analysis with the ITT population. We do not consider this scenario to be reliable because of the high level of uncertainty in the underlying NMA. The scenario also omits relevant comparators (the TNFi drugs), so does not address the specified decision problem. The ERG, therefore, focuses on separate analyses for the two TNFi exposure subgroups which is consistent with committee considerations in the NICE appraisal of vedolizumab (TA342).

Comparator

- *Exclusion of adalimumab in TNFi-exposed sub group*

The company excludes adalimumab, infliximab and golimumab as comparators for patients with prior exposure to a TNFi. Whilst clinical response and remission rates are not available for infliximab or golimumab in this sub group, but they are available for adalimumab. Further, the occurrence of in-class switching is also supported by evidence from the UK IBD Audit: 21% of patients starting adalimumab (17/83) had previously not responded or been intolerant to a TNFi. So, the ERG considers adalimumab as a relevant comparator for at least some patients with prior exposure to a TNFi agent. We therefore include adalimumab in ERG analysis for this subgroup. However, we understand that further treatment with a TNFi may not be appropriate for all patients in this subgroup.

- *Conventional therapy*

The company assumes equal use of 4 drugs in aminosalicylate class (balsalazide, mesalazine

olsalazine & sulfasalazine). However, clinical advice to ERG is that most patients receive mesalazine in UK and the doses for active ulcerative colitis are potentially higher than specified in company base case. We view that the net effect on costs from incorporating the changes in base case is likely to be neutral.

Treatment waning of effects and discontinuation

The company assumes treatment effect to be maintained with ongoing treatment and non-responders are given conventional therapy as second-line. The ERG agrees with company's approach to allow discontinuation for failure to respond in induction or loss of response in maintenance, based on the independent economic analysis in NICE TA329. We note this assumes that in practice, patients who experience exacerbations of symptoms can be assessed and, if appropriate, treatment stopped within 8 weeks. However, the model does not reflect NICE recommendations for annual assessment of benefit and need for continued treatment in previous appraisals TA329 and TA342. Clinical advice suggests that withdrawal of treatment for patients in remission is unlikely in practice, and the effects of this are difficult to quantify given the model structure and limited evidence over long-term maintenance of remission.

The company model applies a constant risk of relapse across each 8-week cycle of maintenance, with treatment stopping immediately when patients lose response. Thus, it assumes that maintenance treatment is stopped within 8 weeks of a loss of response. We consider this assumption to reflect UK practice. However, we have concerns that the costs of monitoring and follow-up in the company's model do not reflect the full cost of ensuring that treatment can be withdrawn within 8 weeks of a relapse. We address this by considering additional costs for outpatient visits to enable treatment cessation within 8 weeks of a relapse in our additional analyses.

Source of clinical effectiveness estimates

- *Choice of NMA models for economic analysis*

In general, we agree with company's approach to use locally-read clinical definitions of response and remission in economic model. Whilst, they state that their choice of NMA models was based on DIC measures of model fit, but they preferred the simpler fixed effect approach when DIC statistics were similar. The ERG has a general preference for the random effect NMA models, as we believe that the fixed effect models may underestimate uncertainty due to

heterogeneity between the studies. We test the impact of different NMA models on cost-effectiveness results in our additional analyses.

- *Combination of TNFi-failed and TNFi-exposed subgroups*

The base case NMAs combine outcomes for subgroups defined as TNFi-failed for vedolizumab with TNFi-exposed subgroups for tofacitinib and adalimumab. We consider that combining results for TNFi-failed and TNFi-exposed subgroups is a potential source of bias in favour of tofacitinib. We conduct a scenario analysis using a more like-for-like comparison between tofacitinib and vedolizumab, using data for the TNFi-failed subgroups from the OCTAVE and GEMINI trials.

- *Transformation of NMA results to transition probabilities*

The company transformed the results of the clinical response/remission NMAs from the probit scale to the natural scale and converted to absolute probabilities for use in the model. They take a simpler approach by assuming constant ratio of patients in remission and response throughout maintenance phase and beyond in extrapolation. Clinical advice to the ERG is that these assumptions might not be realistic as clinical -experience indicates the risk is greatest in the first 6-12 months; and falls thereafter. The proportion of patients with response and remission is likely to increase over time as per our clinical advice. This is because responders (without remission) are more likely to stop or switch therapy whereas those in remission would continue with treatment. However, in the absence of evidence it is difficult to adapt the model. Therefore, we conclude that the model assumption of constant risk of loss of response for patients on maintenance treatment does not reflect clinical experience. Extrapolation of relapse and discontinuation rates from the maintenance trials is likely to underestimate the average duration of treatment and hence both the costs and QALYs of active treatments. However, it is not possible to estimate the net direction of bias in ICERs between comparators, because trends in long-term risks may vary between TNFi drugs, vedolizumab and tofacitinib.

- *Exclusion of other serious adverse events*

The company excluded adverse events other than serious infections We agree that there would have been a risk of double-counting the costs and effects of ulcerative colitis exacerbations had all SAEs had been included in the model. Although, the omission of non-infection SAEs does introduce a risk of bias but given the frequency of these events this is unlikely to change the cost-effectiveness results.

- *NMA method for serious infections*

The company applied a binomial logit NMA model to estimate the risk of serious infections in the induction trials and chose the random effects model for their base case. Whilst the ERG agrees that there is considerable uncertainty associated with the risk of serious infections, we have reservations about the company's approach to estimating this parameter. Our verification checks indicated an even higher level of uncertainty around tofacitinib estimates, and we were unable to replicate the company's base case NMA values. We therefore applied a frequentist NMA approach to estimate the risk of serious infection, which we use as a scenario in ERG analysis

- All-cause mortality

The model adjusted mortality risks for age and gender mix for the general population and applied these to patients in pre-and post-surgery states. They assumed that, except for perioperative deaths, ulcerative colitis and treatment do not influence mortality. In general, we view this approach as reasonable, although there are additional mortality risks not reflected in the model – e.g. for colorectal cancer –although the relative risk estimates are likely to include perioperative deaths already accounted for.

Health Related Quality of life

The company's simple and regression-based analyses of EQ-5D data from the OCTAVE trials are problematic as sources of utility parameters for the economic model. They are relevant to the decision problem and clinical evidence, but the re-randomisation design and lack of intermediate assessments of clinical response and remission between week 8 and week 52 complicate the interpretation of results. We therefore agree with the company that the utility estimates by Woehl et al.[71] provide a more appropriate source for base case parameters that are consistent with previous NICE appraisals for ulcerative colitis. We use these estimates in ERG preferred analyses, but also test scenarios based on the company's OCTAVE analyses and published sources (Swinburn et al.).[72]

Resource use and costs

- *Drug acquisition*

We consider the drug acquisition costs used in the company model to be realistic, although there have been some small changes in NHS prices for included drugs; sulfasalazine (£7.83),

prednisolone (£0.47) and azathioprine (£2.20) (MIMS June 2018). These changes result in a very small reduction in the estimated cost of CT alone (£58.02), with biologic drugs (£51.68) and with tofacitinib (£48.86).

- *Drug administration*

Adalimumab and golimumab are administered by subcutaneous injection. The company assumed at zero cost to the NHS for self-administering these drugs. So, we conduct additional scenario analysis to assess the impact of assuming an initiation of self-administration of subcutaneous injections by adding the cost of a non-consultant led clinic attendance (£107) to the cost of induction for adalimumab and golimumab in our additional analyses.

- *Monitoring and follow up*

The company made health care usage assumptions from Tsai et al. (2008) which are consistent with health state definitions in the model and with previous NICE appraisals for ulcerative colitis (TA329 and TA342). We agree with the use of Tsai et al. as base case. But clinical advice to ERG suggests frequency of outpatient visits and endoscopy exceed current UK practice and additional assumptions about hospital episodes are unrealistic. We test alternative resource use scenario in our additional analyses.

We also question whether the assumption that maintenance treatment will always stop within 8 weeks of a loss of response is consistent with the number of outpatient appointments. To explore this, we conduct two scenario analyses to align the costs of assessing patients on maintenance treatment with the model assumption that treatment will always be discontinued within 8 weeks of a relapse.

The company excludes cost of stoma care and the estimated cost of surgery is low compared with previous appraisals. We test the inclusion of stoma care costs and higher surgery costs in our additional analysis.

8 REFERENCES

1. Dignass A, Eliakim R, Magro F, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis. Part 1: Definitions and diagnosis. *J Crohns Colitis* 2012;6:965-90.
2. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet* 2017;389:1756-70.
3. Mowat C, Cole A, Windsor A, et al. Guidelines for the management of inflammatory bowel disease in adults. *Gut* 2011;60(5):571-607.
4. Magro F, Gionchetti P, Eliakim R, 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. *J Crohns Colitis* 2017;11(6):649-70.
5. Harbord M, Eliakim R, Bettenworth D, et al. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. Part 2: Current management. *J Crohns Colitis* 2017;11:769-84.
6. NICE (National Institute for Health and Care Excellence). Ulcerative colitis: management. Clinical guideline cg166: NICE, 2013:1-36.
7. Burisch J, Jess T, Martinato M, et al. The burden of inflammatory bowel disease in Europe. *J Crohns Colitis* 2013;7(4):322-27.
8. NICE (National Institute for Health and Care Excellence). Infliximab, adalimumab and golimumanb for treating moderately to severely active ulcerative colitis after the failure of conventional therapy. NICE Guidance TA329. London: NICE, 2015.
9. NICE (National Institute for Health and Care Excellence). Vedolizumab for treating moderately to severely active ulcerative colitis. NICE Guidance TA342. London: NICE, 2015.
10. EMA (European Medicines Agency). Summary of Product Characteristics. Xeljanz (tofacitinib) 2018:1-61.
11. Sandborn WJ, Ghosh S, Panes J, et al. Tofacitinib, an oral Janus Kinase inhibitor in active ulcertaive colitis. *NEJM* 2012;367:616-24.
12. EMA (European Medicines Agency). Guideline on the development of new medicinal products for the treatment of ulcerative colitis [draft version]: Committee for Medicinal Products for Human Use (CHMP), 2016:1-16.
13. Mukherjee A, Hazra A, Smith MK, et al. Exposure-response characterization of tofacitinib efficacy in moderate to severe ulcerative colitis: Results from a dose-ranging phase 2 trial. *Br J Clin Pharmacol* 2018;84(6):1136-45.
14. Bonovas S, Lytras T, Nikolopoulos G, et al. Systematic review with network meta-analysis: comparative assessment of tofacitinib and biological therapies for moderate-to-severe ulcerative colitis. *Aliment Pharmacol Ther* 2018;47(4):454-65.
15. Panés J, Vermeire S, Lindsay JO, et al. Tofacitinib in patients with ulcerative colitis: Health-related quality of life in phase 3 randomised controlled induction and maintenance studies. *J Crohns Colitis* 2018;12(2):145-56.
16. Motoya S, Watanabe M, Kim HJ, et al. Tofacitinib induction and maintenance therapy in East Asian patients with active ulcerative colitis: Subgroup analyses from three phase 3 multinational studies. *Intest Res* 2018;16(2):233-45.
17. Sands BE, Moss AC, Armuzzi A, et al. Sa1761 - Efficacy and safety of dose escalation to tofacitinib 10 mg bid for patients with ulcerative colitis following loss of response on tofacitinib 5 mg bid maintenance therapy: Results from OCTAVE Open. *Gastroenterology* 2018;154 (6) Suppl 1:S-385.

18. Sands BE, Taub PR, Feagan BG, et al. Sa1750 - The effect of tofacitinib on serum lipids and cardiovascular safety in patients with ulcerative colitis - results from the tofacitinib ulcerative colitis clinical program. *Gastroenterology* 2018;154 (6) Suppl 1:S-380.
19. Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. *NEJM* 2017;376:1723-36.
20. Pfizer Inc. Interim Clinical Study Report: A Multi-Center, Open-Label Study of CP-690,550 in Subjects With Moderate to Severe Ulcerative Colitis: OCTAVE Open 2016
21. Danese S, Fiocchi C. Ulcerative colitis. *NEJM* 2011;365:1713-25.
22. FDA (Food and Drug Administration). Ulcerative colitis: Clinical trial endpoints. Guidance for industry [draft version]: US Food and Drug Administration, Center for Drug Evaluation and Reserach (CDER), 2016:1-19.
23. Mazzuoli S, Guglielmi FW, Antonelli E, et al. Definition and evaluation of mucosal healing in clinical practice. *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver* 2013;45(12):969-77. doi: 10.1016/j.dld.2013.06.010 [published Online First: 2013/08/13]
24. Khan HM, Mehmood F, Khan N. Optimal management of steroid-dependent ulcerative colitis. *Clinical and experimental gastroenterology* 2015;8:293-302. doi: 10.2147/ceg.s57248 [published Online First: 2015/12/10]
25. Archer R, Tappenden P, Ren S, et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (including a review of TA140 and TA262): clinical effectiveness systematic review and economic model. *Health Technology Assessment* 2016;20(39)
26. Ferrante M, Declerck S, De Hertogh G, et al. Outcome after proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis. *Inflamm Bowel Dis* 2007;14(1):20-28.
27. Feagan BG, Patel H, Colombel JF, et al. Effects of vedolizumab on health-related quality of life in patients with ulcerative colitis: results from the randomised GEMINI 1 trial. *Alimentary Pharmacology & Therapeutics* 2017;45(2):264-75. doi: https://dx.doi.org/10.1111/apt.13852
28. LeBlanc K, Mosli MH, Parker CE, et al. The impact of biological interventions for ulcerative colitis on health-related quality of life. *The Cochrane database of systematic reviews* 2015(9):Cd008655. doi: 10.1002/14651858.CD008655.pub3 [published Online First: 2015/09/24]
29. Stark RG, Reitmeir P, Leidl R, et al. Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. *Inflamm Bowel Dis* 2010;16(1):42-51. doi: 10.1002/ibd.20989 [published Online First: 2009/05/29]
30. Yarlas A, Bayliss M, Cappelleri JC, et al. Psychometric validation of the SF-36((R)) Health Survey in ulcerative colitis: results from a systematic literature review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2018;27(2):273-90. doi: 10.1007/s11136-017-1690-6 [published Online First: 2017/08/30]
31. Ware J E .Jr, Kosinski M, Bjorner JB, et al. User's manual for the SF-36v2 Health Survey. 2nd ed. Lincoln, RI: QualityMetric Incorporated 2007.
32. Reilly M C, Zbrozek A S, Dukes E M. The validity and reproducibility of a work productivity and activity impairment instrument. *PharmacoEconomics* 1993;4(5):353.
33. Travis SP, Feagan BG, Peyrin-Biroulet L, et al. Effect of adalimumab dose escalation on clinical, healthrelated quality of life, treatment satisfaction and work productivity outcomes among patients with ulcerative colitis in a clinical practice setting: Results from inspirada. *Gastroenterology* 2017;152 (5 Supplement 1):S742-S43.
34. Binion DG, Louis E, Oldenburg B, et al. Effect of adalimumab on work productivity and indirect costs in moderate to severe Crohn's disease: a meta-analysis. *Canadian journal*

of gastroenterology = Journal canadien de gastroenterologie 2011;25(9):492-6. [published Online First: 2011/09/14]
35. Mshimesh BR. Efficacy and safety of adalimumab versus infliximab in patients suffered from moderate to severe active ulcerative colitis. *Asian Journal of Pharmaceutical and Clinical Research* 2017;10(3):300-07.
36. Panaccione R, Ghosh S, Middleton S, et al. Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. *Gastroenterology* 2014;146(2):392-400.e3. doi: 10.1053/j.gastro.2013.10.052 [published Online First: 2014/02/12]
37. Suzuki Y, Motoya S, Hanai H, 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. doi: 10.1007/s00535-013-0922-y [published Online First: 2013/12/24]
38. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *The New England journal of medicine* 2005;353(23):2462-76. doi: 10.1056/NEJMoa050516 [published Online First: 2005/12/13]
39. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014;146(1):96-109.e1. doi: 10.1053/j.gastro.2013.06.010 [published Online First: 2013/06/19]
40. Hibi T, Imai Y, Senoo A, et al. 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. doi: 10.1007/s00535-017-1326-1 [published Online First: 2017/03/23]
41. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *The New England journal of medicine* 2013;369(8):699-710. doi: 10.1056/NEJMoa1215734 [published Online First: 2013/08/24]
42. Jiang XL, Cui HF, Gao J, et al. 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. doi: 10.1097/mcg.0000000000000319 [published Online First: 2015/04/07]
43. Kobayashi T, Suzuki Y, Motoya S, 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. doi: 10.1007/s00535-015-1102-z [published Online First: 2015/07/15]
44. Rodgers M, Epstein D, Bojke L, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health technology assessment (Winchester, England)* 2011;15(10):i-xxi, 1-329. doi: 10.3310/hta15100 [published Online First: 2011/02/22]
45. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in medicine* 2004;23(9):1351-75. doi: 10.1002/sim.1761 [published Online First: 2004/04/30]
46. Dias S, Welton NJ, Sutton AJ, et al. 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: University of Sheffield, 2011 (last updated 2016):1-98.
47. Thorlund K, Druyts E, Toor K, et al. Incorporating alternative design clinical trials in network meta-analyses. *Clinical epidemiology* 2015;7:29-35. doi: 10.2147/clep.s70853 [published Online First: 2015/01/08]

48. Takeda. Vedolizumab (Entyvio ®) for the treatment of adults with moderate to severe active ulcerative colitis. Manufacturer submission to NICE TA342., 2014.
49. Thorlund K, Druyts E, Toor K, et al. Comparative efficacy of golimumab, infliximab, and adalimumab for moderately to severely active ulcerative colitis: a network meta-analysis accounting for differences in trial designs. *Expert review of gastroenterology & hepatology* 2015;9(5):693-700. doi: 10.1586/17474124.2015.1024657 [published Online First: 2015/03/13]
50. Sandborn WJ, van Assche G, Reinisch W, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2012;142(2):257-65.e1-3. doi: 10.1053/j.gastro.2011.10.032 [published Online First: 2011/11/09]
51. NICE (National Institute for Health and Care Excellence). Guide to the methods of technology appraisal London, 2013.
52. Buckland A, Bodger K. The cost-utility of high dose oral mesalazine for moderately active ulcerative colitis. *Alimentary pharmacology & therapeutics* 2008;28(11-12):1287-96.
53. Essat M, Tappenden P, Ren S, et al. Vedolizumab for the treatment of adults with moderate-to-severe active ulcerative colitis: an evidence review group perspective of a NICE single technology appraisal. *Pharmacoeconomics* 2016;34(3):245-57.
54. Tappenden P, Ren S, Archer R, et al. A model-based economic evaluation of biologic and non-biologic options for the treatment of adults with moderately-to-severely active ulcerative colitis after the failure of conventional therapy. *Pharmacoeconomics* 2016;34(10):1023-38.
55. Tsai H, Punekar Y, Morris J, et al. 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):1230-39.
56. Wilson MR, Bergman A, Chevrou-Severac H, et al. Cost-effectiveness of vedolizumab compared with infliximab, adalimumab, and golimumab in patients with ulcerative colitis in the United Kingdom. *Eur J Health Econ* 2018;19(2):229-40. doi: https://dx.doi.org/10.1007/s10198-017-0879-5
57. Ali T, Skup M, Yang M, et al. Cost-Effectiveness of adalimumab in moderately to severely active ulcerative colitis. *American Journal of Gastroenterology* 2012;107:S641.
58. Mukhekat V, Ahmad I, Merali Y, et al. PTU-078 The Application Of A Markov Model For Evauating Asa Therapy For Ulcerative Colitis. *Gut* 2014;63(Suppl 1):A73-A73.
59. Wilson M, Kerrigan M, Smyth M, et al. Cost-effectiveness of vedolizumab compared with conventional therapy for treatment of moderately-to-severely active ulcerative colitis in the United Kingdom. *Gut* 2016;65:A260-A60.
60. Wilson M, Mody R, Ursan I, et al. Sa1019 Cost-Effectiveness of Vedolizumab Compared With Conventional Therapy and Biologics for Treatment of Moderately to Severely Active Ulcerative Colitis in the United States. *Gastroenterology* 2015;148(4):S-198-S-99.
61. Yang H, Curry A, Wang S, et al. Cost-effectiveness of adalimumab in moderately to severely active ulcerative colitis. *United European Gastroenterology Journal* 2014;2:A374.
62. Wilson MR, Azzabi Zouraq I, Chevrou-Severac H, et al. Cost-effectiveness of vedolizumab compared with conventional therapy for ulcerative colitis patients in the UK. *ClinicoEcon* 2017;9:641-52. doi: https://dx.doi.org/10.2147/CEOR.S135609
63. Wu B, Wang Z, Zhang Q. Cost-Effectiveness of Different Strategies for the Treatment of Moderate-to-Severe Ulcerative Colitis. *Inflamm Bowel Dis* 2018;27:27. doi: https://dx.doi.org/10.1093/ibd/izy114
64. Royal College of Physicians (RCP). National clinical audit of biological therapies: UK inflammatory bowel disease (IBD) audit. Annual Report September 2016, 2016.
65. Royal College of Physicians (RCP). National clinical audit of biological therapies: UK inflammatory bowel disease (IBD) audit. Adult report. London, 2015.

66. Misra R, Askari A, Faiz O, et al. 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
67. Chhaya V, Saxena S, Cecil E, 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.
68. Frolkis AD, Dykeman J, Negrón ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology* 2013;145(5):996-1006.
69. Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). *Scandinavian journal of gastroenterology* 2009;44(4):431-40.
70. Arai K, Koganei K, Kimura H, et al. Incidence and outcome of complications following restorative proctocolectomy. *The American journal of surgery* 2005;190(1):39-42.
71. Woehl A, Hawthorne A, McEwan P. The relation between disease activity, quality of life and health utility in patients with ulcerative colitis. *Gut* 2008;57:A153-A53.
72. Swinburn P, Elwick H, Bean K, et al. PTU-127 The impact of surgery on health related quality of life in ulcerative colitis. *Gut* 2012;61(Suppl 2):A237-A37.
73. Ara R, Brazier J. Populating an economic model with health state utility values: moving toward better practice. *Value in Health* 2010;13(5):509-18.
74. Diamantopoulos A, Finckh A, Huizinga T, et al. Tocilizumab in the treatment of rheumatoid arthritis: a cost-effectiveness analysis in the UK. *Pharmacoeconomics* 2014;32(8):775-87.
75. Kosmas C, Kerr C, Marsh S, et al. Health Related Quality of Life (Hrqol) In Patients Who Have Undergone Colectomy for Ulcerative Colitis: Impacts of Complications Post-Surgery. *Value in Health* 2015;18(7):A630.
76. NICE (National Institute for Health and Care Excellence). Guide to the methods of technology appraisal 2013. *NICE, London* 2013
77. Hatswell AJ, Porter J, Lee D, et al. The cost of costing treatments incorrectly: errors in the application of drug prices in economic evaluation due to failing to account for the distribution of patient weight. *Value in Health* 2016;19(8):1055-58.
78. Buchanan J, Wordsworth S, Ahmad T, et al. Managing the long term care of inflammatory bowel disease patients: The cost to European health care providers. *Journal of Crohn's and Colitis* 2011;5(4):301-16.

9 APPENDICES

Appendix 1 Additional results tables for subgroup analyses by TNFi-exposure status

Results according to subgroups by TNFi-exposure status for the outcomes of remission (primary outcome), mucosal healing, and Sustained corticosteroid-free remission among patients in remission at baseline are presented below in Table 90 to Table 94.

Remission

Table 90 Proportion of patients in remission in OCTAVE Induction 1 and 2 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup: Prior-TNFi treatment	Tofacitinib 10 mg	Placebo	Difference (95% CI); p-value	p-value for heterogeneity
	n/N (%)	n/N (%)		
OCTAVE 1, week 8				
TNFi-naïve Central read	56/222 (25.2)	9/57 (15.8)	9.4 (−1.6, 20.5); p=0.1328	0.1034
TNFi-exposed Central read	32/254 (12.6)	1/65 (1.5)	11.1 (6.0, 16.1); p=0.0090	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]	
OCTAVE 2, week 8				
TNFi-naïve Central read	43/195 (22.1)	4/47 (8.5)	13.5 (3.7, 23.4); p=0.0352	0.0956
TNFi-exposed Central read	28/234 (12.0)	0/65 (0.0)	12.0 (7.8, 16.1); p=0.0034	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]	
OCTAVE 1 & 2 pooled data, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]	NR
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]	NR
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	NR
TNFi-exposed	[redacted]	[redacted]	[redacted]	NR

Local read				

Source: CS Appendix E Table 121

Table 91 Proportion of patients in remission in OCTAVE Sustain at week 52 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup Prior-TNFi treatment	TOF 5 mg n/N (%))	PBO n/N (%)	Difference vs placebo (95% CI); p-value	TOF 10 mg n/N (%)	Difference vs placebo (95% CI); p-value
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix E Table 125

Mucosal healing

Table 92 Proportion of patients with mucosal healing in OCTAVE Induction 1 and 2 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup: Prior-TNFi treatment	Tofacitinib 10 mg n/N (%)	Placebo n/N (%)	Difference (95% CI); p-value	p-value for heterogeneity
OCTAVE 1, week 8				
TNFi-naïve Central read	88/222 (39.6)	15/57 (26.3)	13.3 (0.2, 26.4); p=0.0630	0.1169
TNFi-exposed Central read	61/254 (24.0)	4/65 (6.2)	17.9 (10.0, 25.7); p=0.0014	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed	[redacted]	[redacted]	[redacted]	

Local read				
OCTAVE 2, week 8				
TNFi-naïve Central read	71/195 (36.4)	9/47 (19.1)	17.3 (4.1, 30.4); p=0.0239	0.3958
TNFi-exposed Central read	51/234 (21.8)	4/65 (6.2)	15.6 (7.8, 23.5); p=0.0040	
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]	
OCTAVE 1 & 2 pooled data, week 8				
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]	NR
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]	NR
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	NR
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]	NR

Source: CS Appendix E Table 122

Table 93 Proportion of patients with mucosal healing in OCTAVE Sustain at week 52 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup Prior-TNFi treatment	TOF 5 mg n/N (%))	PBO n/N (%)	Difference vs placebo (95% CI); p-value	TOF 10 mg n/N (%)	Difference vs placebo (95% CI); p-value
TNFi-naïve Central read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Central read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-naïve Local read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TNFi-exposed Local read	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix E Table 126

Sustained corticosteroid-free remission among patients in remission at baseline

Table 94 Proportion of patients in remission at baseline who had sustained corticosteroid-free remission in OCTAVE Sustain at week 52 according to prior TNFi treatment (FAS, NRI, central and local reads)

Subgroup Prior-TNFi treatment (Yes/No)	TOF 5 mg n/N (%))	PBO n/N (%)	Difference vs placebo (95% CI); p-value	TOF 10 mg n/N (%)	Difference vs placebo (95% CI); p-value
TNFi-naïve Central read					
TNFi-exposed Central read					
TNFi-naïve Local read					
TNFi-exposed Local read					

Source: CS Appendix E Table 129

Appendix 2 NMA sensitivity analyses, additional tables

The tables below are condensed versions of tables that are reported in full in CS Appendix D.1.3.5. Sensitivity analyses were conducted as follows:

- Using centrally read endoscopic subscores for the clinical response, clinical remission and mucosal healing outcomes instead of locally read endoscopic subscores
- Excluding studies in which the majority of participants were Asian. These studies were Suzuki 2014, Mshimesh 2017, Jiang 2015, Kobayashi 2015 and Pursuit J. The CS states that this "*sensitivity analysis is aligned with the base-case assumptions made in the NMA supporting TA329*".
- Limiting the data from the OCTAVE trials and the ULTRA 2 study to patients with prior TNFi failure (i.e. a subset of the base case data which included all patients with prior TNFi-exposure)
- Conducting an overall ITT analysis in which data were not divided into two subgroups by TNFi-exposure status (i.e. combined analysis regardless of prior TNFi-exposure status).

Results for are presented below in Table 95 to Table 102.

Table 95 Clinical response and clinical remission NMA sensitivity analyses – Induction phase, Comparator vs PBO

Comparator	Sensitivity Analyses								
	Centrally read endoscopic subscores			Exclusion of Asian studies			TNFi-exposed using TNFi-failures		
	OR, median (95%CrI)		SUCRA	OR, median (95%CrI)		SUCRA	OR, median (95%CrI)		SUCRA
	Clinical response	Clinical remission		Clinical response	Clinical remission		Clinical response	Clinical remission	
TNFi-naïve subgroup									
PBO			[redacted]			[redacted]			
TOF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
INF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
VED	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
TNFi-exposed subgroup									
PBO			[redacted]						[redacted]
TOF	[redacted]	[redacted]	[redacted]				[redacted]	[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]						
VED	[redacted]	[redacted]	[redacted]				[redacted]	[redacted]	[redacted]

Source: CS Appendix D Table 101; Table 103, Table 104

Table 96 Clinical response and clinical remission NMA sensitivity analyses – Maintenance phase, Comparator vs PBO

Comparator	Sensitivity Analyses								
	Centrally read endoscopic subscores			Exclusion of Asian studies			TNFi-exposed using TNFi-failures		
	OR, median (95%CrI)		SUCRA	OR, median (95%CrI)		SUCRA	OR, median (95%CrI)		SUCRA
	Clinical response	Clinical remission		Clinical response	Clinical remission		Clinical response	Clinical remission	
TNFi-naïve subgroup									
PBO			[redacted]			[redacted]			
TOF 5 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
TOF 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
INF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
GOL 50 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
GOL 100 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
VEDQ8W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
VED Q4W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			
TNFi-exposed subgroup									
PBO			[redacted]						
TOF 5 mg	[redacted]	[redacted]	[redacted]						
TOF 10 mg	[redacted]	[redacted]	[redacted]						
ADA	[redacted]	[redacted]	[redacted]						
VED Q8W	[redacted]	[redacted]	[redacted]						
VED Q4W	[redacted]	[redacted]	[redacted]						

Source: CS Appendix D Table 102, 103, 104

A sensitivity analysis for the overall ITT population (i.e. combining TNFi-naïve and TNFi-exposed participants) was also reported.

Table 97 Overall ITT scenario analysis NMA results – comparative effects and probabilities of achieving clinical response and clinical remission

Comparator	Comparator vs PBO		TOF vs comparator		Absolute probability		SUCRA
	Odds ratio, median (95%CrI)		Odds ratio, median (95%CrI)				
	Clinical response	Clinical remission	Clinical response	Clinical remission	Clinical response	Clinical remission	
Induction phase							
PBO			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

TOF	[redacted]	[redacted]			[redacted]	[redacted]	[redacted]
INF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Maintenance phase (re-randomised responder trials only)							
PBO			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOF 5 mg	[redacted]	[redacted]			[redacted]	[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 50 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL 100 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED Q8W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED Q4W	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix D Table 106

Table 98 Mucosal healing NMA sensitivity analyses – Induction phase, Comparator vs placebo

	Sensitivity Analyses					
	Centrally read endoscopic subscores		Exclusion of Asian studies		TNFi-exposed using TNFi-failures	
Comparator	OR, median (95%CrI)	SUCRA	OR, median (95%CrI)	SUCRA	OR, median (95%CrI)	SUCRA
TNFi-naïve subgroup						
PBO		[redacted]		[redacted]		
TOF	[redacted]	[redacted]	[redacted]	[redacted]		
INF	[redacted]	[redacted]	[redacted]	[redacted]		
ADA	[redacted]	[redacted]	[redacted]	[redacted]		
GOL	[redacted]	[redacted]	[redacted]	[redacted]		
VED	[redacted]	[redacted]	[redacted]	[redacted]		
TNFi-exposed subgroup						
PBO		[redacted]				[redacted]

TOF	[redacted]	[redacted]			[redacted]	[redacted]
ADA	[redacted]	[redacted]				
VED	[redacted]	[redacted]			[redacted]	[redacted]

Source: CS Appendix D Table 112, 114, 115

Table 99 Mucosal healing NMA sensitivity analyses – Maintenance phase, Comparator vs placebo

Comparator	Sensitivity Analyses					
	Centrally read endoscopic subscores		Exclusion of Asian studies		TNFi-exposed using TNFi-failures	
	OR, median (95%CrI)	SUCRA	OR, median (95%CrI)	SUCRA	OR, median (95%CrI)	SUCRA
TNFi-naïve subgroup						
PBO		[redacted]		[redacted]		
TOF 5 mg	[redacted]	[redacted]	[redacted]	[redacted]		
TOF 10 mg	[redacted]	[redacted]	[redacted]	[redacted]		
INF	[redacted]	[redacted]	[redacted]	[redacted]		
ADA	[redacted]	[redacted]	[redacted]	[redacted]		
GOL 50 mg	[redacted]	[redacted]	[redacted]	[redacted]		
GOL 100 mg	[redacted]	[redacted]	[redacted]	[redacted]		
VEDQ8W	[redacted]	[redacted]	[redacted]	[redacted]		
VED Q4W	[redacted]	[redacted]	[redacted]	[redacted]		
TNFi-exposed subgroup						
PBO		[redacted]				[redacted]
TOF 5 mg	[redacted]	[redacted]			[redacted]	[redacted]
TOF 10 mg	[redacted]	[redacted]			[redacted]	[redacted]
ADA	[redacted]	[redacted]			[redacted]	[redacted]
VED Q8W	[redacted]	[redacted]			[redacted]	[redacted]
VED Q4W	[redacted]	[redacted]			[redacted]	[redacted]

Source: CS Appendix D Table 113, 114, 115

Adverse events

Sensitivity analysis was conducted using a network from which the Asian studies (Suzuki 2014, Kobayashi 2015 and Mshimesh 2017) were excluded and also the tofacitinib phase II study (Sandborn 2012). The exclusion of the Asian studies also caused the loss of the UC-SUCCESS trial (azathioprine versus infliximab) from the network as it could no longer be connected to the network of evidence (Figure 14).

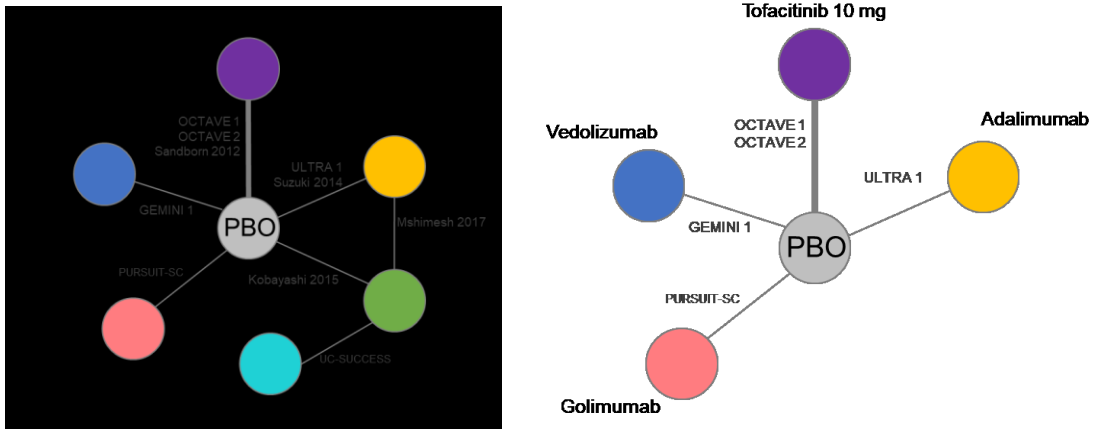


Figure 14 Base-case safety evidence network (left) and sensitivity analysis network (right)

Results are shown in the tables below.

Table 100 Induction phase sensitivity analysis NMA results – comparative effects and probabilities of discontinuing due to AEs

Comparator	Comparator vs PBO		TOF vs comparator	Absolute probability, median (95% CrI)	SUCRA
	Treatment effect (logit scale), median (95% CrI)	Odds ratio, median (95% CrI)	Odds ratio, median (95% CrI)		
PBO			[redacted]	[redacted]	[redacted]
TOF	[redacted]	[redacted]		[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix D Table 116

Table 101 Induction phase sensitivity analysis NMA results – comparative effects and probabilities of serious AEs

Comparator	Comparator vs PBO		TOF vs comparator	Absolute probability, median (95% CrI)	SUCRA
	Treatment effect (logit scale), median (95% CrI)	Odds ratio, median (95% CrI)	Odds ratio, median (95% CrI)		
PBO			[redacted]	[redacted]	[redacted]
TOF	[redacted]	[redacted]		[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix D Table 117

Table 102 Induction phase sensitivity analysis NMA results – comparative effects and probabilities of serious infections

Comparator	Comparator vs PBO		TOF vs comparator	Absolute probability, median (95% CrI)	SUCRA
	Treatment effect (logit scale), median (95% CrI)	Odds ratio, median (95% CrI)	Odds ratio, median (95% CrI)		
PBO			[redacted]	[redacted]	[redacted]
TOF	[redacted]	[redacted]		[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VED	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Appendix D Table 118

Appendix 3 Health economics: results of ERG scenario analyses

Table 103 ERG base case: scenarios on patient age (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Patient age: 28 years	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,644
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£589,024
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,170
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Patient age: 52 years	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£8,019
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£628,794
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,648
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 104 ERG base case: scenarios on weight * (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Patient weight: 70 kgs	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,827
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£607,395
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,401
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Patient weight: 80 kgs	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,819
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£607,504
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,394
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

* Scenario based on “Use average of OCTAVE” option for wastage calculations

Table 105 ERG base case: scenarios on NMA models (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Network meta-analysis (company preferred response/remission)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,793
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£633,458
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,541
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	£8,801,245
TNFi-failed (ved) + TNFi-exposed for tof and ada	***TNFi-exposed***			
	Conventional	[redacted]	[redacted]	£9,669
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£2,521,513
Network meta-analysis (FE for serious infections)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£8,513
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£589,976
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£10,080
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 106 ERG base case: scenarios on utility (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Utility source: Swinburn	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£10,898
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£845,865
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£13,198
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	--
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Utility source: OCTAVE 8 weeks	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£17,764
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£1,360,239
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£21,376
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Utility source: OCTAVE 52 weeks	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£18,256
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£1,373,067
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£21,283
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 107 ERG base case: scenarios on resource use (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Drug stopping rule (6.5 outpatient visit for all patients in maintenance)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£9,090
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	£608,793
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£10,597
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Reduced health state resource use (clinical practice scenario)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£13,938
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	£613,289
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£14,950
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 108 ERG base case: scenarios on drug costs (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
CT drug usage (mesalazine only)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,827
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	£607,576
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,402
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	-
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Drug admin cost (for subcutaneous injection for ada and gol.)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,815
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£607,571
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,389
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 109 ERG base case: scenarios on surgery cost (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Stoma care costs (£81.66 per 8 week cycle TA342)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,804
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£607,561
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,379
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Surgery costs (£13,156 Buchannan et al.)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,764
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£607,522
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,341
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 110 ERG base case: scenarios on surgery risks (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Surgery incidence rate (Chhaya et al.)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,980
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£611,440
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,558
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant
Surgery complications (Tappenden et al.)	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	£7,556
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Golimumab	[redacted]	[redacted]	Tofacitinib dominant
	Infliximab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	£605,226
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	£9,134
	Adalimumab	[redacted]	[redacted]	Tofacitinib dominant
	Tofacitinib	[redacted]	[redacted]	
	Vedolizumab	[redacted]	[redacted]	Tofacitinib dominant

Table 111 ERG base case: drug sequencing scenarios (tofacitinib PAS, others at list price)

	Treatments	Total costs	Total QALYs	Pairwise ICER (£/QALY) (Tof vs comparator)
Treatment sequencing	***TNFi- naive***			
	Conventional	[redacted]	[redacted]	[redacted]
	Inf-Ada-CT	[redacted]	[redacted]	[redacted]
	Inf-Ved-CT	[redacted]	[redacted]	[redacted]
	Inf-Tof-CT	[redacted]	[redacted]	[redacted]
	Tof-Ada-CT	[redacted]	[redacted]	[redacted]
	Ved-Ada-CT	[redacted]	[redacted]	[redacted]
	TNFi- naive			
	Conventional	[redacted]	[redacted]	[redacted]
	Gol-Ada-CT	[redacted]	[redacted]	[redacted]
	Gol-Ved-CT	[redacted]	[redacted]	[redacted]
	Gol-Tof-CT	[redacted]	[redacted]	[redacted]
	Tof-Ada-CT	[redacted]	[redacted]	[redacted]
	Ved-Ada-CT	[redacted]	[redacted]	[redacted]
	TNFi-exposed			
	Conventional	[redacted]	[redacted]	[redacted]
	Tof-Ada-CT	[redacted]	[redacted]	[redacted]
	Ved-Ada-CT	[redacted]	[redacted]	[redacted]