CONFIDENTIAL UNTIL PUBLISHED

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

Ustekinumab for treating moderately to severely active ulcerative colitis
Report version post factual accuracy check

Produced by Southampton Health Technology Assessments Centre

Authors Mrs Neelam Kalita, Research Fellow, SHTAC
Professor Joanne Lord, Director, SHTAC
Dr Karen Pickett, Research Fellow, SHTAC
Professor David A Scott, Director, Diligent Agile Synthesis Limited
Dr Geoff Frampton, Senior Research Fellow, SHTAC

Correspondence to Dr Geoff Frampton
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 16th September 2019

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 12/95/55

Acknowledgements

We thank the following for providing clinical advice and comments on the draft report: Dr Markus Gwiggner, Consultant Gastroenterologist, University Hospital Southampton NHS Foundation Trust; and Dr Rebecca Saich, Consultant Gastroenterologist, Basingstoke and North Hampshire Hospital, Hampshire Hospitals NHS Foundation Trust. We also thank Dr Jill Colquitt (Effective Evidence LLP) for acting as an internal editor for the draft report.

Declared competing interests of the authors and clinical advisors

None from the authors. None from Dr Saich. Dr Gwiggner received a fee for chairing an educational meeting for AbbVie Limited (manufacturer of adalimumb) and a fee for an educational talk from Merck Sharp & Dome Limited (manufacturer of infliximab and golimumab) and Takeda UK Limited (manufacturer of vedolizumab). He received sponsorship for a conference registration, travel and accommodation from AbbVie Limited.

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, Lord J, Pickett K; Scott D; Frampton G. Ustekinumab for treating moderately to severely active ulcerative colitis: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2019.

Contributions of authors

Neelam Kalita critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Joanne Lord critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Karen Pickett critically appraised the clinical effectiveness systematic review and drafted the report; David Scott critically appraised the indirect treatment comparison and

drafted the report; Geoff Frampton critically appraised the clinical effectiveness systematic review, drafted the report, project managed the report and is the project guarantor.

Word count: 57,865

Confidentiality

This report contains confidential information, marked as follows:

[redacted]; [redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADA	Adalimumab
AE	Adverse event
Anti-TNF	TNF-alpha inhibitor therapy (also called TNF agonist)
BNF	British National Formulary
CMU	Commercial Medicines Unit
CODA	Convergence Diagnostics and Output Analysis
CrI	Credible interval
CMU	Commercial Medicines Unit
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
CT	Conventional therapy
DIC	Deviance information criterion
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
ERG	Evidence Review Group
EQ-5D	EuroQol 5-Dimension
FDA	Food and Drug Administration
GOL	Golimumab
HR	Hazard ratio
HRQoL	Health related quality of life
IBDQ	Inflammatory Bowel Disease Questionnaire
ICER	Incremental cost effectiveness ratio
INF	Infliximab
ITC	Indirect treatment comparison
ITT	Intention-to-treat
IV	Intravenous
MIMS	Monthly Index of Medical Specialities
NA	Not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
OS	Overall survival
PAS	Patient access scheme
PBO	Placebo
PD	Progressed disease
PSSRU	Personal Social Services Research Unit
q4w	Every four weeks
q8w	Every eight weeks
q12w	Every twelve weeks
QALY	Quality adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
SC	Subcutaneous
SF-36	Short-form 36 generic quality of life questionnaire
SmPC	Summary of Product Characteristics
SoC	Standard of care
STA	Single Technology Appraisal
TNF-alpha	Tissue necrosis factor alpha
TNF agonist	TNF-alpha inhibitor therapy (also called anti-TNF)
TOF	Tofacitinib
UC	Ulcerative colitis

UST	Ustekinumab
VED	Vedolizumab
WPAI-GH	Work Productivity and Activity Index – General Health

SUMMARY

Scope of the company submission

The NICE scope specifies that the population of interest is people with moderately to severely active ulcerative colitis (UC) who are intolerant of, or whose disease has had an inadequate response, or loss of response, to previous biologic therapy (a TNF-alpha inhibitor or vedolizumab) or a Janus kinase (JAK) inhibitor (tofacitinib), or conventional therapy (oral corticosteroids and/or immunomodulators). The scope specifies that, if evidence allows, subgroups of people who have been previously been treated with one or more biologics, and people who have not received prior biologic therapy should be considered. The company's decision problem and analyses are broadly consistent with the NICE scope. However, whilst the NICE scope defines the prior therapy subgroups in terms of prior treatment exposure, the company define the subgroups in terms of prior treatment failure. The company's subgroups are:

- "non-biologic failure" (i.e. people who have received treatment with 1 or more TNF antagonists or vedolizumab at a dose approved for the treatment of UC, and either did not respond initially, responded initially but then lost response, or were intolerant to the medication.
- "biologic failure" (i.e. people who were biologic-naïve or may have been exposed to biologic therapy but did not demonstrate an inadequate response or intolerance to treatment with a biologic agent (i.e. a TNF antagonist, or vedolizumab). These patients must have demonstrated an inadequate response to, or have failed to tolerate, at least 1 of the specified non-biologic UC therapies.

In the company's pivotal clinical trial the majority of participants in the company's "non-biologic failure" and "biologic failure" subgroups match the respective NICE scope subgroups "people who have not received prior biologic therapy" and "people who have previously been treated with one or more biologics".

Summary of submitted clinical effectiveness evidence

The company submission (CS) includes a review of clinical effectiveness studies, and provides methods and results for:

- The company's pivotal trial (UNIFI) which compared ustekinumab against placebo (placebo reflects background conventional therapy).
- Network meta-analyses (NMAs) comparing ustekinumab, adalimumab, golimumab, infliximab, tofacitinib vedolizumab and placebo.

The UNIFI trial and comparator trials cover the induction and maintenance phases of treatment. In the induction phase of UNIFI the standard dose of ustekinumab was ~6mg/kg (as per the anticipated marketing authorisation), although a lower 130mg fixed dose was also included; in the maintenance phase a standard regimen (90mg q12w) and an escalated-dose regimen (90mg q8w) were compared against the maintenance phase placebo arm.

The company report three sets of NMAs: modelling only the induction phase (approximately 8 weeks); modelling both the induction and maintenance phases (totalling approximately 1 year); and modelling both the induction and maintenance phases (totalling approximately 1 year) for induction responders only, in an approach which they refer to as 1-year NMA conditional on response. The 1-year analyses take into account that some trials (including UNIFI) have a “re-randomised” design whilst others have a “treat-through” design, by adjusting the results of treat-through trials to mimic those that would have been obtained from a re-randomised approach. This is a different NMA approach compared to previous NICE appraisals in moderately to severely active UC.

Both the UNIFI trial results and those from the NMAs are reported separately for non-biologic failure and biologic failure subgroups of patients.

Results of the UNIFI trial

Ustekinumab improved rates of clinical remission and clinical response at induction week 8 and maintenance week 44 compared to the respective placebo arms, both for the non-biologic failure and biologic failure subgroups and for both the q8w and q12w maintenance dose regimens. At the end of induction, rates of remission and response were higher in the non-biologic failure subgroup than the biologic failure subgroup. At the end of maintenance therapy, rates of remission and response were higher in the q8w arm than the q12w arm in the biologic failure subgroup but did not differ between the two dose regimens in the non-biologic failure subgroup. Results for mucosal healing were also favourable for ustekinumab but were not reported by subgroup.

Results of the disease-specific Inflammatory Bowel Disease Questionnaire (IBDQ) are consistent with those of the generic SF-36 and EQ-5D health-related quality of life (HRQoL) measures. These instruments showed that ustekinumab improved patients’ HRQoL in both the induction and maintenance phases of therapy relative to the respective placebo arms, for all dose regimens, and with the differences from placebo exceeding thresholds for being

clinically meaningful. The improvements in HRQoL at week 44 were marginally larger for the q8w maintenance regimen than the q12w regimen, but not reaching the threshold for being clinically meaningful.

Ustekinumab is relatively well tolerated, and although the majority of patients in the UNIFI trial experienced adverse events, fewer than 10% of these were serious.

Results of network meta-analyses

The company identified 18 comparator trials potentially eligible for meta-analysis. This is similar to the set of trials included in NMAs in the recent NICE technology appraisal TA547 (tofacitinib), except that the company has excluded trials that were specifically on Asian populations (included in the TA547 analyses).

Results of the induction NMAs and the 1-year NMAs conditional on response consistently indicate that ustekinumab and all the comparator therapies improved the odds of clinical remission and clinical response both at 8 weeks and 44 weeks compared to the respective placebo arms (i.e. the background conventional therapy). The CS concludes that, in the induction NMAs ustekinumab demonstrated a higher likelihood of response than adalimumab and golimumab in non-biologic failure patients and higher likelihood of response than adalimumab in biologic failure patients. The company also conclude that, in the 1-year NMAs conditional on response, ustekinumab had a higher probability of being more effective than all the comparators (CS section B.2.9.5). The probabilities reported in the CS on which these conclusions are based are subject to uncertainty, but the company have not provided credible intervals for the probabilities.

Summary of submitted cost effectiveness evidence

The company submission includes:

i) a review of published economic evaluations of biologics and JAK targeted therapies for UC; and
ii) An economic evaluation undertaken for the NICE STA process, comparing ustekinumab with other biologics, JAK inhibitors and non-biologic (conventional therapy) for the treatment of adults with moderately to severely active UC.

The company conducted a systematic search of the literature to identify economic evaluations of treatments in patients with moderately to severely active UC. They identified

26 relevant studies; 11 of which were UK based. None of these studies evaluated the cost-effectiveness of ustekinumab in the population of interest.

The company model follows a conventional design for UC, but with some changes to previous Technology Appraisal (TA) models. They developed a hybrid model, consisting of a decision tree (for the induction phase) and a Markov model (for the maintenance phase). The model consists of nine health states: remission; response without remission; active UC; 1st surgery; Post-1st surgery remission; Post-1st surgery complications; 2nd surgery; Post-2nd surgery remission; and death. The company estimate the distribution of the cohort between the health states at each time point by using a set of transition probabilities, obtained from direct trial evidence or NMA of clinical evidence.

Other key features and assumptions of the model are listed below:

- *Model cycle:* induction phase is designed to accommodate induction periods of different lengths for each treatment; maintenance phase: 2 weeks.
- *Time horizon:* 50 years in the base case (effectively lifetime from a starting age of 41 years), with a half-cycle correction.
- *Duration of treatment:* Responders to induction continue maintenance until loss of response or death
- *Treatment stopping rule:* Not applied in the company base case
- *Sequential treatment*: The base case model assumes that after the failure of the initial treatment, all patients switch to conventional therapy alone.
- *Adverse events*: Only serious infections are included; treated as one-off events.
- *Utility and QALY calculations*: The base case company model uses utility estimates from published evidence, as in previous TAs. Utilities are adjusted for age and gender. A utility decrement for the adverse effect of serious infections is incorporated in the company model.
- *Health resource use and costs:* Costs were sourced from published literature, previous NICE TAs, the Monthly index of Medical Specialities (MIMS) and the BNF 2017/2018
- *Discounting:* 3.5% per year for costs and QALYs.
- *Uncertainty:* The model allows for exploration of uncertainty over input parameters using deterministic sensitivity analysis; scenario analyses varying selected model assumptions; and probabilistic sensitivity analysis (PSA) to estimate the joint effects of parameter uncertainty on the estimated costs and QALYs.

Commentary on the robustness of submitted evidence

Strengths

- The company conducted comprehensive searches for clinical effectiveness studies and economic evaluations related to the decision problem, with appropriate eligibility criteria. Their findings are well documented.
- The company's pivotal UNIFI trial was well conducted and judged to be at low risks of the key domains of bias.
- The comparators in the company model reflect the NICE scope.
- The company follow a conventional modelling approach, with a hybrid model: a decision tree for the induction phase of treatment; and a Markov model consisting of nine health states for the maintenance phase.
- The company modelling approach and base case assumptions are mostly reasonable and transparent.
- The model is well implemented with very few errors in inputs or coding.
- The CS gives a realistic view of the limitations of the evidence base and a fair discussion of the uncertainties. The base case uses relatively conservative assumptions and decisions are based on precedent where available, albeit with a few exceptions.

Weaknesses and Areas of uncertainty

- There is heterogeneity in the company's NMAs due to differences between trials, e.g. in central versus local reading of endoscopies; differences in the durations of the induction/maintenance phases; and differences in how non-biologic failure and biologic failure are defined.
- The company excluded Asian trials from their NMAs which is inconsistent with the approach in TA547. A sensitivity analysis including Asian trials was conducted, but due to methodological problems we believe this is invalid.
- The ERG was not able to validate all of the data sources employed by the company in their NMAs.
- A major limitation of the company model structure is the omission of response and remission health states after failure of the initial treatment, implying that all patients follow a chronic active or progressive form of disease, which is inconsistent with previous NICE appraisals and unrealistic.

- In the maintenance phase, the company base case uses absolute response and remission rates from individual treatment arms for their base case analysis. We consider this a major limitation, as there is a high potential for bias due to the lack of control or adjustment for any differences between the trial populations or conduct.
- The company does not include the higher (10mg/kg) dose of infliximab in their economic analysis as it is not recommended in the SmPC. However, clinical advice to the ERG is that dose adjustment for infliximab is common in practice (and the higher dose was included in NMAs).
- The company pool standard and escalated doses in the non-biologic failure subgroup but not in the biologic failure subgroup. They argue that there is an exposure-response relationship for patients with a history of biologic failure, but not for other patients. We consider that the evidence supporting this stance is weak, as it relies on an indirect relationship (exposure-response with/without remission at maintenance baseline) and is based on observations only for ustekinumab.
- The company do not include the cost of concurrent conventional treatment alongside biologic and JAK inhibitors in their analyses. They also use a different mix of conventional treatment drugs compared with the previous NICE TA for UC, TA547. We consider the latter to be more evidence-based, as it is informed by national audit data, rather than expert judgement alone.
- The QALY decrement for serious infections appears to have been overestimated because the disutility of 0.156 is not adjusted for the expected duration of symptoms (assumed to be 28 days in TA329).
- The ERG's clinical advisors considered that the CS may overestimate utility after revision surgery, which on average is expected to be worse than remission after the first phase of surgery.
- The company's probabilistic sensitivity analysis has the following limitations and we believe the results of these analyses should be treated with caution:
    - The company model does not use Convergence Diagnosis and Output Analysis (CODA) samples to reflect uncertainty over NMA results. Thus the PSA does not reflect the joint posterior distribution, with correlations across treatments.
    - The company assign the same random numbers for health state utilities and disease management costs.

Summary of additional work undertaken by the ERG

The ERG identified 7 key aspects of the company base case with which we disagree. We address these issues in our preferred base case:

- *Model structure*: Inclusion of response and remission health states for conventional therapy after failure of the initial treatment: reflecting the chronic intermittent form of disease that some patients experience.
- *Induction:* Whilst we agree with the use of a fixed effects NMA to estimate induction response and remission rates, we found some differences on replication. We use ERG estimates in our preferred analysis.
- *Maintenance:* We prefer an NMA approach to estimation of response and remission rates for the maintenance phase, rather than the company's approach of taking remission and response data directly from individual trial treatment arms and using a pooled placebo.
- *Conventional drug mix:* Cost of CT based on results from the 2016 RCP audit of biologic treatment for IBD, as in TA547
- *Concurrent conventional treatment:* Inclusion of costs for concurrent treatment with conventional therapies alongside biologic or JAK inhibitor treatment, with costs estimated as in TA547.
- *Dose escalation with infliximab:* Same assumptions about dose escalation for infliximab as for other therapies to reflect clinical practice: assume 30% of patients on higher dose.
- *Disutility for serious infection:* Disutility adjusted for duration of symptoms, as in TA329.

The results of the ERG preferred scenarios are presented in Table 1 and Table 2. Compared to the company's base case results, collectively, our preferred assumptions in both the sub groups decrease the total costs of all the treatments and increase their total QALYs thereby decreasing the ICERs and making the treatments more cost-effective. In the full incremental analyses, all the comparators except CT remain dominated or extendedly dominated by ustekinumab. This is consistent with the company's base case. Under our preferred set of assumptions, the ICER for ustekinumab versus CT increases by £9,742 compared to that of the company's base case in the non- biologic failure sub group; and by £10,810 in the biologic failure sub group. However, we note that these results do not take account the PAS discounts for vedolizumab and tofacitinib. Final results, including the company's proposed

Commercial Medicines Unit (CMU) arrangement price for ustekinumab and all PAS discounts for the comparators, are provided in the confidential addendum to this report.

Table 1 ERG preferred scenario: Non-Biologic Failure (Company's proposed CMU arrangement price for ustekinumab and list price for comparators)

Drug	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Company base case (from ERG version of the model)				
Ustekinumab	[redacted]	[redacted]	£23,450	-
Vedolizumab	[redacted]	[redacted]	Dominated	£1,762
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£13,465
Golimumab	[redacted]	[redacted]	Dominated	£12,025
Infliximab	[redacted]	[redacted]	Dominated	£14,710
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£16,606
Adalimumab	[redacted]	[redacted]	Dominated	£18,047
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£19,146
SoC/CT	[redacted]	[redacted]	-	£23,450
ERG preferred base case				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£33,192	-
Infliximab	[redacted]	[redacted]	Dominated	£7,988
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,112
Golimumab	[redacted]	[redacted]	Dominated	£9,672
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£12,540
Adalimumab	[redacted]	[redacted]	Dominated	£23,777
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,807
SoC/CT	[redacted]	[redacted]	-	£33,192

Note: CE results for Biosimilar-Renflexis are excluded from the above table as they are similar as those for biosimilar-inflectra SoC: standard of care; CT: conventional therapy

Table 2 ERG preferred scenario: Biologic Failure (Company's proposed CMU arrangement price for ustekinumab and list price for comparators)

Treatment	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Company base case (from ERG version of the model)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£26,213	-
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£5,394
Adalimumab	[redacted]	[redacted]	Dominated	£18,210
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£19,670
SoC/CT	[redacted]	[redacted]	-	£26,213
ERG preferred base case				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£37,023	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,914

Treatment	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£28,308
SoC/CT	[redacted]	[redacted]	-	£37,023

SoC: standard of care; CT: conventional therapy

Results from the ERG preferred assumptions

The change that has the biggest impact on the cost effectiveness results is the addition of response and remission health states for conventional therapy after initial treatment failure. This decreases total costs and increases total QALYs for all treatments, largely because less time is spent with active disease after the switch to conventional treatment and the incidence of surgery is lower. The net effect of all the ERG preferred assumptions is to increase the ICERs for ustekinumab vs. CT, adalimumab and adalimumab biosimilar, and to decrease the ICERs for ustekinumab vs. other comparators.. We consider that the ERG analysis gives a more realistic representation of the clinical course of UC, with a proportion of patients continuing to experience periods of response and remission despite failure of biologic and conventional treatments. This view is supported by clinical advice to the ERG, and cohort studies.

Results from the scenario analyses conduced on the ERG base case

We performed a range of additional scenario analyses on the ERG base case. The analyses that have the greatest impact are:

- Using health state utilities estimated from the UNIFI trial. In the non-biologic failure subgroup, the ICER for ustekinumab versus CT increases to £110,391 (an increase of £77,199 from the ERG base case); and in the biologic failure subgroup it increases to £122,461 (an increase of £85,438 from the ERG base case). This is caused by the higher utility estimate for active UC ([redacted]) from UNIFI compared with the base case value (0.41) from Woehl et al. (2008).[84]

- Using the ERG ‘maintenance only NMA’. This increases the ICERs for ustekinumab versus CT to £39,903 in the non-biologic subgroup and £44,121 in the biologic failure subgroup. This is driven by different underlying assumptions in the company’s ‘1-year conditional on response NMA’ and our ‘maintenance only NMA’ about the causes of differences in placebo response rates from re-randomised studies (carry-over from induction treatment in re-randomised trials vs. other differences in the trial

populations or conduct). We consider that the truth is likely to lie somewhere between the extremes.

1 INTRODUCTION TO THE ERG REPORT

This report is a critique of the company's submission (CS) to NICE from Janssen-Cilag on the clinical effectiveness and cost effectiveness of ustekinumab (brand name Stelara) for treating patients who have moderately to severely active ulcerative colitis (UC). It identifies the strengths and weaknesses 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 manufacturer by the ERG via NICE on 9th July 2019. A response from the company via NICE was received by the ERG on 31st July 2019 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

The population in the current appraisal is described as people with moderately to severely active UC who "have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a biologic or have medical contraindications to such therapies" (CS section B.1.1 and CS Table 2). This population reflects the indication in the company's anticipated marketing authorisation as specified in the ustekinumab draft Summary of Product Characteristics SmPC[1] (CS Appendix C). Marketing authorisation is expected to be granted in August 2019.

The company's intended marketing authorisation does not mention prior JAK-inhibitor therapy. This contrasts with the NICE scope and company decision problem, which describe the population as: "people with moderately to severely active UC who are intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic therapy (a TNF-alpha inhibitor or vedolizumab), or a JAK inhibitor (tofacitinib), or conventional therapy (oral corticosteroids and/or immunomodulators)." This discrepancy appears to reflect that there is currently a lack of data on prior therapy with tofacitinib in published trials of the intervention and comparators, as discussed in section 2.3 below.

Ustekinumab is a human immunoglobulin monoclonal antibody that specifically binds to the shared p40 protein subunit of the interleukins IL-12 and IL-23, and influences inflammatory processes by down-regulating IL12/13 mediated signalling. The dose regimens in the company's anticipated marketing authorisation (CS Figure 3) are divided into a weight-based intravenous induction regimen (approximating 6 mg/kg) at week 0, followed by a fixed-dose (90 mg) subcutaneous injection maintenance regimen that starts at week 8. Clinical response is assessed around 8 weeks after the start of the maintenance regimen (i.e. by

week 16 after the start of induction). Adequate responders then continue on the maintenance therapy q12w (i.e. once every 12 weeks), inadequate responders continue on the maintenance therapy q8w (i.e. once every 8 weeks), and non-responders discontinue therapy. Patients who lose response whilst on the q12w maintenance regimen are eligible to switch to the more frequent q8w regimen, whilst patients who do not show any therapeutic benefit of the q8w regimen may be considered for discontinuation.

In the company's pivotal trial, delayed responders to ustekinumab induction therapy received the q8w regimen of ustekinumab maintenance therapy (CS Figure 10), and the company state this reflects the expected marketing authorisation (CS section B2.31). However, the SmPC[1] and the ustekinumab treatment pathway (CS Figure 3) do not mention delayed responders. The ERG's clinical experts commented that in clinical practice delayed responders to the induction therapy would receive a q8w ustekinumab maintenance regimen, as in the pivotal trial.

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

As reported in the CS, UC is a chronic inflammatory disease 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.[2,3] UC is classified according to its maximal extent seen on colonoscopy as: proctitis, where disease activity is limited to the rectum (affecting 30% to 60% of patients at diagnosis); left-sided colitis, where disease activity is limited to the left portion of the colon (from the rectum to the splenic flexure (affecting 16% to 45%); or pancolitis, where the entire colon is inflamed (affecting 14% to 47%).[4] These data are from several cohort studies and the wide variation in reported rates might in part reflect differences in how the extent of disease was measured.[4] The studies suggest that disease extends from proctitis to pancolitis in up to 28% of patients after 10 years of disease.[4]

The CS provides a generally clear and accurate overview of moderate to severe UC (CS section B.1.3), with the following provisos:

- The CS cites a study[5] which suggests that people with UC have a more than two-fold increased risk of colorectal cancer compared to the general population. However, a more recent study concluded that the overall relative risk of colorectal cancer is not significantly increased compared with the background population, although people with coexistent primary sclerosing cholangitis, extensive colitis, long duration of

disease, and those aged 60 years and above at diagnosis have a greater risk of developing colorectal cancer.[6]

- The company has misrepresented the published evidence on colonic strictures in CS section B.1.3.1. The CS states that "*in up to 11.2% of patients the disease progresses beyond the mucosal layer and leads to the formation of colonic strictures. This results in severe narrowing of the colon walls and has potential life threatening consequences*", citing reference 14 (Monstad et al.[6]). However, Monstad et al.[6] reported only that up to 11.2% of patients had benign strictures, and they did not mention any sequelae arising from these. According to the ERG's clinical experts, colonic strictures are rare and unlikely to be a problem in the population in which ustekinumab would be used (though they do raise suspicion of malignancy).
- The company have not explicitly listed the known or suspected prognostic factors for UC. According to the literature, age at onset appears to affect the disease course, which is usually more severe in people diagnosed at younger ages compared to those over age 60.[7] There is also evidence that the late proximal spread of colitis, following a period of stable proctitis or left-sided disease, carries a particularly poor prognosis.[8] Patients with pancolitis at diagnosis were found in several cohort studies to have a higher risk of surgery than those with proctitis and left-sided UC at diagnosis.[4] Disease duration and prior treatment history (including failure on conventional or biologic therapy) are likely to be prognostic of subsequent disease severity and response to therapy, and are reported in the CS. The ERG's clinical experts suggested that faecal calprotectin and Mayo endoscopy score (which are also reported in the CS) are useful prognostic markers that may be used in clinical practice.

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

Current treatments for moderately to severely active UC may be pharmacological or surgical, with all patients managed pharmacologically initially, before surgery in some cases. Surgery is usually reserved for patients who are non-responsive to the available drug treatments. Surgery may be carried out earlier if necessary, e.g. if a patient has a high risk of colorectal cancer, or requests surgery to alleviate unpleasant symptoms (such as faecal incontinence) which significantly disrupt their daily living or work.

As stated in CS section B.1.3.3, patients with moderately to severely active UC 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 TNF-alpha inhibitor, the anti-integrin agent vedolizumab), or the Janus kinase (JAK) inhibitor tofacitinib, as summarised in CS Figure 9.

In practice, clinicians often consider sequential treatments, with the choice of next line depending on treatment history, antibody tests, anticipated speed of action and safety profile. According to the ERG's clinical advisors, a common treatment pathway for patients who have failed on, or are intolerant of conventional therapy, would be to start with (biosimilar) infliximab, then escalate the dose or switch to another TNF-alpha inhibitor if antibodies are low, or alternatively try vedolizumab, tofacitinib or (if approved) ustekinumab. The experts commented that vedolizumab has a relatively slow speed of onset, while there are more safety issues to consider with tofacitinib, and clinicians are still learning about which therapies would be best for each specific patient and clinical situation. Although less common, some clinicians do consider 'step-down' treatment, starting with a more effective therapy.

The ERG's clinical experts made the following comments on how the administration of ustekinumab, if licensed, would fit with current service provision:

- The experts agreed with the company that ustekinumab would be considered as an alternative to TNF-alpha inhibitors, tofacitinib, and/or vedolizumab as indicated in CS Figure 9.
- The process of screening of patients for treatment eligibility prior to treatment with ustekinumab would likely be identical to that used for infliximab (i.e. many patients eligible to receive ustekinumab would already have been screened).
- The dosing regimen proposed by the company in their intended licence is the same as that already used in Crohn's disease.
- The initial induction infusion of ustekinumab would likely take place in a nurse-led outpatient infusion clinic (i.e. the same as for other biologic therapies).
- The subcutaneous maintenance injections of ustekinumab would be self-administered by patients at home. The clinical experts envisaged that the existing NHS medicines distribution system for home-use injections of biologic therapies would be employed. That is, a supply of injection pens would be delivered by courier to the patient's home, and the patient would be trained in the use of the injection pen during a nurse home visit (and a second visit if necessary).
- One clinical expert commented that, in their practice, patients in remission would usually see an inflammatory bowel disease nurse for routine consultations whilst

patients who are more ill would see a consultant gastroenterologist. Patients in remission would also see a consultant regularly (e.g. once every three visits).

- The start of the maintenance phase assessment requires patients to be assessed for response as close to the next dose administration date as possible. Patients would need to be evaluated around week 16 to determine whether they would receive the week 16 dose or not, whilst allowing sufficient time after the week 8 dose for this to have had an effect (CS Figure 3). Based on experience in treating Crohn's disease, this very small window is challenging to schedule in clinical practice (e.g. if patients are on holiday or a clinic is cancelled). If in doubt, patients may be given the week 16 dose pending their response assessment.

ERG conclusion: The company's description of current service provision is appropriate. Patients would typically receive one or more TNF-alpha inhibitors before receiving tofacitinib, vedolizumab and/or (if licensed) ustekinumab. However, the ways that therapies are cycled and sequenced is variable in practice, leading to heterogeneity in patients' prior treatment history in clinical trials.

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

The company's decision problem as specified in CS Table 1 is broadly consistent with the NICE scope in terms of the population, intervention, comparators and outcomes, although there are some differences as noted below.

Population: The population stated in the NICE scope is "people with moderately to severely active UC who are intolerant of, or whose disease has had an inadequate response or loss of response to previous biologic therapy (a TNF-alpha inhibitor or vedolizumab), or a JAK inhibitor (tofacitinib), or conventional therapy (oral corticosteroids and/or immunomodulators). The population specified in the decision problem is consistent with the NICE scope, with the following provisos:

- The text describing the company's intended marketing authorisation in CS section B.1.1, CS Table 2 and the draft SmPC (CS Appendix C) does not mention a JAK inhibitor and is therefore inconsistent with the NICE scope and the company's decision problem (CS Table 1). The relevant JAK inhibitor, tofacitinib, was approved by NICE relatively recently,[9] and clinical experts advising the ERG commented that they have had limited experience so far in using tofacitinib. No relevant trials identified by the company or ERG had included populations who had prior exposure to tofacitinib. Thus, the intended marketing authorisation appears to be based on the

availability of evidence, which is currently narrower than the NICE scope. This limitation is specific to considerations of treatment sequencing involving tofacitinib.

- UC can affect people of all ages and the NICE scope and decision problem do not mention any age restrictions. The CS provides effectiveness and safety data only for adults and does not explain this. However, according to the draft SMPC,[1] no data are available on the effectiveness and safety of ustekinumab in people younger than 18 years old and the intended indication is for adults.
- The NICE scope and decision problem imply that the whole population is relevant but that subgroups of people who have been previously treated with one or more biologics, and people who have not received prior biologic therapy, should also be considered if the evidence allows. The CS reports both the whole (intention to treat) population (ITT) and pre-specified subgroup analyses for the company's pivotal ustekinumab trial, but only the subgroup analyses in their network meta-analyses. The ERG agrees that the company's focus on the subgroups is reasonable, as this is consistent TA547 (tofacitinib)[9] where the NICE committee recommendations were based on prior treatment history subgroups rather than the whole population. Subgroup statistical power is not reported; subgroup sample sizes are relatively large for induction, but smaller for maintenance (see section 3.1.6.3).
- The prior treatment experience subgroups reported in the CS are defined differently to those in the NICE scope (the company does not comment on this), but we believe that the NICE and company subgroup definitions are broadly comparable (see Table 3).

Table 3 Prior treatment experience subgroups

Subgroup specified in the NICE scope	ERG comments
People who have not received prior biologic therapy	The NICE subgroup matches the majority (94.3%) of people in the company's "non-biologic failure" subgroup in the pivotal UNIFI trial, but the company's subgroup also includes a small proportion of people (5.7%) who were biologic-exposed and therefore outside of the NICE subgroup (CS Appendix Figures 66 and 72). The non-biologic failure subgroup is defined in the CS as people who were biologic-naïve or exposed to biologic therapy but did <u>not</u> demonstrate an inadequate response or intolerance (CS section B.2.3.2.1). The ERG is unclear why the 5.7% of patients in this

	subgroup who were exposed to biologic therapy but did not demonstrate biologic failure or intolerance would be eligible for ustekinumab; this is not explained in the CS or CSRs.[10,11]
People who have previously been treated with one or more biologics	The NICE subgroup matches all people in the company's subgroup "biologic failure", plus a further 5.7% of people in the company's subgroup "non-biologic failure" (see above description of the non-biological failure subgroup). The biologic failure subgroup is defined in the CS as people who had received treatment with at least one TNF antagonist or vedolizumab at a dose approved for UC and either did not respond, or lost an initial response, or were intolerant to the medication (CS section B.2.3.2.1). Note that tofacitinib is not included in the definition since it was not licensed at the time the company's pivotal trial was conducted.

Intervention: Ustekinumab (as per the NICE scope).

Comparators: The comparators in the NICE scope are adalimumab, golimumab, infliximab, (TNF-alpha inhibitors), vedolizumab (an anti-integrin), tofacitinib (a JAK inhibitor), and conventional therapies (oral corticosteroids and/or immunomodulators), without biological treatments. The comparators included in the CS are consistent with the NICE scope. The company state in CS Appendix section D.1.1.1.2 that conventional therapy was not included as a comparator in the decision problem because it was assumed that it makes up the background treatment received in clinical trials, for both placebo and active arms. The ERG agrees that this approach is appropriate, i.e. placebo reflects conventional therapy in clinical effectiveness trials. Conventional therapy is explicitly modelled as a comparator in the company's economic analysis.

Outcomes: The outcomes specified in the NICE scope are: mortality; measures of disease activity; rates of and duration of response, relapse and remission; rates of hospitalisation; rates of surgical intervention; endoscopic healing; mucosal healing (combined endoscopic and histological healing); corticosteroid-free remission; adverse effects of treatment; and health-related quality of life (HRQoL). The outcomes reported in the CS are consistent with the NICE scope apart from the following differences:

- The CS does not include relapse rate as an outcome in the clinical effectiveness evidence synthesis. Relapse is modelled in the company's economic analysis as loss of response.
- The CS states that disease activity is assessed in clinical trials according to the Mayo score or Partial Mayo score (CS section B.1.3 and CS Table 6). Outcomes based on Mayo scores (i.e. clinical remission and response) are reported in the CS, but not the underlying Mayo or Partial Mayo scores.
- Apart from relapse, all of the listed outcomes are reported in the CS for the company's pivotal clinical trial. However, only a subset of the outcomes were included in the company's clinical effectiveness network meta-analyses (NMAs). These are: clinical response; clinical remission; mucosal healing; and adverse events (all adverse events, serious adverse events, all infections, serious infections, and discontinuations due to adverse events). Of these, clinical response, clinical remission and serious infections are used in the company's cost-effectiveness model.

Equality: The company have not identified any equality issues. The ERG is not aware of any potential limitations in how particular groups of people could access and be treated with ustekinumab.

ERG conclusion: The company's decision problem broadly reflects the NICE scope, with only minor deviations. The population, intervention, comparators and outcomes specified in the decision problem are appropriate for NHS practice.

3 CLINICAL EFFECTIVENESS

3.1 Critique of the company's approach to systematic review

3.1.1 Search strategy

The company conducted searches for the following reviews:

[a] Clinical effectiveness (CS Appendix D1.1)

[b] Economic evaluations (CS Appendix G1.1)

[c] HRQoL, (CS Appendix H1.1)

[d] Costs and resources (CS Appendix I1.1)

The CS Appendices report that search [a] was initially run in August 2018 and searches [b] to [d] were initially run in October 2017. All searches were then updated in January 2019 and March 2019. The overall period covered in each search is January 2006 to March 2019. The results of each search are reported in the CS Appendices separately for each of the three search dates, with a separate PRISMA flow diagram provided for each date.

The search strategies are not structured as efficiently as they could be, but overall appear to be fit for purpose. For the Embase searches there is a discrepancy between the number of hits reported in the search strategies and the number of hits reported in the PRISMA diagrams. This applies to the January 2019 and March 2019 update for reviews [b] to [c] and the January 2019 update for review [d].

The CS Appendices report identical search strategies and search results for review [b] (cost-effectiveness) and for review [d] (costs and resources). The PRISMA flow charts for reviews [b] and [d] are also very similar. It appears that the company has used the same search strategies and search results for these two reviews but applied different study selection criteria in each review, although the CS is not explicit about this.

Given that the searches were reasonably up to date when the CS was received by the ERG (3 months after the searches were conducted) we have not rerun the full search strategies. Instead, we conducted targeted searches in Google Scholar limited to studies published during 2018-2019 as a check for any key study publications since the last NICE technology appraisal of a relevant comparator (TA547, tofacitinib). We conducted broad searches for "ulcerative colitis" combined with the name of each comparator drug. For each search we checked the first 200 hits sorted by relevance (a pilot of more extensive checking did not yield relevant articles, suggesting 200 hits per therapy would be a reasonable pragmatic

number to check). We also checked the studies included in relevant systematic reviews and meta-analyses[12-16] and technology appraisals.[9,17,18] We identified several new abstracts reporting on the UNIFI trial[19-24] and one additional abstract reporting on the VARSITY trial[25] as well as a relevant trial (Mshimesh 2017[26]) that was missed by the company's clinical effectiveness searches but identified in their HRQoL searches (see Appendix 1). We did not identify any key trials that are not reported in the CS.

> **ERG conclusion:** The company's searches were generally up-to-date and broadly appear to be fit for purpose, though with some discrepancies. The ERG and clinical expert advisors did not identify any key missing trials.

3.1.2 Inclusion/exclusion criteria used in the study selection

Eligibility assessment for clinical effectiveness review

The eligibility criteria for the company's clinical effectiveness review are stated in CS Appendix Table 14 (outcome criteria are given in CS Appendix Table 9). These are consistent with the NICE scope and therefore appear appropriate, with the following provisos:

- Endoscopic healing, which is specified as an outcome in the NICE scope, is not listed in the eligibility criteria, although the criteria do include mucosal healing, which is defined as a combination of endoscopic and histological healing.
- The NICE scope specifies HRQoL as an outcome. The company has specifically mentioned the Inflammatory Bowel Disease Questionnaire (IBDQ) in the inclusion criteria but has not named any other HRQoL measures such as other disease-specific measures or EQ-5D. (NB The company does report EQ-5D results for their pivotal trial in CS sections B.2.6.13 and B2.6.2.4 and clarification question response A9).

The reasons for excluding studies at full-text screening are summarised in the PRISMA flow diagrams in CS Appendix Figures 1-3 and listed in CS Appendix Table 31 and appear appropriate.

The CS reports that, following the selection process, 48 publications were identified, referring to 21 clinical trials (CS section B.2.9.1). We note that the PRISMA flow diagrams (CS Figure 25 and CS Appendix Figures 1-3) refer to the number publications included

rather than the number of studies as stated. The identified trials are listed in CS Appendix Tables 15 and 16.

Two trials that the company identified in searches, but excluded (UC-SUCCESS[27] and Mshimesh 2017[26]) appear relevant to the decision problem but are missing from the list of 21 included studies. These trials were excluded by the company without a clear explanation, but we believe that the exclusion of these trials is likely to be inconsequential (explained in Appendix 1). CS Appendix Table 29 lists a reference by Marano 2018 as reporting on the UNIFI trial but this is not included in the reference list and the ERG has been unable to locate it.

The company state that two of their 21 identified trials (Silva 2017[28] and Kobayashi 2019[29]) were excluded for specific reasons as stated in CS section D1.1.6.1. We agree that the reasons for exclusion are appropriate (Appendix 1). The remaining 19 trials were included in the company's clinical effectiveness review, permitting the following seven treatment comparisons:

- Adalimumab versus placebo (NCT00853099, ULTRA1, ULTRA2)
- Adalimumab versus vedolizumab (VARSITY)
- Golimumab versus placebo (PURSUIT-J, PURSUIT-M, PURSUIT-SC)
- Infliximab versus placebo (ACT1, ACT2, Japic CTI-060298, Jiang 2015, Probert 2003)
- Tofacitinub versus placebo OCTAVE 1, OCTAVE 2, OCTAVE Sustain, NCT00787202)
- Ustekinumab versus placebo (UNIFI – the company's pivotal trial)
- Vedolizumab versus placebo (GEMINI 1, NCT02039505)

NB the company refers to the "Japis CTI060297" trial, but the correct name according to the study publication is Japic CTI-060298.

There are a number of referencing discrepancies in the CS and Appendices, which collectively make the matching of publications to studies difficult to follow. We have cross-checked the references, and we provide a list of the publications that report relevant outcomes for the induction and maintenance phases of each trial in Appendix 2.

3.1.3 Identified studies

As described above, the company's clinical effectiveness review identified 19 RCTs of which one (UNIFI) investigated the clinical effectiveness of ustekinumab and 18 investigated the

clinical effectiveness of the comparators (adalimumab, golimumab, infliximab, tofacitinib, vedolizumab). In this section we summarise the key characteristics of the UNIFI trial; key features of the comparator trials that are relevant to the company's meta-analyses are discussed in section 3.1.7 below.

The company's pivotal trial, UNIFI (NCT02407236), compared ustekinumab against placebo for treating patients with moderately to severely active UC. The trial had an induction treatment phase (the 'Induction Study' part of the trial) and a maintenance treatment phase (the 'Maintenance Study'). The company provided NICE and the ERG with two confidential clinical study reports (CSRs) of the trial, describing the Induction Study[10] and the Maintenance Study.[11] The ERG additionally identified a number of abstracts reporting the trial's findings that were published after the company's searches were carried out (see section 3.1.1). As well as reporting adverse events in the UNIFI trial, the CS presents data on the long-term safety of ustekinumab from other studies of its use in psoriasis, psoriatic arthritis and Crohn's disease, as supporting evidence.[30-32]

3.1.3.1 UNIFI trial information provided by the company

Detailed information on the UNIFI trial is reported in the CS and CSRs, including the trial design, patient population, inclusion and exclusion criteria, interventions and comparators, the outcomes assessed and pre-planned subgroup analyses. As described in more detail below, UNIFI had a "re-randomised" design, in which patients were initially randomised to induction ustekinumab therapy or induction placebo. Those who met specified response criteria at the end of the induction phase were either re-randomised to receive maintenance ustekinumab therapy or maintenance placebo, or were allocated to non-randomised maintenance therapy or maintenance placebo groups. Participant flow diagrams are provided in CS Appendix Figures 50, 51 and 52 for the induction phase, randomised maintenance arms, and non-randomised groups respectively. The flow diagrams show the numbers of participants who terminated study participation prior to the end of the induction and maintenance assessments and who discontinued treatment during the maintenance phase, but do not specify the reasons why. Reasons for discontinuation are reported in the maintenance study CSR,[11] and the company subsequently provided further details indicating that the most common reasons for study termination in all groups were withdrawal of consent and adverse events (clarification question response A2). The number of Induction study participants who completed a safety follow-up is also provided in CS Appendix Figure 50. According to the CSR[10] this is the number of participants who discontinued treatment, but who completed the induction study and the safety follow-up around 20 weeks after

receiving their last dose of study treatment. The statistical analyses conducted in the UNIFI trial are summarised in CS Section B.2.4 which refers to CS Appendix L2 for further details, but this is missing from the submission and was provided by the company in clarification question response A3. Details of the statistical power and sample size calculations, definitions of study populations, including the intention to treat (ITT) population, and how missing data were handled are available in the induction and maintenance CSRs.[10,11]

3.1.3.2 Overview of the UNIFI trial

We have summarised the characteristics of the UNIFI trial in Table 4, including the ustekinumab dose regimens used in the induction and maintenance treatment phases. A detailed overview of the "re-randomisation" trial design is shown in CS Figure 10 (reproduced in Figure 1 below). The participants were first randomised to one of three induction treatment arms (fixed-dose ustekinumab 130mg IV, weight-based ustekinumab approximating 6mg/kg IV [the dose in the proposed marketing authorisation], or placebo). At the end of the induction period (8 weeks), responders to ustekinumab, and patients who had not responded to placebo induction treatment at 8 weeks but subsequently responded to ustekinumab induction treatment at 16 weeks, were re-randomised to maintenance treatment with either ustekinumab 90 mg SC q12w, ustekinumab 90 mg SC q8w or a maintenance placebo. Randomisation was stratified by biologic failure status (yes or no) and region (Eastern Europe, Asia or the rest of the world). The primary outcome was clinical remission at week 8 of the Induction Study and week 44 of the Maintenance Study.

Table 4 Summary of the UNIFI trial

Trial overview	Intervention	Comparator
Design: Phase III, double-blind, multicentre re-randomisation RCT with additional non-randomised groups. **Patient population:** Adults who had had a diagnosis of UC for at least 3 months prior to screening, and who had moderately to severely active disease (defined as a Mayo score of 6-12, including an endoscopy score of ≤ 2) at baseline. All patients had had an inadequate response to or failure to tolerate non-biologic or biologic treatment.	**Induction Study (8 weeks) – participants were randomised in a 1:1:1 ratio:**	
	Fixed-dose ustekinumab 130mg IV (N=320) Weight-based ustekinumab (~6 mg/kg IV) (N = 322): • 260 mg if ≤ 55 kg) • 390 mg if > 55 kg but ≤ 85 kg • 520 mg if < 85 kg	Placebo IV (N = 319)
	Maintenance Study (44 weeks) – responders to ustekinumab and patients who had not	

<table>
<tr><td rowspan="4">Sample size:
N randomised to induction treatment = 961 (including [redacted] participants from the UK[10])
N entering maintenance = 783
N re-randomised at maintenance = 523

Length of follow-up: Same as length of treatment periods: outcome assessment took place at week 8 of the induction period and week 44 of the maintenance period.

Concomitant medications for UC permitted during the induction and maintenance studies: Oral corticosteroids, oral 5-aminosaliclaye compounds, or the immunomodulators 6-mercaptopurine, azathioprine or methotrexate. To be permitted, all had to be maintained at a stable dose until the end of induction treatment. If patients were receiving oral corticosteroids on entry to the maintenance study, tapering was started</td><td colspan="2">responded to placebo induction treatment but subsequently responded to ustekinumab induction treatment were re-randomised in a 1:1:1 ratio:</td></tr>
<tr><td>Ustekinumab 90 mg SC every 12 weeks (N= 172)

Ustekinumab 90 mg SC every 8 weeks (N= 176)</td><td>Placebo SC (N = 175)</td></tr>
<tr><td colspan="2">Non-randomised maintenance groups:</td></tr>
<tr><td colspan="2">Participants who had responded to placebo at week 8 of the induction period were not re-randomised but instead continued to receive placebo as maintenance treatment.
‘Delayed responders’ to ustekinumab entered a non-randomised group for maintenance treatment, in which they received ustekinumab 90 mg SC q8W. See Figure 1 for the full details of the study design.</td></tr>
<tr><td colspan="3">Sources: CS section B.2 summary; CS section B.2.3, CS Tables 6 and 8, CS Figure 10, CS Appendices section D4.2 and Figures 50 and 51, and Induction Study CSR.[10]</td></tr>
</table>

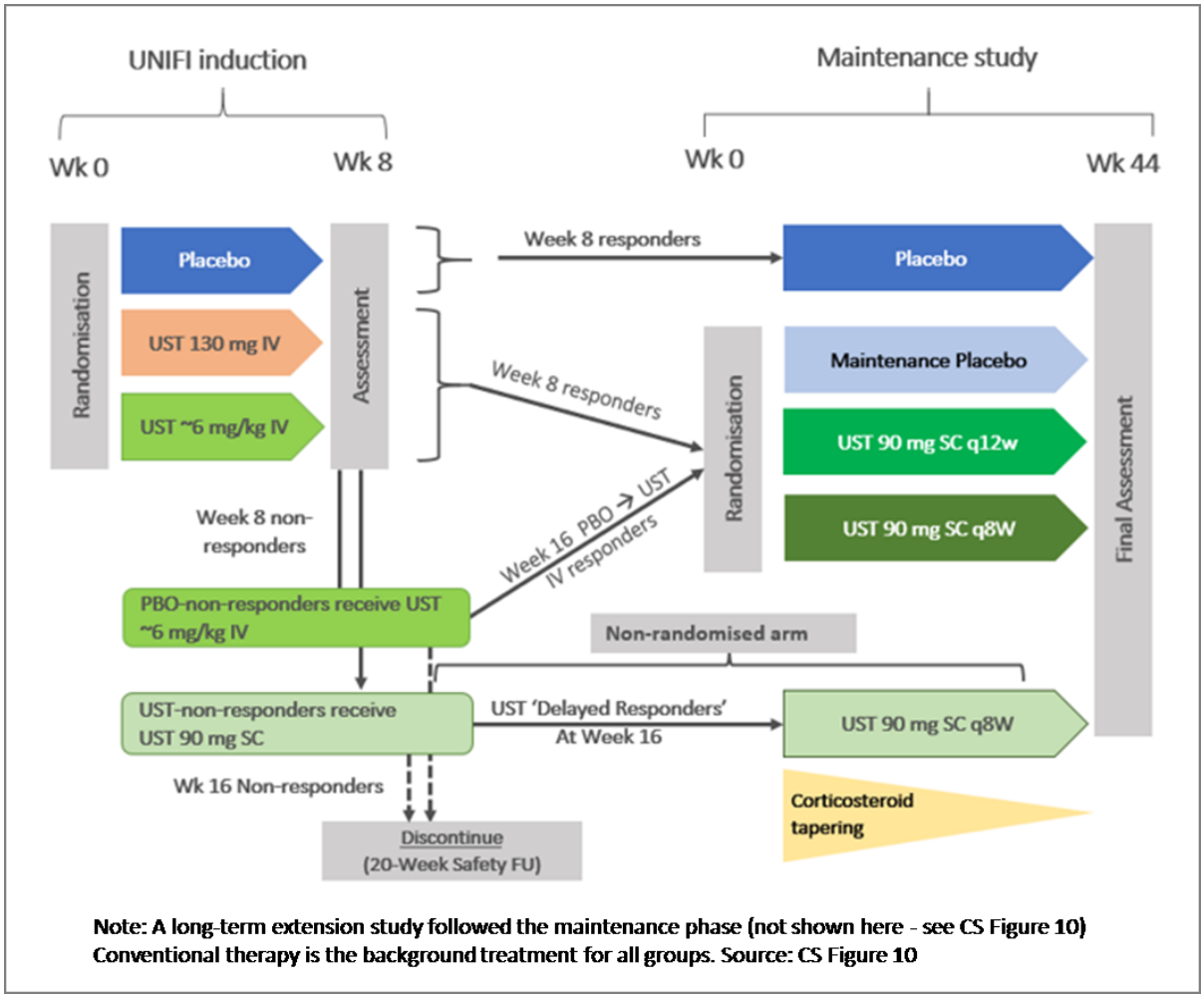


Figure 1 Overview of the UNIFI trial design

3.1.3.3 Overview of how the UNIFI trial addresses NICE's final scope, the decision problem and the draft SmPC

The UNIFI trial patient population matches that specified by NICE in the final scope, the company's decision problem and the draft SmPC (provided in CS Appendix C). The ustekinumab weight-based 6 mg/kg IV induction intervention matches the posology stated in the draft SmPC,[1] but the SmPC does not specify a fixed-dose 130 mg IV induction regimen, and therefore efficacy and safety results from this arm of the Induction Study are not directly relevant to the current appraisal. In the trial, participants who received the 130mg dose were re-randomised at maintenance along with those who had received the weight-range-based dose approximating 6mg/kg, which ranged from 260mg to 520 mg. This means some of the re-randomised patients had been under-dosed at induction, compared to the posology in the draft SmPC and therefore the expected use of ustekinumab in clinical practice. The ERG's clinical experts agreed this would have a conservative impact on the treatment effects found for ustekinumab in the trial.

A draft SmPC for the maintenance regimen of ustekinumab is not available. However, CS Table 2 suggests that the ustekinumab maintenance treatment strategy for UC would be the same as that employed for Crohn's disease.[33] That is, a 90 mg SC dose of ustekinumab would be administered at week 8 after the IV induction dose, and subsequent 90 mg SC doses are then recommended every 12 weeks (q12w). Patients who have not responded 8 weeks after the first subcutaneous dose may receive another dose (i.e. at 16 weeks) to allow for delayed response. Those who lose response on the q12w regimen may be escalated to a q8w regimen. After this, clinicians may use their judgement to determine if a patient should continue on the q12w or q8w regimen. The maintenance dosing pattern in the UNIFI trial does not follow this expected use in clinical practice. In practice, this dose may be more likely to be used in patients who have lost response to the q12w regimen, while in the trial, participants treated with this regimen were randomised to it following responding to induction treatment. This may mean that the efficacy seen in clinical practice with the q8w regimen will differ to that found in the trial, as it is likely to be used with a different subgroup of patients.

3.1.3.4 Participant baseline characteristics

The CS provides a summary of the baseline characteristics of the participants randomised to the induction and maintenance studies in CS Table 10. A table of trial baseline characteristics, Table "TSIDEM02", is missing from the versions of the induction and maintenance CSRs provided by the company and was provided in response to clarification question A1. Table TSIDEM02 reports means for C-reactive protein, faecal lactoferrin and faecal calprotectin concentrations (CS Table 10 reports only medians); and reports baseline clinical remission, endoscopic healing, and IBDQ data that are missing from CS Table 10. We have summarised the key participant baseline characteristics of the participants in the UNIFI trial in Table 5. Baseline characteristics for both the randomised and non-randomised maintenance arms of UNIFI are reported in Table TSIDME02.

The participant baseline characteristics presented in the CS are generally well balanced across the treatment arms in both the Induction and Maintenance Studies, with a few exceptions (highlighted in bold in Table 5). Proportionally more participants treated with ustekinumab ~6 mg/kg had an endoscopy score of 3 (indicative of severe disease) compared with those treated with placebo at baseline in the Induction Study. In the Maintenance Study, proportionally more participants treated with ustekinumab 90 mg q8w had abnormal faecal calprotectin and abnormal faecal lactoferrin than those treated with placebo. The ustekinumab q8w group also had higher median concentrations of these two

markers than the placebo group. These differences are noted by the company in the CS. They suggest that the differences indicate participants treated with ustekinumab ~6 mg/kg at induction and ustekinumab 90 mg q8w at maintenance had a higher inflammatory burden. The company also state that "These higher inflammatory markers indicate a more difficult and harder to treat population in the ustekinumab arm than the maintenance placebo arm" (CS section B.2.3.3). Clinical experts advising the ERG commented that faecal calprotectin is a good marker of inflammation and is a key prognostic factor in UC, but that higher levels of this marker do not necessarily mean patients are harder to treat. There are some differences in C-reactive protein (CRP) evident between the groups in Table 5 but CRP is a nonspecific inflammatory marker that is not clinically meaningful or prognostic in UC as it can vary considerably among patients who have a similar extent of inflammation. The clinical experts felt that the key prognostic factors for UC are covered in CS Table 10, with the most important being faecal calprotectin concentration and Mayo endoscopy subscore.

The baseline characteristics of the non-randomised delayed responders maintenance arm (Figure 1) were similar to those of participants in the randomised maintenance arms, except that proportionally fewer were in clinical remission and proportionally fewer demonstrated endoscopic healing (clarification questions response Appendix Table 2, Table TSIDEM02).

Table 5 Key baseline characteristics of participants in the UNIFI trial

Induction Study	**Placebo (N=319)**	**UST 130 mg (N=320)**	**UST ~6 mg/kg (N=322)**
Male sex, %	61.8	59.4	60.6%
White race, %	77.7	74.7	75.5
Age, years – mean (SD)	41.2 (13.50)	42.2 (13.94)	41.7 (13.67)
Duration of disease, years – mean (SD)	8.01 (7.19)	8.13 (7.18)	8.17 (7.82)
Moderate UC (6≤ Mayo score ≤10), %	82.4	84.7	86.0 (N=321)
Severe UC (Mayo score >10), %	16.9	15.0	14.0 (n=321)
Endoscopy subscore of 3, %[a]	**67.7** [a]	**65.9** [a]	**74.8** [a]
Biologic failure status – yes, %	50.5	51.3	51.6
Biologic failure status – no, %	49.5	48.8	48.4
Maintenance Study	**Placebo (N=175)**	**UST q12w (N=172)**	**UST q8w (N=176)**
Male sex, %	61.1	55.8	53.4

White race, %	71.4	78.5	72.2
Age, years – mean (SD)	42.0 (13.85)	40.7 (13.47)	39.5 (13.32)
Abnormal CRP (>3 mg/L), %	**34.5** (n=174)	**28.8** (n=170)	**36.9**
Faecal lactoferrin, µg/g, mean (SD)	**142 (229)** (n=167)	**125 (200)** (n=161)	**147 (218)** (n=163)
Abnormal faecal lactoferrin (>7.24 µg/g), %	**73.1** (n=167)	**72.7** (n=161)	**82.2** (n=163)
Faecal calprotectin, µg/g, mean (SD)	**909 (1842)** (n=168)	**945 (1423)** (n=160)	**1147 (2083)** (n=161)
Abnormal faecal calprotectin (> 250 mg/kg), %	**55.4** (n=168)	**60.0** (n=160)	**64.0** (n=161)
Corticosteroid use, %	**54.3** [a]	**48.3** [a]	**54.0** [a]

Source: CS section B.2.3.3, CS Table 10 and Table TSIDEM02 in clarification response A1
[a] number of participants not reported

The ERG's clinical experts confirmed that the UNIFI trial population matches the patients who would likely be seen in NHS clinical practice. The average disease duration of around eight years implies that the trial participants would be less responsive to treatment than those newly-diagnosed, but is reflective of the NHS population.

3.1.3.5 Ongoing studies

In CS Section B.2.11, the company identifies one ongoing study, which is an extension of the UNIFI trial, stating that "After completion of the maintenance phase, eligible patients are being followed for an additional three years in a long-term extension, under the same protocol." The CS says that the methods of the long-term extension study are reported in Appendix D. However, no methods or interim results from this study are reported in the CS or Appendices. The ERG's searches (section 3.1.1) did not identify any other ongoing studies of the clinical effectiveness or safety of ustekinumab in moderately to severely active UC.

> **ERG conclusion:** A single multi-national, placebo-controlled, RCT with a re-randomised design (UNIFI trial) has investigated the clinical effectiveness ustekinumab in the population and indication of interest. The trial design covers both the induction and maintenance phases of therapy and the population and design are generally applicable to NHS practice. Exceptions are that the lower of the two ustekinumab induction doses is not relevant to clinical practice, and the patient

population who received maintenance ustekinumab q8w may not fully represent those who would receive it in clinical practice.

3.1.4 Approach to validity assessment

The CS includes a tabulated quality (risk of bias) assessment of the UNIFI trial (CS Table 11; CS section B.2.5). The company do not report how many reviewers conducted the assessment or provide a rationale for their judgements. However, the ERG agrees with the company's assessment (Table 6).

Table 6 Company and ERG assessment of trial quality

NICE assessment criteria (applied to UNIFI Induction and Maintenance studies)	CS judgement	ERG judgement
1. Was the method used to generate random allocations adequate?	Yes	Yes (a computer-generated randomisation schedule was used)
2. Was the allocation adequately concealed?	Yes	Yes (performed centrally)
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Yes	Yes (some baseline imbalances in prognostic factors noted – see Section 3.1.3 – but ERG's clinical experts felt that these were not sufficient to introduce bias)
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Yes	Yes (confirmed in clarification response A7)
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No	No (for both the Induction and Maintenance Studies)
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No (results are reported either in the CS or the CSRs[10,11] for the key outcomes)
7. Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes and yes (ERG determined from information in the CSRs[10,11] that the 'primary efficacy analysis set' presented in the CS is equivalent to the ITT population; conservative methods were used to account for missing data; see Section 3.1.6).

ERG conclusion. The CS reports an appropriate assessment of the risks of bias in the UNIFI trial and we agree with their assessment. Overall, the company and ERG agree that the trial is at low risks of performance, detection, selection, reporting and attrition biases for the primary outcome.

3.1.5 Outcome selection

The outcomes in the CS are consistent with those specified in the NICE scope and the company's decision problem (CS section 2.3) and are appropriate for assessing the efficacy of treatments for UC. The CS reports UNIFI trial results for all outcomes specified in the NICE scope except for rates of and duration of relapse. We checked the trial CSRs,[10,11] and the rate of relapse outcome does not appear to have been measured in the UNIFI trial. However, relapse is modelled in the company's economic analysis as loss of response during maintenance treatment (see Section 4.3.4.2) – we discuss this further below under 'loss of response'. No clinical efficacy data were available for this outcome in the CS.

Clinical response, clinical remission, endoscopic healing, mucosal healing and disease activity are based on the Mayo Index, which is scored 0 (normal) to 12 (severe disease) based on four subscales, each scored 0 to 3 (Table 7). The definitions of response and remission in the CS (see Table 8) are consistent with those employed in recent NICE technology appraisals and clinical experts advising the ERG confirmed they are clinically appropriate.

Table 7 Mayo Index subscales and scores

Score Subscale	0	1	2	3
Stool frequency	Normal	1-2/day more than normal	3-4/day more than normal	>4/day more than normal
Rectal bleeding	None	Streaks	Obvious	Mostly blood
Mucosal appearance at endoscopy	Normal or inactive disease	Mild disease (erythema, decreased vascular pattern, mild friability)	Moderate disease (marked erythema, absent vascular pattern, friability, erosions)	Severe disease (spontaneous bleeding, ulceration)
Physician's global assessment of disease activity	Normal	Mild	Moderate	Severe
Source: CS Table 4 with additional explanation added by ERG from https://www.mdcalc.com/mayo-score-disease-activity-index-dai-ulcerative-colitis				

The company provides definitions of some of the trial efficacy outcomes in CS Table 9 (reproduced in Table 8 below, with some adaptations). Rates of response and remission,

and HRQoL outcomes (specifically, EQ-5D-5L data directly collected from the UNIFI trial) inform the company's economic model. We did not identify any issues with how any of the other clinical effectiveness outcomes had been defined or measured.

Table 8 Definitions of clinical effectiveness outcomes used in the UNIFI trial

Outcome	Definition
Clinical remission – global definition	Mayo score ≤2 points, with no individual subscore >1
Clinical response	A decrease from induction baseline in the Mayo score by ≥30% and ≥3 points, with either a decrease in the rectal bleeding subscore ≥1 or a rectal bleeding subscore of 0 or 1.
Endoscopic healing	Mayo endoscopy subscore of 0 or 1.
Histologic healing	Based on features of the Geboes score,[34] defined as neutrophil infiltrations in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue.
Mucosal healing	Both endoscopic healing (Mayo endoscopy subscore of 0) and histologic healing (neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue).
Disease activity	Based on the Mayo score and Partial Mayo score (CS Table 6). The Partial Mayo score uses the three non-invasive components of the full Mayo Score (stool frequency, rectal bleeding and physician's global assessment) and has a possible score ranging from 0 to 9.
Source: CS Tables 6, 7 and 9	

3.1.5.1 Rates of and duration of response and remission

The CS states that the primary outcome in the UNIFI trial was clinical remission at week 8 of the Induction Study and week 44 of the Maintenance Study. Secondary outcomes included (among others listed in CS Table 8) clinical response at week 8 of the Induction Study and maintenance of clinical response through to week 44 of the Maintenance Study.

Two definitions of remission were employed in the UNIFI trial: the "global" definition and "US" definition. The global definition (Table 8) is consistent with that used in other trials and is the definition applied in the company's NMAs. The US definition (which is defined in CS Appendix section D1.1.8.1) is not used in any of the NMAs.

EMA guidelines[35] on the development of medicinal products to treat UC recommend that endoscopic assessments of disease activity in trials should ideally be independently verified, preferably by central assessment of the endoscopic examinations. CS Table 7 confirms that clinical remission outcomes at week 8 of induction and week 44 of maintenance in the UNIFI trial were based on centrally read endoscopic subscores, which is in line with the guidance.

However, CS Appendix D1.1.8.1 states that local endoscopic readings were also taken during the UNIFI trial and it was these locally-read endoscopy scores that were used for efficacy endpoints in the company's NMAs. This was to ensure comparability of the methods across trials included in the NMAs, since all but one of the other trials included in the NMAs employed only locally-read endoscopies (CS Appendix Table 23).

3.1.5.2 Loss of response

We note that there is no consensus in the literature about how secondary loss of response is defined (that is, loss of response during maintenance treatment), but commonly an assessment of this is based on Mayo scores in UC: if patients experience substantial improvements in these scores but then experience clinical relapse, they would be classified as having had a secondary loss of response to treatment.[36] Based on this, we suggest that loss of response may adequately reflect relapse. In the model base case, loss of response was calculated, using UNIFI trial data, as: "1 minus the ratio of the proportion of patients responding to treatment at the end of the induction phase and the proportion of patients responding to treatment at the end of the maintenance phase of the trials (among the intention-to-treat [ITT] population) and adjusting this for the length of the maintenance period" (CS section B.3.3.1.2.1).

3.1.5.3 Health-related quality of life

Health-related quality of life was measured in the UNIFI trial primarily using the IBDQ, SF-36 and EQ-5D (5L version) (CS section B.3.4.1). A further patient-reported outcome, The Work Productivity and Activity Impairment Questionnaire-General Health (WPAI-GH), is also briefly mentioned in the CS. The IBDQ and SF-36 have been validated in populations with UC.[37,38] The IBDQ, SF-36 and EQ-5D were also the key patient-reported HRQoL instruments employed in the recent technology appraisals TA342 (vedolizumab) and TA547 (tofacitinib). The IBDQ evaluates disease-specific HRQoL across 4 dimensional scores: bowel, systemic, social, and emotional. Scores range from 32 to 224, with higher scores indicating better HRQoL.

The CS provides minimum thresholds for clinically meaningful changes in the IBDQ and SF-36 measures (i.e. changes that are meaningful to patients and for which a clinician would consider a change in the patient's care):

- IBDQ: A widely used threshold for clinically meaningful change in the total IBDQ score, that has been used in some trials of biologics in UC, is >16 points. However, a recent study has established that a more stringent >20 point change in IBDQ score is

an appropriate threshold for clinically meaningful improvement in UC[37] (clarification question response A6). The CS reports IBDQ results for both thresholds.

- SF-36: The CS states that a ≥5-point change in the SF-36 Physical Component Summary (PCS) and Mental Component Summary (MCS) subscales indicates a clinically important change, but does not provide a reference to justify this (CS Section B.2.6.1.4). We note that the threshold for a clinically important change in UC has previously been established as >3.1 for the PCS and >3.8 for the MCS[38] and therefore the company's threshold of 5 appears reasonable.

The trial EQ-5D data are used to estimate utilities for some of the health states in a scenario analysis in the economic model (but were not used in the base case). IBDQ, SF-36 and WPAI-GH data are not used in the company's economic model. However, we summarise results from these instruments alongside those of the EQ-5D in section 3.3.4 below for comparison, as a check for consistency among these disease-specific (IBDQ) and generic (SF-36, EQ-5D) HRQoL measures. Very little information is reported in the CS for the WPAI-GH, so we summarise this only briefly in section 3.3.4.

ERG conclusion. The company selected and presented appropriate outcomes in the CS that addressed those specified in NICE's final scope and the company's decision problem and provided results from the UNIFI trial for these in the CS or accompanying submission documents. The only NICE scoped outcome for which there was no trial evidence available was rate of and duration of relapse.

3.1.6 Approach to trial statistics

When reporting results, the company provides the unit of measurement, size of effect, measures of variance (where applicable; an exception is that ranges were not provided where median results are reported for the EQ-5D HRQoL findings from the induction phase) and the numbers included in the analyses, with some exceptions; see 'Analysis populations' below.

Most of the trial results were presented in terms of the proportion (%) of participants in each study group in the Induction and Maintenance Studies achieving a particular outcome. In the statistical analyses, each ustekinumab dose group in the induction and maintenance studies is compared to placebo. The ustekinumab groups are not compared to each other. The statistical analyses were stratified by biologic failure status (yes or no) and region (Eastern Europe, Asia or rest of world). No interim data are presented in the CS.

3.1.6.1 Power calculations

The induction CSR[10] reports the sample size required to provide statistical power of 90% to detect a significant difference for the primary outcome of clinical remission at week 8 between the ustekinumab and placebo groups using a chi-squared test. The sample size calculations were different for the US and global definitions of clinical remission, to support submissions in the US and elsewhere, although both the global and the US definitions of clinical remission were applied to all trial participants, regardless of location. The power calculations assumed the clinical remission rate was 12% (US definition) or 7% (global definition) in the placebo group; and 25% (US definition) or 19% (global definition) in each ustekinumab group. This gave a required sample size of 220 subjects per arm (660 in total) based on the US definition; and 135 subjects per arm (405 in total) based on the global definition. In practice, the actual sample size (N=961) exceeds these numbers, and the observed differences in clinical remission rates between arms are smaller than those assumed in the power calculations. We therefore believe that the UNIFI induction phase analyses for clinical remission in the whole (ITT) population are adequately powered.

The maintenance CSR[11] reports the sample size required to provide statistical power of 90% to detect a significant difference for the primary outcome of clinical remission at week 44 of the Maintenance Study between the ustekinumab 90mg q8w and placebo groups using a chi-squared test. Based on clinical remission rates in two other similarly designed trials of golimumab and vedolizumab for UC, the company assumed clinical remission at week 44 would be 40% in the the US and global definitions). This gave a required sample size of 109 subjects per arm (327 in total). In practice, the actual sample size (N=523) exceeds this number. We therefore believe that the UNIFI maintenance phase analyses for clinical remission in the whole (ITT) population are adequately powered.

The CS and CSRs do not report any power calculations for the non-biologic failure and biologic failure subgroups (section 3.1.6.3 below).

3.1.6.2 Analysis populations

The CS refers to Appendix L2 for further information about the study groups and data handling. This appendix is missing from the submission but was provided by the company as Appendix O in response to clarification question A3.

ITT population

CS Section B.2.4.2 and the CSRs[10,11] report that all analyses of the efficacy outcomes were based on the primary efficacy analysis set which is synonymous to the ITT population. The CSRs also provide information about how missing data were handled and we note that conservative, appropriate methods were used (the CSRs report that sensitivity analyses were conducted on different imputation methods, although results of these are not presented). People in whom treatment had failed prior to week 8 (induction phase) and prior to week 44 (maintenance phase) were considered not to be in clinical remission and not to have had a clinical response. Participants who had all four Mayo subscales missing in either the induction or maintenance phases were considered not to be in clinical remission nor clinical response. Otherwise, generally, the last observation carried forward (LOCF) approach was used for continuous endpoints and where there was missing data for dichotomous endpoints, participants were considered not to have achieved these (clarification question response Appendix O).

The ERG has checked the numbers of participants stated to be included in the analyses of the results for each outcome presented in CS Section B.2.6. The sample sizes, where provided, match the numbers randomised or re-randomised to each trial arm in both the induction and maintenance phases, confirming that these were based on the primary efficacy analysis set. However, the numbers included in the analyses are not provided for the following outcomes:

- mucosal healing at induction week 8 and maintenance week 44
- histologic healing at induction week 8 and maintenance week 44
- disease-related hospitalisations and surgeries in the induction and maintenance phases
- UC disease-related hospitalisations and surgery at induction week 8
- the IBDQ results from the maintenance phase

This means that it was not possible for us to verify that results for these outcomes were based on the primary efficacy analysis set (ITT analysis population), which introduces some uncertainty in interpreting the results.

Safety analysis set

The safety analysis set consists of participants who had received at least one dose (including a partial dose) of the study treatment. Analyses were based on the treatment that participants actually received.

3.1.6.3 Subgroup analyses

Pre-specified subgroup analyses were conducted for the biologic and non-biologic failure participants (for a description of these subgroups see Section 2.3). Data from the biologic and non-biologic subgroups rather than the whole ITT population were used to inform the company's NMAs of clinical response, clinical remission, and mucosal healing, and their economic model. HRQoL results, including EQ-5D results, are not reported separately for the biologic failure and non-biologic failure subgroups (the EQ-5D data are provided in clarification response Appendix Q, in response to clarification question A9). The economic model assumes the same utility values for the biologic and non-biologic subgroups in the scenario analysis that uses the trial's EQ-5D results. It is unclear why the company has not provided HRQoL findings for the biologic and non-biologic subgroups.

Table 9 below shows the numbers of participants included in the non-biologic failure and biologic failure subgroups according to the trial arms in the UNIFI Induction and Maintenance Studies. The CS and CSRs do not report any power calculations for these subgroups and so it is unclear whether they would have been adequately powered to detect effects on the primary outcome of clinical remission. We note that the sample sizes of the induction subgroups (N=156 to 166) are close to the size required in the power calculations for the number per treatment arm in the ITT population (N=135 or N=220, depending on which calculation is used) (section 3.1.6.1 above). It is plausible (though not certain) that these induction subgroup sample sizes have reasonable power for detecting differences between induction ustekinumab and placebo arms. However, the Maintenance Study subgroups (which are arguably the more important ones in the context of long-term clinical effectiveness), are notably smaller (N=70 to 102) and less likely to be adequately powered to detect differences between ustekinumab and placebo in clinical remission rates.

Table 9 Sample sizes for the non-biologic failure and biologic failure subgroups by trial arm

Induction Study	**Placebo (trial ITT N = 319)**	**Ustekinumab 6mg/kg (trial ITT N = 322)**	**Ustekinumab 130 mg (trial ITT N = 320)**
Non-biologic failure	158	156	156
Biologic failure	161	166	164
Maintenance study	**Placebo (trial ITT N = 175)**	**Ustekinumab q8w (trial ITT N = 176)**	**Ustekinumab q12w (trial ITT N = 172)**
Non-biologic failure	87	85	102
Biologic failure	88	91	70
Sources: CS Figures 17, 18, 19 and 20			

Other subgroup analyses (reported in CS Appendix E) are not directly of interest to NICE's final scope, the company's decision problem or the economic model, so we have not detailed them here.

> **ERG conclusion.** The statistical analysis approaches in the UNIFI trial appear appropriate, with conservative imputations employed for missing data. The whole population (ITT) analyses of the primary outcome of clinical remission in the UNIFI trial are adequately statistically powered. It is plausible (but not certain) that the Induction Study subgroup analyses based on biologic failure status would also be adequately powered for this outcome. However, biological failure status subgroups analyses in the Maintenance Study are based on smaller sample sizes and are less likely to be adequately powered.

3.1.7 Approach to the evidence synthesis

The company presents the results of the UNIFI trial, which compared ustekinumab against placebo (see section 3.1.3 above). A further 18 trials of comparator therapies were identified by the company (section 3.1.2 above) but no direct head-to-head comparisons between ustekinumab and the comparator therapies have been conducted. The company therefore ran a series of NMAs, described in detail below.

The company also conducted direct pairwise meta-analyses for each active comparator versus placebo where sufficient data were available (CS Appendix Tables 63 to 66). These analyses were only feasible for the non-biologic failure group, apart from a single comparison of tofacitinib against placebo in the biologic failure group (CS Appendix Table 66); they relate only to the induction phase; and they do not inform the company's economic analysis. We therefore considered these direct meta-analyses to have low priority and we have not checked their validity.

3.1.7.1 Risk of bias assessments for trials included in NMAs

The CS reports risk of bias assessments for the 19 included trials based on standard NICE questions (CS Appendix Tables 24 and 85) but does not discuss whether specific trials should be included in or excluded from meta-analyses based on these assessments. We have briefly compared the company's risk of bias assessments to those made by ERGs in previous NICE technology appraisals and we consider that overall the included trials were well conducted and likely to be at low (or in some cases unclear) risks of bias, with no individual trials definitively being at high risk (see Appendix 3). The main issue identified by these assessments is that several trials had relatively high rates of drop-out, with drop-out

rates being higher in placebo than active comparator arms. There is potential for attrition bias influencing NMA results if unbalanced drop-outs in the individual studies are not handled appropriately in analyses. The risk of attrition bias is reduced by using ITT data from the trials in NMAs, provided that missing data are imputed appropriately. The company do not discuss the integrity of the ITT populations within individual trials so there is some uncertainty around the potential for attrition bias affecting NMA results.

3.1.7.2 Trial eligibility assessment for NMAs

In addition to the eligibility criteria for their systematic review of clinical effectiveness (see section 3.1.2) the company employed a further set of eligibility criteria to assess the eligibility of trials for NMAs. These criteria are reported in CS section B.2.9.1 and CS Appendix D1.8 and summarised in section 3.1.7.2

The NMA eligibility criteria are similar to the overall systematic review criteria, with the following exceptions:

- Asian trials are excluded from the main NMAs but included in sensitivity analyses (discussed further in section 3.1.7.2.1 below).
- Dose regimens of ustekinumab and comparator therapies mainly follow EMA licensed doses, for induction as stated in CS Appendix Table 34 and maintenance as stated in CS Appendix Table 35. An exception is that unlicensed doses of infliximab are permitted (discussed in section 3.1.7.2.2 below).
- The duration of trials is restricted to those that had an induction period of 6-8 weeks and those that had a maintenance period of 44-54 weeks (discussed further in section 3.1.7.2.3 below).

3.1.7.2.1 Trials on Asian populations

Four of the 19 trials were conducted only in Asian (i.e. Chinese or Japanese) populations (Japic CTI-060298; Jiang 2015; NCT00853099; NCT02039505) and the company excluded these from their main NMA analyses but included them in sensitivity analyses. This differs from the recent technology appraisals TA342 (vedolizumab) and TA547 (tofacitinib) in which companies included Asian trials in their NMAs (with a sensitivity analysis excluding the Asian trials in TA547). The CS does not give any specific reasons for excluding Asian trials, other than to “increase comparability of the trials and include patients more reflective of the UK setting” (CS section B.2.9.1). Clinical experts advising the ERG noted that Asian patients are treated in the NHS and that there is no specification in the NICE scope to exclude Asian populations. According to the draft SmPC (CS Appendix C), clearance of ustekinumab in

Crohn's disease differs between Asian and non-Asian populations although it is unclear whether this is sufficient to warrant Asian populations being treated as a separate subgroup.

We agree that the approach of conducting a sensitivity analysis to test the impact of Asian trials on NMA results is appropriate. However, the company appears to have misinterpreted the Japic CTI-060298 trial which the CS claims had a re-randomised design whilst the trial publication suggests it had a treat-through design (Table 10). The company also state that both induction responders and non-responders in Japic CTI-060298 received maintenance therapy (CS Appendix Tables 19 and 32) but according to the trial publication only induction responders received the maintenance infliximab or placebo.[39] These discrepancies cast some doubt on the reliability of the company's NMA sensitivity analysis on the Asian trials.

Apart from the Asian trials, all trials included in the company's clinical effectiveness review were multinational (CS Appendix Table 32).

3.1.7.2.2 Dose regimens

The NMA eligibility criteria reported in CS Appendix D1.8 restrict trials to those using EMA licensed dose regimens, but permit the inclusion of unlicensed maintenance doses of infliximab, without an explanation. The company's response to clarification question A15 explains that inclusion of the higher (i.e. escalated) unlicensed maintenance infliximab dose is necessary to enable comparisons of standard and escalated regimens across therapies in the NMAs. The ERG's clinical advisors confirmed that the escalated maintenance dose of infliximab is used in clinical practice and therefore we agree with the company's approach.

3.1.7.2.3 Duration of induction and maintenance

The company's NMA eligibility criteria permitted the inclusion of trials with induction assessments at 6-8 weeks and maintenance assessments at 44-54 weeks (CS Appendix D1.8.1).

All trials met the 6-8 week induction duration criterion except the Asian trial NCT02039505 which had an induction period of 10 weeks (CS Appendix Table 17). The CS does not specifically discuss the exclusion of this trial, although, as noted above (section 3.1.7.2.1), being an Asian trial, it would not be eligible for inclusion in the main NMA analyses.

Two trials did not meet the 44-54 week maintenance duration criterion (CS Appendix Table 18). These were ACT2 which had a maintenance assessment at 30 weeks, and

NCT02039505 which had a maintenance assessment at 60 weeks. The company excluded the ACT2 trial as they considered the 30-week maintenance assessment unrepresentative of maintenance 1-year outcomes. However, the CS does not mention exclusion of the Asian NCT02039505 trial. The company consider that the 44-54 week range of maintenance assessment times is a reasonable reflection of 1-year maintenance outcomes which their NMAs were aiming to model.

The ERG agrees that differences in trial duration can introduce heterogeneity into an NMA and therefore it is appropriate to exclude the outlier trials, although the CS does not discuss the implications of this. However, we note that after applying the eligibility criteria there is still residual variation in trial duration within the NMAs that could potentially introduce bias, as discussed in section 3.1.7.3.4 below.

3.1.7.3 Heterogeneity of studies in the NMAs

The company considered several sources of potential heterogeneity across the trials included in their NMAs, as summarised in sections 3.1.7.3.1 to 3.1.7.3.4 below.

3.1.7.3.1 Definitions of outcome assessments

Clinical remission

Most of the trials included in the NMAs used a definition consistent with the 'global definition' in the UNIFI trial (see Table 8). However, OCTAVE 1, OCTAVE 2, OCTAVE Sustain, and Probert 2003 employed different definitions (CS Appendix D1.1.8.1). The company do not explain how these differences were addressed or interpreted in the NMAs. The ERG's clinical experts suggested that the definitions used across the different studies are sufficiently similar that they can be ignored when considering the eligibility of the studies for NMA.

Clinical response

This was defined consistently across all trials included in the NMAs (CS Appendix D1.1.8.1).

Mucosal healing

The UNIFI trial used a different definition of mucosal healing compared to all other trials (CS Appendix D1.1.8.1). However, the "endoscopic healing" outcome in UNIFI was defined in the same way as mucosal healing in the other trials (i.e. Mayo endoscopic subscore of 0 or 1) (Table 8). Therefore the company used the endoscopic healing outcome from UNIFI in the

mucosal healing NMAs, and used the term “mucosal healing” to refer to the endoscopic healing outcome from UNIFI when referring to the NMAs in the rest of the CS.

Central versus local endoscopy reading

Most of the trials available for NMA had employed local endoscopy reading (or the method of reading was not reported), while OCTAVE1, OCTAVE2 and OCTAVE Sustain employed central endoscopy reading, and UNIFI employed both methods (CS Appendix Table 23 and company clarification Table 14). The company report that they used centrally-read scores from the OCTAVE trials and locally-read scores from all other trials in their NMAs. Presumably this is because it is the only way that connected evidence networks could be formed that included the OCTAVE tofacitinib trials, although the company are not explicit about this (a further tofacitinib trial is available, NCT00787202, but this reported only response, not remission, and is excluded from CS Appendix Table 23).

Centrally-read endoscopy scores are usually less variable than locally-read scores, although this may depend on a number of factors, including the training and experience of the readers as well as the protocol used.[40] In the OCTAVE trials clinical outcomes based on both centrally-read and locally read endoscopy data are reported, but these are for the whole (ITT) population only, not the non-biological failure and biological failure subgroups of interest in the company’s NMAs and economic model. In TA547 (tofacitinib) central reading was consistently associated with lower rates of clinical remission in the ITT population, for both the tofacitinib and placebo groups in all three OCTAVE trials, although this difference was not evident for the clinical response outcome. The company’s inclusion of remission outcomes based on centrally-read endoscopies in the OCTAVE trials and locally-read endoscopies in all other trials could introduce bias against tofacitinib in the NMAs.

3.1.7.3.2 Variation in prior biologic therapy subgroup definitions

The prior therapy subgroups reported in the UNIFI trial (for definitions see Table 3) are compared against similar subgroups, where available, in the comparator trials, in CS Appendix Table 21, although not all of the 19 trials included in the company’s clinical effectiveness review are listed in the table. CS Appendix Table 21 shows that the trials can be grouped into whether they used biologic-exposure subgroups (as specified in the NICE scope (see Table 3) or biologic failure subgroups as defined in the UNIFI trial. The company state that “to allow meaningful comparisons to be made accounting for population heterogeneity” they consistently employed the following subgroup definitions to the trials (CS Appendix D1.1.7):

- **non-biologic failure:** either biologic-naïve patients (including anti-TNF naïve), or biologic-experienced (including anti-TNF experienced) patients without previous anti-TNF failure;
- **biologic failure:** biologic-experienced patients (including anti-TNF experienced) who failed their previous biologic treatment (including failing anti-TNF treatment)

As we have shown in for the UNIFI trial in Table 3, there was good, but not perfect, quantitative concordance between the proportions of trial participants who met the biologic exposed/naive definitions in the NICE scope and the biologic failure/non-biologic failure subgroup definitions in the UNIFI trial. However, the company do not discuss the quantitative degree of concordance between the subgroup definitions employed in the comparator trials and those of the UNIFI trial. Imprecise matching of the subgroup definitions when combining the trials in NMAs would introduce heterogeneity into the NMA results but the CS does not discuss this explicitly as a source of uncertainty.

3.1.7.3.3 Variation in trial population demographic and disease characteristics

Most of the trials included in the company's clinical effectiveness review were also included in the NMAs in TA547 (tofacitinib) and so similar issues of trial heterogeneity apply. Mean disease duration ranged from 4.3 to 10.9 years across the trials (CS Appendix Figure 10) although, despite this being a 7-year range, CS Appendix D1.5.1 interprets this as "no major variabilities in disease duration". Mayo scores at induction baseline ranged from 8.0 to 9.1 (CS Appendix Figure 14). Use of concomitant steroids ranged from 27.0% to 84.2% (clarification Appendix Table 12). The proportion of patients who received previous anti-TNF therapy ranged from 28% to 58% (clarification Appendix Table 13). CRP levels were also variable across the trials (2.2 to 18.8 mg/L) (CS Appendix Figure 12), although the ERG's clinical experts suggested CRP is not a reliable prognostic factor. We note that in TA547 (tofacitinib), baseline IBDQ scores ranged from 114 to 167, which would exceed the threshold for a clinically meaningful difference (see section 3.1.5.3), although IBDQ was not reported for all trials in the current appraisal. As acknowledged in CS Appendix D1.2.1, patients' age, gender and weight were generally evenly balanced across the trials.

The data summarised above clearly indicate there is considerable heterogeneity across the trials included in the NMAs, and there may also have been unobserved heterogeneity in population characteristics that were not measured. Standard approaches to account for heterogeneity in NMAs are to break the data down into subgroups so that heterogeneity can be tested and accounted for, e.g. in sensitivity analyses by including/excluding outlier trials;

and to employ random effects statistical models (although the former may reduce sample size and fragment evidence networks). As discussed in the NMA methods (sections 3.1.7.5.1 to 3.1.7.5.3 below), the company mainly rely on random effects models to deal with heterogeneity, although these were not always feasible.

3.1.7.3.4 Trial duration

The company applied eligibility criteria to limit trials included in the NMAs to those which had induction assessments in the range 6-8 weeks and those that had maintenance assessments in the range 44-54 weeks (section 3.1.7.2.3 above). Thus, there is still some heterogeneity in trial duration remaining after application of the eligibility criteria.

The trials can be divided into those that had a treat-through design and those that had a re-randomised design (see section 3.1.7.4). As noted in the ERG report for TA547 (tofacitinib), differences in the duration of induction phases in re-randomised trials could bias against studies with a shorter induction period (e.g. a 6-week trial would miss any remission or response events that occur at 8 weeks). Differences in the duration of the maintenance phases in re-randomised trials could also introduce bias, but in favour of trials with shorter maintenance phases (e.g. if fewer responders lose response in the shorter time frame).

As in TA547, these differences in trial durations are not adjusted for in the NMAs. It is therefore possible that there may be bias in favour of ustekinumab (UNIFI 8 weeks) in the induction phase against golimumab (PURSUIT-J 6 weeks, although this is an Asian trial) and vedolizumab (GEMINI1 6 weeks). It is also possible that there may be bias in favour of ustekinumab versus all the maintenance phase comparators in re-randomised trials (golimumab, tofacitinib, vedolizumab), since the UNIFI trial had the shortest maintenance assessment time among the re-randomised trials (44 weeks in UNIFI, 46 weeks in GEMINI1, all other trials 52-54 weeks).

3.1.7.4 Evidence available for clinical effectiveness NMAs

Of the 19 trials included in the company's clinical effectiveness review, 15 covered the induction phase and 14 covered the maintenance phase (10 covered both induction and maintenance periods, five covered induction only, and four covered maintenance only) (Table 10).

A fundamental consideration when conducting the NMAs is that the maintenance trials employed two contrasting methodological approaches:

- **Treat-through trials**: patients were randomised to placebo and comparator at baseline and outcomes were assessed at the end of an induction phase (8 weeks) and at the end of a maintenance phase (30 to 54 weeks).
- **Re-randomised trials**: patients who responded to induction therapy (6 to 10 weeks) were re-randomised to new placebo and comparator arms for the maintenance therapy and outcomes were assessed at the end of the maintenance phase (44 to 60 weeks).

Table 10 Overview of induction and maintenance trials

Comparison	Trial	Induction	Maintenance	Maintenance design
ADA vs placebo	NCT00853099 [41]	●	●	Treat-through
	ULTRA1 [42,43]	●	●	Treat-through
	ULTRA2 [44]	●	●	Treat-through
ADA vs VED	VARSITY [25,45 a]	NA	●	Treat-through
GOL vs placebo	PURSUIT-SC [46]	●	NA	
	PURSUIT-J [47]	NA	●	Re-randomised
	PURSUIT-M [48]	NA	●	Re-randomised
INF vs placebo	ACT1 [49]	●	●	Treat-through
	ACT2 [49]	●	●	Treat-through
	Japic CTI-060298 [39]	●	●	Treat-through [b]
	Jiang 2015 [50]	●	●	Treat-through
	Probert 2003 [51]	●	NA	
TOF vs placebo	NCT00787202 [52]	●	NA	
	OCTAVE1[53]	●	NA	
	OCTAVE2 [53]	●	NA	
	OCTAVE Sustain [53]	NA	●	Re-randomised
UST vs placebo	UNIFI [10,11]	●	●	Re-randomised
VED vs placebo	GEMINI1 [54]	●	●	Re-randomised
	NCT02039505 [55]	●	●	Re-randomised

NA: not applicable
[a] The VARSITY trial included induction and maintenance therapy but only maintenance period outcomes are reported in the abstracts[25,45] (response at week 14, all other outcomes at week 52).
[b] Japic CTI-060298 is reported by the company as a re-randomised trial (CS Appendix Tables 19 and 32) but the trial publication[39] indicates it was a treat-through trial. The company also incorrectly refers to this trial as “Japis CTI060297”.

3.1.7.4.1 Treat-through trials

Eight of the 14 maintenance trials had a treat-through design. According to the trial publications, in ULTRA1,[43] VARSITY,[25] ACT1,[49] ACT2,[49] and Jiang 2015[50] all patients who

received induction therapy (i.e. both induction responders and non-responders) continued in the trial and received maintenance therapy. As noted above, in Japic CTI-060298 only induction responders received maintenance therapy.[39] In NCT00853099[41] and ULTRA2,[44] patients who had an inadequate response after the induction period could enter an alternative open-label arm, meaning that non-responders would not have received the maintenance therapy in their originally randomised arm, although the time points at which these switches occurred during the trials' maintenance phases were not reported.

3.1.7.4.2 Re-randomised trials

The six re-randomised maintenance trials (Table 10) can be divided into three groups, according to whether induction placebo responders were re-randomised:

- Responders from only the active therapy induction arm were re-randomised: PURSUIT-J,[47] PURSUIT-M,[48] GEMINI1,[54] NCT02039505.[55]
- Responders from the induction active therapy arms and delayed responders from the induction placebo arm were re-randomised: UNIFI (CS Figure 10).
- Responders from both the active therapy and placebo induction arms were re-randomised: OCTAVE Sustain.[53]

In the trials that re-randomised only the active therapy responders, placebo responders went on to receive further maintenance placebo in a non-randomised arm (apart from the PURSUIT-J trial which only had a single active therapy induction arm).

In the UNIFI trial, patients who were delayed responders to IV ustekinumab induction therapy at week 8 but had responded to a subcutaneous dose of ustekinumab by week 16 then received ustekinumab q8w maintenance therapy in a non-randomised arm (Figure 1).

Carry-over effect in re-randomised trials

The company argue that an induction carry-over effect is present in the maintenance placebo arm of the UNIFI trial and has also been observed in the appraisal of ustekinumab in Crohn's disease (TA456[33]). The company believe this carry-over effect differs between UNIFI and comparator re-randomised trials (CS Appendix section D10.2).

The company suggest that the carry-over effect might be explained by the mode of action and half-life of ustekinumab, although the ERG's clinical experts were unconvinced that this would cause a different effect compared to the other biologic treatments. Previous reviews by Macaluso et al.[56] and Jairath et al.[57] identified heterogeneity in placebo arms of UC trials but attributed this to an imbalance of prognostic factors rather than carry-over effects. The

prognostic factors included concomitant steroids at baseline, disease duration, naïvety to anti-TNF therapy, central reading of endoscopy, and the time point of assessment.[56,57] The company acknowledge in response to clarification question A25 that the carry-over effect is likely to be multifactorial. We agree that the pattern of Partial Mayo scores in maintenance placebo arms shown in CS Appendix Figures 38 to 40 differ between UNIFI and other trials and could plausibly reflect a carry-over effect, but evidence appears to be sparse.

3.1.7.5 NMA methods

The company formed connected evidence networks for subsets of the 19 identified trials to conduct three main sets of NMAs (Table 11):

- Induction NMAs (0 to ~8 weeks*)
- "1-year NMA" (induction plus maintenance, 0 to ~52 weeks*), with re-randomised trials adjusted to mimic the treat-through approach
- "1-year NMA" conditional on response

*For discussion of the duration of the induction and maintenance in the trials included in the NMAs see section 3.1.7.3.4.

Table 11 Overview of NMAs conducted and their role in the economic model

Outcomes Included in NMA	Induction NMA		1-year NMA		1-year NMA conditional on response	
	NMA conducted	Informs model	NMA conducted	Informs model	NMA conducted	Informs model
Clinical remission	Yes [a]	Base case	Yes	No	Yes [a]	Scenario [c]
Clinical response	Yes [a]	Base case	Yes	No	Yes [a]	Scenario [c]
Mucosal healing	Yes [b]	No	No	--	No	--
Overall AEs	Yes [b]	No	No	--	No	--
Serious AEs	Yes [b]	No	No	--	No	--
Overall infections	Yes [b]	No	No	--	No	--
Serious infections [d]	Yes [b]	No	No	--	No	--

AEs: adverse events; -- : not applicable
[a] Key analysis, validated by ERG
[b] Subordinate analysis, not validated by ERG
[c] Model base case informed by direct trial data (active arms only), not NMA
[d] Serious infections inform the model but taken from observational study, not NMA

The company analysed clinical response and clinical remission separately, although these are correlated outcomes. The CS explains that a multinomial probit analysis approach, which was used in TA547 (tofacitinib) to jointly model remission and response to account for their correlation, was precluded due to differences in the placebo arms. The handling of correlations in the company's economic analysis is discussed in section 4.4.2 below.

The clinical effectiveness NMAs were each conducted for the non-biologic failure and biologic failures subgroups, but not for the overall (ITT) trial populations. This is consistent with the economic modelling approach which utilises clinical remission and response results from the non-biologic failure and biologic failure subgroups (sections 4.3.4.1 and 4.3.4.2).

These different NMA approaches employed by the company, and an exploratory additional scenario analysis conducted by the ERG, are described further below. A general overview of the approaches is shown in Table 12.

NMA models were run in WinBUGS using logistic regression for binary outcomes. The NMA WinBUGS code is not included in the company's submission but has been provided in response to clarification question A12.

Table 12 Overview of the NMA methods employed by the company and ERG

	Induction NMA	**1-year NMA [a]**	**1-year NMA conditional on response [b]**	**Maintenance only NMA (ERG scenario analysis)**
Description	Standard NMA approach according to NICE DSU methods	Captures whole induction + maintenance pathway using ITT population. Mimics an ITT treat-through approach based on Thorlund et al. [58] by re-calculating data from response-based trials to correspond to a treat-through design, maintaining the initial randomisation.	Captures whole induction + maintenance pathway using ITT population. Mimics an ITT re-randomised approach using only responders to induction therapy.	Captures maintenance pathway following re-randomisation of responders to induction therapy. Mimics an ITT re-randomised approach following TA547.
How implemented	Standard NMA based on RCTs; takes remission or response data at end of induction as NMA inputs	Takes remission or response data for active treatment or placebo at end of maintenance period as NMA inputs depending	Takes remission or response data for active treatment or placebo at end of maintenance period based on induction responders.	Takes remission or response data for re-randomised active treatment or placebo at end

		upon initial randomisation.		of maintenance period.
Population modelled	Induction responders	Includes induction non-responders (i.e. delayed) responders so maintenance therapy can be given to late responders	Excludes induction non-responders (i.e. delayed) responders	Excludes induction non-responders (i.e. delayed) responders
Key considerations	Subject to standard NMA assumptions of heterogeneity and consistency	Imputation required in recalculating data from the re-randomised trials to mimic threat through trials. Imputation of placebo maintenance data where missing for induction responders and non responders. Imputations are based on existing relationships between the data to impute missing subgroups.	Does not use the post-re randomisation placebo arm due to differences in carry-over effect. Imputation required re-calculating data from treat-through trials to correspond to the re-randomised design. Imputations are based on existing relationships between the data to impute missing subgroups.	Assumes re-randomised placebo arms are similar thus no carryover effect. Imputation required re-calculating data from treat-through trials to correspond to the re-randomised design. Imputations are based on existing relationships between the data to impute missing subgroups.

[a] The company refer to this as their "base case" NMA. To avoid confusion with the economic model base case we avoid using this terminology to describe the NMA.

[b] The company refer to this as a NMA "sensitivity analysis". To avoid confusion with the economic model sensitivity analyses we avoid using this terminology to describe the NMA.

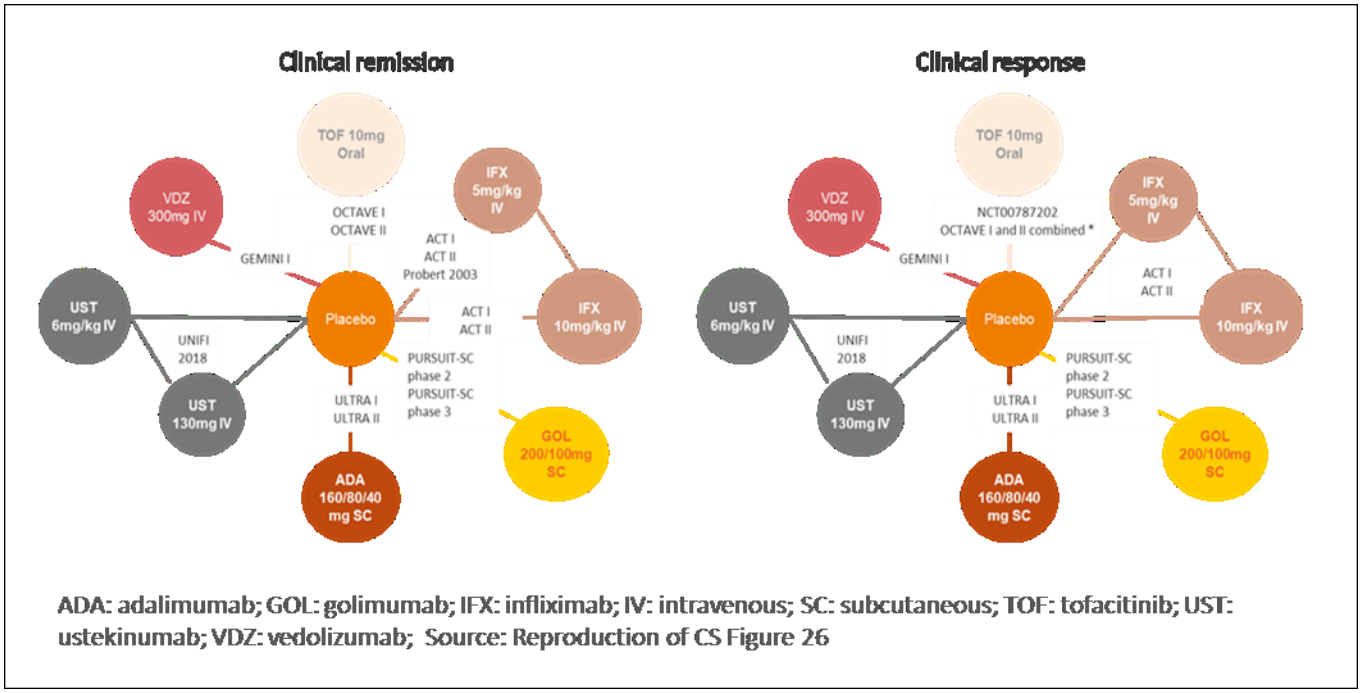
3.1.7.5.1 Induction NMAs

The induction trials were standard RCTs and therefore the company applied standard NMA methods[59] to analyse these. The network diagrams for clinical remission and response are shown in CS Figure 26 for the non-biologic failure subgroup and in CS Figure 27 for the biologic failure subgroup, reproduced below in Figure 2 and Figure 3 respectively.

Both fixed and random effects analyses were conducted. Model fit, assessed using the deviance information criterion (DIC) was similar across fixed-effects and random-effects models for the induction analyses but the company preferred the fixed effects model which assumes there is no heterogeneity between studies. The company's economic analysis base

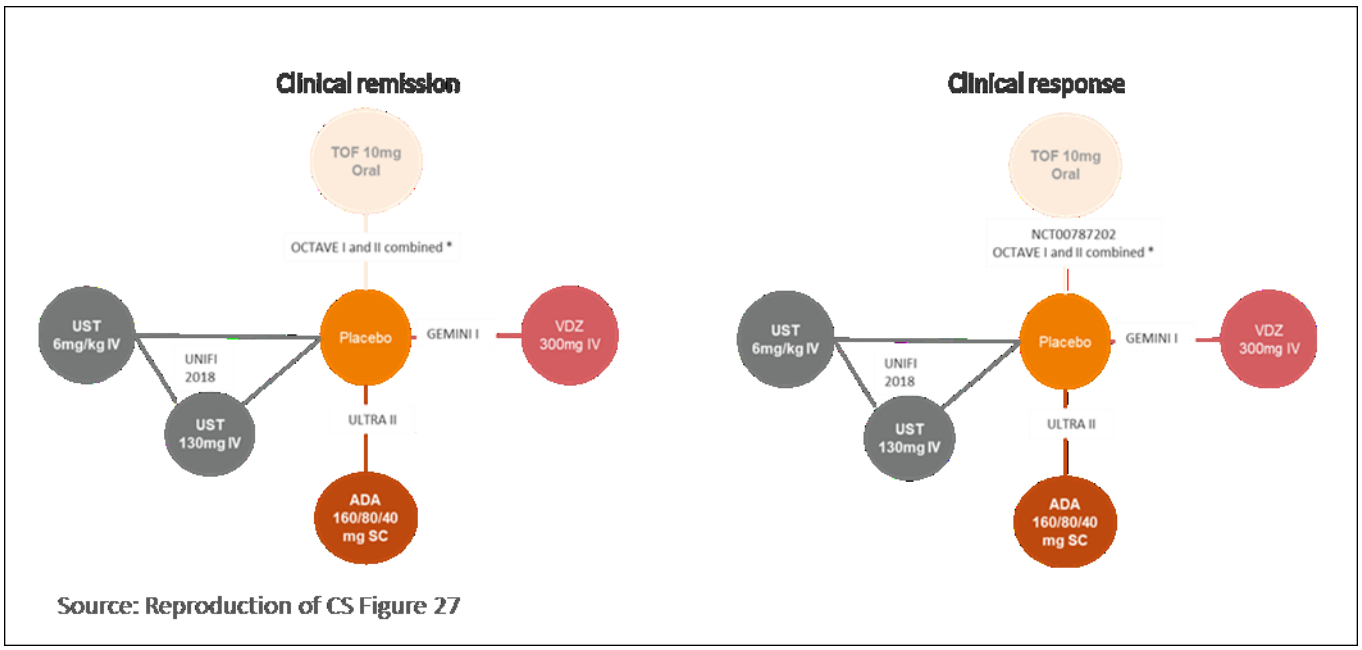
case uses induction response and remission NMA results based on a fixed-effects model, with random-effects NMA results used in a scenario (section 4.3.4.1).

Results of the company's NMAs for response and remission for the non-biologic failure and biologic failure subgroups are reported in CS section B.2.9.4.



ADA: adalimumab; GOL: golimumab; IFX: infliximab; IV: intravenous; SC: subcutaneous; TOF: tofacitinib; UST: ustekinumab; VDZ: vedolizumab; Source: Reproduction of CS Figure 26

Figure 2 Evidence network for induction phase clinical remission and response in non-biologic failure patients



Source: Reproduction of CS Figure 27

Figure 3 Evidence network for induction phase clinical remission and response in biologic failure patients

The ERG has validated the company's induction NMAs, and our results are compared with those of the company in section 3.3.6.1 below. We identified some discrepancies in the induction response and remission outcomes data for the UNIFI and OCTAVE trials between the input data listed in CS Appendix Table 60, the company's NMA code, and the trial publications (Appendix 4) and we have corrected these in our analyses.

3.1.7.5.2 One-year NMA

Meta-analysis of the maintenance trials is not straightforward, as the different treat-through and re-randomised designs cannot be included in a standard NMA. In CS section B.2.9.3.1 the company state that two possible alternative approaches were considered to enable NMA to be conducted on the treat-through and re-randomised trials:

- Adjusting the treat-through trial responder outcomes data so that they mimic those that would have been obtained in a re-randomised trial, e.g. using an approach employed by the company in TA547 (tofacitinib).
- Adjusting the re-randomised trial responder outcomes data so that they mimic those that would have been obtained in a treat-through trial, based on an approach reported by Thorlund et al. (2015).[58]

The first approach involves NMA only of the maintenance phase, and assumes that, in the treat-through trials, responders at the end of induction were the same as responders at the end of maintenance. The company considered the TA547 maintenance NMA approach to be "severely limited for several important methodologic reasons" (CS section B.2.3.9.1).

The second approach captures both the induction and maintenance phases, and the company refer to this as a "1-year NMA" (CS section B.2.9.3.1). The company argue that this approach reflects clinical practice, allowing delayed responders to induction therapy to be included. They also suggest that the 1-year NMA approach overcomes methodological issues of non-comparable placebo arms in re-randomised trials (CS section B.2.9.3.6). The company therefore preferred the 1-year NMA approach over the maintenance-only approach employed in TA547, and they conducted 1-year NMAs for the clinical remission and response outcomes.

An overview of the maintenance-phase NMA approach was provided by the company in response to clarification question A16, reproduced below in Figure 4, and an overview of the 1-year NMA approach is presented in CS Appendix D10.1, reproduced in Figure 5 below.

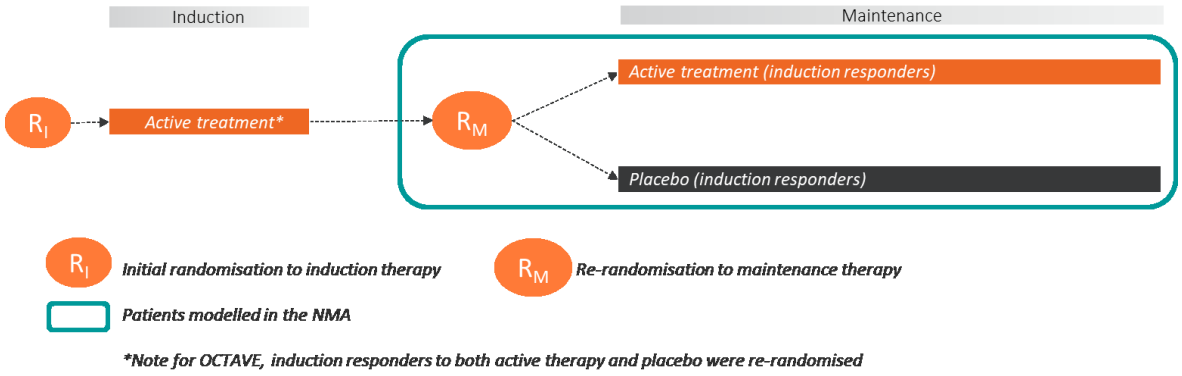


Figure 4 Overview of TA547 maintenance-phase NMA approach (mimics re-randomised trial design)

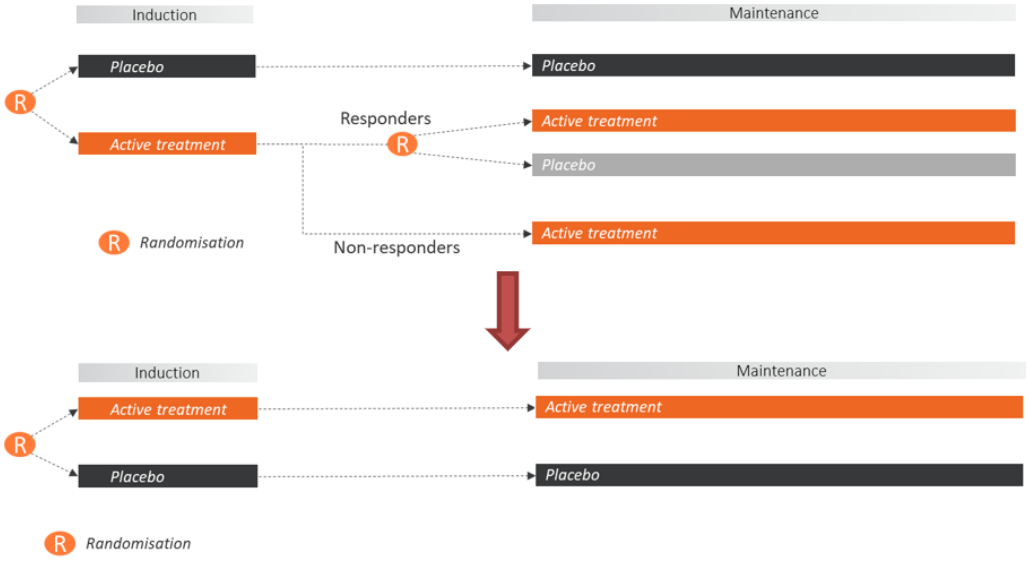


Figure 5 Overview of 1-year NMA approach (mimics treat-through trial design)

The CS reports that the 1-year NMAs were based on the method reported by Thorlund et al. (2015) whereby the re-randomised trials were converted to mimic threat-through trials.[58] The calculations are presented in CS Appendix sections D10.3.3 to D10.3.8 and CS Appendix Tables 58,59, and 61, but these are not adequately clear and the ERG was unable to verify whether the Thorlund approach had been correctly and reasonably applied. Nor is it clear in the CS which data were imputed and which were taken directly from the clinical trials. The company provided further detail in Appendix R in response to clarification question A13. However, we were still unable to verify many data sources.

Results of these 1-year NMAs are reported in CS section B.2.9.4. But, despite the company's claimed advantages of the 1-year NMA approach, the clinical remission and response outcomes from the 1-year NMAs are not used in the economic analysis, and no explanation for this is provided in the CS. The company say in their response to clarification question A21 that their main concern is heterogeneity in the maintenance phase placebo populations, although, according to CS section B.2.9.3.6, one of the advantages of the 1-year NMA approach is that it overcomes problems of non-comparability of maintenance placebo arms.

In the economic model, the company employed a loss of response analysis as their model base case, which takes clinical remission and response data directly from the individual trial arms (sections 4.3.4.1 and 4.3.4.2 below).

Given that the company's 1-year NMAs do not inform the economic analysis the ERG has not validated them and we do not discuss them further in this report. Instead, we focus our critique on a further NMA approach employed by the company employed which does inform the economic analysis. This is referred to as a "1-year NMA conditional on response."

3.1.7.5.3 One-year NMA conditional on response

The company conducted what they refer to as a "1-year NMA: ITT approach conditional on response to induction" (CS section B.2.9.4.3) which, for brevity, we refer to as a 1-year NMA conditional on response. Results from this NMA approach inform a scenario in the economic model, but do not inform the model base case.

Note that the company also refer to the 1-year NMA conditional on response as a "sensitivity analysis" (CS section B.2.9.4.3); to avoid the risk of confusion we avoid this terminology in the current report.

The methods of the 1-year NMA conditional on response are mentioned only very briefly in the CS (section B.2.9.3.1) and are unclear, and the CS does not provide a rationale for using this approach. The company's response to clarification question A16 confirms that the 1-year NMA conditional on response is similar to the 1-year NMA but does not include delayed responders (Table 12).

The company provide an overview of the 1-year NMA conditional on response approach in their response to clarification question A16, reproduced below in Figure 6.

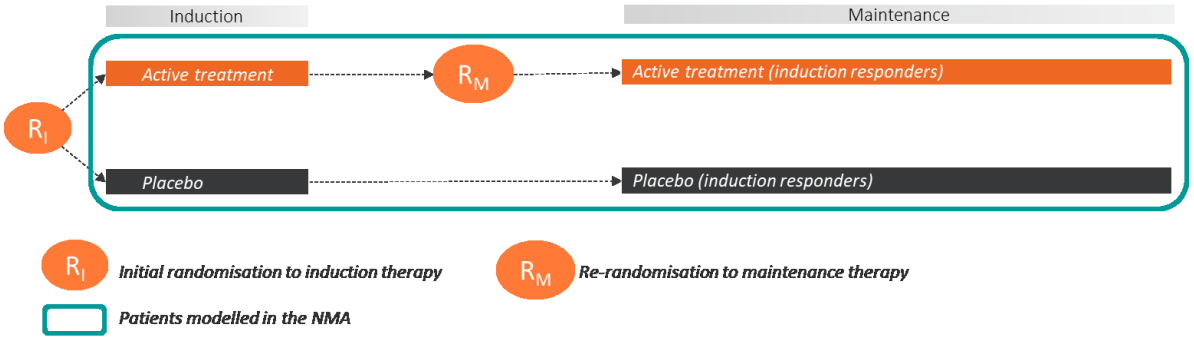


Figure 6 Overview of 1-year NMA conditional on response approach

As an attempt to address their concerns about a carry-over effect of induction therapy into the maintenance placebo arm in re-randomised trials (section 3.1.7.4.2), the company pooled the maintenance placebo arms across trials when conducting the 1-year NMAs conditional on response (Table 13).

A summary comparison of how the maintenance-phase active therapy and placebo arms are formed for each of the NMA approaches is provided in Table 13. The underlying calculations that support the NMA approaches are given in CS Appendix Table 40, and the assumptions and adjustments necessary to implement these calculations for each trial are reported in CS Appendix sections D10.3.2 to D10.3.8.

Table 13 Source of maintenance-phase active treatment and placebo groups in the different NMA approaches

NMA approach	Source data for maintenance ACTIVE arm		Source data for maintenance PBO arm	
	Treat-through trials	Re-randomised trials	Treat-through trials	Re-randomised trials
TA547	Mimics re-randomised active therapy arm by assuming number of induction responders is a proxy for the number of patients entering maintenance	Takes data directly from the active therapy trial arm	Mimics re-randomised PBO arm by assuming number of induction responders is a proxy for the number of patients entering maintenance	Takes data directly from the active therapy trial arm
1-year NMA	Takes data directly from the active therapy trial arm	Takes remission or response data at end of the maintenance period based on induction responders and non-responders	Takes data directly from the active therapy trial arm	Takes remission or response data at end of the maintenance period based on induction responders and non-responders

1-year NMA conditional on response	Mimics re-randomised active therapy arm by assuming number of induction responders is a proxy for the number of patients entering maintenance	Takes remission or response data at end of the maintenance period based on induction responders	Mimics re-randomised active therapy arm by assuming number of induction responders is a proxy for the number of patients entering maintenance	Imputed based on average response across placebo arms

The CS does not explicitly discuss the relative strengths and weaknesses of the two different 1-year NMA approaches, but instead reiterates the advantages of the overall 1-year NMA approach over the maintenance-only approach that was employed in TA547 (CS section B.2.9.3.4) (see section 3.1.7.5.2 above).

Network diagrams for the 1-year NMAs conditional on response (not reported in the CS) were provided by the company in response to clarification question A17, and are reproduced below for the non-biologic and biologic failure subgroups for clinical remission (Figure 7 and Figure 8) and clinical response (Figure 9 and Figure 10).

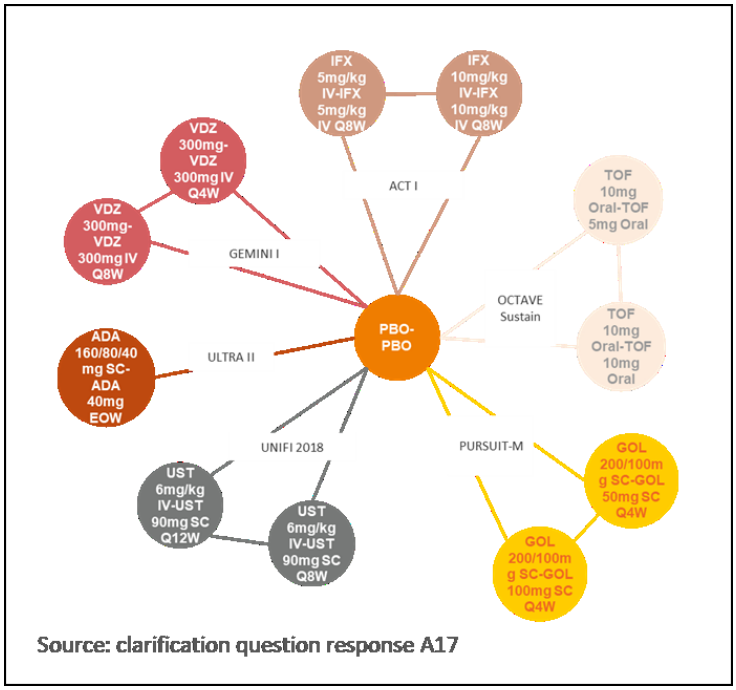


Figure 7 Evidence network for clinical remission in non-biologic failure patients, 1-year NMA conditional on response

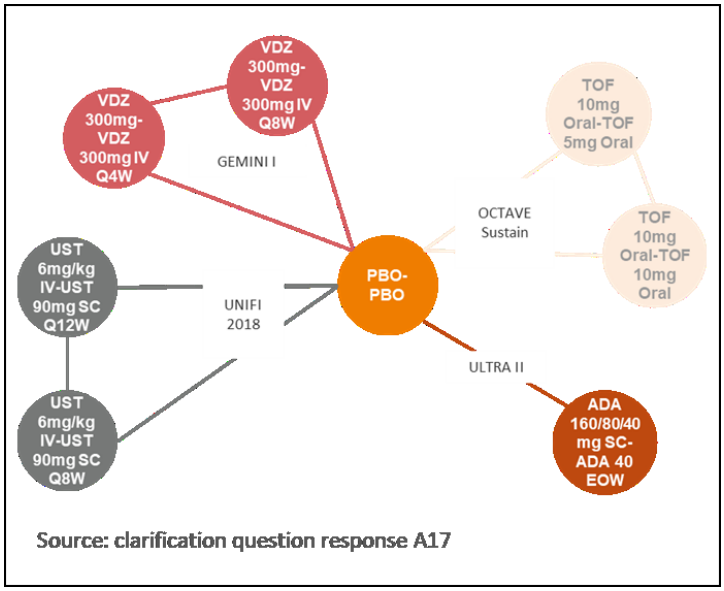


Figure 8 Evidence network for clinical remission in biologic failure patients, 1-year NMA conditional on response

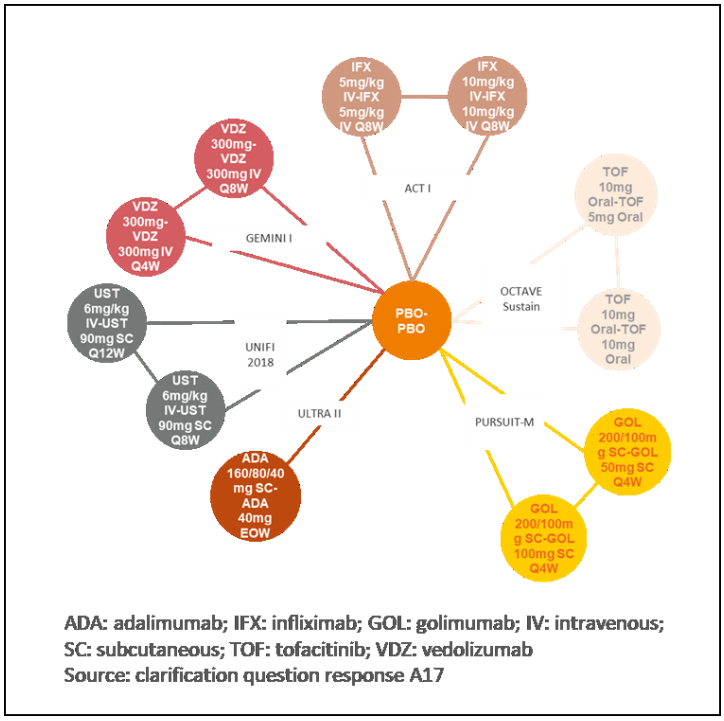


Figure 9 Evidence network for clinical response in non-biologic failure patients, 1-year NMA conditional on response

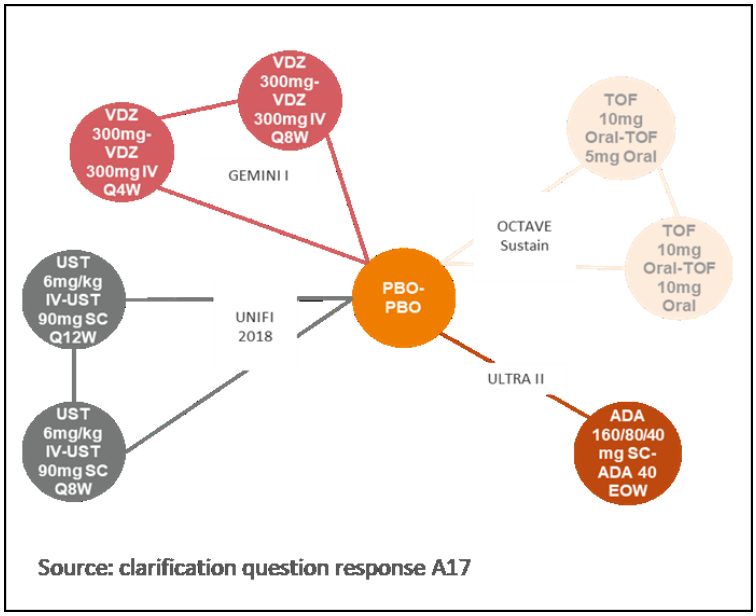


Figure 10 Evidence network for clinical response in biologic failure patients, 1-year NMA conditional on response

The CS states that the NMA conditional on response did not "allow for the inclusion of head-to-head data from the VARSITY trial as only treat-through data are available from this trial" (CS Appendix section D10.1). The rationale for this is unclear.

CS Tables 29 and 30 summarise the results of the 1-year NMA sensitivity analysis (ITT conditional on response) but only provide head-to-head comparisons against ustekinumab. The company provided a table of comparisons for each treatment versus placebo used in the model in response to clarification question A17.

The imputed calculations presented in CS Appendix sections D10.3.3 to D10.3.8 and CS Appendix Table 62 are not fully clear and we were unable to verify whether the methodology had been correctly and reasonably applied. The company provided further granularity in response to clarification question A13 and although the methodology is clearer (the company calculates maintenance responders as a proportion of induction responders to mimic a response-based design) and less complex than the 1-year NMA, we were still unable to verify some of the data sources and calculations.

The ERG agrees with the company that there is little difference in DIC (model fit) between fixed and random effects models. Total residual deviance is referred to in the methods (CS Appendix D1.11.2.1) but it is not reported in the model fit statistics (CS Tables 22 and 23)

nor included in the model code (which was provided in response to clarification question A12).

There were insufficient data to inform a random effects model. Given the potential for heterogeneity as noted above, the ERG requested the company to run the random effects model with an informative prior (clarification question A14). The company re-ran the NMA conditional on response sensitivity analysis random effects with a weakly informative prior but did not provide the comparisons against placebo as needed by the economic model. The ERG therefore reran these analyses (results are reported in section 3.3.6.2).

3.1.7.5.4 NMA sensitivity analyses including Asian trials

The company conducted a series of NMAs in which the Asian trials were included, for the induction phase and for the combined induction and maintenance phases using the 1-year NMA conditional on response approach. No specific methods are reported for these NMA sensitivity analyses, so it is unclear whether they used fixed effects or random effects models. Network diagrams have not been provided for these analyses. The NCT02039505 trial had longer duration of the induction and maintenance phases than all other trials (see section 3.1.7.3.4) but the eligibility of this trial for inclusion in the sensitivity analyses is not discussed. The company do not discuss whether adding the Asian trials increased or reduced heterogeneity, or whether there was any inconsistency in the networks. The 1-year NMA conditional on response analyses involved pooling doses of comparators, but the rationale for this is not explained.

The induction phase results are reported in CS Appendix Tables 74 to 79 for clinical remission, clinical response and mucosal healing in non-biologic and biologic failure subgroups. The 1-year NMA conditional on response results are reported in CS Tables 80 to 82 for the same three outcomes, but only in the non-biologic failure subgroup. The company do not discuss the results of any of these analyses.

The ERG believes that these sensitivity analyses including Asian trials are unlikely to be valid, as the company misclassified the Japic CTI-060298 trial (see section 3.1.7.2) and so presumably would have applied inappropriate assumptions for this trial in their NMA calculations. We assume that the errors identified in the main 1-year NMAs conditional on response, noted in the sections above, would also affect these analyses. It was not feasible for us to check and rerun these analyses in the time available. We suggest that the results presented in CS Appendix Tables 74 to 82 are unreliable and could be misleading.

3.1.7.5.5 Additional NMA analyses conducted by the ERG

The company's economic model base case takes absolute data on clinical remission and response directly from the individual arms of the clinical trials (see section 4.3.4.2 below). This circumvents the within-trial group randomisation of the RCTs meaning that the data effectively become observational in nature and potentially prone to selection bias. It is preferable to use NMA results to inform the model where possible to protect within-trial randomisation and minimise risks of bias.

The ERG explored a scenario in the NMA and economic model which assumes there is no relative difference in the carry-over effect between treatments. For brevity, we refer to this as the ERG maintenance-only NMA. Note that this scenario does not assume that there is no carry-over effect, but by using the re-randomised placebo maintenance arms the ERG scenario assumes any carry-over is similar across placebo arms.

To be able to include both re-randomised and treat-through trials, the ERG's maintenance-only NMA scenario followed the methodology described in TA547. The maintenance data from re-randomised trials for patients who responded to induction therapy (for both active treatment and placebo) were used directly from the trials without adjustment, whilst the data from treat through trials were imputed based on the assumption that the number of induction responders is a proxy for the number of patients entering maintenance. Calculations and assumptions are described in Appendix 7. Data included in the model are reported in Appendix 8. The VARSITY abstracts did not report a split between non-biologic failure and biologic failure and this trial was therefore not included.

The ERG maintenance-only NMA scenario pooled doses across treatments and used a random effects model with the same weakly informative prior used by the company for consistency. Whilst the use of an informative prior is not ideal, this was a trade-off between its use or fixed effects to adequately capture uncertainty in a heterogeneous set of studies. The evidence networks are shown in Figure 11 and Figure 12.

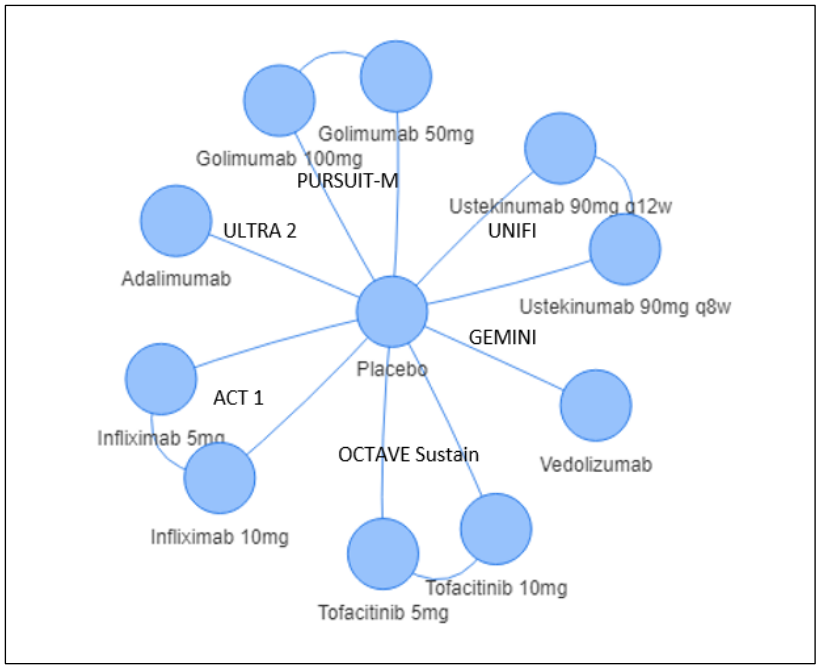


Figure 11 Non-biologic failure evidence network for maintenance-only scenario

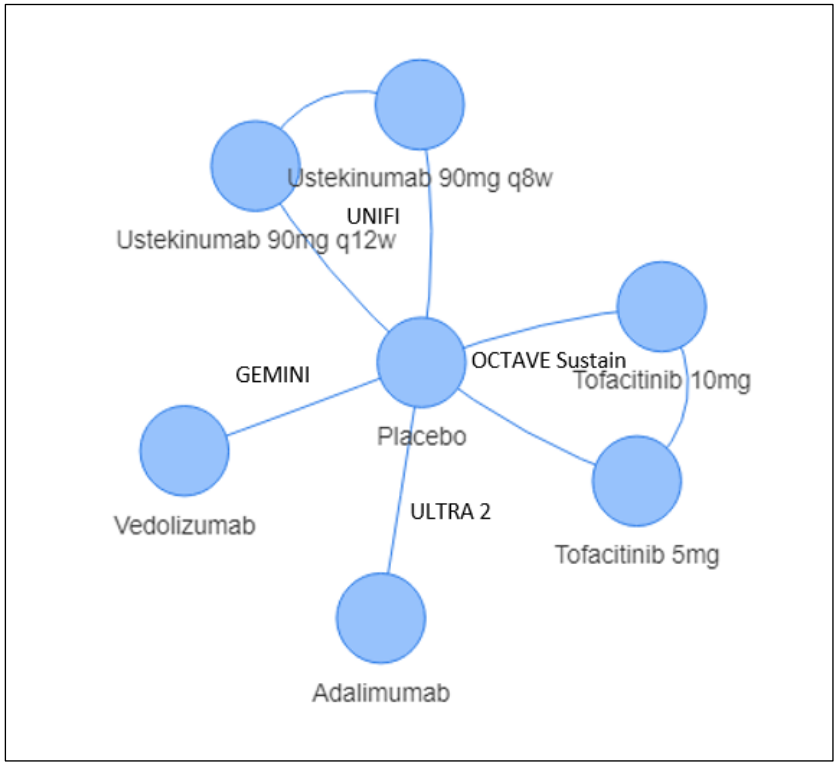


Figure 12 Biologic failure evidence network for maintenance-only scenario

This should be interpreted as an extreme scenario whereby placebo arms are equivalent inferring no relative differences in carry-over effects between treatments. Results are presented below in section 3.3.6.3 (Table 32 and Table 33) and these inform an ERG maintenance-only NMA scenario in the economic model (section 4.4.3).

3.1.7.5.6 Dose regimen pooling in the maintenance phase

For the maintenance phase in NMAs the CS states that the standard and escalated doses (i.e., for ustekinumab, q8w and q12w) were pooled in the non-biologic failure subgroup to increase statistical power as no dose-response relationship was observed. However, doses were not pooled for the biologic failure subgroup as a potential dose-response relationship was observed (CS section B.2.9). The CS does not explain how a dose-response effect was defined and does not explicitly say which therapies the pooling was applied to. The company explain in response to clarification question A22 that the dose-relationship was determined by comparing the proportions of patients with clinical remission and symptomatic remission at the end of maintenance across the quartile serum ustekinumab average trough concentrations through week 44 in the UNIFI Maintenance Study. These comparisons, when separated for patients who were in clinical remission at maintenance baseline and those who were not in clinical remission at maintenance baseline suggest a dose-response relationship was present in the latter group only (Figure 18 provided in clarification response A22). The company use these findings to argue, indirectly, that since the biologic failure population is more refractory, *"it is anticipated that there are more subjects with a lower response to treatment in this population, and thus the exposure-response (and dose-response) relationships are more pronounced in the biologic failure population*". The ERG considers that this assumption is uncertain since no direct evidence has been provided to support it, and there appears to be no objective cut-off for deciding when a dose-response relationship would be sufficiently strong to preclude dose pooling. The company do not discuss whether their interpretation for ustekinumab would also apply to the standard and escalated maintenance doses for the comparator therapies.

Given the uncertainty around the company's assumption the ERG would prefer that the NMAs are run using both pooled and unpooled doses, or at least that the same approach (pooling or not pooling) is applied consistently to both the biologic and non-biologic failure subgroups. Clinical remission and response NMA results have been provided by the company based on both pooled and unpooled doses in the non-biologic failure subgroup (Tables 10 and 11 in response to clarification question A22), but these apply to the 1-year NMA model, not the 1-year NMA conditional on response.

3.1.7.6 Summary of the ERG's NMA critique

A summary of the ERG's critique of the company's NMA approach is provided in the checklist in Table 14. The company followed standard NMA procedures, supported with

additional assumptions and calculations to enable treat-through and re-randomised trials to be included in the NMAs. The main issues encountered by the ERG were lack of transparency in how calculations had been performed, lack of clarity around source data for the NMAs and heterogeneity of the trials included in the NMAs.

Table 14: ERG appraisal of the NMA approach

NMA methodology component	ERG response (yes/no)
Does the MS present an NMA?	Yes, a number of NMAs were run for different outcomes, population subgroups and trial phases
Are the NMA results used to support the evidence for the clinical effectiveness of the intervention?	Yes, for clinical response, clinical remission and mucosal healing, but mucosal healing results are not discussed in detail
Are the NMA results used to support the evidence for the cost-effectiveness of the intervention	Partly. The main 1-year NMAs were not used in the cost-effectiveness analysis, which instead was informed by 1-year NMAs conditional on response. Results for clinical response and clinical remission, but not mucosal healing, informed the economic analysis.
Homogeneity	
1. Is homogeneity considered?	Yes. This is considered in CS section B.2.9.3.4.4 and CS Appendix sections D1.1.7 and D1.1.8, and summarised in CS Appendix D1.5.1
2. Are the studies homogenous in terms of patient characteristics and study design?	No. The trials varied considerably in design (treat-through versus re-randomised), prior treatment exposure (handled as subgroups), duration, method of outcome assessment (central/local read), induction-to-maintenance placebo carry-over effect, etc. Some of the heterogeneity is accounted for in the analytical approach but other residual sources of heterogeneity are not
3. Is the method used to determine the presence of statistical heterogeneity adequate? (e.g. Chi-squared test, I-squared statistic)	Partly. The CS does not report assessments of heterogeneity for the induction,1-year, and 1-year conditional on response NMAs. However, CS Table 24 does report p-values for chi-squared tests of heterogeneity among the maintenance placebo arms of the four included re-randomised trials (GEMINI1, OCTAVE, PURSUIT-M, UNIFI).
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)	Partly. Some sources of heterogeneity are accounted for, e.g. by adjustments to match different trial designs, or analyses conducted by prior biologic failure subgroups, but residual sources of heterogeneity remain.
Consistency	
1. Does the analysis explicitly assess consistency?	No. Not discussed in the CS. However , the company's response to clarification question A23 indicates consistency between the direct and indirect trial evidence.
2. Does the method described include a description of the analyses/ models/ handling of potential bias/ inconsistency/ analysis framework?	Not applicable
3. Are patient or trial characteristics compared between direct and indirect evidence trials?	Not applicable

	4. If Q3 is yes, and inconsistency is reported, is this accounted for by not combining the direct and indirect evidence?	Not applicable

3.1.7.7 NMA limitations and uncertainties

As noted above, there are a number of methodological limitations with the company's NMAs and these are reported in various places in the CS and CS Appendices which make it difficult to get a clear oversight of what the key issues are and whether they have been adequately resolved. For clarity, we have summarised these issues, and their implications in the overall clinical effectiveness summary (section 3.4.3) below (Table 36 below).

3.2 Summary statement of the company's approach

Overall, we consider the company's approach to the clinical effectiveness data identification and selection to be generally appropriate (Table 15). The company's searches were fit for purpose and reasonably up-to-date and we do not believe any key trials have been missed. The selection process for including studies in NMAs is generally appropriate, conducted by independent reviewers. The number of reviewers conducting the risk of bias assessments is not reported, although we concur with most of the company's assessments. The main issue encountered by the ERG when interpreting the company's clinical effectiveness review is that the trials are not summarised as clearly as they could be, meaning that it was difficult for us to verify the sources of data used in NMAs. Additionally, the company misreported a treat-through trial as being a re-randomised trial which has implications for the validity of their analyses on Asian trials.

Table 15 Quality assessment (CRD criteria) of the CS clinical effectiveness review

Question	ERG response
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Yes. The CS reports a set of eligibility criteria for their clinical effectiveness review (CS Appendix Table 14) and a further set of more specific eligibility criteria for their NMA (summarised in section 3.1.7.2 above). The ERG agrees that the eligibility are generally appropriate (with some provisos noted in section 3.1 above), although the company has not stated whether the criteria were pre-specified or developed post-hoc.
2. Is there evidence of a substantial effort to search for all relevant research, i.e. all studies identified?	Yes. The company conducted extensive searches in appropriate bibliographic databases as well as agency websites, meeting proceedings and clinical trial registers. There are some issues with the searches and reporting of the search results (see section 3.1) but the ERG does not believe that any key trials or publications have been missed.

3. Is the validity of included studies adequately assessed?	Yes. The company assessed the risk of bias in the intervention and comparator studies using standard criteria. The company have not explained their judgements but we agree that the judgements made by the company appear broadly appropriate (discussed for UNIFI in section 3.1.4 and for comparator studies in section 3.1.7.1). An exception is that for the comparator studies there is uncertainty as to whether appropriate approaches were employed for handling missing data (Appendix 3)
4. Is sufficient detail of the included studies presented?	Yes. The individual studies are generally well reported, although some baseline characteristics data for the UNIFI trial were missing from the CSRs (provided in response to clarification question A1).
5. Are the included studies summarised appropriately?	No. Overall the included studies are well summarised. However, there are some inaccuracies in trial data reported in the CS; the company have misclassified a treat-through trial as a randomised trial; and the link between data reported in trials and those employed in company analyses is obscure for a number of analyses.

3.3 Summary of the submitted evidence

In this section we have summarised the clinical effectiveness outcomes from the UNIFI trial (sections 3.3.1 to 3.3.5) and the company and ERG NMAs (section 3.3.6), focusing on outcomes specified in the NICE scope and the company's decision problem, and those that inform the company's economic model. Where available we have presented results for the non-biologic failure and biologic failure subgroups and the whole trial (ITT) population for comparison, although only the subgroups inform the company's economic analysis. In addition to the biologic failure status subgroups, the company reported several other subgroup analyses for the UNIFI trial and these are summarised in section 3.3.5.

3.3.1 Clinical remission

As noted above in section 3.1.5, the company employed two definitions of clinical remission in the UNIFI trial – the global and US definitions. Almost all of the clinical remission results in the CS are based on the global definition, and this was the definition used in the company's NMAs. Clinical remission results presented here are based on the global definition.

Rates of clinical remission at the end of induction were statistically significantly higher in the ustekinumab ~6mg/kg and 130mg groups than the placebo group, for both the non-biologic failure and biologic failure subgroups and the ITT population (Table 16). Rates of remission were higher for non-biologic failure participants than those with biologic failure, but did not

differ between the two ustekinumab doses within each group (~6 mg/kg is the regimen relevant to the company's proposed marketing authorisation in the draft SmPC).

Table 16 UNIFI: clinical remission at end of induction (week 8)

Trial population	Placebo	Ustekinumab ~6mg/kg	Ustekinumab 130mg
Non-biologic failure subgroup, % (n/N)	9.5 (15/158)	18.6 (29/156); p=0.022	19.9 (31/156); p=0.009
Biologic failure subgroup, % (n/N)	1.2 (2/161)	12.7 (21/166); p<0.001	11.6 (19/164); p<0.001
Primary efficacy analysis set (ITT population), % (n/N)	5.3 (17/319)	15.5 (50/322); p<0.001	15.6 (50/320); p<0.001
P-values are for chi-squared test versus placebo. Source: CS Figures 12 and 17			

At week 44 of the maintenance phase, a statistically significant greater proportion of participants treated with both ustekinumab maintenance doses, in both the non-biologic failure and biologic failure subgroups, and in the overall ITT population, were in clinical remission than those treated with maintenance placebo (Table 17). As noted in CS section B.2.7.2.1, the biologic failure patients treated with maintenance ustekinumab q8w had higher rates of remission than those treated with q12w, while such a pattern is not evident for the non-biologic failure patients.

Table 17 UNIFI: clinical remission at end of maintenance (week 44)

Trial population	Placebo	Ustekinumab 90 mg SC q8w	Ustekinumab 90 mg SC q12w
Non-biologic failure subgroup, % (n/N)	31 (27/87)	48.2 (41/85); p=0.024	49.0 (50/102); p=0.020
Biologic failure subgroup, % (n/N)	17 (15/88)	39.6 (36/91); p<0.001	22.9 (16/70); p=0.044
Primary efficacy analysis set (ITT population), % (n/N)	24 (42/175)	43.8 (77/176); p<0.001	38.4 (66/172); p=0.002
Subgroup analyses of clinical remission at maintenance week 44			
Maintenance of clinical remission through week 44 among patients who had achieved clinical remission at maintenance baseline,[a] %	37.8	57.9; p=0.069	65.0; p=0.011
P-values where reported are for chi-squared test versus placebo [a] Denominators and numerators not reported. Source: CS Table 16, CS Figures 14 and 19			

Among participants who had clinical remission at maintenance baseline, proportionally more of those treated with both ustekinumab maintenance doses maintained clinical remission at

the end of the maintenance period than those treated with placebo, although only the ustekinumab q12w arm reached statistical significance (Table 17). The CS reports that among the delayed responders to ustekinumab, who were all treated with the ustekinumab q8w regimen in the trial's non-randomised arm, [redacted] achieved clinical remission at maintenance week 44; however this was based on the US definition of clinical remission (CS Table 19).

3.3.2 Clinical response

Rates of clinical response at the end of induction were statistically significantly higher in the ustekinumab ~6mg/kg and 130mg groups than in the placebo group, for both the non-biologic failure and biologic failure subgroups and the ITT population (Table 18). The clinical response rates were slightly higher in the ~6mg/kg group than the 130mg group and slightly higher in the non-biologic failure than the biologic failure subgroup.

Table 18 UNIFI: clinical response at end of induction (week 8)

Trial population	Placebo	Ustekinumab ~6mg/kg	Ustekinumab 130mg
Non-biologic failure subgroup, % (n/N)	35.4 (56/158)	66.7 (104/156); p<0 .001	57.7 (90/156); p<0 .001
Biologic failure subgroup, % (n/N)	27.3 (44/161)	57.2 (95/166); p<0 .001	45.1 (74/164); p<0 .001
Primary efficacy analysis set (ITT population), % (n/N)	31.3 (100/319)	61.8 (199/322); p<0.001	51.3 (164/320); p<0.001
P-values are for chi-squared test versus placebo. Source: CS Figures 13 and 18			

A statistically significant greater proportion of participants treated with each ustekinumab maintenance dose had experienced a clinical response at the end of maintenance treatment at week 44 than those treated with placebo, in both the non-biologic and biologic subgroups, and in the ITT population (Table 19). As in the induction phase, response rates were higher in the non-biologic failure than the biologic failure subgroup. As noted in CS section B.2.7.2.1, the biologic failure patients treated with maintenance ustekinumab q8w had a better response rate than those treated with q12w, but this pattern is not evident for the non-biologic failure patients.

Table 19 UNIFI: clinical response at end of maintenance (week 44)

Trial population	Placebo	Ustekinumab 90 mg SC q8w	Ustekinumab 90 mg SC q12w
Non-biologic failure subgroup, % (n/N)	50.6 (44/87)	77.6 (66/85); p<0.001	76.5 (78/102); p<0 .001
Biologic failure subgroup, % (n/N)	38.6 (34/88	64.8 (59/91); p<0 .001	55.7 (39/70); p=0.008

Primary efficacy analysis set (ITT population), % (n/N)	44.6 (78/175)	71.0 (125/176); p<0.001	68.0 (117/172); p < 0.001
Delayed responders to UST induction, % (n/N)	Not applicable	[redacted]	Not applicable
P-values are for chi-squared test versus placebo. Source: CS Table 19, CS Figures 15 and 20			

The CS reports that among the delayed responders to ustekinumab (who were treated with the ustekinumab q8w regimen in the trial's non-randomised arm), [redacted] had maintained a clinical response to the ustekinumab maintenance treatment at maintenance week 44.

3.3.3 Other secondary outcomes

Table 20 shows the UNIFI Induction Study results for the other measured secondary outcomes in the trial that are relevant to the NICE scope and the company's decision problem. Rates of endoscopic and histologic healing, and mucosal healing (which combines endoscopic and histologic healing) were statistically significantly higher in both the ~6mg/kg and 130mg ustekinumab arms than in the placebo arm, but were similar for the two ustekinumab doses (not tested statistically). As would be expected, rates of hospitalisations and surgery related to UC were relatively low and were more frequent in the placebo group, with no surgery occurring up to 8 weeks in the ustekinumab groups.

Table 20 UNIFI: other secondary outcomes at end of induction (week 8)

Outcome	Placebo	Ustekinumab ~6mg/kg	Ustekinumab 130mg
Endoscopic healing[a], %	13.8%	27.0%; p<0.001	26.3%; p<0.001
Mucosal healing (combined endoscopic and histological healing)[b], %	8.9	18.4; p<0.001	20.3%; p<0.001
Histologic healing[b]	20.4	32.6; p<0.001	35.3; p<0.001
UC -related hospitalisations[b], %	4.4%	1.6; p = 0.0348	0.6; p = 0.002
UC -related surgery[b], %	0.6	0	0
Corticosteroid free clinical remission, %	Not reported	Not reported	Not reported
P-values where reported are for chi-squared test versus placebo [a] Primary efficacy analysis set (ITT population) [b] Analysis population unclear Source: CS Table 12; CS sections B.2.6.1.3 and B.2.6.1.5			

At maintenance week 44 the rates of endoscopic, histologic and mucosal healing, as well as corticosteroid-free remission, were higher for both the q8w and q12w ustekinumab regimens than for the placebo group, with the differences for endoscopic healing and corticosteroid-free remission being statistically significant (p-values for histologic and mucosal healing are

not reported) (Table 21). The results for UC-related hospitlisations show the pooled rate for both q8w and q12w ustekinumab groups was lower than for the placebo group, but not reaching statistical significance (sample size is small). It is unclear why the company have pooled the two ustekinumab regimens for this outcome for the Maintenance Study, but reported them separately for the Induction Study. It is also unclear why rates of UC-related surgery have been reported for the Induction Study but not the Maintenance Study.

Table 21 UNIFI: other secondary outcomes at end of maintenance (week 44)

Outcome	Placebo	Ustekinumab 90 mg SC q8w	Ustekinumab 90 mg SC q12w
Endoscopic healing[a], %	28.6	51.1; p<0.001	43.6; p=0.002
Mucosal healing (combined endoscopic and histological healing)[b], %	23.4	44.9	38.4
Histologic healing[b], %	31.4	56.3	51.2
Corticosteroid free clinical remission[b], %	23.4	42.0; p<0.001	37.8; p=0.002
UC-related hospitalisations[b], %	5.7 (n=10)	2.3 (n=8); p=0.071	
UC-related surgery	Not reported	Not reported	Not reported

P-values where reported are for chi-squared test versus placebo
[a] Primary efficacy analysis set (ITT population)
[b] Analysis population unclear Source: CS Table 16; CS section B.2.6.2.3

The CS does not report the Mayo or Partial Mayo scores, which CS Table 6 states are measures of disease activity (i.e. relevant to the NICE scope and the company's decision problem). However, the induction and maintenance CSRs[10,11] present results for these outcomes, which show [redacted].

CS Section B.2.7.1 states that subgroup analyses by biologic failure status (yes or no) were conducted for the endoscopic healing and mucosal healing outcomes at induction week 8 and for the endoscopic healing, corticosteroid-free clinical remission, maintenance of clinical response and mucosal healing at maintenance week 44. Neither the CS nor CS Appendix E (subgroup analyses) provide the results for the subgroup analyses by biologic failure status at induction week 8 for these outcomes. Although CS Section B.2.7.2.1 provides a brief overall summary of these subgroup results at maintenance week 44, this does not mention the individual outcomes. It states that, generally, proportionally more participants in both subgroups who were treated with each maintenance dose of ustekinumab achieved each outcome than those treated with maintenance placebo. The CS also notes that there was a

trend across outcomes for the biologic failure patients treated with maintenance ustekinumab q8w to do better than those treated with q12w, while no such trend was observed for the non-biologic failure patients.

> **ERG conclusion:** Ustekinumab improved rates of clinical remission and clinical response at induction week 8 and maintenance week 44 compared to the respective placebo arms, both for the non-biologic failure and biologic failure subgroups and for both the q8w and q12w maintenance dose regimens. At the end of induction, rates of remission and response were higher in the non-biologic failure subgroup than the biologic failure subgroup. At the end of maintenance, rates of remission and response were higher in the q8w arm than the q12w arm in the biologic failure subgroup but did not differ between the two dose regimens in the non-biologic failure subgroup. Results for mucosal healing were also favourable for ustekinumab but were not reported by subgroup.

3.3.4 Health related quality of life

Three measures of health-related quality of life were taken in the UNIFI trial: EQ-5D-5L, IBDQ and SF-36. The EQ-5D-5L results inform the utility values for a scenario analysis in the company's economic model, while the IBDQ and SF-36 results do not inform the economic model.

3.3.4.1 EQ-5D (5L)

Changes in the overall EQ-5D index and health state scores on the EQ-5D visual analogue scale (VAS) during the UNIFI trial induction and maintenance phases are summarised in Table 22. The company provided some p-values in the source tables for these data, but it is unclear to which comparisons they relate, so we do not comment on the statistical significance of the findings here.

At end of induction (week 8), all groups had experienced improvements (i.e. increases) in their mean and median index EQ-5D scores from induction baseline levels, with the smallest improvement being in the placebo group and the largest in the ustekinumab ~6mg/kg group. Mean and median VAS scores also improved in all groups, with the largest improvement being in both ustekinumab groups compared to placebo.

At end of maintenance (week 44), the maintenance placebo group had experienced a decrease in their mean EQ-5D index scores from maintenance baseline values, while the q8w maintenance ustekinumab group experienced a slight improvement (0.025) and the

q12w group experienced a marginal improvement (0.008). The mean VAS scores improved in the ustekinumab q8w group but decreased in the q12w and placebo groups, with the largest decrease being in the placebo group. The median values of the EQ-5D index and VAS scores also decreased (worsened) in the placebo group, but showed no change from baseline in the ustekinumab groups.

In summary, these results suggest that both the ~6mg/kg and 130mg induction dose regimens of ustekinumab improved the trial participants' HRQoL at 8 weeks compared to placebo, with no clear difference between the regimens. As would be expected, in the maintenance phase the higher-dose regimen (q8w) had a larger positive impact on participants' HRQoL at 44 weeks than the lower-dose regimen (q12w), with both ustekinumab regimens being better than the placebo..

Table 22 EQ-5D scores during UNIFI trial induction and maintenance

EQ-5D measure	Placebo N=319	Ustekinumab ~6mg/kg N=322	Ustekinumab 130mg N=320	Placebo N=175	Ustekinumab q8w N=176	Ustekinumab q12w N=172	Combined ustekinumab groups N=348
EQ-5D index mean (SD), [median]	**Induction Baseline**			**Maintenance baseline**			
	0.66 (0.208) [0.71]	0.67 (0.195) [0.71]	0.67 (0.204) [0.71]	0.820 (0.1516) [0.837] [a]	0.801 (0.1588) [0.795] [b]	0.810 (0.1563) [0.795]	0.806 (0.1574) [0.795] [b]
	Change, baseline to week 8 [c]			**Change, maintenance baseline to week 44**			
	0.04 (0.182) [0.01]	0.11 (0.172) [0.06]	0.09 (0.182) [0.06]	-0.048 (0.1587) [-0.019] [a]	0.025 (0.1674) [0.000] [b]	0.008 (0.1656) [0.000]	0.017 (0.1665) [0.000] [b]
Health state VAS, mean (SD) [median]	**Induction baseline**			**Maintenance baseline**			
	55.11 (20.815) [60]	55.76 (19.333) [55 [d]]	54.14 (20.545) [55 [d]]	75.2 (13.57) [78] [a]	73.2 (16.24) [80] [b]	75.7 (16.28) [80]	74.4 (16.28) [80] [b]
	Change, baseline to week 8 [c]			**Change, maintenance baseline to week 44**			
	5.71 (19.584) [5]	13.51 (18.447) [10 [d]]	13.64 (20.394) [10 [d]]	-7.7 (18.75) [-5.0] [a]	2.4 (17.28) [0.0] [b]	-2.2 (19.87) [0.0]	0.1 (18.72) [0.0] [b]

[a] Sample size n=2 less than the ITT population

[b] Sample size n=1 less than the ITT population

[c] p<0.001. [d] p≤0.001. Source: CS Table 15 and CS Appendix Table 145 (induction); company clarification response Appendix Q Table 6 (maintenance)

3.3.4.2 IBDQ

The company report changes in the IBDQ using two thresholds for a minimum clinically important difference (16 or 20 points). As explained in response to clarification question A6, the 16-point threshold has been employed in some recent trials of biologics in UC, but a recent study concluded that a more stringent 20-point threshold is appropriate when applying the IBDQ to UC.

Changes in IBDQ scores at week 8 of the Induction Study are summarised in Table 23. The median IBDQ score, and the proportion of participants with a clinically meaningful improvement in the IBDQ score, assessed according to both the 16-point and 20-point thresholds, increased from baseline to week 8 and the increase was statistically significantly larger for both the ~6mg/kg and 130mg ustekinumab groups than for the placebo group. These changes indicate a greater improvement in HRQoL in the ustekinumab groups than the placebo group. There were no clear differences in these outcomes between the two ustekinumab induction dose groups (these comparisons were not tested statistically).

Table 23 also shows that the proportion of patients with a clinically meaningful improvement in their IBDQ scores from maintenance baseline to week 44 of the Maintenance Study was statistically significantly larger in the ustekinumab q8w and q12w groups than the maintenance placebo group.

Table 23 Changes in IBDQ scores during UNIFI trial induction and maintenance

Measurement time	**IBDQ overall score**	**Induction placebo N=319**	**Induction ustekinumab ~6mg/kg N=322**	**Induction ustekinumab 130mg N=320**
Induction study Change from induction baseline to week 8	Participants with 20-point improvement, %	37.0	62.1; p<0.001	61.3; p<0.001
	Participants with 16-point improvement, %	44.2	68.6; p<0.001	66.6%; p<0.001
	Median score change	10.0 (n=317) [a]	31.0; p<0.001 (n=316) [a]	31.5; p<0.001 (n=321) [a]
Maintenance Study		**Maintenance placebo [b]**	**Ustekinumab q8w [b]**	**Ustekinumab q12w [b]**
Change from induction baseline to maintenance week 44	Participants with 20-point improvement, %	42.9	69.9; p<0.001	66.3; p<0.001
	Participants with 16-point improvement, %	47.4	73.3; p<0.001	68.6; p<0.001
P-values refer to ANCOVA and chi-square tests of comparison against placebo				

[a] Not the full ITT population: Data are from CS Table 12 where the company have mixed up the N-values for the two ustekinumab arms; n=321 for the 130mg arm is not correct as it exceeds the number randomised; unclear whether n=316 for ~6mg arm is correct.
[b] Sample sizes not reported.
Sources: CS Tables 12 and 13; CS Section B.2.6.2.4; CS Appendix Tables 142 and 143; company clarification response A4

3.3.4.3 SF-36

The company report results for the physical and mental subscales of the SF-36 (PCS and MCS respectively) but not the overall SF-36 scores. As shown in Table 24, results for the PCS and MCS subscales of the SF-36 show a similar pattern to those of the IBDQ, for both the induction and maintenance phases of the UNIFI trial. A statistically significant higher proportion of participants achieved clinically important improvements of ≥5 points on each SF-36 subscale in the ustekinumab ~6mg/kg and 130mg induction dose groups, and in the q8w and q12w maintenance regimen groups, than those in the respective induction and maintenance placebo groups.

Table 24 Changes in SF-36 scores during UNIFI trial induction and maintenance

Measurement time	SF-36 score	Induction placebo N=319	Induction ustekinumab ~6mg/kg N=322	Induction ustekinumab 130mg N=320
Induction Study Change from induction baseline to week 8	Participants with ≥ 5- point improvement in PCS, %	26	45.3; p<0.001	48.3; p<0.001
	Participants with ≥ 5- point improvement in MCS, %	31.3	44.4; p<0.001	43.9; p<0.001
Maintenance Study		**Maintenance placebo [a]**	**Ustekinumab q8w [a]**	**Ustekinumab q12w [a]**
Change from induction baseline to maintenance week 44	Participants with ≥ 5- point improvement in PCS, %	30.3	53.4; p<0.001	50.0; p<0.001
	Participants with ≥ 5- point improvement in MCS, %	28.6	54.0; p<0.001	47.1; p<0.001
Maintenance of improvement at maintenance week 44 among those with a ≥ 5-point improvement at	Participants with ≥ 5- point improvement in PCS, %	38.3%	62.4; p=0.002	59.5; p=0.004
	Participants with ≥ 5- point improvement in MCS, %	36.1	59.8; p=0.001	58.3; p=0.002

maintenance baseline				
MCS: mental component summary; PCS: physical component summary [a] sample sizes not reported Note: induction baseline PCS and MCS scores (not shown) are reported in CS Appendix Table 144. Sources: CS Table 14 CS Figure 16; CS section B.2.6.2.4				

3.3.4.4 Work Productivity and impairment (WPAI) scale

Only brief results from the WPAI-GH are reported by the company, in CS section B.2.6.1.5 and (for induction only) in CS Appendix Table 146. At induction week 8, participants treated with each dose of ustekinumab showed greater decreases in their scores on this measure (indicating improvement) than participants treated with placebo. At maintenance week 44, improvements were maintained for the ustekinumab groups, with some additional improvements found for the q8w group on some domains, while the placebo group experienced worsened (increased) scores on all four domains of this measure.

> **ERG conclusion:** Results of the disease-specific IBDQ are consistent with those of the generic SF-36 and EQ-5D HRQoL measures in showing that ustekinumab improved patients' HRQoL in both the induction and maintenance phases of therapy relative to the respective placebo arms, for all dose regimens, and with the differences from placebo exceeding thresholds for being clinically meaningful. The improvements in HRQoL at week 44 were marginally larger for the q8w maintenance regimen than the q12w regimen, but not reaching the threshold for being clinically meaningful.

3.3.5 Other sub-group analyses

Subgroup analyses of clinical remission by biologic failure status, which separate failures on vedolizumab from failures on other anti-TNF therapies, are reported in CS Appendix E for the UNIFI Induction Study (not reported whether these were post-hoc). Rates of remission were larger for ustekinumab than for placebo in all the subgroups tested, with no statistically significant differences between the subgroups (95% confidence intervals for the ORs overlap) (CS Appendix Figures 62 to 65).

A brief narrative synthesis of the results of subgroup analyses by induction treatment received in given in CS Section B.2.7.2. The CS comments that participants in the Maintenance Study (particularly those on the q12w regimen) who had received the 130mg

ustekinumab induction treatment or induction placebo followed by ~6mg/kg ustekinumab had a lower maintenance treatment effect. However, quantitative data are not reported and the company cautions that these analyses were based on small subgroups of participants.

CS Appendix E also reports sub-group analyses of clinical remission at induction week 8 based on participants' baseline demographic characteristics, baseline UC clinical disease characteristics, baseline ulcerative-related concomitant medication and UC-related medication history. Across the subgroups, results generally favoured treatment with ustekinumab as compared to placebo. Aside from a very brief summary statement, no subgroup analysis results are reported for the Maintenance Study in CS Appendix E.

3.3.6 NMA results

3.3.6.1 Induction NMAs

The ERG have rerun the company's induction NMA results, correcting the discrepancies noted in section 3.1.7.5.1 above (Table 25 to Table 27). Whilst the majority of our results are consistent with the company's, we identified a number of differences.

In the non-biological failure subgroup, the ERG clinical remission results are less favourable to tofacitinib compared to those in the CS. This pattern is seen using both fixed effects (Table 25) and random effects models (Table 26). In the biological failure subgroup, the ERG and Company submission results are comparable (Table 27). A random effects NMA on the biological failure population resulted in considerable uncertainty and we considered it unreliable (not presented here).

Ustekinumab and the comparators all had significantly better odds of achieving remission and response compared to placebo (i.e. background conventional therapy alone), but credible intervals are wide, overlapping for all therapies. The CS concludes that, in the induction NMAs ustekinumab ~6mg/kg demonstrated a higher likelihood of response than adalimumab and golimumab in non-biologic failure patients and higher likelihood of response than adalimumab in biologic failure patients (CS section B.2.9.5). The probabilities reported in the CS on which these conclusions are based are subject to uncertainty, but the company have not provided credible intervals for the probabilities.

Table 25 ERG and company results for induction NMA, non-biologic failure subgroup, fixed effects

Comparator	Median OR[CrI], comparator vs. PBO			
	Clinical remission		Clinical response	
	Company	ERG	Company	ERG
UST 6mg/kg	2.19 [1.14; 4.39]	2.22 [1.15; 4.42]	3.66 [2.31;5.88]	3.68 [2.32; 5.91]
UST 130mg	2.38 [1.24; 4.78]	2.41 [1.26; 4.80]	2.49 [1.58;3.96]	2.50 [1.58; 3.96]
ADA 160/80/40mg	2.21 [1.37 ; 3.67]	2.22 [1.37; 3.68]	1.89 [1.35 ; 2.65]	1.89 [1.35; 2.65]
GOL 200/100mg	2.97 [1.73 ; 5.24]	2.95 [1.74; 5.19]	2.29 [1.63 ; 3.22]	2.29 [1.64; 3.22]
INF 5mg/kg	4.44 [2.84 ; 7.10]	4.41 [2.85; 7.02]	4.11 [2.82 ; 6.02]	4.10 [2.83; 6.02]
INF 10mg/kg	3.40 [2.13 ; 5.54]	3.3 [2.14; 5.50]	3.81 [2.63 ; 5.57]	3.82 [2.62; 5.57]
TOF 10mg	2.43 [1.33 ; 4.80]	**2.25 [1.23; 4.45]**	2.70 [1.81 ; 4.04]	2.69 [1.82; 4.07]
VED 300mg[3]	4.54 [1.76 ; 14.24]	4.47 [1.77; 13.92]	3.21 [1.75 ; 6.05]	3.20 [1.76; 6.03]

ADA: adalimumab; GOL: golimumab; INF: infliximab; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

Table 26 ERG and company results for induction NMA, non-biologic failure subgroup, random effects

Comparator	Median OR[CrI], comparator vs. PBO			
	Clinical remission		Clinical response	
	Company	ERG	Company	ERG
UST 6mg/kg	2.20 (0.56)	2.21 [0.73; 6.92]	3.67 [0.47]	3.68 [1.47; 9.18]
UST 130mg	Not reported	2.40 [0.80; 7.50]	Not reported	2.50 [1.01; 6.22]
ADA 160/80/40mg	2.23 (0.40)	2.23 [1.00; 5.04]	1.88 [0.33]	1.88 [0.98; 3.60]
GOL 200/100mg	2.90 (0.45)	2.87 [1.14; 6.73]	2.22 [0.35]	2.22 [1.06; 4.25]
INF 5mg/kg	4.33 (0.36)	4.31 [2.02; 8.55]	4.12 [0.34]	4.12 [2.12; 8.09]
INF 10mg/kg	Not reported	3.41 [1.58; 7.52]	Not reported	3.82 [1.98; 7.53]
TOF 10mg	2.49 (0.45)	**2.30 [0.98; 5.99]**	2.61 [0.37]	2.61 [1.20; 5.19]

VED 300mg	4.54 (0.69)	4.42 [1.24; 19.28]	3.22 [0.51]	3.21 [1.20; 8.76]
ADA: adalimumab; GOL: golimumab; INF: infliximab; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab				

Table 27 ERG and company results for induction NMA, biologic failure subgroup, fixed effects

Comparator	Median OR[CrI], comparator vs. PBO			
	Clinical remission		Clinical response	
	Company	ERG	Company	ERG
UST 6mg/kg	13.41 [3.62; 94.58]	13.80 [3.61; 94.92]	3.58 [2.27; 5.74]	3.59 [2.28; 5.77]
UST 130mg	12.12 [3.24; 86.24]	12.42 [3.22; 85.37]	2.20 [1.39; 3.53]	2.20 [1.39; 3.55]
ADA 160/80/40mg	1.37 [0.48 ; 4.07]	1.37 [0.49; 4.12]	1.45 [0.8; 2.65]	1.44 [0.80; 2.64]
TOF 10mg	22.33 [4.04 ; 633.0]	23.06 [4.07, 801.91]	3.41 [2.23; 5.38]	3.42 [2.24; 5.34]
VED 300mg	3.76 [0.85 ; 28.67]	3.87 [0.85; 29.96]	2.52 [1.19; 5.51]	2.51 [1.20; 5.47]
ADA: adalimumab; tofacitinib; UST: ustekinumab; VED: vedolizumab				

3.3.6.2 One-year NMAs conditional on response

The ERG was able to replicate the company's models (Table 28 and Table 29). Our results are similar to those of the company, except that the ustekinumab clinical remission odds ratio is lower for the biological failure population. However, there is considerable uncertainty around these estimates.

Results of the 1-year NMAs conditional on response consistently indicate that ustekinumab and all the comparator therapies improved the odds of clinical remission and clinical response both at 8 weeks and 44 weeks compared to the respective placebo arms (i.e. the background conventional therapy). The CS concludes that, in the 1-year NMAs conditional on response, ustekinumab had a higher probability of being more effective than all the comparators (CS section B.2.9.5). The probabilities reported in the CS on which these conclusions are based are subject to uncertainty, but the company have not provided credible intervals for the probabilities.

Table 28 ERG and company results for 1-year NMA conditional on response, non-biologic failure subgroup, fixed effects model

Comparator		Median OR [CrI], comparator vs. PBO			
		Clinical remission		Clinical response	
Induction	Main-tenance	Company	ERG	Company	ERG
VED 300mg	VED 300mg pooled	4.83 [1.83; 15.2]	4.76 [1.82; 15.24]	4.17 [1.81; 10.65]	4.18 [1.82; 10.68]
INF pooled	INF pooled	3.18 [1.75; 6.16]	3.18 [1.76; 6.12]	3.8 [2.18; 6.98]	3.82 [2.18; 7.06]
GOL 200/100mg	GOL pooled	1.63 [1.03; 2.61]	1.63 [1.03; 2.59]	2.47 [1.59; 3.85]	2.47 [1.58; 3.85]
ADA 160/80/40 mg	ADA 40mg EOW	2.66 [1.33; 5.59]	2.65 [1.31; 5.57]	2.11 [1.21; 3.75]	2.11 [1.21; 3.74]
TOF 10mg	TOF pooled	3.49 [1.84; 7.26]	3.51 [1.83; 7.34]	3.46 [2; 6.27]	3.46 [2.00; 6.31]
UST 6mg/kg	UST 90mg pooled	5.57 [2.91; 11.13]	5.59 [2.92; 11.21]	6.20 [3.57; 11.04]	6.21 [3.59; 11.05]

ADA: adalimumab; EOW: every other week; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

Table 29 ERG and company results for 1-year NMA conditional on response, biologic failure subgroup, fixed effects model

Comparator		Median OR [CrI], comparator vs. PBO			
		Clinical remission		Clinical response	
Induction	Main-tenance	Company	ERG	Company	ERG
VED 300mg	VED 300mg q8w	9.53 [1.38; 148.4]	**8.88 [1.32; 144.60]**	2.97 [0.74; 12.55]	2.99 [0.75; 12.24]
VED 300mg	VED 300mg q4w	8.79 [1.19; 138.8]	**8.28 [1.15; 135.37]**	2.64 [0.6; 11.53]	2.64 [0.61; 11.43]
ADA 160/80/40 mg	ADA 40mg EOW	6.74 [1.5; 58.85]	6.77 [1.50; 58.44]	2.97 [1.13; 8.8]	2.98 [1.13; 9.01]
TOF 10mg	TOF 5mg	6.18 [1.96; 28.75]	6.17 [1.94; 27.94]	3.42 [1.65; 7.65]	3.43 [1.68; 7.77]
TOF 10mg	TOF 10mg	10.24 [3.43; 46.35]	10.25 [3.40; 45.06]	5.05 [2.51; 11.08]	5.07 [2.57; 11.26]
UST 6mg/kg	UST 90mg q12w	7.76 [2.49; 25.89]	7.89 [2.52; 26.60]	5.21 [2.33; 11.72]	5.21 [2.33; 11.65]
UST 6mg/kg	UST 90mg q8w	10.23 [3.90; 30.98]	10.33 [3.87; 31.22]	5.26 [2.64; 10.68]	5.24 [2.64; 10.54]

ADA: adalimumab; EOW: every other week; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

Table 30 ERG analysis results for 1-year NMA conditional on response, non-biologic failure subgroup, random-effects model using half-normal prior

Comparator		Median OR [CrI], comparator vs. PBO	
Induction	Maintenance	Clinical remission	Clinical response
VED 300mg	VED 300mg pooled	4.82 [1.50; 17.71]	4.20 [1.47; 12.86]
INF pooled	INF pooled	3.21 [1.34; 7.93]	3.83 [1.65; 9.14]
GOL 200/100mg	GOL pooled	1.63 [0.75; 3.56]	2.46 [1.14; 5.32]
ADA 160/80/40mg	ADA 40mg EOW	2.65 [1.04; 6.99]	2.11 [0.91; 4.94]
TOF 10mg	TOF pooled	3.51 [1.42; 9.08]	3.47 [1.50; 8.20]
UST 6mg/kg	UST 90mg pooled	5.60 [2.27; 14.15]	6.22 [2.69; 14.48]

ADA: adalimumab; EOW: every other week; GOL: golimumab; INF: infliximab; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

Table 31 ERG analysis results for 1-year NMA conditional on response, biologic failure subgroup, random-effects model using half-normal prior

Comparator		Median OR [CrI], comparator vs. PBO	
Induction	Maintenance	Clinical remission	Clinical response
VED 300mg	VED 300mg q8w	9.03 [1.19; 136.32]	2.97 [0.66; 14.04]
VED 300mg	VED 300mg q4w	8.38 [1.05; 128.12]	2.62 [0.53; 12.95]
ADA 160/80/40mg	ADA 40mg EOW	6.72 [1.30; 62.55]	2.98 [0.94; 10.37]
TOF 10mg	TOF 5mg	6.25 [1.66; 32.49]	3.43 [1.31; 9.42]
TOF 10mg	TOF 10mg	10.40 [2.87; 52.51]	5.07 [1.98; 13.72]
UST 6mg/kg	UST 90mg q12w	7.90 [2.15; 30.88]	5.21 [1.88; 14.54]
UST 6mg/kg	UST 90mg q8w	10.37 [3.24; 36.74]	5.24 [2.07; 13.48]

ADA: adalimumab; EOW: every other week; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

3.3.6.3 Additional NMA analyses by the ERG

Results for the ERG's maintenance-only NMA scenario are provided below in Table 32 (non-biologic failure) and Table 33 (biologic failure). As the networks are star-shaped the median relative effects closely resemble those from the trial data, with ustekinumab being less favourable given the high placebo response rate. This should be interpreted as an extreme scenario whereby placebo arms are equivalent inferring no relative differences in carry-over effects between treatments. These NMA results are used to inform an ERG maintenance-only scenario analysis in the economic model (section 4.4.3).

Table 32 ERG maintenance-only NMA scenario analysis, non-biologic failure, random effects model using half-normal prior

Comparator		Median OR [CrI], comparator vs. PBO	
		Clinical remission	Clinical response
Induction	**Maintenance**		
VED 300mg	VED 300mg pooled	3.86 [1.57; 9.64]	4.34 [1.83; 10.43]
INF pooled	INF pooled	1.80 [0.67; 5.07]	2.29 [0.91; 5.85]
GOL 200/100mg	GOL pooled	1.79 [0.83; 3.89]	2.08 [0.98; 4.40]
ADA 160/80/40mg	ADA 40mg EOW	1.47 [0.55; 3.97]	1.31 [0.52; 3.31]
TOF 10mg	TOF pooled	6.25 [2.56; 15.94]	4.67 [2.08; 10.58]
UST 6mg/kg	UST 90mg pooled	2.13 [0.93; 4.89]	3.30 [1.44; 7.59]

ADA: adalimumab; EOW: every other week; GOL: golimumab; INF: infliximab; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

Table 33 ERG maintenance-only NMA scenario analysis, non-biologic failure, random effects model using half-normal prior

Comparator		Median OR [CrI], comparator vs. PBO	
		Clinical remission	Clinical response
Induction	**Maintenance**		
VED 300mg	VED 300mg pooled	12.16 [2.72; 96.06]	4.53 [1.46; 15.58]
ADA 160/80/40mg	ADA 40mg EOW	3.17 [0.70; 18.38]	2.85 [0.80; 10.98]
TOF 10mg	TOF pooled	3.61 [1.39; 9.85]	6.59 [2.69; 16.83]
UST 6mg/kg	UST 90mg pooled	2.37 [0.97; 5.93]	2.50 [1.10; 5.71]

ADA: adalimumab; EOW: every other week; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab

3.3.7 Adverse events

The company provide data on the incidence of adverse events in the UNIFI trial safety analysis population in CS Tables 32 and 33, and in CS Appendix F, summarised in section 3.3.7.1 below. The company also conducted four induction-phase safety NMAs for overall adverse events, serious adverse events, overall infections, and serious infections (CS section D2.2). These safety NMAs do not inform the economic analysis (see section 3.3.7.2 below).

The only adverse event that informs the company's economic model is serious infections, on the grounds of the high costs associated with treating these. However, the serious infections data from the UNIFI trial and from the company's serious infections induction NMA do not inform the economic model. Instead, the company has taken serious infections data from a real-world observational study of serious infections in people with psoriasis treated with ustekinumab (PSOLAR). The company's rationale for this is discussed and critiqued in section 3.3.7.3 below.

3.3.7.1 Summary of adverse events in the UNIFI trial

CS Table 32 (reproduced in Table 34 below) summarises the adverse events that occurred during the induction and maintenance treatment phases of the UNIFI trial. The incidence of adverse events was largely comparable between the ustekinumab and placebo arms, or higher in the placebo arms than the ustekinumab arms. Overall, proportionally more participants treated with maintenance ustekinumab 90 mg q8w experienced an adverse event than those treated with ustekinumab 90 mg q12w, particularly any infection. One death occurred during the trial, in the induction ustekinumab ~6 mg/kg group.

Table 34 Summary of adverse events in UNIFI induction and maintenance phases (safety analysis set)

Events, n or n(%)	Induction			Maintenance		
	Placebo	UST 130 mg	UST ~6 mg/kg	Placebo	UST 90mg q12w	UST 90mg q8w
Any AE	153 (48.0)	133 (41.4)	160 (50.0)	138 (78.9)	119 (69.2)	136 (77.3)
Serious AE	22 (6.6)	12 (3.7)	10 (3.1)	17 (9.7)	13 (7.6)	15 (8.5)
Most frequent AE						
Worsening of UC	18 (5.6)	9 (2.8)	7 (2.2)	50 (28.6)	19 (11.0)	18 (10.2)
Nasopharyngitis	NR	NR	NR	28 (16.0)	31 (18)	26 (14.8)
Headache	14 (4.4)	22 (6.9)	13 (4.1)	7 (4.0)	11 (6.4)	18 (10.2)
Arthralgia	2 (0.6)	3 (0.9)	6 (1.9)	15 (8.6)	15 (8.7)	8 (4.5)

Infections, n (%)						
Any infection [a]	48 (15.0)	51 (15.9)	49 (15.3)	81 (46.3)	58 (33.7)	86 (48.9)
Serious infection [a]	4 (1.3)	2 (0.6)	1 (0.3)	4 (2.3)	6 (3.5)	3 (1.7)
AE of special interest						
Malignancies (excluding non-melanoma skin cancer)	0	0	0	0	1	1
Possible anapyhlatic and possible delayed hypersensitivity	1	0	0	0	0	0
Cardiovascular events [b]	1	0	0	0	0	0
Death [c]	0	0	1	0	0	0
AE leading to discontinuation [d]	NA	NA	NA	20 (11.4)	9 (5.2)	5 (2.8)
Abnormal laboratory results	NR	NR	NR	1	0	0

AE: adverse events; MACE: major adverse cardiovascular events; NA: not applicable as patients received a single IV infusion at week 0 and therefore could not be discontinued from further induction drug administration; NR: not reported; UST: ustekinumab

[a] Infection as assessed by the investigator.

[b] Among all treated patients, serious MACE (ie, nonfatal myocardial infarction, nonfatal stroke, and cardiovascular death)

[c] There was 1 death reported for a patient who was a delayed ustekinumab induction responder and who was receiving ustekinumab q8w. The cause of death was attributed to acute respiratory failure that occurred during thyroid surgery for a multinodular goiter.

Source: Direct reproduction of CS Table 32 with ERG edits

The company do not mention whether any longer-term safety data for ustekinumab in UC would be available from the UNIFI ongoing long-term extension study (which is mentioned only briefly, see section 3.1.3.5 above).

3.3.7.2 Induction NMAs of adverse events

The company ran four induction-phase NMAs, for overall adverse events, serious adverse events, overall infections, and serious infections (CS Appendix D2.2). Analyses were based on the whole safety population (i.e. not distinguishing non-biologic failure and biologic failure subgroups) which is reasonable given the overall rarity of many adverse events.

A key limitation of the induction phase NMAs is the short duration of the induction phase (6-8 weeks). However, the company considered that 1-year safety NMAs that cover both induction and maintenance would not be appropriate, due to different definitions of the

placebo safety population across trials, differences in trials' eligibility criteria, and lack of information to correct for these factors (CS Appendix D2.2; with further explanation given in clarification question response A21). The CS does not discuss whether adverse event NMAs based only on the maintenance phase of trials would be feasible or appropriate. We note that, due to differences in the treat-through and re-randomised study designs, adverse event rates in some trials are not separable for the induction and maintenance phases. There are thus insufficient data for infliximab to be included in the induction serious infections NMA (CS Appendix Table 68). Maintenance-only or 1-year serious infections NMAs would also not be able to include all the relevant comparators (unless data are adjusted or imputed). Furthermore, NMAs of serious infections are problematic because the low incidence of events, including zero event rates in some trial arms, inflates the statistical heterogeneity (also identified as a problem in the NICE TA547 appraisal of tofacitinib[9]). Overall, we agree with the company that results of safety NMAs that requiring relative comparisons against placebo are not straightforward to interpret. The company's economic model requires data on serious infections (section 4.3.4.5), but these are not taken from the serious infections NMA. Instead the company has sourced data on the incidence of serious infections from an observational study, as discussed and critiqued below (section 3.3.7.3).

Due to the limitations of the company's four adverse event induction NMAs and the fact that they do not inform the economic model we have not attempted to check or validate the results of these NMAs reported in CS Appendix Tables 67 and 68.

3.3.7.3 Serious infections – observational data

The company provide a brief qualitative summary of some observational studies that report safety of ustekinumab, including the incidence of serious infections, in Crohn's disease and psoriasis (CS section B.2.10.7). The CS reports, without providing a rationale, that serious infections data for their economic model were sourced from the PSOLAR registry study[32] in psoriasis (CS section B.3.3.3). We note that most participants in the PSOLAR registry (90%) were enrolled in North America and Canada. A British registry study (BADBIR) [60] also reports serious infections for psoriasis patients who received ustekinumab, but the CS does not mention this or discuss whether it would be an appropriate source of data (the length of follow-up is not reported in the BADBIR publication but it appears that at least 50% of patients completed 2 years).

The company do not explain why they have not used serious infections data directly from the clinical trials included in their clinical effectiveness review, nor why they preferred serious infections data from ustekinumab-treated patients with psoriasis rather than Crohn's disease.

As we show in Table 35 below, most of the UC trials reported serious infections. The ERG's clinical experts suggested that psoriasis is a more appropriate reference for serious infections than Crohn's disease since Crohn's disease patients are prone to rectal infections. However, we note that while the anticipated licensed dose of ustekinumab in UC is the same as in Crohn's disease, it is usually lower for psoriasis (variable in clinical trials but often 45mg at 12-week intervals as a maintenance regimen)[61]. This lower dosing might lead to underestimation of the rate of serious infections compared to the dose regimen used in UC. The PSOLAR registry does have a longer follow-up (median 1.6 years, i.e. 83 weeks) compared to the UC trials (Table 35), but this is still short in relation to the chronic nature of UC. The rates of serious infections among patients treated with ustekinumab for psoriasis were 0.83 per 100 patient-years in the PSOLAR study[32] and 15.1 per 1000 patient-years in the BADBIR study.[60]

Table 35 Serious infections reported in trials compared with company estimates of serious infections reported in CS Table 49

Drug	Trial	Regimen	N	Serious infections in trial		Serious infections in CS Table 49
				Induction	Maintenance	
ADA	ULTRA1 [42]	80/40 mg	130 [a]	1.5% (8 wk)	- - - -	1.97% (PSOLAR [32])
		160/80/40 mg	130 [a]	0% (8 wk)	- - - -	
	ULTRA1 [43]	Any dose (160, 80, 40 mg)	557 [b]	3.1% (0-51 wk)		
	ULTRA2 [44]	160/80/40 mg	257	1.6% (0-52 wk) [c]		
GOL	PURSUIT-SC [46]	100/50 mg	71	0% (6 wk)	- - - -	2.49% (assumed)
		200/100 mg	331	0.3% (6 wk)	- - - -	
		400/200 mg	332	0.9% (6 wk)	- - - -	
	PURSUIT-M [48]	50 mg	154	- - - -	3.2% (54 wk)	
		100 mg	154	- - - -	3.2% (54 wk)	
INF	ACT1 [49]	5 mg/kg	122	2.5% (0-54 wk)		2.49 (PSOLAR [32])
		10 mg/kg	122	6.6% (0-54 wk)		
	ACT2 [49]	5 mg/kg	121	1.7% (0-30 wk)		
		10 mg/kg	120	2.5% (0-30 wk)		
	Japic CTI-060298 [39]	5 mg/kg	104	1.0 (0-38 wk)		
	Jiang 2015 [50]	5 mg/kg	41	2.4% (0-30 wk)		
	Probert 2003 [51]	5 mg/kg	23	0% (8 wk) [d]	- - - -	
TOF	OCTAVE 1 [53]	10 mg	476	1.3% (8 wk)	- - - -	0.83 (assumed)
	OCTAVE 2 [53]	10 mg	429	0.2% (8 wk)	- - - -	
	NCT00787202 [52]	10mg	33	6.0% (8 wk)	- - - -	
	OCTAVE	5 mg	197	- - - -	1.0% (52 wk)	

	Sustain [53]	10 mg	198	- - - -	0.5% (52 wk)	
UST	UNIFI (CS Table 32)	130 mg	320	0.6% (8 wk)	- - - -	0.83 (PSOLAR [32])
		6 mg/kg	322	0.3% (8 wk)	- - - -	
		90 mg q12w	172	- - - -	3.5% (44 wk)	
		90 mg q8w	176	- - - -	1.7% (44 wk)	
VED	GEMINI1 [54]	300 mg	225	0.4% (6 wk)	- - - -	0.83 (assumed)
		300 mg q8w	122	- - - -	2.5% (52 wk)	
		300 mg q4w	125	- - - -	1.6% (52 wk)	
	NCT02039505 [55] both	300 mg	164	0.6% (6 wk)	- - - -	
		300 mg	41	- - - -	2.4% (54 wk)	

ADA: adalimumab; GOL: golimumab; INF: infliximab; TOF: tofacitinib; UST: ustekinumab; VED: vedolizumab;

- - - -: Induction or maintenance phase not reported in the trial; wk: weeks

[a] Patients randomised after protocol amendment 3 in ULTRA1 trial ("ITT-A3" population).

[b] All patients randomised in ULTRA1 trial who received any dose of ADA before and after protocol amendment 3 ("ITT-E" population).

[c] Reported as "serious infectious adverse events"

[d] Serious infections not explicitly reported but paper states there were no serious adverse events in this group

Data from the trials and CS show that for adalimumab, golimumab, ustekinumab and vedolizumab rates of serious infections were higher in maintenance/full study than induction, so looking at induction-only rates would underestimate serious infection rates. As shown in Table 35, the PSOLAR data underestimate the rate of serious infections in the maintenance and 1-year trials for golimumab, ustekinumab and vedolizumab.

> **ERG conclusion:** Adverse events data from the UNIFI trial show that ustekinumab is relatively well-tolerated, and although the majority of patients experienced adverse events, fewer than 10% of these were serious. To inform the economic model, the company uses serious infections data from patients receiving ustekinumab in a psoriasis registry instead of from the UC trials. The registry data provide marginally longer follow-up but appear to underestimate the rate of serious infections in the maintenance phase for ustekinumab and several comparators. A general limitation is the short-term nature of the safety data for ustekinumab (<2 years).

3.4 Summary of the clinical effectiveness evidence

3.4.1 UNIFI trial results

Ustekinumab improved rates of clinical remission and clinical response at induction week 8 and maintenance week 44 compared to the respective placebo arms, both for the non-biologic failure and biologic failure subgroups and for both the q8w and q12w maintenance

dose regimens. At the end of induction, rates of remission and response were higher in the non-biologic failure subgroup than the biologic failure subgroup. At the end of maintenance, rates of remission and response were higher in the q8w arm than the q12w arm in the biologic failure subgroup but did not differ between the two dose regimens in the non-biologic failure subgroup. Results for mucosal healing were also favourable for ustekinumab but were not reported by subgroup.

Results of the disease-specific IBDQ are consistent with those of the generic SF-36 and EQ-5D HRQoL measures in showing that ustekinumab improved patients' HRQoL in both the induction and maintenance phases of therapy relative to the respective placebo arms, for all dose regimens, and with the differences from placebo exceeding thresholds for being clinically meaningful. The improvements in HRQoL at week 44 were marginally larger for the q8w maintenance regimen than the q12w regimen, but not reaching the threshold for being clinically meaningful.

Ustekinumab is relatively well-tolerated, and although the majority of patients in the UNIFI trial experienced adverse events, fewer than 10% of these were serious.

3.4.2 NMA results

Results of the induction NMAs and the 1-year NMAs conditional on response consistently indicate that ustekinumab ~6mg/kg and all the comparator therapies improved the odds of clinical remission and clinical response both at 8 weeks and 44 weeks compared to the respective placebo arms (i.e. the background conventional therapy). The CS concludes that, in the induction NMAs ustekinumab ~6mg/kg demonstrated a higher likelihood of response than adalimumab and golimumab in non-biologic failure patients and higher likelihood of response than adalimumab in biologic failure patients. The company also conclude that, in the 1-year NMAs conditional on response, ustekinumab had a higher probability of being more effective than all the comparators (CS section B.2.9.5). The probabilities reported in the CS on which these conclusions are based are subject to uncertainty, but the company have not provided credible intervals for the probabilities.

3.4.3 Limitations and uncertainties

A general limitation of the evidence base is the short-term nature of the clinical effectiveness and safety data for ustekinumab (<2 years).

There are a number of uncertainties, mainly arising from the NMA methods, but also some related to the UNIFI trial. A summary of these is provided in Table 36.

Table 36 Limitations and uncertainties in the company's analyses and their implications

Limitation	Where discussed	Implications
Possible directional biases in NMAs		
Trial duration heterogeneity in NMAs	Section 3.1.7.3.4	Unresolved possible bias in favour of ustekinumab against some induction comparators and all maintenance comparators for remission and response outcomes
Central/local endoscopic read inconsistency in NMAs	Section 3.1.7.3.1	Unresolved possible bias in NMAs against tofacitinib for remission outcomes
UNIFI induction UST 130mg outside licence but combined with 6mg/kg when recruiting the maintenance re-randomised population	Section 3.1.3.3	Dilution of ustekinumab effects in the population re-randomised to maintenance therapy, likely conservative against ustekinumab for remission and response (ERG clinical expert opinion)
Frequency of serious infections in maintenance phase underestimated by using observational psoriasis data rather than UC trial data	Section 3.3.7.3	Possible biases introduced but direction unclear due to heterogeneity; however overall serious infections rates low. Considered unlikely to be important in ERG critique of the economic model (section 4.3.4.5)
Carry-over effect of previous induction therapy in maintenance placebo arms	Section 3.1.7.4.2	Plausible larger carry-over effect in ustekinumab maintenance placebo arm than comparator placebo arms could bias against ustekinumab for remission and response. This is explored in an ERG scenario analysis.
Residual uncertainties in NMAs (including biases of unknown direction)		
Heterogeneity across trials in definition of non-bio failure and bio-failure subgroups	Section 3.1.7.3.2	Possible unquantifiable error of unknown direction introduced into NMA results
Not all data used in NMAs could be validated by ERG for 1-year NMAs conditional on response	Section 3.1.7.5.3	Possible unquantifiable error of unknown direction introduced into NMA results
Possible attrition bias risk in some studies in NMAs due to possibly inappropriate handling of missing data	Section 3.1.7.1	Possible unquantifiable error of unknown direction introduced to NMA results
Asian trials NMA sensitivity analysis likely invalid	Section 3.1.7.2.1	There are no reliable analyses that include Asian-only trials, in contrast to TA547
Other issues		
Statistical power of non-biologic failure and biologic failure subgroups	Section 3.1.6.3	Induction subgroups likely adequately powered, maintenance subgroups probably underpowered
Maintenance regimen pooling of standard and escalated doses for the non-biologic failure subgroup but not the biologic-failure subgroup	Section 3.1.7.5.6	Company provided pooled and un-pooled data in clarification response but for 1-year NMA not 1-year NMA conditional on response. The ERG prefers pooled analysis in both subgroups because of high uncertainty over the exposure-response

		relationships, so use this approach in our base case economic analysis.
Issues of applicability (generalisability)		
UNIFI delayed responders management not quite reflective of clinical practice	Section 3.1.3.1	Probably a minor issue; clinical practice may itself be variable

Heterogeneity in NMAs due to variation in the duration of trial induction and maintenance phases, and heterogeneity due to inclusion of both centrally-read and locally-read endoscopies were both issues that were identified, but remained unresolved, in TA547 (tofacitinib).

As shown in Table 36, whilst some of the limitations could lead to bias in favour of ustekinumab, others could lead to bias against ustekinumab, and in some cases the most likely direction of any possible bias is unclear. It is plausible, but not certain, that some of the potential biases would cancel each other out. Overall, it is not possible to conclude with any certainty that the NMA limitations summarised in Table 36 would, collectively, definitively bias for or against ustekinumab, although the inherent residual heterogeneity in the NMAs reduces certainty of the results, as reflected in relatively wide credible intervals for some analyses. Given the uncertainty around the possibility of a carry-over effect, the ERG conducted a maintenance-only NMA as a scenario, which is described below in section 4.4.3.

4 COST EFFECTIVENESS

4.1 Overview

The company submission includes:

- A systematic review of published economic evaluations of biologics and JAK targeted therapies for UC (CS B.3.1 and Appendix G);
- A description of the company's *de novo* model developed to assess the cost-effectiveness of ustekinumab compared with other biologics, JAK inhibitors and non-biologic (conventional therapy) for the treatment of adults with moderately to severely active UC (CS B.3.2 to B.3.11 and Appendices H to L).

We summarise and critique these elements of the CS in sections 4.2 and 4.3 below. Additional ERG work, including model validation and alternative scenarios are presented in section 4.4.

The cost-effectiveness results presented in this report include a confidential company's proposed Commercial Medicines Unit (CMU) arrangement price discount for ustekinumab (CS Table 2) but not existing PAS discounts for some of the comparators (golimumab, tofacitinib and vedolizumab). This means that the estimated costs and ICERs may be misleading, as they do not reflect actual prices paid by the NHS. Results including all agreed PAS discounts for comparators as well as the company's proposed CMU arrangement price discount for ustekinumab are presented in a confidential addendum to this ERG report.

4.2 Company's review of published economic evaluations

The company conducted a search to identify studies assessing the cost, healthcare use and cost-effectiveness of interventions for the treatment of moderately to severely active UC. The methods and results of the review of cost-effectiveness studies are described in section B.3.1 and Appendix G of the CS. The review of cost and healthcare use is described in section B.3.5 and Appendix I of the CS. We consider that the company's search strategy and inclusion/exclusion criteria were appropriate. As the searches were conducted in March 2019, we conducted a focused literature search to identify any more recent relevant publications but did not identify any that assessed the cost-effectiveness of ustekinumab for patients with moderate to severe UC.

The company identified 26 relevant studies (21 were cost-utility studies; 3 cost-effectiveness analyses; and 2 budget impact analyses), described in CS Table 96 (Appendix G.1.3). Eleven of these studies were UK based, of which three were informed previous NICE TAs. [62-66] No studies evaluating the cost-effectiveness of ustekinumab in the population of interest were identified. The company state that they used these studies to inform the model structure and model parameters.

ERG conclusion*:* We view the company's search strategy and eligibility criteria for their review of cost-effectiveness studies as appropriate. This did not identify any economic evaluations of ustekinumab in the population of interest and the ERG did not identify any other relevant studies.

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

Table 37 NICE reference case

Criterion	Included?	Comment
Decision problem as in scope	Y	
Comparators as listed in scope	Y	
Perspective on costs: NHS and PSS	Y	
Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Y	
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	Y	
Cost utility analysis with fully incremental analysis	Y	
Synthesis of evidence on outcomes based on a systematic review	Y	
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Y	
Health effect expressed in QALYs. EQ-5D is preferred measure of health-related quality of life	Y	
Health related quality of life reported directly by patients and/or carers.	Y	
Preference data from representative sample the UK population	Y	
An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	Y	
Discount rate: 3.5% pa for costs & health effects	Y	

ERG conclusion: The ERG considers that the submitted economic evaluation meets NICE reference case requirements.

4.3.2 Modelled decision problem

4.3.2.1 Population and subgroups

The population in the company's model is defined in CS section B.3.2.1. This is appropriate for the NICE scope, given the proposed marketing authorisation and UNIFI trial population (see 2.3 above).

The model does not produce results for the whole population, but only for the subgroups:

- **Biologic Failure**: patients previously treated with one or more biologic agent at a dose approved for the treatment of UC who did not respond initially, responded initially but then lost response or were intolerant to the medication.
- **Non-biologic failure**: all other members of the population, including people not previously exposed to a biologic (biologic-naïve) as well as those previously exposed to a biologic but not having demonstrated inadequate response or intolerance.

Age and gender affect mortality and quality of life in the model; and weight influences drug dosage and hence costs. Baseline characteristics for the modelled subgroups are based on those in the UNIFI Induction trial (see Table 38). Mean age, body weight and the gender mix were similar for the two UNIFI subgroups. These characteristics were also similar in UNIFI and overall for comparator induction trials (see Table 39), although there were large differences between individual trials.

Reported demographics from the Royal College of Physicians (RCP) UK IBD audit suggest that the modelled subgroups are similar to the wider population starting treatment with a biologic for UC.[67] Clinical experts consulted by the ERG agreed that the UNIFI trial population is reasonably reflective of NHS patients who would be suitable for ustekinumab if it were to be recommended.

The subgroups in the company model are defined by failure of previous biologic treatment, not by prior exposure to biologics as requested in the scope. In practice, this would be unlikely to affect results, as only a small proportion of the 'non-biologic failure' subgroup in UNIFI (5.7%) had previously been exposed to a biologic (see Table 3). We note some differences in the subgroup definitions for comparator trials (section 3.1.7.3.2 above).

Previous technology appraisals have focussed on results for subgroups defined by treatment history. In TA342 (vedolizumab), results were presented for biologic-failure and biologic-naïve subgroups.[17] The committee concluded that it was useful to consider these subgroups

as separate populations and that ICERs were higher for the biologic-failure subgroup than for the biologic-naïve subgroup. In TA547 (tofacitinib), biologic-exposure and biologic-naïve subgroups were considered, and ICERs were higher for the former than the latter.[9] Both TA342 and TA547 committees noted high uncertainty over the network meta-analysis (and hence economic) results based on whole ITT populations. They therefore focussed on cost-effectiveness results for the biologic exposure/failure subgroups.

Table 38 Patient baseline characteristics used in model (UNIFI Induction trial)

Characteristic		Non-biologic failure (n = 470)	Biologic failure (n= 491)	Whole population (n=961)
Age, mean (sd)	years	41.4 (NR)	41.9 (NR)	41.7 (13.7)
Male	n (%)	282 (60.0%)	300 (61.1%)	582 (60.6%)
Weight, mean	kg	73.6	72.8	73.2 (17.6)
<55kg	n (%)	70 (14.9%)	57 (11.6%)	NR
55-85kg	n (%)	293 (62.4%)	334 (68.0%)	NR
>85kg	n (%)	107 (22.8%)	100 (20.4%)	NR
Source: CS Table 34 and Clarification response Appendix M Table 1				

Table 39 Baseline characteristics for the UC population

Characteristic		All induction trials in NMA [a] 16 trials (n=6,607)	UK IBD Audit 2016 [b] (n=903)
Age	years	Mean 40 (range 34 to 44)	Median 39 (IQR: 28 to 52)
Male	%	Mean 60 (range 48 to 73)	529 (59%)
Weight	kg	Mean 71 (range 58 to 80)	NR
Source: [a] Estimated by ERG from Clarification response Appendix R Table 12 [b] Adults with UC at initial biologic tretment, Royal College of Physicians 2016 [67]			

ERG conclusion: The model population is appropriate for the scope, the anticipated marketing authorisation and UNIFI trial population. We agree with the decision to present results for the subgroups only and not for the whole ITT population (due to heterogeneity and TA precedent). Although the subgroups are defined by biologic failure, rather than biologic exposure as requested in the scope, this is unlikely to affect the results. Baseline demographics of the modelled subgroups are broadly reflective of the ustekinumab and comparator trial populations and similar to patients starting biologic treatment for UC in the UK. There were variations in mean age, body weight and the proportion of men between trials, but the ERG has confirmed that model is not sensitive to these parameters.

4.3.2.2 Intervention and comparators

The CS states that all comparators are modelled for both patient subgroups (B.3.2.3), although the NMA and economic model do not include infliximab and golimumab for the biologic-failure subgroup.

The model also includes biosimilar versions of infliximab and adalimumab, with the same assumed clinical effects and safety profile as the original licensed brands but at lower cost. The CS reports cost-effectiveness results for both original and biosimilar infliximab and adalimumab. In 2016, the RCP National Audit found that 44% (292/520) of adults with UC starting biologic treatment for the first time with infliximab had a biosimilar product.[67] Since then, initiation of treatment with biosimilar products is likely to have increased, supported by RCP guidance and NHS England advice.[67,68]

> **ERG conclusion**: The model includes all comparators in the scope except infliximab and golimumab in the biologic failure subgroup. This omission is unavoidable because the infliximab and golimumab trials excluded people with previous biologic treatment (CS Appendix Table 20). The modelling of available biosimilars for infliximab and adalimumab is appropriate, with the assumption of equal effects and safety profile but lower costs compared with the original products. We anticipate increasing use of biosimilars, but presentation of results for the original biologic drugs as well is useful for comparison.

Induction regimens

Modelled dose regimens for the biologics and tofacitinib reflect SmPC recommendations (Table 40). There is a standard induction phase for all these treatments, with defined duration and dosing. If patients do not have an adequate response during this time, induction may be extended to check for a delayed response (except for adalimumab). The company base case assumes use of extended induction when patients do not respond within the standard induction period and that the loss of response rate in maintenance therapy is the same for delayed and early responders. Two scenarios for delayed responders are presented: loss of response based on trial data (Scenario 9); and no extended induction (Scenario 10).

> **ERG conclusion**: The model appropriately reflects recommended induction regimens, including extended induction for delayed response. The company scenario without extended induction illustrates the effect of possible variations in clinical practice. Maintenance efficacy may well differ for initial and delayed responders, but

evidence is sparse, so the company's base case assumption of equal loss of response rates for initial and late responders is reasonable.

Maintenance regimens

Patients with an initial or delayed response to induction proceed to maintenance treatment with the same drug (Table 40). Maintenance starts with a standard regimen, but all drugs except infliximab also have escalated regimens that can be used when response declines or is lost. The CS states that clinicians are likely to consider dose escalation before surgery (CS B.3.2.3). Clinical experts consulted by the ERG agreed that this is the case, and noted that the decision to adjust the dose or frequency of biologic treatments would be informed by drug level and antibody testing.

The company excludes the higher dose of infliximab as an option in the model, on the basis that this is not specified in the marketing authorisation. We acknowledge this, but note that, clinical advice to the ERG is that dose escalation for infliximab is common in practice.

The model applies a fixed dose mix throughout maintenance treatment, with 30% of patients on the escalated regimens in the base case and 10% and 50% scenarios (Scenarios 7 and 8). These estimates are based on retrospective studies.[69-71] The largest and most relevant study for the UK is a retrospective case note review in Europe and Canada for patients who started anti-TNF therapy between 2009 and 2013.[72] This concluded that for UC, 26% of patients without prior anti-TNF treatment and 17% of patients with prior anti-TNF treatment required dose escalation. The assumption of 30% dose escalation therefore appears to be reasonable, with scenario analysis to test the impact on results.

The dose escalation percentage is used in the model to adjust the cost of maintenance therapy and, for the biologic-failure subgroup only, also its effectiveness. For the non biologic-failure subgroup, the model uses pooled estimates of effectiveness for the standard and escalated regimens. The company justify this difference in dose pooling by arguing that there is an exposure-response relationship for people with previous biologic failure, but not otherwise (Clarification Response questions A22 and B2). As discussed above in section 3.1.7.5.6, the evidence presented for this claim is indirect: based on a lower incidence of remission at the start of maintenance in the biologic-failure subgroup and a clear exposure-response relationship for ustekinumab without (but not with) clinical remission at maintenance baseline (Clarification Response Figure 18). Direct evidence of a difference in exposure-response (or dose-response) between the subgroups is not presented from UNIFI or other trial data.

The implementation of dose-pooling for the non biologic failure subgroup is done by taking a simple unweighted mean of direct trial results for the two regimens in the base case, and pooled estimates in the company's maintenance NMA scenario. The former is a simplification (Clarification Response question B2), but as there were similar numbers of patients in higher and lower dose arms in the relevant trials, this will make little difference in practice.

ERG conclusion: The model appropriately reflects recommended maintenance regimens, including escalation to higher dose or more frequent treatment when indicated. The assumption that 30% of patients on maintenance have the escalated regimen is reasonable, with exploration of uncertainty through scenario analysis.

The company does not include the higher (10mg/kg) dose of infliximab because it is not recommended in the SmPC. However, clinical advice to the ERG is that dose adjustment for infliximab is common in practice. This suggests that the same dose escalation assumptions should be made for infliximab as for other comparators.

The company argues that there is an exposure-response relationship for patients with a history of biologic failure, but not for other patients. Consequently, they pool standard and escalated doses in the non-biologic failure subgroup but not in the biologic failure subgroup. The ERG considers that evidence supporting this stance is weak, as it relies on an indirect relationship (exposure-response with/without remission at maintenance baseline) and only for ustekinumab. We therefore think that the same dose pooling approach should be used in both subgroups. We prefer pooled effect estimates, because of high uncertainty over the exposure-response relationships, so use this approach in our base case analysis. Additional ERG scenarios explore separate effect estimates: 1) unpooled estimates for both subgroups; and 2) standard regimen (which may be realistic as patients only have the escalated regimen after failure of standard treatment). However, we have not had time to run these scenarios for the company or ERG maintenance NMA versions of the economic model.

Table 40 Recommended dose regimens for ustekinumab, other comparator biologics and tofacitinib

Drug	Induction		Maintenance	
	Standard dose (duration)	Extended dose (duration)	Standard dose	Escalated dose
Infliximab [a]	5 mg/kg IV at weeks 0, 2 & 6 (8 weeks)	Discontinue if no response after 3 doses (+6 weeks)	5 mg/kg IV every 8 weeks	Not recommended in SmPC
Golimumab	200 mg SC at week 0; 100 mg at week 2 (6 weeks)	Reassess if no response after 12-14 weeks (+8 weeks)	50 mg SC every 4 weeks	100 mg every 4 weeks if ≥80 kg or inadequate response
Adalimumab [a]	160 mg SC at week 0; 80mg at week 2; 40 mg at weeks 4 & 6 (8 weeks)	Discontinue if no response within 8 weeks (no extended induction)	40 mg SC every 2 weeks	40 mg once per week if necessary
Vedolizumab	300 mg IV at weeks 0, 2 & 6 (6 weeks)	300 mg IV at week 6 discontinue if no response by week 10 (+ 4 weeks)	300 mg IV every 8 weeks	Consider 4-weekly if decrease in response
Tofacitinib	10 mg oral twice daily (8 weeks)	10 mg oral twice daily discontinue if no response by week 16 (+ 8 weeks)	5 mg oral twice daily	Consider 10 mg twice daily if necessary
Ustekinumab	6 mg/kg IV at week 0 (8 weeks)	90 mg SC week 8 consider stopping if no evidence of benefit by week 16 (+8 weeks)	SC 90 mg every 12 weeks	May reduce to 8 weekly if response is lost.

IV intravenous administration; SC subcutaneous injection

a Available biosimilars are included in the company's model, with the same regimens, effects and safety parameters.

Source: Adapted from CS Table 38 (B.3.2.3), additional information from MIMS [73]

Stopping rule

CS analyses assume that responders to induction continue maintenance until loss of response or death. The model includes a stopping rule option but this is not used. The model option allows discontinuation at a defined time, with subsequent (constant) loss of response based on either: i) trial data for responders to active induction re-randomised to placebo (UST, GOL, VED and TOF only); or ii) the same rate as for CT (trial data for responders to placebo induction, PBO-PBO). TA329 and TA342 recommend annual assessment of benefit and need. Clinical advice to the ERG suggests that one-year assessment and trial of treatment withdrawal is variable: with some centres routinely planning a trial of withdrawal and others rarely considering this option.

> **ERG conclusion**: Given uncertainty over routine use of a 'stopping rule' for biologics in UC, we think it is appropriate to assume continued treatment until loss of response in the base case. We use the 'stopping rule' option in the model to illustrate the impact of discontinuation at one-year, but note uncertainty over this scenario. It is not clear if the assumed post-discontinuation loss of response rates are accurate or whether the scenario reflects trial of discontinuation in practice: which is usually restricted to patients with remission, with re-initiation of treatment after relapse.

Sequential treatment

The base case model assumes that after the failure of the initial treatment, all patients switch to conventional therapy alone. However, the model includes an option to add a second-line of treatment and a scenario is presented with patients switching to vedolizumab after all other treatments, or adalimumab after vedolizumab (Scenario 6). The rationale for this choice of second-line treatments is not stated. In practice, clinicians often consider sequential treatments, with the choice of next line depending on treatment history, antibody tests, anticipated speed of action and safety profile. Clinicians consulted by the ERG stated that a common treatment pathway was to start with (biosimilar) infliximab, escalate dose or switch to another anti-TNF drug if antibodies are low, or alternatively to try vedolizumab, tofacitinib or (if recommended) ustekinumab. They noted that vedolizumab was considered to have a slow speed of onset, while there were more safety issues to consider with tofacitinib. Although less common, some clinicians do consider 'step-down' treatment, starting with a more effective (and expensive) treatment.

> **ERG conclusion**: Many patients who might be considered for ustekinumab would not have exhausted all other treatment options. Sequential use of therapies is

common in practice, but variable, and cost-effectiveness is potentially sensitive to the choice of subsequent treatment.

Conventional therapy

Conventional therapy (CT) is included in the model as a comparator at the induction phase and as the initial default treatment after failure of ustekinumab or comparators (including CT). The modelled doses and proportions of patients using drugs that make up the CT are shown in CS Table 39. Concurrent use of conventional treatments alongside the biologics and tofacitinib is also routine in current practice, but the company's model does not include concurrent treatment costs. See section 4.3.6.1 below for further details and discussion.

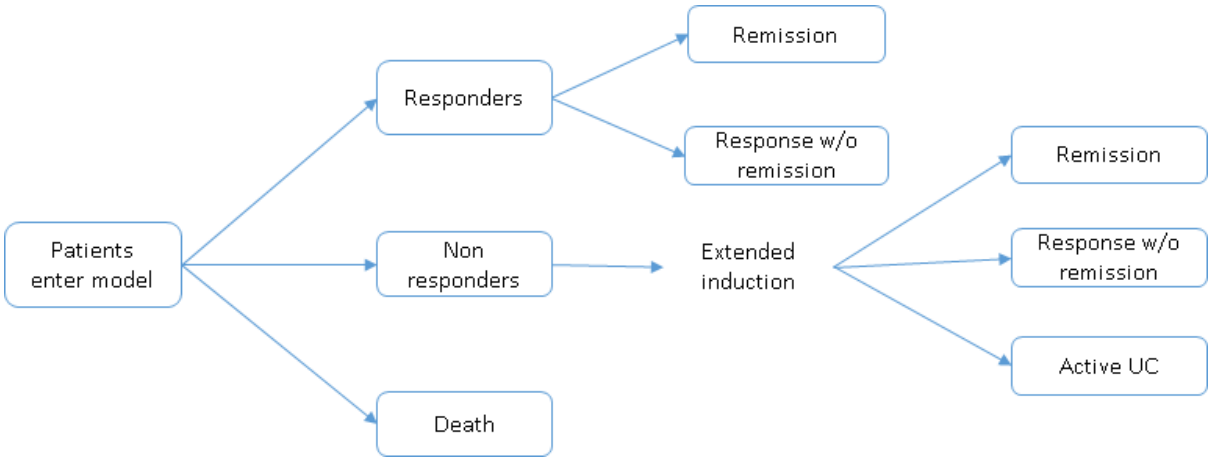
4.3.3 Model structure

4.3.3.1 Overview

The company describes the structure and key features of their model in CS Section B.3.2.2. They summarise assumptions in CS Tables 59 and 61, the parameters in CS sections B.3.3 to 3.5 and CS Table 60. The model follows a conventional design for UC, but with some changes to previous TA models, which we discuss below. The model is a hybrid, consisting of a decision tree (for the induction phase) and a Markov model (for maintenance and ongoing care) in Microsoft Excel®: see Figure 13. The Markov has a cycle length of 2 weeks, designed to accommodate induction periods of different lengths. The model uses a 50 year time horizon (effectively lifetime from a starting age of 41 years), with a half-cycle correction. Costs and QALYs are discounted at an annual rate of 3.5%.

ERG conclusion: The overall model structure is appropriate, consistent with previous TA models and accurately implemented. The only major exception is the omission of response and remission health states after failure of the initial treatment (see below). The 2-week Markov cycle is short (e.g. 8 weeks was used in TA547). This will cause some underestimation of costs if symptom recurrence is not always detected and treatment discontinued within 2 weeks. Experts have advised the ERG that clinics provide fast access on request, but this may not be consistent at all times throughout the NHS. However, delays in treatment discontinuation are unlikely to have a significant impact on costs.

Decision Tree for the Induction Phase (ERG's illustration)



Markov model for the Maintenance Phase (CS Figure 38)

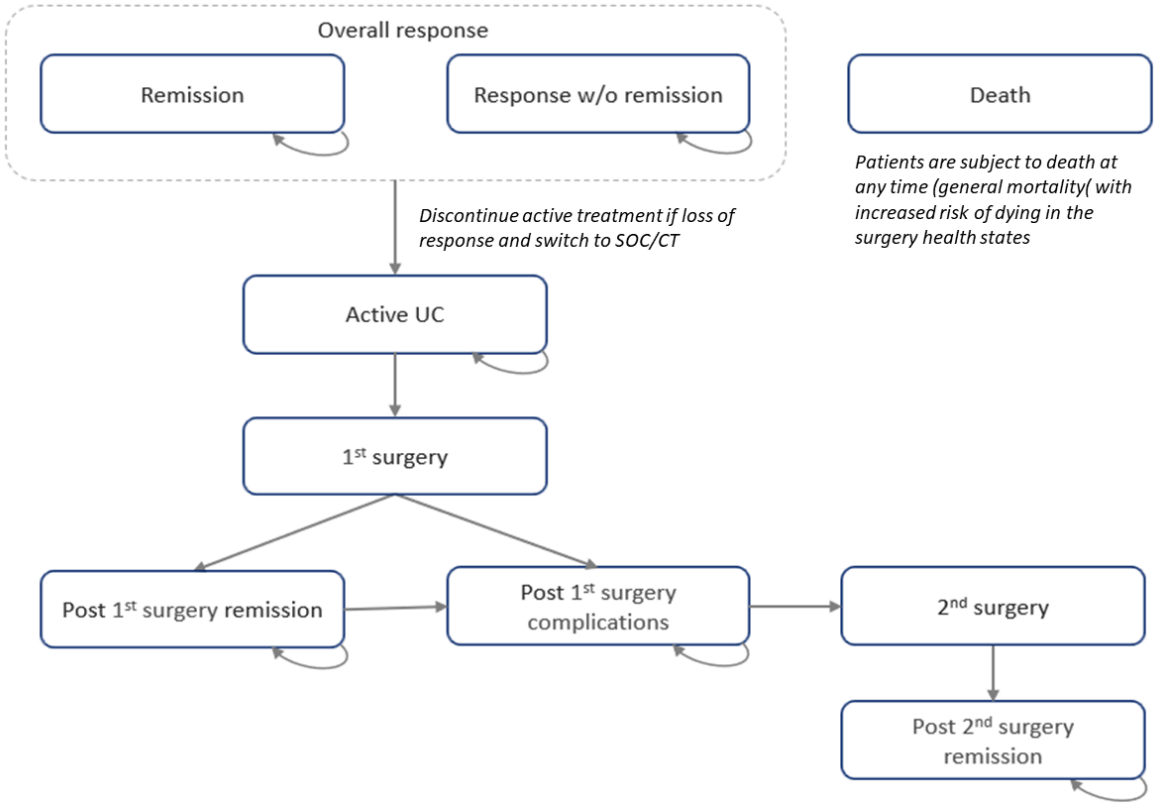


Figure 13 Illustration of the model structure
(Source: CS Figure 37 (adapted) and Figure 38, CS B.3.2.2)

4.3.3.2 Induction phase

A decision tree is used to represent induction. This includes two stages of variable length to reflect the standard and extended induction regimens (see Table 40 above). Patients enter the model in the *Active UC* health state at initiation of ustekinumab or one of the comparator treatments. Patients with a clinical response by the end of standard induction transition to either the *Remission* or *Response without remission* health state. Those who do not respond stay on induction for an additional time to assess for delayed response. At the end of extended induction, delayed responders transition to remission or response without remission and people without a response remain in Active UC.

4.3.3.3 Maintenance phase

Patients who respond to induction (including delayed responders) enter the Markov model in the remission or response without remission health state and start maintenance treatment, which continues as long as patients retain response. A proportion of patients (30% in the base case) are assumed to require a higher dose or more frequent treatment to maintain a clinical response (dose escalation). The model includes an option to add a stopping rule, after a defined duration of treatment, but this is not used.

4.3.3.4 Conventional treatment

Patients who do not achieve response after extended induction and those who lose response to maintenance treatment enter the Markov model in the *Active UC* health state on conventional therapy alone. Subsequently, patients can continue with Active UC, have surgery or die. This approach differs from models in previous NICE TAs (TA547 and TA342), which also included transitions from Active UC to Remission and Response without Remission after switching to conventional treatment alone. This is more realistic as UC is not always a progressive disease and many people with UC have ongoing periods of relapse and remission[74] In response to a clarification question (B1), the company argue that the impact of introducing response and remission health states after failure of initial treatment would be negligible, as it would affect all treatments in a similar manner. However, we note that the effect of omitting these states is to exaggerate the benefits of inducing and retaining clinical response or remission, introducing a bias in favour of the more effective interventions. For this reason, we consider it important that the model should more accurately reflect long-term UC epidemiology. We address this issue in ERG additional analyses in section 4.4.3.

ERG conclusion: The omission of response and remission health states after failure of the initial treatment option is a major limitation. This implies that all patients follow

a chronic active or progressive form of disease, which is inconsistent with previous NICE appraisals and unrealistic. For face validity, the model should reflect long-term patterns of disease. This is also necessary for accurate estimation of the downstream benefits of inducing and retaining initial response.

4.3.3.5 Surgery

The company's approach to modelling surgery and its related complications differs from previous TAs. The model includes surgery as an option for patients with active UC after failure of initial therapy. Once patients commence surgery, they are assumed to stop all drug treatments (including CT) for the remaining time horizon. Two phases of surgery are modelled, each lasting for six months to allow for staged procedures. The first phase comprises subtotal colectomy with ileostomy followed by either IPAA (pouch) surgery or by permanent ileostomy (*1st surgery*). If the first phase is successful, patients stay in remission until death (*Post 1st surgery remission*). However, some patients have chronic complications after surgery (*Post 1st surgery complications*), including pouch failure which may require a second phase of surgery for revision (*2nd surgery*). The model assumes that all patients achieve remission after revision surgery (*Post 2nd surgery remission*).

ERG conclusion: The model includes two phases of surgery, each lasting for six months to allow for staged procedures. This approach differs from previous appraisals (TA547 and TA342), which treated surgery as a one-off event. However, we consider that the current model better reflects the usual process of staged procedures: subtotal colectomy with ileostomy followed by either IPAA (pouch) surgery or permanent ileostomy (phase 1); and potential revision surgery due to pouch failure (phase 2). The model assumes that all patients who have revision surgery reach remission with no chronic complications. This is a reasonable simplification; although it will not be true for all patients, the number of people affected and hence the impact on overall costs and QALYs will be small.

4.3.3.6 Mortality

The model includes death as an absorbing state and death can occur from any of the health states at any time. Mortality rates are assumed to be the same as for the general population, except for a small mortality risk associated with surgery. The company cites evidence of elevated standardised mortality rates for UC[75] and state that their approach is a simplification for the model (CS B.1.3 and B.3.3.4). This approach is consistent with previous TAs.

4.3.4 Clinical parameters

4.3.4.1 Response and remission: induction phase

The base case parameters for response and remission at the end of standard induction are estimated from the induction NMA: CS Table 40 (reproduced in Table 41 below for convenience). A weighted average of the trial placebo arms is taken for CT, and adjusted for other comparators using odds ratios: Fixed Effects (FE) in the base case and Random Effects (RE) in a scenario (Scenario 1). As might be expected the deterministic results for the FE and RE models are similar, but there is more uncertainty over the RE results. See section 3.1.7 above for the ERG critique of the company's induction NMAs.

Table 41 Effects of standard induction (fixed effects NMA)

Treatment	Remission		Overall Response (including remission)		Response without remission
	OR	Percent (calculated)	OR	Percent (calculated)	Percent (calculated)
Non-biologic failure subgroup					
Ustekinumab	2.19	18.7%	3.67	66.6%	47.9%
Infliximab	4.44	31.9%	4.11	69.1%	37.2%
Golimumab	2.97	23.8%	2.29	55.4%	31.6%
Adalimumab	2.21	18.9%	1.89	50.6%	31.7%
Vedolizumab	4.54	32.4%	3.21	63.5%	31.1%
Tofacitinib	2.43	20.4%	2.70	59.4%	39.0%
CT	1.00	9.5%	1.00	35.2%	25.7%
Biologic failure subgroup					
Ustekinumab	13.41	26.9%	3.58	55.5%	28.6%
Adalimumab	1.37	3.6%	1.45	33.6%	30.0%
Vedolizumab	3.76	9.4%	2.52	46.8%	37.4%
Tofacitinib	22.33	38.0%	3.41	54.3%	16.3%
CT	1.00	2.7%	1.00	25.9%	23.2%
NB: identical clinical efficacy rates were used for the biosimilars of infliximab and adalimumab, for all efficacy outcomes in the model.					
Source: reproduced from CS Table 40					

ERG conclusion: Base case response and remission rates for standard induction are based on the company's fixed effects induction NMA. The ERG prefers the random effects model, which gives similar results but with more uncertainty. ERG replication of the company's induction NMAs found some discrepancies (see section 3.3.6.1 above). We use ERG estimates in scenario analysis.

4.3.4.2 Response and remission: maintenance phase

Constant loss of response risk

The model assumes a constant risk of loss of response (both with and without remission) during maintenance treatment. This applies within the initial year of maintenance for which there are data, and for extrapolations over the time horizon (although the Markov trace graphs in the model show that few patients retain response over more than 5-10 years on any treatment). The company conducted a scenario analysis to illustrate the possible impact of declining loss of response risk (Scenario 3): this assumed a one-off 25% reduction in the loss of response after the first two years of treatment.

The company explains their assumption of constant loss of response in CS B.3.3.1.2.1. This approach was taken, and accepted, in TA547 (tofacitinib) due to a lack of intermediate data on clinical response and remission within one-year maintenance trials, or in longer-term follow up. There is some other data for infliximab. As reported in TA329, 6-month response and remission data indicated that loss of response risk declined over time.[62] Ferrante et al. (2008)[76] reported longer follow-up in 81 people with refractory UC treated with infliximab. The Kaplan-Meier curve for sustained clinical response (see Figure 14) suggests an increasing risk in the first year, but the rate appears relatively constant after that. However, these data are sparse and the risk may well change in different ways for other treatments.

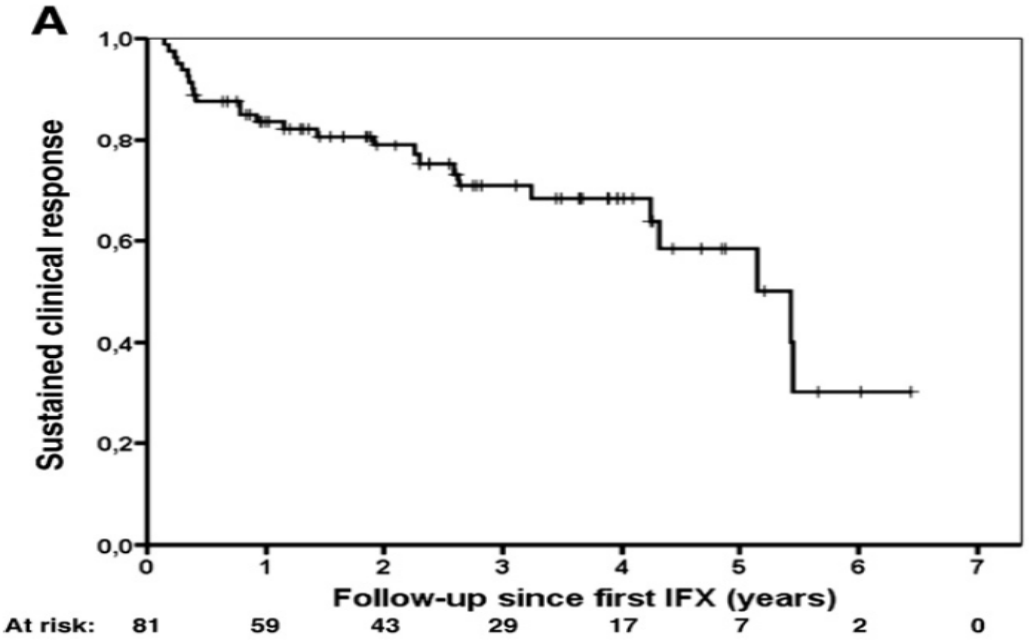


Figure 14 Sustained clinical response in 81 outpatients with refractory UC treated with infliximab (Ferrante et al 2008)[76]

ERG conclusion: In the absence of interim response/ remission data for the clinical trials or longer-term follow-up it is difficult to predict how the absolute or relative loss of response changes. We therefore agree with the assumption of a constant risk over time. This is consistent with the assumption in TA547.

Base case (direct trial estimates)

In their base case analysis, the company use direct trial data to estimate the proportion of induction responders who lost response between the end of standard induction and the end of maintenance follow up: CS Table 43 (adapted in Table 42 below). The company justify their use of direct trial data by arguing that it avoids the problem of response and remission differences between the maintenance placebo arms, which they ascribe to carry-over effects for patients who received active treatment during induction. They also argue that this provides a more realistic reflection of clinical practice, in which patients who respond to induction treatment, continue with the same treatment for maintenance.

Table 42 Base case maintenance loss of response (direct trial data)

	52 week Remission	**52 week response including remission**		**52 week response without remission**	
	% of induction responders	% of induction responders	Loss of response (2 weeks)	% of induction responders	Loss of response (2 weeks)
Non-biologic failure subgroup					
UST (pooled doses)	53.6%	81.5%	0.009	28.0%	0.042
IFX (5mg/kg q8w)	42.7%	55.9%	0.025	13.2%	0.059
GOL (pooled)	23.5%	48.6%	0.026	25.1%	0.030
ADA (40mg q2w)	33.0%	51.1%	0.030	18.1%	0.055
VED (pooled doses)	46.9%	60.8%	0.021	13.9%	0.053
TOF (pooled doses)	43.0%	60.5%	0.019	17.5%	0.050
CT	26.7%	40.2%	0.041	13.5%	0.074
Biologic failure subgroup					
UST (90mg q12w)	37.5%	70.8%	0.016	33.3%	0.020
UST (90mg q8w)	46.2%	71.8%	0.015	25.6%	0.031
ADA (40mg q2w)	25.7%	45.7%	0.035	20.0%	0.066
VED (300mg q8w)	37.2%	46.5%	0.033	9.3%	0.089
VED (300mg q4w	35.0%	42.5%	0.037	7.5%	0.098
TOF (5mg BID)	24.1%	44.6%	0.031	20.5%	0.031
TOF (10mg BID)	36.6%	59.1%	0.020	22.5%	0.020
CT	13.0%	34.6%	0.047	21.6%	0.063
Source: Adapted by ERG from CS Table 43					

For the active arms, the analysis used data for induction responders only from the maintenance trials UNIFI, ACT1, PURSUIT-M, ULTRA2, GEMINI and OCTAVE Sustain. As

discussed above (4.3.2.2), standard and escalated dose results were pooled (by taking simple means for the two regimens) for ustekinumab, golimumab, vedolizumab and tofacitinib in the non-biologic failure subgroup. In the biologic failure subgroup, the standard and escalated regimens for these drugs were modelled separately (with 30% of patients assumed to have the escalated regimen in the base case). In both subgroups, escalated regimens for infliximab and adalimumab were excluded: because the higher dose is not recommended for infliximab; and because of lack of data for adalimumab.

Loss of response rates for CT were taken as a weighted mean for induction responders who had received placebo during both induction and maintenance (PBO-PBO). This restricted the data source for CT to UNIFI, ACT1, PURSUIT-M and ULTRA (PBO-PBO results were not available for GEMINI or OCTAVE). Consequently, the sample sizes for the CT response and remission 'direct trial' estimates are small: for response 281 and 75 in the non-biologic failure and biologic-failure subgroups respectively (model sheet 'Data Storage (Direct Trial)').

Loss of response over the maintenance period was adjusted for the duration of the Markov cycle, to provide 2-week loss of response probabilities (with and without remission). Loss of response probabilities were estimated separately for the 'Remission' and 'Response without remission' health states. Note that the model does not explicitly allow for transitions between the 'Remission' and 'Response without remission' health states.

Maintenance NMA scenario (1-year conditional on response)

The company also present a scenario based on their NMA sensitivity analysis (1 year ITT, conditional on response, fixed effects) (CS Tables 29 and 30). In this scenario, a pooled placebo loss of response rate (weighted average for trial control arms) is adjusted for comparators using the NMA odds ratios. We summarise the resulting remission, response and loss of response rates in Table 43. Note that although the absolute proportions in response or remission at 52 weeks appear much less favourable compared with the base case (Table 42), this is because the results are reported with respect to a different denominator (induction responders only for the base case and ITT at the beginning of induction for the NMA scenario).

See 3.1.7.5.3 for ERG discussion of this NMA sensitivity analysis. We replicated the analysis, with some moderate differences from the company's analysis (Table 28 and Table 29). At the request of the ERG, the company conducted a random effects version of this analysis, using a weakly informative prior (Clarification Response question A14), which we

replicated to obtain odds ratios in a format that could be used in the economic model (Table 30 and Table 31).

Table 43 Maintenance NMA scenario (one-year ITT, conditional on response)

	52 week remission	*52 week response including remission*		*52 week response without remission*	
	% of ITT	*% of ITT*	*Loss of response (2 weeks)*	*% of ITT*	*Loss of response (2 weeks)*
Non-biologic failure subgroup					
UST (pooled doses)	35.2%	49.8%	0.013	14.7%	0.052
IFX (5mg/kg q8w)	23.6%	37.8%	0.026	14.2%	0.041
GOL (pooled)	13.7%	28.4%	0.025	14.7%	0.028
ADA (40mg q2w)	20.6%	25.3%	0.031	4.7%	0.083
VED (pooled doses)	32.0%	40.0%	0.020	8.1%	0.057
TOF (pooled doses)	25.4%	35.7%	0.019	10.3%	0.050
CT	8.9%	13.8%	0.042	4.9%	0.072
Biologic failure subgroup					
UST (90mg q12w)	18.6%	35.6%	0.020	16.9%	0.024
UST (90mg q8w)	23.2%	35.8%	0.020	12.6%	0.037
ADA (40mg q2w)	16.6%	23.9%	0.015	7.3%	0.062
VED (300mg q8w)	22.0%	23.9%	0.029	2.0%	0.120
VED (300mg q4w	20.6%	21.9%	0.033	1.2%	0.138
TOF (5mg BID)	15.4%	26.6%	0.027	11.2%	0.027
TOF (10mg BID)	23.2%	34.9%	0.017	11.6%	0.017
CT	2.9%	9.6%	0.044	6.7%	0.055

Source: Estimates extracted from company model by ERG

ERG conclusion: We have strong concerns over the use of absolute response and remission rates from individual treatment arms, as in the company's base case analysis. We acknowledge the difficulties in integrating treat-through and re-randomised trial data, and the potential for bias due to 'carry over' effects for maintenance placebo patients who had active treatment in induction. However, there is also a high potential for bias in the company's "direct trial" analyses, which take data directly from individual trial arms. This approach ignores the original randomisation, meaning that any differences between the trial populations or conduct are not adjusted for. Given these reservations, the ERG has a preference for the company's maintenance NMA scenario over their base case; and because of potential heterogeneity, we prefer the random effects version of the NMA scenario.

However, we do also question the validity of attributing all of the differences between maintenance placebo arms to 'carry over' effects from induction. It is more likely that

other differences between the trials also contribute to these differences. Furthermore, we could not verify all of the sources of data and imputations in the company NMA scenario. We therefore conducted an alternative NMA following the methods applied in the TA547 appraisal (see section 3.3.6.3, Table 32 and Table 33). We conducted a scenario analysis using this ERG maintenance only (no carry over) NMA for consistency with TA547 and to illustrate the range of uncertainty associated with carry over (see section 4.4.3 below).

4.3.4.3 Response and remission: delayed responders

The probabilities of response and remission at the end of extended induction for non-responders to standard induction are shown in CS Table 41. These estimates were derived directly from trial data, using results for individual treatment arms ('breaking randomisation'). Direct trial data is also used to estimate loss of response rates during maintenance treatment for responders to extended induction (delayed responders), CS Table 44.

ERG conclusion: There is high uncertainty over the direct trial estimates of response and remission for extended induction and loss of response rates for delayed responders. The company's scenario excluding extended induction tests the impact of assumptions about delayed response.

4.3.4.4 Incidence of surgery and surgery related complications

The CS states that a focused literature search was conducted to inform the surgery parameters (CS Section B.3.3.2). Table 44 below shows the clinical inputs used in the model. For simplicity, the company used the same set of estimates for both subgroups.

Table 44 Model inputs for surgery related parameters

Parameters	Values	Source
Annual probability of first surgery	0.47%	Misra et al 2016 [77]
Proportion of post-surgery chronic complications (%)	33.5%	RCP National clinical audit of inpatient care for adults with UC, National report 2014.[78]
Annual probability from post-surgery remission to chronic complications	3.25%	Segal et al. 2018[79]
Annual probability of second (revision) surgery	0.47%	Assumed to be the same as first surgery (Misra et al 2016) [77]
Source: CS Tables 45 to 48 and model Sheet 'Clinical_Inputs'		

Of the 8 studies identified, the company chose Misra et al. (2016)[77] as the source for the initial incidence of surgery (CS Table 45). They argue that this is appropriate as it was a large UK-based study and had informed the economic analysis in TA547. Misra et al. analysed Hospital Episode Statistics (HES) data for 73,318 people admitted with a diagnosis of UC over a 15-year period (1997 to 2012), of whom 5,044 (6.9%) had a colectomy. This gives an annual rate of 0.47%, which is similar to the estimate of 0.59% from the only other UK study (Chhaya et al. 2015).[80] Other estimates were higher (1.03% to 13.93%) but were based on smaller samples and may not be representative of UK practice.

The company also uses the same estimates as in TA547 to inform the proportion of people who developed chronic complications within 6 months of first surgery. These estimates were based on the 2013 national clinical audit for inpatient care for adults with UC, which reported complication rates of 32% and 35% for elective and non-elective surgery (33.5% used in the model).[78] Patients who survived the first phase of surgery without complications could subsequently develop late chronic complications. Five studies reporting on late complications were identified (CS Table 47). The company selected the estimate of 3.25% per year based on Segal et al. (2018)[79], despite its small sample size (39 patients), because this was the only UK study. TA547 used an alternative source, Ferrante et al. (2008)[81]: 9.04% per year. We note that the ICERs are not sensitive to this higher estimate.

The company assumes that the probability of a second phase of revision surgery is the same as for the initial surgery. The CS reports a study by Loftus et al. (2008)[82] but notes that the follow up was short (6 months) and that the proportion of patients having second surgery was unrealistically very high (79%). We note that this statistic appears to relate to any follow up surgery including IPAA and permanent ileostomy after initial subtotal colectomy, which are part of the six-month first surgery phase in the model. Thus, the Loftus et al. estimate is not appropriate for the model structure. Previous appraisals did not explicitly include a second stage of surgery.

The CS assumes that all the patients undergoing second surgery attain remission and transition to post-second surgery remission health state.

> **ERG conclusion:** We agree with the company's use of UK estimates for the incidence of first surgery (Misra et al. 2016) [77] and rates of early (RCP audit 2013)[78] and late complications (Segal et al. 2018)[79]. The first two of these sources were also used in TA547. A different source was used for late complications in TA547 (Ferrante et al. 2008), but the model is not sensitive to this difference. The company's

assumption that the incidence of revision surgery for patients with chronic complications is the same as that for initial surgery is arbitrary. However, this only affects a small proportion of the cohort and the model is not sensitive to this assumption. Use of the same set of parameters to characterise the incidence and complications of surgery for patients with and without prior biologic failure is a reasonable simplification.

4.3.4.5 Adverse events: serious infection rates

Only serious infections are included in the company's model, which is consistent with TA547. Discontinuation due to adverse events is not explicitly modelled and serious infection is treated as a one-time event. These are reasonable simplifying assumptions.

The annual serious infection rates used in the model are presented in CS Table 49. Note that although the table is titled 'induction phase serious infections', these rates are applied in the model to induction and maintenance treatments, as well as conventional medical treatment after failure of the first-line.

The serious infection rates in the model are based on a multinational registry for systemic treatment of psoriasis: the PSOLAR study.[32] This included 7,300 patients treated with ustekinumab, infliximab or adalimumab over a total of 13,349 person years (mean follow up 22 months): annual risks 0.83%, 2.49% and 1.97% respectively. Due to a lack of data for other comparators, the company assume that the risk of serious infections with vedolizumab, tofacitinib and CT are the same as for ustekinumab; and that golimumab and the infliximab biosimilar have the same risk as infliximab. The company conducts a scenario analysis with the same rate of serious infections (0.83%) for all treatments (Scenario 11).

We discuss clinical opinion on the relevance of the psoriasis data to the UC population and compare reported rates of serious infections in the ustekinumab and comparator trials with those from PSOLAR in section 3.3.7.3 (Table 35) above. On balance, we concur with the use of PSOLAR. It is a large 'real-world' study and the results are of the same order of magnitude as observed rates from the trials. There is uncertainty due to the use of data for a psoriasis population, the assumptions used to infer rates for comparators not included in PSOLAR and the still limited follow up (just under two years) compared with the model time horizon. However, the ICERs are not sensitive to the company's scenario or to wider scenario analysis conducted by the ERG.

ERG conclusion: Overall, the rates of serious infections used in the model appear reasonable. Despite uncertainties over use of the PSOLAR data and assumptions, this is still the best available source of evidence and the model is not sensitive to plausible changes in serious infection rates.

4.3.4.6 Mortality rates

The model uses general population all-cause mortality rates adjusted for age and gender from UK Life tables. The only excess mortality for UC was a relative risk of 1.3 for surgery from a meta-analysis by Jess et al. (2007)[83] which was applied during the six-month first and second surgery health states. This approach is similar to that in TA547 and TA329, although in TA342 excess mortality was assumed for all active UC and post-operative health states. The company comments that their approach is a simplifying assumption for the model, although patients with UC have a higher standardised mortality rate than the general population (CS B.1.3). We note that Jess et al. concluded that "The overall risk of dying in patients with UC did not differ from that of the background population". The model is not sensitive to the relative risk of mortality for surgery.

ERG conclusion: The company's assumptions about mortality are reasonable, with an excess risk for surgery, but otherwise the same risks as for the general population. We note that model is not sensitive to the relative risk assumed during surgery.

4.3.5 Utilities

The company model includes the following parameters for utility:

- A baseline utility, adjusted for age and gender, for patients without UC;
- Utility multipliers to reflect reduced utility for the UC and surgery health states; and
- A utility decrement for the adverse effect of serious infections.

Parameter estimates in the base case model were obtained from a systematic review of the literature on utility in UC (CS B.3.4.2 and Appendix H). The company also present a scenario analysis based on EQ-5D data from the UNIFI trial (CS B.3.4.1).

Utilities from published sources

The company conducted a systematic search for utility estimates, described in CS Appendix H). We consider that the search strategy was satisfactory. They included 26 studies in their review, 6 of which reported EQ-5D utilities (Table 115, CS Appendix H). In the main submission, the company use three published studies for their base case: Woehl et al. (2008)[84], Arseneau et al (2006)[85] and Stevenson et al. (2016)[86]. See Table 45. We note that the disutility of 0.156 for serious infections appears to have been misapplied in the model, as

it was not adjusted for the duration of illness (assumed 28 days in the TA329 analysis). This only makes a small difference to the estimated ICERs because of the rarity of serious infections. The company presents a scenario using utilities for surgery, post-surgery remission and post-surgery complications from the study by Swinburn et al (2012)[87] (Scenario 5).

Table 45 Utility estimates used in the company's base case

Health state	Value	Source	ERG comments
No disease	Equivalent to general population	Ara and Brazier (2010)[88]	Adjusted for age and gender of cohort. Formula derived from Health Survey for England 2003 and 2006 EQ-5D-3L (n=25,080).
Remission	0.87	Woehl et al. (2008)[84]	UK EQ-5D-3L study of 180 UC patients. Source is consistent with TA329, TA342 and TA547. In the base case, utility multipliers calculated with respect to remission were used to adjust the 'no disease' general population values.
Response without remission	0.76		
Active UC	0.41		
Surgery (first and second)	0.61	Arseneau et al. (2006)[85]	This US based study reported utility weights using TTO for ileostomy and J pouch, from a sample of 48 UC patients. The CS uses a weighted average of the utilities for ileostomy (0.57) and J- pouch (0.68) assuming 60% of the patients undergo ileostomy and 40% J pouch. The base case used the same utility multiplier for both six-month phases of surgery.
Post- surgery remission (first and second)	0.72	Woehl et al. (2008)[84]	The same utility multiplier was applied for the remission state after both phases of surgery.
Post-first surgery complications	0.34	Arseneau et al. (2006)[85]	Estimated as a weighted average of the utilities for chronic pouchitis (0.40), obstruction (0.21) and post-colectomy CD (0.41) with respective weights 54.82%, 32.14% and 13.04%.
Serious infection	-0.156	Stevenson et al. (2016)[86]	The utility decrement of 0.156 derives from a company model for TA329, as reported by Stevenson et al. However, in that appraisal the value was applied to an assumed duration of 28 days, equating to a QALY loss of 0.012 (0.156*28/365) (Stevenson et al page 213). In the current appraisal, the company subtracted 0.156 QALYs for each serious infection.

Source: Adapted from CS Table 51 and CS section B.3.4.2

Utility data collected in the UNIFI trial

EQ-5D outcomes from the UNIFI trial are outlined in CS B.2.6.1.3, B.2.6.2.4 and K.2.4, with further information provided in response to clarification question A9. We discuss EQ-5D results from the UNIFI trial in section 3.3.4.1 above. EQ-5D-5L data was collected from patients randomised in UNIFI at baseline, 8 and 16 weeks in the induction phase and at baseline, 20 and 44 weeks in the maintenance phase. Utility scores were calculated using the van Hout et al. (2012) cross-walk method [89] as recommended by NICE (CS K.2.4). Mean utility estimates were obtained for *remission, response without remission* and *active UC* health states (see Table 46 below), with classification of disease severity at the time of assessment based on Mayo and Partial Mayo scores as discussed in CS section B.3.4.1.

Table 46 Utility values estimated from the UNIFI trial using EQ-5D-3L

Health state	N	Average	Standard deviation	Minimum	Maximum
Remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response without remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Active UC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: CS Table 50					

The company use these results in a scenario analyses (Scenario 4), presented in CS Tables 69 and 70. This set of utility estimates is a major driver of cost-effectiveness results, as the ICERs for ustekinumab versus all the comparators (except vedolizumab) rise significantly above the NICE's willingness-to-pay threshold of £30,000 per QALY. The company justify not using the utilities from the UNIFI trial in the base case in CS section B.3.4.1. Briefly, they state that there are differences in active UC in the modelled health state and the UNIFI trial as: i) patients in the trial continue to receive ustekinumab, unlike in the model where they are assumed to switch to CT on loss of response; ii) inconsistency in the summary results from the UNIFI trial and published literature; and iii) insufficient duration of trial follow up to assess the change of utility over time. They also argue that the trial does not provide any information on the surgery states and that there were uncertainties as assumptions were made for patients with missing EQ-5D response and remission data.

> **ERG conclusion:** We consider that the utilities in the company's base case are generally reasonable, but with two exceptions. First, the QALY decrement for serious infections appears to have been overestimated because the disutility of 0.156 is not adjusted for the expected duration of symptoms (assumed to be 28 days in TA329). Second, clinical advice to the ERG is that the CS may overestimate utility after

revision surgery, which on average is expected to be worse than remission after the first phase of surgery. The impact of these issues is tested in ERG scenario analysis.

We agree with the company's decision not to use utility estimates from the UNIFI EQ-5D data: primarily because they are inconsistent with the values used in previous NICE appraisals for UC. However, the number of observations in the three severity health states is large and the analysis appears to have been well-conducted. The ERG therefore considers the scenario analysis with UNIFI utility estimates to be important.

4.3.6 Resource use and costs

The CS reports a systematic literature review conducted to identify resource use and costs (Appendix I). The model includes estimates of costs for drug acquisition and administration, monitoring and follow-up care and the treatment of serious infections (CS section 3.5).

4.3.6.1 Drug acquisition costs

The base case unit costs and total costs for the biologic and JAK inhibitor treatments are summarised in Table 34 below (see Table 40 above for regimens). In addition to the standard induction and maintenance, we show costs for extended induction and escalated maintenance regimens. As on the CS, this table includes the company's proposed CMU arrangement price for ustekinumab but list prices for all other drugs. Thus, these costs do not reflect the NHS price paid for other drugs with agreed PAS discounts (golimumab, vedolizumab and tofacitinib).

Conventional therapy costs used in the base case are summarised in Table 49 below. The assumptions about the percentage of patients using each drug were based on TA342, resulting in an estimated cost of £37 per 8 weeks (£235 per year). We note that the usage assumptions were updated in TA547, using results from the 2016 RCP audit of biologic treatment for IBD[67]: 50.3% aminosalicylates, 47.9% corticosteroids and 46.4% azathioprine. These result in a higher estimated cost of CT: about £59 per 8 weeks (£385 per year). Based on clinical advice to the ERG, we consider the TA547 estimates to be more realistic. We also note that the company's base case does not include costs for concomitant treatment with conventional drugs alongside biologics, which is standard practice. TA547 estimated the cost of concomitant conventional therapies at £52 with biologics and £49 with tofacitinib.

ERG conclusion: Changes to assumptions about the use and costs of CT are unlikely to be influential in the model because of their low cost and similar impact on cost-effectiveness of comparators. Nevertheless, for face validity we update the assumptions about use of conventional therapy drugs as a comparator and concurrent with other treatments as per TA547.

4.3.6.2 Drug administration costs

The cost per intravenous drug administration was estimated at £142, the cost of an outpatient visit: assuming a weighted average of consultant-led and non-consultant led, non-admitted, face-to-face follow-up appointment, 2017/18 NHS Reference Costs. Self-administered subcutaneous injections were assumed not to incur an NHS cost. Clinical advice to the ERG is that patient education and home delivery is provided by biologic drug companies without charge.

4.3.6.3 Other healthcare costs

Assumptions about resource use for monitoring and follow-up care are reported in CS Tables 57 and 58: summarised in Table 47 below.

Table 47 Health state and adverse event costs

Health state	Unit	Mean cost	Costing assumptions
Remission	Per year	£380	Tsai et al. (2008)[64] for outpatient visits, blood tests, emergency and elective endoscopies and care without colectomy
Response (without remission)	Per year	£1,021	
Active UC	Per year	£2,500	
Surgery	Per year	£2,500	Assumed equal to Active UC
Post-surgery remission	Per year	£1,398	Tsai et al. (2008)[64] with stoma care as per TA547
Post-surgery complications	Per year	£8,507	Tsai et al. (2008)[64]
First phase surgery	Per event	£15,311	Buchanan et al 2011[90] assuming 40% IPAA and 60% ileostomy, with one acute complication
Second surgery for pouch failure	Per event	£10,998	Assumed same cost as ileostomy

Serious infections	Per event	£2,674	NHS reference costs 2016-2017, HRG data. Average of 5 different types of serious infections: sepsis, pneumonia, urinary tract infection, respiratory infection and bronchitis
Source: Adapted from CS Table 56			

These originate from a panel of UK gastroenterologists reported by Tsai et al. (2008)[64] and were used in TA329, TA342 and with some adjustments in TA547. Pre-surgery admission rates were estimated from Sandborn et al. (2016).[91] Costs for surgery were based on a European study reported by Buchanan et al 2011[90]. Unit costs were based on NHS Reference Costs: inflated to 2019 prices using CPI.

> **ERG conclusion**: Estimates of health state, surgery and adverse event costs are reasonably consistent with previous UC appraisals.

4.3.7 Model validation

The company describes their approach to model validations in CS section B.3.10. They state that they engaged a clinical key opinion leader, three biostatisticians and four health economists to validate their approach to the NMA, cost-effectiveness model structure and model inputs in an advisory board meeting.

The key conclusions that the company drew from the validation exercise were:

- The experts are reported to agree with the company's 1-year NMA approach
- The CS stated that the model structure aligned with the advisory board's understanding of the natural history of the disease, and that it was consistent with previous TAs
- For input parameters, the board recommended the use of the study by Woehl et al. to inform base case utilities.
- The economic model was quality checked by an independent health economist.

Whilst the company has conducted internal validity checks (as outlined above), they have not reported any face validity checks such as comparing the proportion of patients in response and remission predicted by the model against the estimated values from the NMA. Further, they also do not provide any comparison of the model results in the current appraisal with those from previous TAs. We discuss the ERG approach to model validation in section 4.4.1 below.

Table 48 Drug acquisition costs: biologics and JAK inhibitors (CMU price for ustekinumab, other drugs at list price)

Treatment	Unit	Cost per unit	Induction (per period)				Maintenance (per year)			
			Standard period		*Extended period*		*Standard dose*		*Escalated dose*	
			Units	Cost	Units	Cost	Units	Cost	Units	Cost
Ustekinumab	130 mg	[redacted]	3.08	[redacted]	-	-	-	-	-	-
	90mg	[redacted]	-	-	1.00	[redacted]	4.33	[redacted]	6.50	[redacted]
Infliximab	100mg	£419.62	12.00	£5,035	0.00	0.00	26.00	£10,910	52.00	£21,820
- biosimilar	100mg	£377.66	12.00	£4,532	0.00	0.00	26.00	£9,819	52.00	£19,638
Golimumab	50 mg	£762.97	6.00	£4,578	4.00	£3,052	13.00	£9,919	26.00	£19,837
Adalimumab	40 mg	£352.14	8.00	£2,817	-	-	26.00	£9,156	52.00	£18,311
- biosimilar	40 mg	£308.13	8.00	£2,465	-	-	26.00	£8,011	52.00	£16,023
Vedolizumab	300 mg	£2,050.00	2.00	£4,100	1.00	£2,050	6.50	£13,325	13.00	£26,650
Tofacitinib	5 mg	£12.32	-	-	-	-	730.50	£9,001	-	-
	10 mg	£24.64	112.00	£2,760	112.00	£2,760	-	-	730.50	£18,002

Source: Adapted by ERG from CS Tables 52 and 53, with additional information from model sheet "Cost&MRU Inputs_UK

Table 49 Drug acquisition costs: conventional therapies

Treatment	Dose	Unit	Cost per unit	Base case (per 8 weeks)			Usage (% patients) in TA547		
				% patients	Units	Cost	CT alone	With biologic	With tofacitinib
Azathioprine	2.5mg/kg/day	50 mg	£0.04	39%	206	£8.28	46.4%	37.2%	0%
Mercaptopurnine	1.5mg/kg/day	50 mg	£1.97	15%	124	£243.16	-	-	-
Methotrexate	17mg/kg/day	2.5 mg	£0.06	9%	55	£3.38	-	-	-
Mesalazine	1g/week	750 mg	£0.31	13%	21	£6.56	12.6%	11.6%	11.6%
Balsalazide	1.5 g bid	750 mg	£0.23	-	-	-	12.6%	11.6%	11.6%
Olsalazine	500mg bid	500 mg	£2.68	-	-	-	12.6%	11.6%	11.6%
Sulfasalazine	500mg bid	500 mg	£0.06	-	-	-	12.6%	11.6%	11.6%
Prednisone	20mg/day	20 mg	£0.03	36%	14	£0.49	44.1%	19.9%	19.9%
Budesonide	3mg tid	3 mg	£0.75	1%	168	£126.08	-	-	-
Total cost (per 8 weeks)						£37.43	£59.30	£52.18	£49.40

Source: Adapted by ERG from CS Tables 54 and 55, with additional information from model sheet "Cost&MRU Inputs_UK

4.3.8 Company cost effectiveness results

4.3.8.1 Base case deterministic results

The company present their base case results in CS section B.3.7. These incorporate the confidential company's proposed CMU arrangement price for ustekinumab, but not for the comparator arms. We report results including the company's proposed CMU arrangement price for ustekinumab and all available PAS discounts for the comparators in a confidential addendum to this report.

Results for the people without previous failure of biologic treatment are shown in Table 50.

- Adalimumab, adalimumab biosimilar, golimumab, tofacitinib, infliximab, infliximab biosimilar and vedolizumab are dominated by ustekinumab;
- Ustekinumab gives a mean QALY gain of [redacted] for a mean additional cost of [redacted] compared with conventional therapy: giving an incremental cost-effectiveness ratio (ICER) of £23,446 per QALY gained.

Table 50 Cost effectiveness: Company base case, non-biologic failure

Technologies	Total Discounted costs (£)	Total Discounted QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	-	£23,446
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£19,146
Adalimumab	[redacted]	[redacted]	Dominated	£18,047
Infliximab biosimilar	[redacted]	[redacted]	Extended Dominated	£16,606
Infliximab	[redacted]	[redacted]	Dominated	£14,710
Golimumab	[redacted]	[redacted]	Dominated	£12,025
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£13,465
Vedolizumab	[redacted]	[redacted]	Dominated	£1,762
Ustekinumab	[redacted]	[redacted]	£23,446	-
Source: reproduced from CS Table 62				

Company base case results for the biologic failure subgroup are shown in Table 51. The company appropriately omits golimumab and infliximab as comparators in this subgroup due to the lack of effectiveness evidence.

- Ustekinumab dominated adalimumab, adalimumab biosimilar, tofacitinib and vedolizumab;
- Compared with conventional therapy, ustekinumab gives a mean QALY gain of [redacted] for an additional cost of [redacted]; hence, an ICER of £26,205 per QALY gained.

Table 51 Cost effectiveness: Company base case, biologic failure subgroup

Technologies	Total Discounted costs (£)	Total Discounted QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	-	£26,205
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£19,670
Adalimumab	[redacted]	[redacted]	Dominated	£18,210
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£5,394
Ustekinumab	[redacted]	[redacted]	£26,205	-
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Source: reproduced from CS Table 63				

4.3.8.2 Deterministic sensitivity analyses

The company briefly summarises the parameters and ranges included in their Deterministic Sensitivity Analysis (DSA) in CS section B.3.8.1.1. Results of the DSA for the non-biologic failure and biologic-failure subgroups are tabulated in CS Tables 64 and 65 and presented as tornado plots in CS Figures 39 and 40. The tornado plots for both subgroups show that the health state utility values, discount rates and disease management costs are key drivers of model results. Other parameters such as model starting age, time horizon and response/remission odds ratio for induction also influence the base case results, but to a lesser extent.

4.3.8.3 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 Table 60. Briefly, the company assigns lognormal distribution for efficacy and safety parameters for the induction phase and beta distribution for maintenance phase. Health state utilities are assigned beta distributions; and gamma distributions are used for adverse event costs and surgery related costs. Probabilistic results are presented in CS Tables 66 and 67; scatter plots in CS Figures 42 and 44; and cost effectiveness acceptability curves (CEACs) are in CS Figures 41 and 43. The PSA results

are similar to the base case results. The CS states that at a willingness-to-pay threshold of £30,000 per QALY, ustekinumab had 100% probability of being cost-effective compared to CT in the non-biologic failure group; and 95% probability of being cost-effective in the biologic failure group respectively.

The company provided a revised version of their model with corrections to the random number sampling in response to clarification question B7. We consider that the PSA still has limitations and does not reflect uncertainty over the input parameters. In particular, it does not preserve the joint posterior distribution for NMA parameters and the same random numbers are used to sample sets of health state utilities and disease management costs.

> **ERG conclusion:** We consider that the PSA has limitations that mean that it may not appropriately reflect uncertainty over the input parameters.

4.3.8.4 Scenario Analysis

The company conducted a range of scenario analyses to assess the impact of key variables on the model outcomes. We reproduce a summary of the scenarios in Table 52 and Table 53 below (from CS Tables 69 and 70). The company concluded that the cost effectiveness results in both sub-groups were predominantly influenced by: the efficacy source for the maintenance phase (the 1-year NMA conditional on response, rather than direct trial data), health state utilities from UNIFI trial (rather than estimates from the literature) and including subsequent treatment upon loss of response.

We highlight in particular the large increase in the ICERs for ustekinumab with UNIFI utility estimates. This is driven by the high utility for active UC, which reduces the QALY gain from inducing and retaining response and remission.

We extend the range of scenario analyses in ERG additional analyses below (section 4.4.3).

Table 52 Company scenario analyses, non-biologic failure (ustekinumab vs comparators)

Scenario	Infliximab	Infliximab biosimilar	Golimumab	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
Base Case	***£14,710***	***£16,606***	***£12,025***	***£18,047***	***£19,146***	***£1,762***	***£13,465***	***£23,446***
1) Induction NMA	£14,705	£16,603	£12,025	£18,051	£19,147	£1,755	£13,427	£23,446
2) Maintenance NMA	£10,665	£13,648	£6,294	£17,198	£18,785	Dominant	£7,625	£24,575
3) Non-constant loss of response	£15,647	£17,312	£13,159	£18,379	£19,349	£3,888	£14,361	£23,053
4) Utilities from UNIFI trial	£48,809	£55,103	£39,980	£60,069	£63,726	£5,879	£45,136	£78,091
5) Utility values from Swinburn et al 2012[87]	£14,658	£16,548	£11,984	£17,984	£19,079	£1,756	£13,419	£23,363
6) Subsequent treatment	£13,953	£15,889	£11,245	£17,359	£18,480	£7,474	£12,708	£27,785
7) Dose escalation 10%	£12,261	£14,158	£11,319	£17,078	£18,055	£2,703	£13,152	£21,701
8) Dose escalation 50%	£17,158	£19,055	£12,731	£19,017	£20,238	£821	£13,778	£25,191
9) Delayed responder loss of response	£11,767	£14,475	£9,496	£16,903	£18,200	Dominant	£8,599	£23,297
10) Exclude delayed responders	£7,953	£10,521	£9,339	£13,869	£15,446	Dominant	£11,762	£21,870
11) Serious infection	£14,823	£16,726	£12,103	£18,084	£19,184	£1,762	£13,465	£23,446
Source: CS Table 69								

Table 53 Company scenario analyses, biologic failure (ustekinumab vs comparator)

Scenario	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
Base Case	***£18,210***	***£19,670***	***Dominant***	***£5,394***	***£26,205***
1) Induction NMA	£18,316	£19,783	Dominant	£5,590	£26,334
2) Maintenance NMA	£14,194	£20,355	Dominant	Dominant	£28,018
3) Non-constant loss of response	£18,680	£19,985	£2,471	£7,388	£25,711
4) Utility values from UNIFI trial	£60,278	£65,111	Dominant	£18,037	£86,723
5) Utility values from Swinburn et al 2012 [87]	£18,142	£19,597	Dominant	£5,375	£26,106
6) Dose escalation set to 10%	£17,530	£18,878	Dominant	£6,590	£24,733
7) Dose escalation set to 50%	£18,934	£20,505	Dominant	£3,338	£27,705
8) Delayed responder loss of response	£15,805	£17,637	Dominant	Dominant	£25,880
9) Exclude delayed responders	£11,068	£13,261	Dominant	£5,488	£23,525
10) Serious infection	£18,253	£19,714	Dominant	£5,394	£26,205

Source: CS Table 7

4.4 Additional work undertaken by the ERG

4.4.1 ERG model validation

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 response to ERG Clarification question A12. We conducted a range of 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 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).

The company model was generally well-implemented, with no substantive errors in parameter inputs or coding. We consider that there were problems with the PSA calculations and the disutility for adverse events, as discussed in sections 4.3.8.3 and 4.3.5 above. See section 4.4.2 for further detail.

We compare the modelled QALY estimates from the current appraisal with those from two previous NICE appraisals for UC (TA342 and TA329) and the study by Wu et al.[92] (see Table 54). QALY results from the NICE appraisal on Tofacitinib (TA547) are not available, as they were commercial in confidence. Despite methodological differences between the models, they provide some means of cross-validation. We highlight that the QALY estimates from the ustekinumab company model are lower than those from the other available lifetime models: e.g. QALY estimates were around 10.5 with conventional treatment in TA329 and Wu et al. but only about 8.5 in the current model.

Table 54 Comparison of modelled outcomes

Source (time horizon)	QALYs	
Current appraisal (Lifetime)	**Non-biologic failure**	**Biologic failure**
	CT: [redacted]	CT: [redacted]
	Infliximab: [redacted]	
	Adalimumab: [redacted]	Adalimumab: [redacted]
	Golimumab: [redacted]	
	Vedolizumab: [redacted]	Vedolizumab: [redacted]
	Tofacitinib: [redacted]	Tofacitinib: [redacted]
	Ustekinumab: [redacted]	Ustekinumab: [redacted]
TA342 (10 years)	CT: 4.28	CT: 5.37
	Infliximab: 5.82	
	Adalimumab: 5.76	
	Golimumab: 5.79	
	Vedolizumab: 5.90	Vedolizumab: 5.46
	Surgery: 4.28	Surgery: 4.28
TA329 (Lifetime)	**Moderate to severe UC who failed at least 1 prior therapy**	
	CT: 10.47	
	Infliximab:10.81	
	Adalimumab: 10.82	
	Golimumab: 10.63	
Wu et al. (lifetime)	**Moderate to severe UC**	
	CT:10.49	
	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	

Source: ERG Table 76 in NICE TA547

4.4.2 ERG corrections to company model

We summarise ERG comments on errors in the company model in Table 55. Due to the limitations of the PSA, we consider that it does not appropriately reflect uncertainty over the model parameters. We would like to have corrected the PSA by including CODA output and revising the utility sampling, but we did not consider this a priority as it would not impact on the base case and scenario results. However, we do urge caution in interpreting the PSA. We did correct the QALY decrement for serious infections to adjust for their duration as well disutility (see ERG additional analysis in section 4.4.3). This is a small change that is unlikely to have a significant impact on cost-effectiveness.

Table 55 ERG comments on errors in the company model

Aspect of model	Problem	ERG comment
Probabilistic sensitivity analysis	The submitted model used a single random number per PSA iteration to sample response and remission values for all treatments. This underestimates uncertainty over relative treatment effects and correlations between response and remission probabilities.	The company addressed this issue in their response to clarification question B7 and supplied a corrected version of the model.
	The model did not use CODA samples to reflect uncertainty over NMA results. Thus, the PSA does not reflect the joint posterior distribution, with correlations across treatments.	The post-clarification version of the model did not use CODA samples for the PSA. The company argued this was not feasible, given that 200,000 NMA iterations were required for the model to converge.
	The company assign the same random numbers for health state utilities and disease management costs.	The company samples absolute health state utilities, rather than one baseline utility and the utility multipliers. We consider the latter approach to be better, as it avoids inconsistent values.
Adverse event disutility	The company do not adjust the utility decrement associated with serious infections for the duration of symptoms.	We adjust this utility in our additional analyses (see section 4.4.3).

4.4.3 ERG additional analyses

A full summary of ERG observations on key aspects of the company's economic model is provided in Appendix 9. Based on this critique, we have identified 7 key aspects of the company base case with which we disagree. Our preferred approach is:

1. **Recurrent response and remission**; To include response and remission health states for conventional therapy after failure of the initial treatment: reflecting the chronic intermittent form of disease that some experience [74] (see section 4.3.3.4). In our base case, we assume an overall response rate of 5.5% per 8 weeks (4.0% response without remission): converted to 2-week probabilities (with and without remission) and applied at each cycle to patients in the active UC health state. This was chosen to be lower than the response rate for CT as a comparator (Table 41) and to produce a lifetime discounted QALY estimate of 10.5, similar to TA329 and Wu et al. (see Table 54).[63,66] We assumed the same rate of loss of response as for maintenance CT.
2. **Induction NMA**: We agree with the use of a fixed effects NMA to estimate induction response and remission rates, but found some differences on replication (section 4.3.4.1). We use ERG estimates in our preferred analysis.
3. **Maintenance NMA**: Use of an NMA approach to estimate response and remission on maintenance, rather than individual treatment arms from trials with a pooled placebo (section 4.3.4.2). In our base case, we use the company's 1-year NMA conditional on response (ERG replicated random effect model with pooled doses), which pools placebo arms across trials to adjust for potential carry over of effects from induction. The ERG maintenance only NMA (no carry over) is used in scenario analysis. We consider that the true effect is likely to lie somewhere between these extremes.
4. **Conventional drug mix**: Cost of CT based on results from the 2016 RCP audit of biologic treatment for IBD, as in TA547 (section 4.3.6.1).[67,93]
5. **Concurrent conventional treatment**: Inclusion of costs for concurrent treatment with conventional therapies alongside biologic or JAK inhibitor treatment, with costs estimated as in TA547. [67,93]
6. **Dose escalation with infliximab**: Same assumptions about dose escalation for infliximab as for other therapies to reflect clinical practice: assume 30% of patients on higher dose (section 4.3.2.2).
7. **Disutility for serious infection**: Disutility adjusted for duration of symptoms, as in TA329.[63]

Table 57 shows the cumulative effect of these changes to the company base case. We observe a minor difference in the costs and QALYs for CT when results in the company's model are pulled from the 'Markov_SOC' sheet, rather than the 'Markov_UK' sheet, which is used for all other comparators: see Table 56. We used the latter Markov sheet for all comparators (including CT) to add response and remission health states after discontinuation of the initial treatment. This explains why the results with company base assumptions in the first section of Table 57 differ slightly from those in the CS. A comparison of the scenario results calculated with the company model and the ERG version also shows small differences (Appendix 10).

Table 56 Comparison with CT from Markov-SoC and Markov_UK sheets

Drugs	CT results from 'Markov_SoC' sheet			CT results from 'Markov_UK' sheet		
	Total Discounted costs (£)	Total Discounted QALYs	ICER (£/QALY) vs comparator	Total Discounted costs (£)	Total Discounted QALYs	ICER (£/QALY) vs comparator
Non biologic failure subgroup						
CT	[redacted]	[redacted]	£23,446	[redacted]	[redacted]	£23,450
Adalimumab biosimilar	[redacted]	[redacted]	£19,146	[redacted]	[redacted]	£19,146
Adalimumab	[redacted]	[redacted]	£18,047	[redacted]	[redacted]	£18,047
Infliximab biosimilar	[redacted]	[redacted]	£16,606	[redacted]	[redacted]	£16,606
Infliximab	[redacted]	[redacted]	£14,710	[redacted]	[redacted]	£14,710
Golimumab	[redacted]	[redacted]	£12,025	[redacted]	[redacted]	£12,025
Tofacitinib	[redacted]	[redacted]	£13,465	[redacted]	[redacted]	£13,465
Vedolizumab	[redacted]	[redacted]	£1,762	[redacted]	[redacted]	£1,762
Ustekinumab	[redacted]	[redacted]	-	[redacted]	[redacted]	-
Biologic failure subgroup						
CT	[redacted]	[redacted]	£26,205	[redacted]	[redacted]	£26,213
Adalimumab biosimilar	[redacted]	[redacted]	£19,670	[redacted]	[redacted]	£19,670
Adalimumab	[redacted]	[redacted]	£18,210	[redacted]	[redacted]	£18,210
Tofacitinib	[redacted]	[redacted]	£5,394	[redacted]	[redacted]	£5,394
Ustekinumab	[redacted]	[redacted]	-	[redacted]	[redacted]	-
Vedolizumab	[redacted]	[redacted]	Dominant	[redacted]	[redacted]	Dominant

Table 57 Cumulative impact of ERG preferred assumptions: Non-biologic failure (company's proposed CMU arrangement price for ustekinumab and list price for comparators)

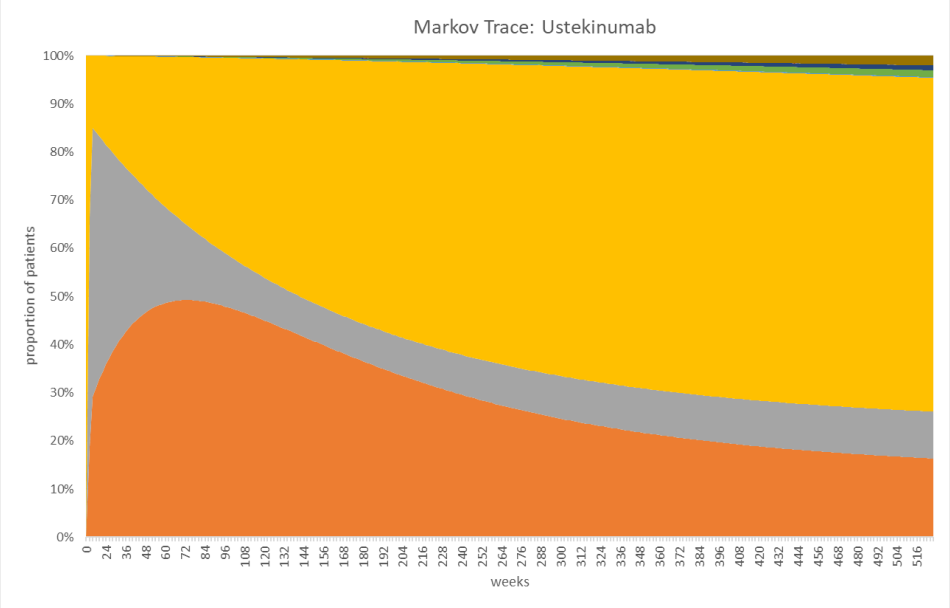
Drug	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Company base case (from ERG version of the model)				
Ustekinumab	[redacted]	[redacted]	£23,450	-
Vedolizumab	[redacted]	[redacted]	Dominated	£1,762
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£13,465
Golimumab	[redacted]	[redacted]	Dominated	£12,025
Infliximab	[redacted]	[redacted]	Dominated	£14,710
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£16,606
Adalimumab	[redacted]	[redacted]	Dominated	£18,047
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£19,146
SoC/CT	[redacted]	[redacted]	-	£23,450
+ Response and remission after initial treatment failure; 8 week response on CT 0.055				
Ustekinumab	[redacted]	[redacted]	£31,609	-
Vedolizumab	[redacted]	[redacted]	Dominated	£3,782
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£18,581
Golimumab	[redacted]	[redacted]	Dominated	£16,706
Infliximab	[redacted]	[redacted]	Dominated	£20,328
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£22,760
Adalimumab	[redacted]	[redacted]	Dominated	£24,664
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£26,076
SoC/CT	[redacted]	[redacted]	-	£31,609
+ Induction NMA, fixed effects (ERG replication)				
Ustekinumab	[redacted]	[redacted]	£31,602	-
Vedolizumab	[redacted]	[redacted]	Dominated	£3,790
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£18,563
Golimumab	[redacted]	[redacted]	Dominated	£16,704
Infliximab	[redacted]	[redacted]	Dominated	£20,323
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£22,754
Adalimumab	[redacted]	[redacted]	Dominated	£24,660
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£26,072
SoC/CT	[redacted]	[redacted]	-	£31,602
+ 1 year NMA conditional on response, random effects (ERG replication)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£32,813	-
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£10,853
Golimumab	[redacted]	[redacted]	Dominated	£9,384
Infliximab	[redacted]	[redacted]	Dominated	£14,835

Drug	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£18,647
Adalimumab	[redacted]	[redacted]	Dominated	£23,491
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,519
SoC/CT	[redacted]	[redacted]	-	£32,813
+ TA547 assumptions on mix of treatments for CT				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£33,037	-
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,027
Golimumab	[redacted]	[redacted]	Dominated	£9,555
Infliximab	[redacted]	[redacted]	Dominated	£15,011
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£18,823
Adalimumab	[redacted]	[redacted]	Dominated	£23,674
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,702
SoC/CT	[redacted]	[redacted]	-	£33,037
+ TA547 assumptions on concomitant treatments				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£33,200	-
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,115
Golimumab	[redacted]	[redacted]	Dominated	£9,625
Infliximab	[redacted]	[redacted]	Dominated	£15,041
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£18,854
Adalimumab	[redacted]	[redacted]	Dominated	£23,744
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,772
SoC/CT	[redacted]	[redacted]	-	£33,200
+ Dose escalation for Infliximab (30% same as other treatments)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£33,200	-
Infliximab	[redacted]	[redacted]	Dominated	£7,941
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,115
Golimumab	[redacted]	[redacted]	Dominated	£9,625
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£12,466
Adalimumab	[redacted]	[redacted]	Dominated	£23,744
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,772
SoC/CT	[redacted]	[redacted]	-	£33,200
+ Adjusted utility decrement for serious infections (as in TA329)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£33,192	-
Infliximab	[redacted]	[redacted]	Dominated	£7,988

Drug	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,112
Golimumab	[redacted]	[redacted]	Dominated	£9,672
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£12,540
Adalimumab	[redacted]	[redacted]	Dominated	£23,777
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,807
SoC/CT	[redacted]	[redacted]	-	£33,192
ERG preferred base case				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£33,192	-
Infliximab	[redacted]	[redacted]	Dominated	£7,988
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,112
Golimumab	[redacted]	[redacted]	Dominated	£9,672
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£12,540
Adalimumab	[redacted]	[redacted]	Dominated	£23,777
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,807
SoC/CT	[redacted]	[redacted]	-	£33,192

Note: CE results for Biosimilar-Renflexis are excluded from the above table as they are similar to those for biosimilar-Inflectra

ERG base case



Company base case

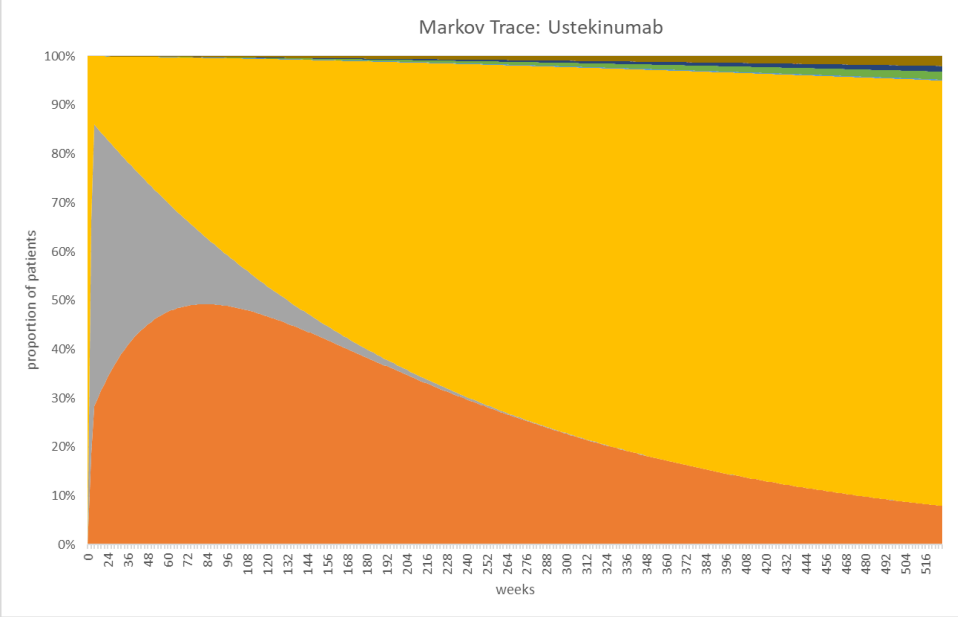
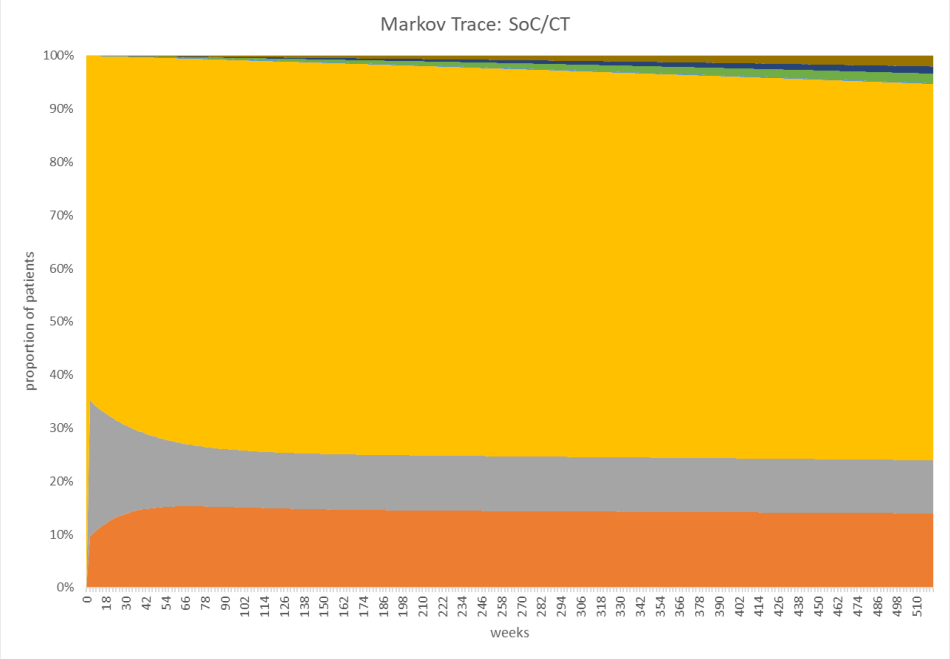


Figure 15 Comparison of Markov Traces for ustekinumab: Proportion of cohort in each Health State over time, non-biologic failure subgroup

ERG Base case



Company Base case

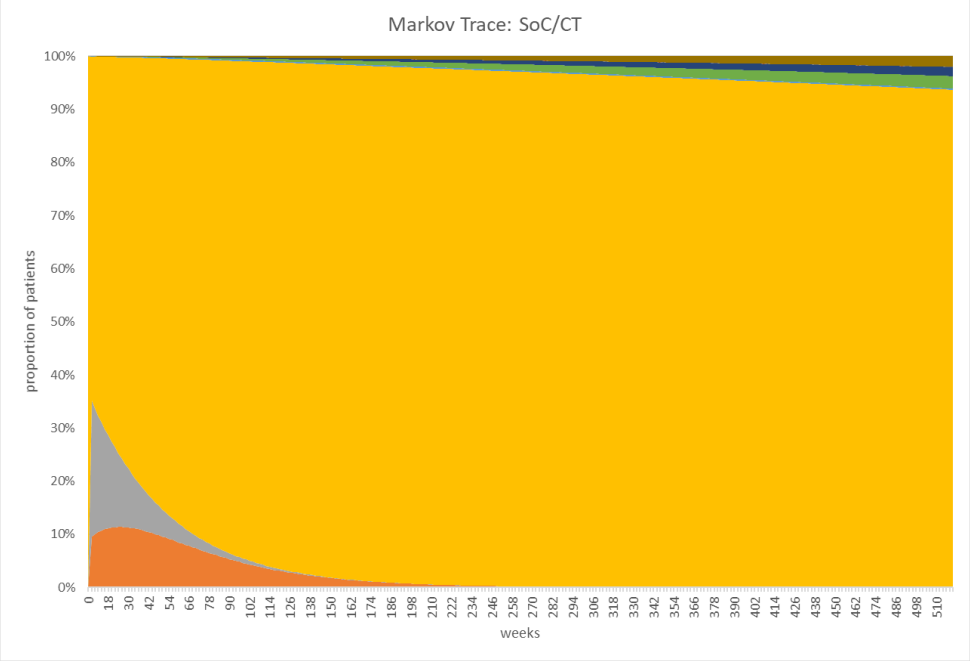


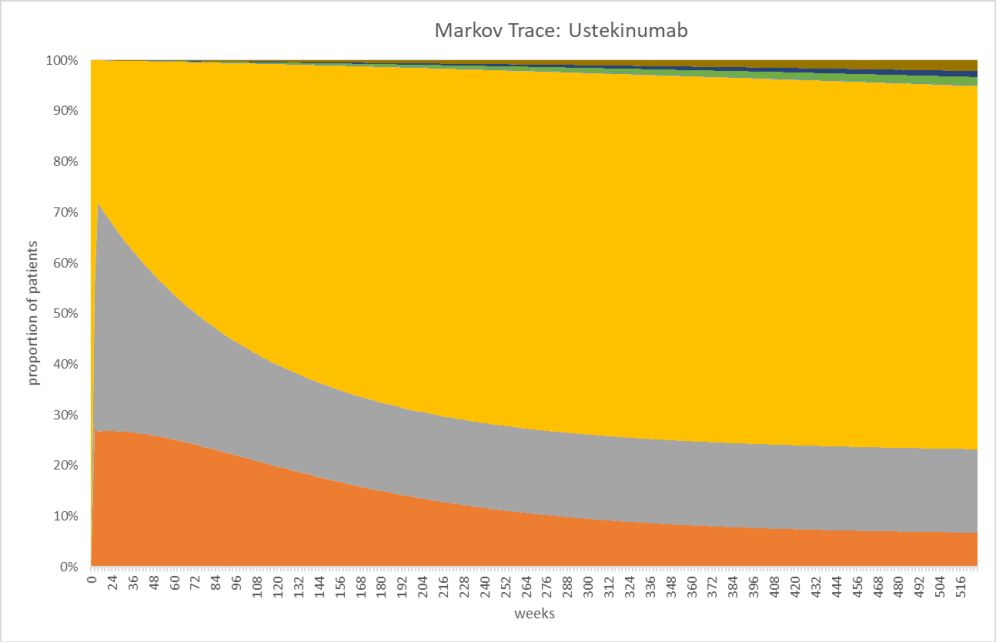
Figure 16 Comparison of Markov Traces for SoC/CT: proportion of cohort in each Health State over time, non-biologic failure subgroup

Table 58 Cumulative impact of ERG preferred assumptions: Biologic Failure subgroup, company's proposed CMU arrangement price for ustekinumab and list price for comparators

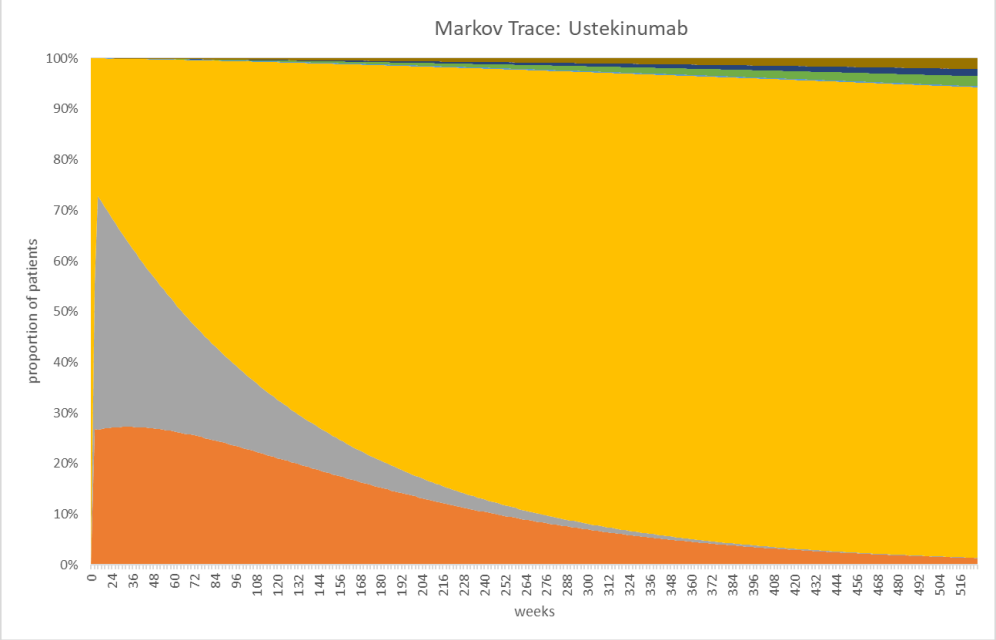
Treatment	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Company base case (from ERG version of the model)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£26,213	-
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£5,394
Adalimumab	[redacted]	[redacted]	Dominated	£18,210
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£19,670
SoC/CT	[redacted]	[redacted]	-	£26,213
+ Response and remission after initial treatment failure; 8 week response on CT: 0.055				
Ustekinumab	[redacted]	[redacted]	£33,879	-
Vedolizumab	[redacted]	[redacted]	Dominated	£766
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£7,818
Adalimumab	[redacted]	[redacted]	Dominated	£23,978
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,799
SoC/CT	[redacted]	[redacted]	-	£33,879
+ Induction NMA, fixed effects (ERG replication)				
Ustekinumab	[redacted]	[redacted]	£33,972	-
Vedolizumab	[redacted]	[redacted]	Dominated	£823
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£7,970
Adalimumab	[redacted]	[redacted]	Dominated	£24,064
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£25,883
SoC/CT	[redacted]	[redacted]	-	£33,972
+ 1 year NMA conditional on response, random effects (ERG replication)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£36,560	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,527
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£27,863
SoC/CT	[redacted]	[redacted]	-	£36,560
+ TA547 assumptions on mix of treatments for CT				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant

Treatment	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£36,808	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,737
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£28,073
SoC/CT	[redacted]	[redacted]	-	£36,808
+ TA547 assumptions on concomitant treatments				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£37,033	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,778
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£28,114
SoC/CT	[redacted]	[redacted]	-	£37,033
+ Dose escalation for Infliximab (30% same as other treatments)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£37,033	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,778
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£28,114
SoC/CT	[redacted]	[redacted]	-	£37,033
+ Adjusted utility decrement for serious infections (as in TA329)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£37,023	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,914
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£28,308
SoC/CT	[redacted]	[redacted]	-	£37,023
ERG preferred base case				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£37,023	-
Adalimumab	[redacted]	[redacted]	Dominated	£19,914
Adalimumab biosimilar	[redacted]	[redacted]	Extended Dominated	£28,308
SoC/CT	[redacted]	[redacted]	-	£37,023

ERG base case



Company base case



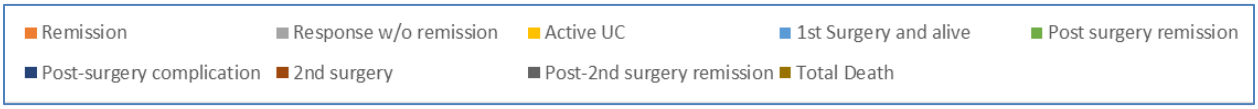
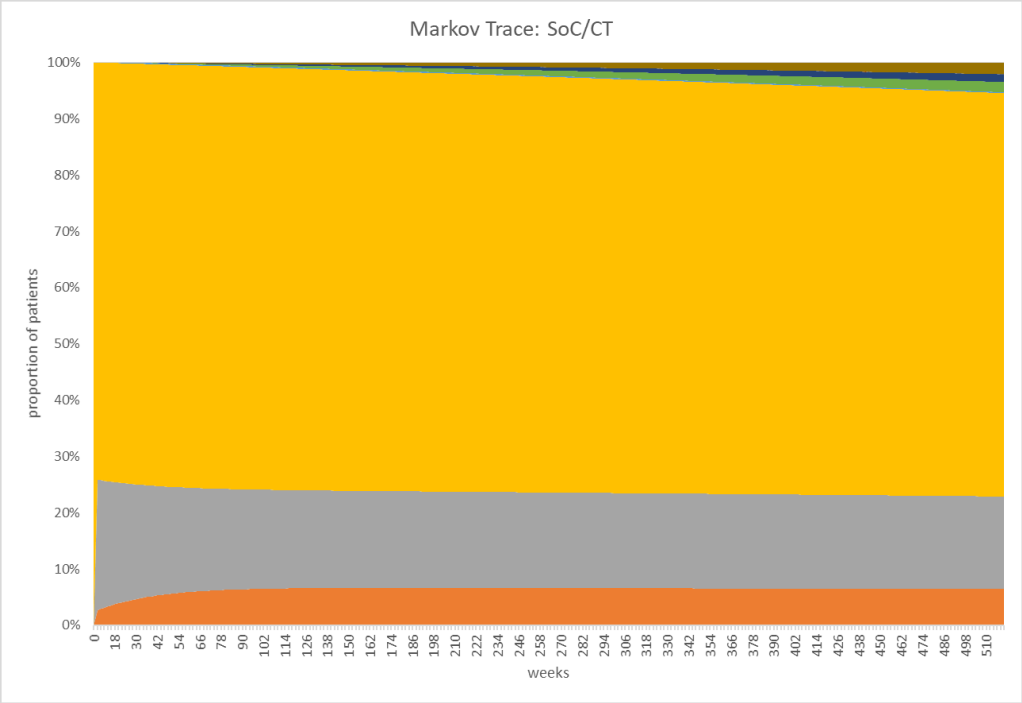


Figure 17 Comparison of Markov Traces for ustekinumab: proportion of cohort in each Health State over time, biologic failure subgroup

ERG base case



Company base case

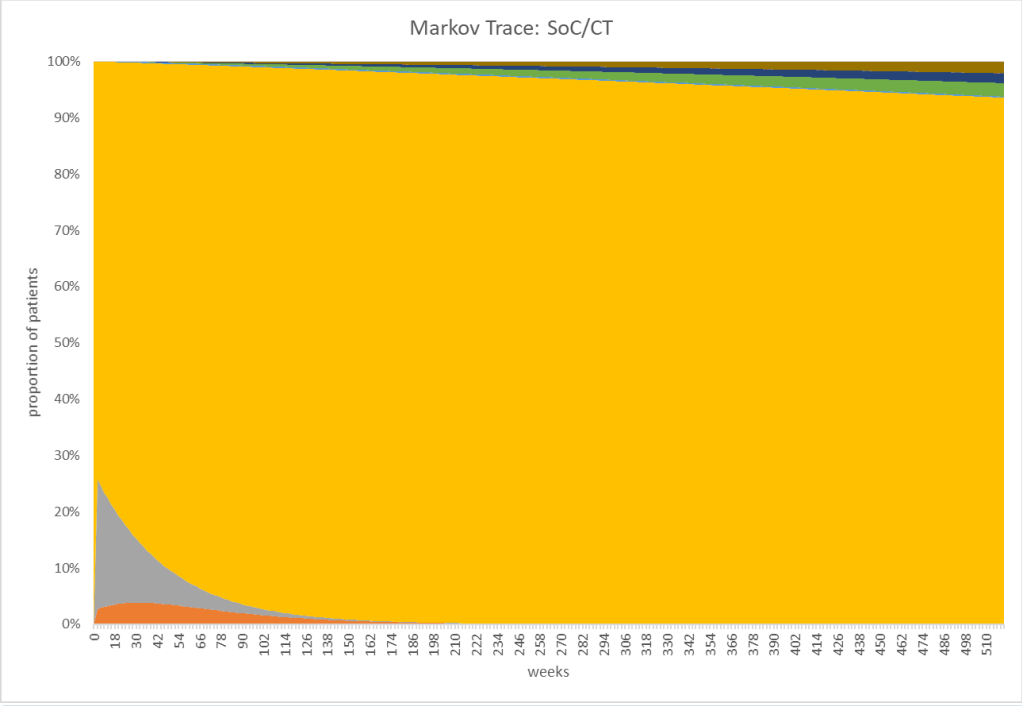


Figure 18 Comparison of Markov Traces for SoC/CT: proportion of cohort in each Health State over time, biologic failure subgroup

Table 59 Additional scenario analyses conducted on the ERG base case, non-biologic failure (ustekinumab vs comparators), company's proposed CMU arrangement price for ustekinumab; list prices for comparators

Scenario	Infliximab	Infliximab biosimilar	Golimumab	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
ERG Base Case	**£7,988**	**£12,540**	**£9,672**	**£23,777**	**£25,807**	**Dominant**	**£11,112**	**£33,192**
1) 8-week response on CT: 0.03	£6,721	£10,835	£8,289	£20,865	£22,691	Dominant	£9,585	£29,387
2) 8-week response on CT: 0.08	£9,254	£14,244	£11,045	£26,716	£28,953	Dominant	£12,630	£37,021
3) Induction NMA – ERG Random effects model	£8,122	£12,643	£9,422	£23,796	£25,821	Dominant	£10,519	£33,180
4) Maintenance only NMA- ERG scenario of no carry over effect	£2,990	£8,983	Dominant	£24,583	£27,249	Dominant	Less costly and less effective	£39,903
5) Utilities from UNIFI trial	£26,604	£41,766	£32,252	£78,989	£85,735	Dominant	£37,225	£110,391
6) Utility values from Swinburn et al 2012 [87]	£7,961	£12,498	£9,641	£23,694	£25,718	Dominant	£11,076	£33,079
7) Dose escalation 10% for all treatments	£9,091	£13,165	£9,628	£22,874	£24,682	Dominant	£12,092	£30,920
8) Dose escalation 50% for all treatments	£6,885	£11,915	£9,716	£24,680	£26,933	Dominant	£10,132	£35,465
9) Delayed responder loss of response	£1,173	£7,634	£7,038	£23,079	£25,368	Dominant	£854	£33,609
10) Exclude delayed responders	Dominant	Dominant	£5,290	£16,854	£19,906	Dominant	£8,827	£31,783
11) Serious Infection	£8,121	£12,676	£9,780	£23,816	£25,847	Dominant	£11,112	£33,192

Table 60 Additional scenario analyses conducted on the ERG base case, biologic failure (ustekinumab vs comparators), company's proposed CMU arrangement price for ustekinumab; list prices for comparators

Scenario	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
ERG Base Case	**£19,914**	**£28,308**	**Dominant**	**Dominant**	**£37,023**
1) 8-week response on CT: 0.03	£17,218	£24,707	Dominant	Dominant	£33,071
2) 8-week response on CT: 0.08	£22,709	£32,039	Dominant	Dominant	£40,991
3) Induction NMA –ERG Random effects model	£20,216	£28,392	Dominant	Dominant	£37,065
4) Maintenance only NMA-ERG scenario of no carry over effect	£3,830	£28,027	Less costly, less effective	Less costly, less effective	£44,121
5) Utilities from UNIFI trial	£65,017	£92,421	Dominant	Dominant	£122,461
6) Utility values from Swinburn et al 2012 [87]	£19,831	£28,189	Dominant	Dominant	£36,888
7) Dose escalation 10% for all treatments	£23,459	£31,069	Dominant	Dominant	£35,180
8) Dose escalation 50% for all treatments	£16,659	£25,832	Dominant	Dominant	£38,909
9) Delayed responder loss of response	£12,003	£24,448	Less costly, less effective	Less costly, less effective	£37,149
10) Exclude delayed responders	Dominant	Dominant	Dominant	Dominant	£34,219
11) Serious infection	£20,078	£28,476	Dominant	Dominant	£37,023

4.4.4 Summary of ERG additional analysis results

Results from the ERG preferred assumptions

We show the cumulative impact of applying the ERG preferred assumptions to the company's base case model in Table 57 and Table 58. We observe the following:

- The change that has the biggest impact on the cost effectiveness results is the addition of response and remission health states for conventional therapy after initial treatment failure. Introducing these additional health states in the model increases the ICERs for ustekinumab vs comparators; particularly the ICER for ustekinumab versus SoC/CT. In the non-biologic failure subgroup, the ICER increases from £23,450 (company base case) to £31,609 in the ERG scenario; an increase of £8,159. In the biologic-failure subgroup, the ICER for ustekinumab versus SoC/CT increases from £26,213 (company base case) to £33,879; an increase of £7,666. In both the subgroups, the ICERs for ustekinumab versus all other comparators increase slightly, although they remain below £30,000 (in this analysis, which includes the company's proposed CMU arrangement price for ustekinumab but not for all comparators).

- We present a comparison of the Markov traces for the ERG and company base cases showing the proportion of the cohort in each health state over time in Figure 15 (ustekinumab, non-biologic failure), Figure 16 (SoC/CT, non-biologic failure), Figure 17 (ustekinumab, biologic failure), and Figure 18 (SoC/CT, biologic failure).

- As expected, the proportions of patients in remission and response without remission health states are higher for both the subgroups in the ERG base case compared with the company's base case. We consider that the ERG analysis gives a more realistic representation of the clinical course of UC, with a proportion of patients continuing to experience periods of response and remission despite failure of biologic and conventional treatments. This view is supported by clinical advice to the ERG, and cohort studies cited in the CS.[6]

- Using the NMA results from the ERG replication for the induction phase has minimal impact on the cost-effectiveness results in both the subgroups. This is consistent with the company's scenario analysis (Scenario 1 in CS Table 69 and CS Table 70).

- The scenario using the company's 1–year NMA conditional on response (using the ERG replicated random effects model with pooled doses), causes a modest increase in the ICER for ustekinumab vs SoC/CT. In the non-biologic failure subgroup, the ICER increases to £32,813, and in the biologic failure sub-group, it increases to £36,560. In both the subgroups, all other comparators remain dominated or extendedly dominated in full incremental analyses (without PAS discounts for some comparators).

- Using similar assumptions as NICE TA547 on treatment mix for CT and the use of concomitant treatments, in both the subgroups the ICERs for ustekinumab versus comparators increase minimally, without changing the direction of the overall cost-effectiveness results.

- Using the same dose escalation for infliximab as other treatments (i.e. 30%) decreases the ICER for ustekinumab versus infliximab slightly, making the latter slightly less cost-effective, as might be expected. This scenario is only applicable in the non-biologic failure subgroup and does not influence the results in the biologic failure subgroup.

- Adjusting the utility decrement for serious infections similar to the approach in NICE TA329 has a minimal impact on the overall cost-effectiveness results in both the subgroups.

Compared with the company's base case, our preferred assumptions collectively decrease the total costs of all the treatments and increase their total QALYs: this is largely because the addition of response and remission health states after patients revert to standard care reduces mean time spent with active disease and the incidence of surgery. In the full incremental analyses, all the comparators except CT remain dominated or extendedly dominated by ustekinumab. This is consistent with the company's base case, although under our preferred set of assumptions, the ICER for ustekinumab versus CT increases by £9,742 in the non- biologic failure subgroup; and by £10,810 in the biologic failure subgroup. However, we note again that these results do not take account the PAS discounts for vedolizumab and tofacitinib. Final results including the company's proposed CMU arrangement price for ustekinumab and all PAS discounts for the comparators are provided in the confidential addendum to this report.

Results from the scenario analyses conduced on the ERG base case

We performed a range of additional scenario analyses on the ERG base case, as summarised in Table 59 for non-biologic failure subgroup and Table 60 for biologic failure subgroup, respectively. We note:

- Of all the scenarios, using health state utilities estimated from the UNIFI trial had the greatest impact on cost-effectiveness. In the non-biologic failure subgroup, the ICER for ustekinumab versus CT increases to £110,391 (an increase of £77,199 from the ERG base case); and in the biologic failure subgroup it increases to £122,461 (an increase of £85,438 from the ERG base case). This is caused by the higher utility estimate for active UC ([redacted]) estimated from UNIFI compared with the base case value (0.41) from Woehl et al. (2008)[84], which reduces the QALY gain from better induction and maintenance of response and remission.

- For all other scenarios, the ICERs for ustekinumab versus all the comparators (except SoC/CT) remain under £30,000 and are dominated or extendedly dominated in full incremental analyses. This is true for both the subgroups. However, the ICERs for ustekinumab versus SoC/CT range between £29,387 *(Scenario: 8-week response rate on CT: 0.03)* and £39,903 *(Scenario: NMA maintenance only- ERG scenario of no carry over effect)* in the non-biologic subgroup. In the biologic subgroup, the ICERs for ustekinumab versus SoC/CT ranges between £33,071*(Scenario: 8-week response rate on CT: 0.03)* and £44,121 *(Scenario: NMA maintenance only- ERG scenario of no carry over effect)* respectively. The ERG maintenance-only NMA scenario is less favourable to ustekinumab than the 1-year conditional on response NMA that we use in our base case. This is driven by different underlying assumptions about the causes of differences in placebo response rates from re-randomised studies (carry-over effects from active induction treatment or other differences between trial populations or conduct).

The ERG have also conducted scenario analyses on the company's base case, see Table 64 and Table 65 in Appendix 11. We note that none of the scenarios, except for using a 1-year stopping rule for the treatments, has any significant impact on the overall cost-effectiveness results.

5 End of life

End of life considerations are not applicable to this technology appraisal.

6 Innovation

The company list a number of points in support of the innovative nature of ustekinumab in CS section B.2.12. Most of the points listed by the company refer to aspects of the UNIFI trial or the population treated rather than features of ustekinumab or its use that would make it innovative. In the opinion of the ERG, the key point in support of innovation made by the company is that ustekinumab provides a new mechanism of action for the treatment of UC.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The NICE scope specifies prior therapy subgroups based on exposure whereas the company define prior therapy subgroups according to treatment failure. There is reasonable concordance between the subgroups as defined in the UNIFI trial and those in the NICE scope, but the agreement between subgroup definitions across the comparator trials included in the company's NMAs is less clear. Overall, the UNIFI trial was well conducted and is reflective of clinical practice, with two provisos:

- One of the induction doses (130mg) is not relevant to the intended marketing authorisation
- In the maintenance phase patients were randomised to the standard and escalated dose regimens which is not fully reflective of clinical practice.

The statistical power of the UNIFI trial subgroups is not reported, but we believe the subgroups are adequately powered for induction clinical remission but may be under-powered for maintenance clinical remission.

There are a number of sources of heterogeneity in the company's NMAs which in some cases could introduce bias. These are summarised and discussed briefly in section 3.4.3 above, with links to the specific sections where they are discussed in detail (see Table 36). In summary, the key issues with the NMAs are:

- There is heterogeneity in the company's NMAs due to differences between trials, e.g. in central versus local reading of endoscopies; differences in the durations of the induction/maintenance phases; and differences in how non-biologic failure and biologic failure are defined.
- The company excluded Asian trials from their NMAs which is inconsistent with the approach in TA547. A sensitivity analysis including Asian trials was conducted, but due to methodological problems we believe this is invalid.
- The ERG was not able to validate all of the data sources employed by the company in their NMAs.

7.2 Summary of cost effectiveness issues

7.2.1 Model structure

- The company model structure is accurately implemented and generally consistent with previous TA models in UC, there is one major exception: omission of response and remission health states after failure of the initial treatment. We consider this a major limitation, as it implies that all patients follow a chronic active or progressive form of disease, which is inconsistent with previous NICE appraisals and unrealistic. We address this issue in the ERG additional analyses.

- The company model includes two phases of surgery, each lasting for six months to allow for staged procedures. This approach is different from previous appraisals (TA547 and TA342), which treated surgery as a one-off event. However, we consider that the current model better reflects the usual process of staged procedures: subtotal colectomy with ileostomy followed by either IPAA (pouch) surgery or permanent ileostomy (phase 1); and potential revision surgery due to pouch failure (phase 2). The model assumes that all patients who have revision surgery reach remission with no chronic complications. We accept this assumption as a reasonable simplification. As the number of people affected will be small, we expect the impact on overall costs and QALYs to be minimal.

7.2.2 Response and remission rates

Induction phase

The company's base case response and remission rates for standard induction are based on their fixed effects induction NMA. We prefer the random effects model, which gives similar results but with more uncertainty.

Maintenance phase

We have strong concerns over the use of absolute response and remission rates from individual treatment arms in the company's base case analysis. This introduces a high potential for bias by ignoring the original trial randomisation, meaning that any differences between the trial populations or conduct are not adjusted for. The ERG, therefore, prefers the company's maintenance NMA scenario over their base case; and because of potential heterogeneity, we prefer the random effects version of the NMA scenario.

We also question the validity of attributing all of the differences between maintenance placebo arms to 'carry over' effects from induction. It is more likely that other differences between the trials also contribute to these differences. Furthermore, we could not verify all of the sources of data and imputations in the company NMA scenario. We therefore conducted an alternative 'maintenance only' NMA following the methods applied in the TA547 appraisal, which we use in a scenario analysis on the ERG base case economic analysis.

We agree with the company's assumption of a constant risk over time. This approach is consistent with the assumption in NICE TA547.

7.2.3 Dose regimens

The model accurately reflects the recommended induction and maintenance regimens, including extended induction for delayed response and escalation to higher dose or more frequent treatment when indicated. We agree with the company's base case assumption of equal loss of response rates for initial and later responders. We also view the company's base case assumption that 30% of patients on maintenance have the escalated regimen as reasonable. However, we note some limitations of the company's approach:

- The company does not include the higher (10mg/kg) dose of infliximab as it is not recommended in the SmPC. However, clinical advice to the ERG is that dose adjustment for infliximab is common in practice. This suggests that the same dose

escalation assumptions should be made for infliximab as for other comparators. We test this assumption in the ERG additional analyses.

- The company pools standard and escalated doses in the non-biologic failure subgroup but not in the biologic failure subgroup. They argue that there is an exposure-response relationship for patients with a history of biologic failure, but not for other patients. We consider that the evidence supporting this stance is weak, as it relies on an indirect relationship (exposure-response with/without remission at maintenance baseline) and only for ustekinumab. We therefore think that the same dose pooling approach should be used in both subgroups. We prefer pooled effect estimates, because of high uncertainty over the exposure-response relationships, so use this approach in our base case analysis.

7.2.4 Resource use and costs

The company do not include the cost of concurrent treatment with conventional drugs alongside biologics and JAK inhibitors in their analyses. We add this cost in ERG analysis, with usage assumptions for conventional drugs as in TA547.

Further, they use a different treatment mix for CT compared to previous TA547. Whilst we acknowledge that changes to assumptions about the use and costs of CT are unlikely to be influential in the model because of their low cost and similar impact on cost-effectiveness of comparators, nevertheless, for face validity we update the assumptions about use of conventional therapy drugs as a comparator and concurrent with other treatments as per TA547. Estimates of health state, surgery and adverse event costs are reasonably consistent with previous UC appraisals.

7.2.5 Utilities

We consider that the utilities in the company's base case are generally reasonable, with some exceptions.

- The QALY decrement for serious infections appears to have been overestimated because the disutility of 0.156 is not adjusted for the expected duration of symptoms (assumed to be 28 days in TA329). We add this to the ERG base case.

- Clinical advice to the ERG is that the CS may overestimate utility after revision surgery, which on average is expected to be worse than remission after the first phase of surgery. We examine this in our additional analyses.

- Whilst we agree with the company's decision not to use utility estimates from the UNIFI EQ-5D data due to inconsistency with the values used in previous NICE appraisals for UC, we note that the number of observations in the three severity health states is large and the analysis appears to have been well-conducted. The ERG therefore considers the scenario analysis with UNIFI utility estimates to be important and we repeat this scenario on our base case.

7.2.6 Other issues

Other uncertainties of the company's cost effectiveness are summarised below. We consider these to have lower impact on the overall cost-effectiveness analyses.

Population

The population in the company's economic model reflects the NICE scope, the anticipated marketing authorisation and UNIFI trial population. The company appropriately presents the results for the subgroups only and not for the whole ITT population. We note that the subgroups are defined by biologic failure, rather than biologic exposure as requested in the scope. However, we do not anticipate this to affect the results. Baseline demographics of the modelled subgroups are broadly reflective of the ustekinumab and comparator trial populations and similar to patients starting biologic treatment for UC in the UK.

Intervention and comparators

The modelled intervention and comparators are consistent with the NICE scope and reflective of current clinical practice, except for the exclusion of infliximab and golimumab in the biologic failure subgroup. We agree with the company's omission of these two drugs in the biologic failure subgroup because the infliximab and golimumab trials excluded people with previous biologic treatment. We also agree with the company for including biosimilars for infliximab and adalimumab, assuming equal effects and safety profile but lower costs compared with the original products, as they provide helpful comparisons.

Stopping rule

We agree with the company's base case approach to assume continued treatment until loss of response due to uncertainty over routine use of a 'stopping rule' for biologics in UC.

Response and remission: delayed responders

We think that there is high uncertainty over the direct trial estimates of response and remission for extended induction and loss of response rates for delayed responders.

Incidence of surgery and surgery related complications

We agree with the company's use of UK estimates for the incidence of first surgery and rates of early and late complications. The first two of these sources were also used in TA547. A different source was used for late complications in TA547, but the model is not sensitive to this difference. Although we view the company's assumption that the incidence of revision surgery for patients with chronic complications is the same as that for initial surgery, as arbitrary; this does not affect the overall cost-effectiveness results because the model is not sensitive to this assumption.

Adverse events: serious infection rates

We view the rates of serious infections used in the model as reasonable. Despite uncertainties over use of the PSOLAR data and assumptions, this is still the best available source of evidence and the model is not sensitive to plausible changes in serious infection rates.

Mortality rates

We view the company's assumptions about mortality are reasonable, with an excess risk for surgery, but otherwise the same risks as for the general population. We note that model is not sensitive to the relative risk assumed during surgery.

8 REFERENCES

1. European Medicines Agency. Draft Summary of Product Characteristics (SmPC) - ustekinumab (provided in CS Appendix C), 2018.
2. 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.
3. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet* 2017;389:1756-70.
4. Magro F, Rodrigues A, Vieira AI, et al. Review of the disease course among adult ulcerative colitis population-based longitudinal cohorts. *Inflammatory bowel diseases* 2012;18(3):573-83.
5. Jess T, Gamborg M, Munkholm P, et al. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. *The American journal of gastroenterology* 2007;102(3):609-17.
6. Monstad I, Hovde O, Solberg IC, et al. Clinical course and prognosis in ulcerative colitis: results from population-based and observational studies. *Ann Gastroenterol* 2014;27(2):95-104.
7. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet* 2017;389(10080):1756-70.
8. Torres J, Billioud V, Sachar DB, et al. Ulcerative colitis as a progressive disease: the forgotten evidence. *Inflamm Bowel Dis* 2012;18(7):1356-63.
9. National Institute for Health and Care Excellence. *Tofacitinib for moderately to severely active ulcerative colitis TA547*. https://www.nice.org.uk/guidance/ta547.
10. Janssen Inc. STELARA (ustekinumab) Induction Clinical Study Report- Protocol CNTO1275UCO3001 (Data on File), 2018.
11. Janssen Inc. STELARA (ustekinumab) Maintenance CSR-Protocol CNTO127UCO3001 (Data on File). 2018.
12. Cholapranee A, Hazlewood GS, Kaplan GG, et al. Systematic review with meta-analysis: comparative efficacy of biologics for induction and maintenance of mucosal healing in Crohn's disease and ulcerative colitis controlled trials. *Alimentary pharmacology & therapeutics* 2017;45(10):1291-302.
13. Dignass AU, Siegmund B, Goertz R, et al. Indirect comparison of vedolizumab and adalimumab for biologic-naive patients with ulcerative colitis. *Scandinavian journal of gastroenterology* 2019;54(2):178-87.

14. Jairath V, Lasch K, Chan K, et al. P325 Integrating efficacy and safety of vedolizumab and other advanced therapies for the treatment of ulcerative colitis: Results from a network meta-analysis. *Journal of Crohn's and colitis* 2019;13(Supplement_1):S263-S64.
15. Bonovas S, Lytras T, Nikolopoulos G, et al. Systematic review with network neta-analysis: comparative assessment of tofacitinib and biological therapies for moderate-to-severe ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 2018;47:454-65.
16. Singh S, Fumery M, Sandborn WJ, et al. Systematic review with network meta-analysis: first- and second-line pharmacotherapy for moderate-severe ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 2018;47:162-75.
17. National Institute for Health and Care Excellence. *Vedolizumab for treating moderately to severely active ulcerative colitis [TA342]*. https://www.nice.org.uk/guidance/ta342.
18. National Institute for Health and Care Excellence. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (TA329). 2015.
19. Sands BE, Han C, Zhang H, et al. Ustekinumab therapy induced clinically meaningful improvement and remission as measured by the Inflammatory Bowel Disease Questionnaire: Results from the phase 3 UNIFI induction and maintenance studies. *Journal of Crohn's and colitis* 2019;13 supplement 1:S460.
20. Van Assche G, Targan SR, Baker T, et al. OP47 Sustained remission in patients with moderate to severe ulcerative colitis: Results from the Phase 3 UNIFI maintenance study. Proceedings, ECCO 2019 Conference. *ECCO 2019*: European Crohn's and Colitis Organisation, 2019.
21. Sands BE, Sandborn WJ, Panaccione R, et al. 833–Efficacy and Safety of Ustekinumab As Maintenance Therapy in Ulcerative Colitis: Week 44 Results from Unifi. *Gastroenterology* 2019;156(6):S-181.
22. Danese S, Sands B, O'Brien C, et al. DOP54 Efficacy and safety of ustekinumab through Week 16 in patients with moderate-to-severe ulcerative colitis randomised to ustekinumab: results from the UNIFI induction trial. *Journal of Crohn's and colitis* 2019;13(Supplement_1):S061-S62.
23. Sands B, Peyrin-Biroulet L, Marano C, et al. P312 Efficacy in biologic failure and non-biologic-failure populations in a Phase 3 study of ustekinumab in moderate–severe ulcerative colitis: UNIFI. *Journal of Crohn's and colitis* 2019;13(Supplement_1):S256-S57.

24. Li K, Friedman J, Marano C, et al. DOP71 Effects of ustekinumab induction therapy on endoscopic and histological healing in the UNIFI Phase 3 study in ulcerative colitis. *Journal of Crohn's and colitis* 2019;13(Supplement_1):S073-S73.
25. Sands BE, Peyrin-Biroulet L, Loftus EV, et al. 416a–Vedolizumab Shows Superior Efficacy Versus Adalimumab: Results of Varsity—The First Head-To-Head Study of Biologic Therapy for Moderate-To-Severe Ulcerative Colitis. *Gastroenterology* 2019;156(6):S-81.
26. 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.
27. 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.
28. Silva R BG, et al. Infliximab versus adalimumab: clinical and endoscopy response in ulcerative colitis patients. A prospective study [Abstract]. *Conference: 12th congress of the european crohn's and colitis organisation, ECCO 2017. Spain* 2017.
29. Kobayashi T. Suzuki Y. Motoya S. Hirai F. Ogata H. Ito H. Sato N. Ozaki K. Watanabe M. Hibi T. A Phase 1, Multiple-Dose Study of Vedolizumab in Japanese Patients With Ulcerative Colitis. *Journal of Clinical Pharmacology* 2019;59(2):271-79.
30. Ghosh S, Mitchell RJJoCs, Colitis. Impact of inflammatory bowel disease on quality of life: Results of the European Federation of Crohn's and Ulcerative Colitis Associations (EFCCA) patient survey. 2007;1(1):10-20.
31. Papp K, Gottlieb AB, Naldi L, et al. Safety Surveillance for Ustekinumab and Other Psoriasis Treatments From the Psoriasis Longitudinal Assessment and Registry (PSOLAR). *J Drugs Dermatol* 2015;14(7):706-14.
32. Kalb RE, Fiorentino DF, Lebwohl MG, et al. Risk of Serious Infection With Biologic and Systemic Treatment of Psoriasis: Results From the Psoriasis Longitudinal Assessment and Registry (PSOLAR). *JAMA Dermatol* 2015;151(9):961-9.
33. National Institute for Health and Care Excellence. *Ustekinumab for moderately to severely active Crohn's disease after previous treatment*. https://www.nice.org.uk/guidance/ta456/resources/ustekinumab-for-moderately-to-severely-active-crohns-disease-after-previous-treatment-pdf-82604848449733.
34. Geboes K, Riddell R, Öst A, et al. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* 2000;47:404-09.

35. European Medicines Agency. *Guideline on the development of new medicinal products for the treatment of ulcerative colitis*. https://www.ema.europa.eu/documents/scientific-guideline/guideline-development-new-medicinal-products-treatment-ulcerative-colitis-revision-1_en.pdf.
36. Roda G, Jharap B, Neeraj N, et al. Loss of response to anti-TNFs: definition, epidemiology, and management. *Clinical and Translational Gastroenterology* 2016;7(e135):1-5.
37. Higgins PDR, Schwartz M, Mapili J, et al. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. *Gut* 2005;54(6):782-88.
38. Yarlas A, Bayliss M, Cappelleri JC, et al. Psychometric validation of the SF-36 Health Survey in ulcerative colitis: results from a systematic literature review. *Quality of LIfe Research* 2018;27:273-90.
39. 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.
40. Panés J, Feagan BG, Hussain F, et al. Central endoscopy reading in inflammatory bowel diseases. *Journal of Crohn's and colitis* 2016;10(Suppl 2):S542-S47.
41. 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.
42. Reinisch W, Sandborn WJ, Hommes DW, et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. *Gut* 2011;60(6):780-7.
43. Reinisch W, Sandborn WJ, Panaccione R, et al. 52-week efficacy of adalimumab in patients with moderately to severely active ulcerative colitis who failed corticosteroids and/or immunosuppressants. *Inflamm Bowel Dis* 2013;19(8):1700-9.
44. 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. e3.
45. Schreiber S, Peyrin-Biroulet L, Loftus EV, et al. OP34 VARSITY: A double-blind, double-dummy, randomised, controlled trial of vedolizumab versus adalimumab in patients with active ulcerative colitis. Proceedings, ECCO 2019. *ECCO 2019*: European Crohn's and Colitis Organisation, 2019.

46. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014;146(1):85-95; quiz e14-5.
47. 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.
48. 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.
49. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005;353(23):2462-76.
50. 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.
51. Probert CS, Hearing SD, Schreiber S, et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. *Gut* 2003;52(7):998-1002.
52. Sandborn WJ, Ghosh S, Panes J, et al. Tofacitinib, an oral Janus kinase inhibitor, in active ulcerative colitis. *N Engl J Med* 2012;367(7):616-24.
53. Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* 2017;376(18):1723-36.
54. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2013;369(8):699-710.
55. Motoya S, Watanabe K, Ogata H, et al. Vedolizumab in Japanese patients with ulcerative colitis: A Phase 3, randomized, double-blind, placebo-controlled study. *PLoS ONE* 2019;14(2):e0212989.
56. Macaluso FS, Maida M, Ventimiglia M, et al. Factors affecting clinical and endoscopic outcomes of placebo arm in trials of biologics and small molecule drugs in ulcerative colitis: a meta-analysis. *Inflammatory bowel diseases* 2019;25(6):987-97.
57. Jairath V, Zou GY, Parker CE, et al. Placebo response and remission rates in randomised trials of induction and maintenance therapy for ulcerative colitis. *Cochrane Database of Syetematic Reviews* 2017;Issue 9. Art No. : CD011572:1-140.
58. Thorlund K, Druyts E, Tor K, et al. Incorporating alternative design clinical trials in network meta-analyses. *Clinical Epidemiology* 2015;7:9-35.

59. Dias S, Sutton AJ, Ades A, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* 2013;33(5):607-17.
60. Yiu ZZN, Smith CH, Ashcroft DM, et al. Risk of serious infection in patients with psoriasis receiving biologic therapies: a prospective cohort study from the British Association of Dermatologists Biologic Interventions Register (BADBIR). *Journal of Investigative Dermatology* 2018;138:534-41.
61. Ghosh S, Gensler LS, Yang Z, et al. Ustekinumab Safety in Psoriasis, Psoriatic Arthritis, and Crohn's Disease: An Integrated Analysis of Phase II/III Clinical Development Programs. *Drug Saf* 2019;42(6):751-68.
62. 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).
63. 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.
64. Tsai HH, Punekar YS, 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 and Therapeutics* 2008;28(10):1230-39.
65. 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. *ClinicoEconomics and Outcomes Research* 2017;9:641-52.
66. Wu B, Wang Z, Zhang Q. Cost-Effectiveness of Different Strategies for the Treatment of Moderate-to-Severe Ulcerative Colitis. *Inflammatory bowel diseases* 2018;24(11):2291-302.
67. Royal College of Physicians. National clinical audit of biological therapies: annual report. UK inflammatory bowel disease (IBD) audit. London, 2016.
68. NHS England and NHS Improvement. What is a biosimilar medicine?, 2019.
69. Fausel R, Afzali A. Biologics in the management of ulcerative colitis - comparative safety and efficacy of TNF-alpha antagonists. *Ther Clin Risk Manag* 2015;11:63-73.

70. Patel H, Lissoos T, Rubin DT. Indicators of suboptimal biologic therapy over time in patients with ulcerative colitis and Crohn's disease in the United States. *PLoS ONE* 2017;12(4):e0175099.
71. Rostholder E, Ahmed A, Cheifetz AS, et al. Outcomes after escalation of infliximab therapy in ambulatory patients with moderately active ulcerative colitis. *Aliment Pharmacol Ther* 2012;35(5):562-7.
72. Lindsay JO, Armuzzi A, Gisbert JP, et al. Indicators of suboptimal tumor necrosis factor antagonist therapy in inflammatory bowel disease. *Dig Liver Dis* 2017;49(10):1086-91.
73. MIMS. Monthly Index of Medical Specialities (MIMS) database. 2018.
74. 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). *Scand J Gastroenterol* 2009;44(4):431-40.
75. Bewtra M, Kaiser LM, TenHave T, et al. Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. *Inflammatory bowel diseases* 2013;19(3):599.
76. Ferrante M, Vermeire S, Fidder H, et al. Long-term outcome after infliximab for refractory ulcerative colitis. *J Crohns Colitis* 2008;2(3):219-25.
77. 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. 2016;2016.
78. Royal College of Physicians. National clinical audit of inpatient care for adults with ulcerative colitis UK inflammatory bowel disease (IBD) audit, 2014.
79. Segal JP, McLaughlin SD, Faiz OD, et al. Incidence and Long-term Implications of Prepouch Ileitis: An Observational Study. *Dis Colon Rectum* 2018;61(4):472-75.
80. 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. 2015;41(1):87-98.
81. Ferrante M, Declerck S, De Hertogh G, et al. Outcome after proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis. *Inflamm Bowel Dis* 2008;14(1):20-8.
82. Loftus EV, Jr., Delgado DJ, Friedman HS, et al. Colectomy and the incidence of postsurgical complications among ulcerative colitis patients with private health insurance in the United States. *Am J Gastroenterol* 2008;103(7):1737-45.
83. Jess T, Rungoe C, Peyrin–Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clinical Gastroenterology and Hepatology* 2012;10(6):639-45.

84. Woehl A, A.B. Hawthorne, and P. McEwan. The Relation Between Disease Activity, Quality of Life and Health Utility in Patients With Ulcerative Colitis. *Gut* 2008;57:A1-A172.
85. Arseneau KO, Sultan S, Provenzale DT, et al. Do patient preferences influence decisions on treatment for patients with steroid-refractory ulcerative colitis? *Clinical Gastroenterology and Hepatology* 2006;4(9):1135-42.
86. Stevenson M, Archer R, Tosh J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health Technol Assess* 2016;20(35):1-610.
87. Swinburn P, Elwick H, Bean K, et al. PTU-127 The impact of surgery on health related quality of life in ulcerative colitis. 2012;61(Suppl 2):A237-A37.
88. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;13(5):509-18.
89. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health.* 2012;15(5):708-15. doi: 10.1016/j.jval.2012.02.008. Epub 12 May 24.
90. 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. *J Crohns Colitis* 2011;5(4):301-16.
91. Sandborn W, al. e. Infliximab Reduces Ulcerative Colitis-Related Hospitalizations Requiring High-Dose Corticosteroids *American College of Gastroenterology, Las Vegas, Nevada, USA*; 2016, October.
92. Xu J, Lin H, Feng X, et al. Different therapeutic approaches on quality of life in patients with inflammatory bowel disease. *BMC Gastroenterol* 2014;14:199.
93. Kalita N, et al. Tofacitinib for previously treated active ulcerative colitis [ID1218]. *Health Technology Assessment TA547* 2018.
94. Sandborn WJ, Colombel JF, D'Haens G, et al. One-year maintenance outcomes among patients with moderately-to-severely active ulcerative colitis who responded to induction therapy with adalimumab: subgroup analyses from ULTRA 2. *Aliment Pharmacol Ther.* 2013;37(2):204-13. doi: 10.1111/apt.12145. Epub 2012 Nov 23.
95. Panaccione R, Colombel JF, Reinisch W, et al. Durable clinical remission and response in adalimumab-treated patients with ulcerative colitis. *Journal of Crohn's and colitis* 2015;9 Suppl 1:S256-S57.

96. Janssen. A Phase 3, Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Study to Evaluate the Safety and Efficacy of Ustekinumab Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis. *Clinical Study Report CNTO1275UCO3001 Induction* 2018.

97. Philip G, Marano C, Adedokun J, et al. Early Dose Optimization in Nonresponders to Golimumab Induction Treatment for Ulcerative Colitis is Supported by Pharmacokinetic Data through 1 Year. Proceedings, United European Gastroenterology Meeting 2018. *United European Gastroenterology*. Vienna, Austria, 2018.

98. Sandborn WJ, Lichtenstein GR, Colombel JF, et al. Infliximab therapy reduces hospitalizations in ulcerative colitis patients. *American Journal of Gastroenterology* 2005;100 Supplement (70th Annual Scientific Meeting, American College of Gastroenterology) S312.

99. Suzuki Y, Motoya S, Hirai F, et al. Infliximab therapy for Japanese patients with ulcerative colitis: efficacy, safety, and association between serum infliximab levels and early response in a randomized, double-blind, placebo-controlled study. *Journal of Crohn's and colitis* 2015;9:S372-S73.

100. Feagan BG, Vermeire S, Reinisch W, et al. Tofacitinib for induction therapy in patients with active ulcerative colitis in two phase 3 clinical trials: Results by local and central endoscopic assessments. *United European Gastroenterology Journal* 2016;4(5S):A262-A63.

101. Sandborn WJ, Sands BE, D'Haens G, et al. Efficacy and safety of oral tofacitinib as induction therapy in patients with moderate-to-severe ulcerative colitis: Results from 2 phase 3 randomised controlled trials. *Journal of Crohn's and colitis* 2016;10:S15.

102. Dubinsky M, Peyrin-Biroulet L, Melmed G, et al. Efficacy of tofacitinib in patients with ulcerative colitis by prior tumor necrosis factor inhibitor treatment status: results from OCTAVE Induction and Maintenance studies. Proceedings, American College of Gastroenterology Annual Scientific Meeting
2017.

103. Panes J, Su C, Marren A, et al. Improvement in patient-reported outcomes in 2 Phase 3 studies of tofacitinib in patients with moderately to severely active ulcerative colitis. *Journal of Crohn's and colitis* 2016;10:S283-S84.

104. 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.

105. Hanauer S, Rubin DT, Gionchetti P, et al. P712 Tofacitinib efficacy in patients with moderate to severe ulcerative colitis: Subgroup analyses of OCTAVE Induction 1 and 2 and OCTAVE Sustain by 5-aminosalicylates use. Proceedings, ECCO 2019 *European Crohn's and Colitis Organisation (ECCO) 2019*, 2019.
106. Feagan BG, Rubin DT, Danese S, et al. Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. *Clinical Gastroenterology and Hepatology* 2017;15(2):229-39 e5.
107. Panés J, Rubin DT, Vermeire S, et al. Maintenance of quality of life improvement in a phase 3 study of tofacitinib for patients with moderately to severely active ulcerative colitis. *Gastroenterology* 2017;152(5 Supplement 1):S601-S02.
108. Panés J, Su C, Bushmakin AG, et al. Randomized trial of tofacitinib in active ulcerative colitis: analysis of efficacy based on patient-reported outcomes. *BMC Gastroenterol* 2015;15:14:1-10.
109. Panés J, Su C, Bushmakin AG, et al. Direct and indirect effects of tofacitinib on treatment satisfaction in patients with ulcerative colitis. *Journal of Crohn's and colitis* 2016;10(11):1310-15.
110. Feagan BG, al. e. Induction and maintenance therapy with vedolizumab, a novel biologic therapy for ulcerative colitis. *Gastroenterol Hepatol (N Y)* 2014;10(1):64-6.
111. 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 and Therapeutics* 2017;45(2):264-75.
112. Loftus EV, Sands BE, Colombel JF, et al. Sustained corticosteroid-free remission with vedolizumab in moderate-to-severe ulcerative colitis: a post hoc analysis of GEMINI 1. Proceedings, ECCO 2018. *Journal of Crohn's and Colitis. 13th Congress of European Crohn's and Colitis Organisation, ECCO 2018.* . Austria, 2018:S137-38.
113. Rutgeerts P. SW, Feagan b., et al., . Infliximab for Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* 2005;353:2462-76.
114. Rutgeerts P, Feagan BG, Marano CW, et al. Randomised clinical trial: a placebo-controlled study of intravenous golimumab induction therapy for ulcerative colitis. *Aliment Pharmacol Ther.* 2015;42(5):504-14. doi: 10.1111/apt.13291. Epub 2015 Jun 29.
115. 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.

116. Feagan BG, Rubin DT, Danese S, et al. Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. *Clin Gastroenterol Hepatol* 2017;15(2):229-39.e5.

9 APPENDICES

Appendix 1 Trials excluded from meta-analyses

Trial	Relevant outcomes	Reason for exclusion	ERG comments
Silva 2017[28] (adalimumab versus infliximab)	Clinical response and clinical remission	Abstract with unclear dose, unclear timing of outcome assessment, and very small sample size (N=10 in infliximab arm) (CS section D1.1.6.1).	Exclusion reasons are appropriate
Kobayashi 2019[29] (2 doses of vedolizumab compared)	Clinical response and clinical remission	No placebo arm (and very small sample size) (CS section D1.1.6.1).	Exclusion reasons are appropriate
UC-SUCCESS Panaccione et al. 2017[27] (azathioprine versus infliximab)	Mucosal healing and serious infections	Trial is not discussed in the CS. Stated "not intervention of interest" in CS Appendix Table 31, without a more detailed reason given.	Exclusion reason unclear but induction NMAs for mucosal healing and serious infections do not inform the company's economic analysis and so the omission of this trial would be inconsequential
Mshimesh 2017[26] (adalimumab versus infliximab)	Clinical response and clinical remission	Trial is not discussed in the CS. Identified in HRQoL searches but not in clinical effectiveness searches. Stated "not study type of interest" in CS Appendix Table 108, without a more detailed reason given.	Exclusion reason unclear, but it appears appropriate to exclude this trial because (i) population specifically Iraqi patients, unlikely to reflect UK setting; (ii) small sample size (N=25 per arm); (iii) the adalimumab-infliximab path in the NMA network would have limited influence on overall results; and (iv) limited to induction only

Appendix 2 Trials included in the company's clinical effectiveness review and NMAs

Therapy	Trial	Induction outcomes	Maintenance outcomes	Outcomes not used in NMAs
ADA vs placebo	NCT00853099	Suzuki (2014) [41]	Suzuki (2014) [41]	
	ULTRA 1	Reinisch (2011) [42]	Reinisch (2013) [43]	
	ULTRA 2	**Sandborn (2012) [44]** Sandborn (2013) [94] Panaccione (2015) [95]	**Sandborn (2012) [44]** Panaccione (2015) [95] Sandborn (2013) [94]	
ADA vs VED	VARSITY		Sands (2019)* [25] Schreiber (2019) [45] (abstracts only)	
GOL vs placebo	PURSUIT-J		Hibi (2017) [47]	
	PURSUIT-M		**Sandborn (2014) [48]** Colombel (2016) [96] (post-hoc)	
	PURSUIT-SC	Sandborn (2014) [46]	Colombel (2016) [96] (post-hoc)	
	PURSUIT	Philip (2018) [97]	Philip (2018) [97]	
INF vs placebo	ACT1 ACT2	Rutgeerts (2005) [49]	Rutgeerts (2005) [49]	Sandborn (2016) [98] hospitalisations
	Japic CTI-060298	**Kobayashi (2016) [39]** Suzuki (2015) [99]	**Kobayashi (2016) [39]** Suzuki (2015) [99]	
	Jiang 2015	Jiang (2015) [50]	Jiang (2015) [50]	
	Probert 2003	Probert (2003) [51]		
TOF vs placebo	OCTAVE 1 OCTAVE 2	**Sandborn (2017) [53]** Feagan (2016) [100] Sandborn (2016) [101] Dubinsky (2017) [102]		Panes (2016) [103] PROs Panes (2018) [104] HRQoL Hanauer (2019) [105] 5-aminosalicylate subgroups
	OCTAVE Sustain		**Sandborn (2017) [53]** Feagan (2017)[106] Sandborn (2017)[53] Dubinsky (2017) [102]	Panes (2017) [107] HRQoL Panes (2018) [104] HRQoL Hanauer (2019) [105] 5-aminosalicylate subgroups
	NCT00787202	Sandborn (2012) [52]		Panes (2015) [108] Panes (2016) [109] Both IBDQ
UST vs placebo	UNIFI	**CS** **CSR [10]** Danese (2019)* [22] Sands (2019)* [23] Sands (2019)* [21]	**CS** **CSR [11]** Sands (2019)* [23] Sands (2019)* [21] Van Aasche (2019)* [20]	Li (2019)* [24] endoscopic & histological healing Sands (2019)* [19] IBDQ
VED vs placebo	GEMINI 1	**Feagan (2013) [54]** Feagan (2014) [110]	**Feagan (2013) [54]** Feagan (2014) [110]	Feagan (2017) [106] post hoc efficacy subgroups

				Feagan (2017) [111] HRQoL Loftus (2018) [112] post hoc corticosteroid-free remission
	NCT02039505	Motoya (2019) [55]	Motoya (2019) [55]	
*reference identified by ERG, not provided in CS				

Appendix 3 Risk of bias assessments for trials included in NMAs

The company conducted a risk of bias assessment for each of the trials included in the NMAs, based on standard NICE criteria (CS Appendix Tables 24 and 85). We note that most of the trials have also been subject to independent assessments of their risks of bias by ERGs in previous technology appraisals. We therefore compared the company's assessments against the following independent ERG assessments to gauge whether the company's assessments are generally appropriate:

- NICE TA329:[18] ACT1, ACT2, NCT00853099, Probert 2003, Pursuit-M, PURSUIT-SC, ULTRA1, ULTRA2
- NICE TA547:[9] NCT00787202, OCTAVE1, OCTAVE2, OCTAVE Sustain
- NICE TA342:[17] GEMINI1
- Current report, section 3.1.4: UNIFI

The concordance between the company's risk of bias assessments (CS Appendix Table 24) and those of previous NICE TA ERG reports,[9,17,18] is summarised in Table 61.

Table 61 Summary of company risk of bias assessments for trials included in NMAs compared to previous technology appraisals

Risk of bias question (CS Appendix Table 24)	Interpretation	Comments
Was randomisation carried out appropriately?	Yes answer suggests low risk of bias	Company and independent NICE TA ERG reports agree that this risk of bias is low for all these trials
Was the concealment of allocation adequate?	Yes answer suggests low risk of bias	Some minor discrepancies; the previous NICE TA ERG reports suggest that this risk of bias is low for all trials except NCT00853099 (unclear risk)
Were groups similar at the outset in terms of prognostic factors	Yes answer suggests low risk of bias	Not consistently assessed in the previous NICE TA ERG reports. CS Appendix Figures 4, 6, 8 and 12 suggest patients' age, gender, weight and CRP levels were balanced across arms within trials. Within-trial differences in Mayo score were generally within 0.4 points (CS Appendix Figure 14). The largest within-trial differences in disease duration were 2-3 years, in 2 trials (NCT00787202: tofacitinib 10mg 10.9 years, tofacitinib 5mg 8.0 years; ACT1: infliximab 10mg 8.4 years, placebo 6.2 years). We believe the company's yes answer is appropriate, with the proviso that there was some within-trial variation in disease duration.

Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes answer suggests low risk of bias	Some minor discrepancies; the previous NICE TA ERG reports suggest risk of bias is low for all trials except unclear for ACT1 & ACT2 and for unclear outcome assessors in NCT00853099. The CS is not clear about whether “double blind” covers care providers, participants and/or outcome assessors.
Were there any unexpected drop-outs between groups?	Yes answer suggests high risk of bias, unless appropriate ITT analysis is conducted	The company has answered “no” for all trials except ULTRA1. The previous NICE TA ERG reports identified that, especially in the maintenance phase, all trials except ULTRA1 had large and unbalanced differences in the proportion of drop-outs between placebo and active arms. The company has not explained their “no” responses so it is unclear whether they have interpreted that there were no within-trial imbalances or that there were imbalances but these were not unexpected. The latter interpretation would appear appropriate, as most dropouts were usually due to lack of efficacy, consistent with expectation.
Did the analysis include an ITT analysis?	Yes answer suggests low risk of bias, provided that missing data are accounted for appropriately	The company and previous NICE TA ERG reports agree that ITT analysis was conducted in most trials (the company say “no” for the OCTAVE trials which disagrees with the TA547 ERG assessment). The company and independent NICE TA ERG reports agree that ITT analysis was not reported for Probert 2003 or PURSUIT-M. Variation in the judgements about ITT appear to reflect that some assessments (e.g. the company’s interpretation in CS Appendix D.1.9) are based on both induction + maintenance periods in re-randomised trials although strictly a separate ITT assessment should be made for each outcome (i.e. in re-randomised trials for the induction outcome and the maintenance outcome, as these are based on different randomised populations).

Overall, the company’s risk of bias assessments appear to be broadly comparable with those of ERGs in the previous NICE appraisals, and those we have made in the current ERG report for the UNIFI trial (section 3.1.4), with some exceptions noted above. The main issue identified by the risk of bias assessments conducted by previous ERGs and ourselves, but not discussed by the company, is that several trials had a relatively high drop-out rate in

the maintenance phase which was consistently higher in the placebo than active comparator arms (not the case in UNIFI; section 3.1.4).

The company do not discuss whether any specific trials should have been included in or excluded from meta-analysis (e.g. in sensitivity analyses) based on their risk of bias assessments. The CS includes all trials in the analyses (subject to meeting the eligibility criteria). It is unclear whether this is appropriate because the issue of unbalanced dropouts is not discussed in the CS. The risk of attrition bias could be mitigated in the NMAs by ensuring that only ITT data are included in NMAs with missing data imputed using conservative methods. Whilst the company do utilise ITT data from the trials, the imputations and assumptions used to generate the ITT population in each trial are not discussed.

The assessments summarised above cover 14 of the 19 trials included in the company's NMAs, as listed above. The remaining trials were on Asian populations (Jiang 2015, Japic CTI-060298, NCT02039505) and the VARSITY trial. We did not investigate risks of bias in the Asian trials since these are excluded from the base case NMA analyses. It is not possible to assess the risks of bias in the VARSITY trial as insufficient information is reported in the available abstracts.

Appendix 4 ERG corrections made to discrepancies in company induction NMA data inputs

Trial Subgroup	**Arm**	**Outcome**	**CS Appendix Table 60**	**Company NMA code**	**Trial publication**	**Data used in ERG analyses**
UNIFI biologic failure	PBO	Response, n/N	33/160	33/160	44/161 [23]	44/161
UNIFI non-biologic failure	PBO	Response, n/N	57/159	57/159	56/158 [23]	56/158
UNIFI non-biologic failure	PBO	Remission, n/N	15/159	15/158	15/158 [23]	15/158
OCTAVE1 non-biologic failure	TOF	Remission, n/N	56/222	61/233	56/222 [53]	56/222
OCTAVE2 non-biologic failure	PBO	Remission, n/N	4/47	4/52	4/47 [53]	4/47
OCTAVE2 non-biologic failure	TOF	Remission, n/N	45/195	45/207	43/195 [53]	43/195
PBO: placebo; TOF: tofacitinib						

Appendix 5 Data calculations and sources for non-biologic failure 1-year NMAs conditional on response (red data ERG unable to validate)

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
Clinical remission	UNIFI	UST 6mg - UST pooled	66.7%	UNIFI CSR	53.9%	UNIFI IPD	(A x B)= **36.0%**	39.87	111
		PBO-PBO	35.4%	UNIFI CSR	26.3%	UNIFI IPD	(A x B)= **9.3%**	14.72	158
	ACT I	IFX pooled-IFX pooled	65.4%	Rutgeerts 2005[113]	44.7%	Imputation	(A x B)= **29.2%**	71.07	243
		PBO-PBO	37.2%	Rutgeerts 2005[113]	31.4%	Imputation	(A x B)= **11.7%**	14.13	121
	PURSUIT	GOL pooled - GOL pooled	52.3%	Sandborn 2014[46]	23.5%	Sandborn 2014[48]	(A x B)= **12.3%**	56.07	457
		PBO-PBO	31.6%	Sandborn 2014[46] Rutgeerts 2015[114]	25.2%	Sandborn 2014[48]	(A x B)= **8.0%**	31.25	393
	ULTRA II	ADA 160/80/40mg – ADA 40mg EOW	59.3%	Sandborn 2012[115]	33%	Sandborn 2013[94]	(A x B)= **19.6%**	29.35	150
		PBO-PBO	38.6%	Sandborn 2012[115]	22.1%	Imputation	(A x B)= **8.5%**	12.37	145
	OCTAVE	TOF 10mg - TOF pooled	64.5%	Dubinsky 2017[102]	42.9%	Dubinsky 2017[102]	(A x B)= **27.7%**	81.34	294
		PBO-PBO	39.1%	Dubinsky 2017[102]	25.8%	Imputed	(A x B)=**10.1%**	11.10	110
	GEMINI I	VDZ 300-VDZ 300 pooled	53.1%	Feagan 2017[116]	46.9%	Feagan 2017[116]	(A x B)= **24.9%**	20.97	84
		PBO-PBO	26.3%	Feagan 2017[116]	25.8%	Imputed	(A x B)= **6.8%**	5.16	76
Clinical response	UNIFI	UST 6mg - UST pooled	66.7%	UNIFI CSR	82.9%	UNIFI IPD	(A x B)= **55.3%**	61.26	111
		PBO-PBO	35.4%	UNIFI CSR	47.4%	UNIFI IPD	(A x B)= **16.8%**	26.49	158
	ACT I[113]	IFX pooled-IFX pooled	65.4%	Rutgeerts 2005[113]	NR	-	**37.8%***	91.97	243
		PBO-PBO	37.2%	Rutgeerts 2005[113]	NR	-	**14.0%***	16.94	121
	PURSUIT	GOL pooled - GOL pooled	50.0%	Sandborn 2014[46]	48.6%	Sandborn 2014[48]	(A x B)= **24.3%**	51.04	210
		PBO-PBO	31.6%	Sandborn 2014[46] Rutgeerts 2015[114]	36.6%	Sandborn 2014[48]	(A x B)= **11.5%**	45.38	393

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
	ULTRA II[115]	ADA 160/80/40mg – ADA 40mg EOW	59.3%	Sandborn 2012[115]	51.1%	Sandborn 2013[94]	**29.3%***	44	150
		PBO-PBO	38.6%	Sandborn 2012[115]	NR	-	**16.6%***	24	145
	OCTAVE	TOF 10mg - TOF pooled	64.5%	Dubinsky 2017[102]	60.3%	Dubinsky 2017[102]	(A x B)= **38.9%**	114.22	294
		PBO-PBO	39.1%	Dubinsky 2017[102]	40.2%	Imputed	(A x B)= **15.7%**	17.29	110
	GEMINI I	VDZ 300-VDZ 300 pooled	53.1%	Feagan 2017[116]	60.7%	Feagan 2017[116]	(A x B)= **32.2%**	27.13	84
		PBO-PBO	26.3%	Feagan 2017[116]	40.2%	Imputed	(A x B)= **10.6%**	8.04	76

Appendix 6 Data calculations and sources for biologic failure 1-year NMAs conditional on response (highlighted data ERG unable to validate)

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
Clinical remission	UNIFI	UST 6mg/kg – UST q8w	57.2%	UNIFI CSR	46.2%	UNIFI IPD	(A x B)= **26.4%**	16.92	64
		UST 6mg/kg – UST q12w	57.2%	UNIFI CSR	37.5%	UNIFI IPD	(A x B)= **21.5%**	8.45	39
		PBO-PBO	27.3%	UNIFI CSR	13.0%	UNIFI IPD	(A x B)= **3.6%**	5.73	161
	ULTRA II	ADA 160/80/40mg – ADA 40mg EOW	36.7%	Sandborn 2012[115]	25.7%	Sandborn 2013[94]	(A x B)= **9.4%**	9.24	98
		PBO-PBO	28.7%	Sandborn 2012[115]	6.2%	Imputed	(A x B)= **1.8%**	1.80	101
	OCTAVE	TOF 10mg - TOF 5mg BID	51.0%	Dubinsky 2017[102]	24.1%	Dubinsky 2017[102]	(A x B)= **12.3%**	17.90	146
		TOF 10mg - TOF 10mg BID	51.0%	Dubinsky 2017[102]	36.6%	Dubinsky 2017[102]	(A x B)= **18.7%**	30.46	163
		PBO-PBO	23.4%	Dubinsky 2017[102]	10.4%	Imputed	(A x B)= **2.4%**	3.02	124
	GEMINI I	VDZ 300mg IV – VDZ q4w	39.0%	Feagan 2017[116]	35.0%	Feagan 2017[116]	(A x B)= **13.7%**	3.70	27
		VDZ 300mg IV – VDZ q8w	39.0%	Feagan 2017[116]	37.2%	Feagan 2017[116]	(A x B)= **14.5%**	4.22	29
		PBO-PBO	20.6%	Feagan 2017[116]	10.4%	Imputed	(A x B)= **2.1%**	1.35	63
Clinical response	UNIFI	UST 6mg/kg – UST q8w	57.2%	UNIFI CSR	71.8%	UNIFI IPD	(A x B)= **41.1%**	26.32	64
		UST 6mg/kg – UST q12w	57.2%	UNIFI CSR	70.8%	UNIFI IPD	(A x B)= **40.5%**	15.96	39
		PBO-PBO	27.3%	UNIFI CSR	43.5%	UNIFI IPD	(A x B)= **11.9%**	19.11	161
	ULTRA II[115]	ADA 160/80/40mg – ADA 40mg EOW	36.7%	Sandborn 2012[115]	45.7%	Sandborn 2013[94]	**15.3%***	15	98
		PBO-PBO	28.7%	Sandborn 2012[115]	NR	-	**5.9%***	6	101
	OCTAVE	TOF 10mg - TOF 5mg BID	51.0%	Dubinsky 2017[102]	44.6%	Dubinsky 2017[102]	(A x B)= **22.7%**	33.12	146
		TOF 10mg - TOF 10mg BID	51.0%	Dubinsky 2017[102]	59.1%	Dubinsky 2017[102]	(A x B)= **30.1%**	49.19	163

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
		PBO-PBO	23.4%	Dubinsky 2017[102]	34.6%	Imputed	(A x B)= **8.1%**	10.04	124
	GEMINI I	VDZ 300mg IV – VDZ q4w	39.0%	Feagan 2017[116]	42.5%	Feagan 2017[116]	(A x B)= **16.6%**	4.49	27
		VDZ 300mg IV – VDZ q8w	39.0%	Feagan 2017[116]	46.5%	Feagan 2017[116]	(A x B)= **18.1%**	5.28	29
		PBO-PBO	20.6%	Feagan 2017[116]	34.6%	Imputed	(A x B)= **7.1%**	4.49	63

Appendix 7 Imputed treat-through data included in ERG maintenance-only NMA scenario

Trial	Arm	N [a]	Induction responders		Maintenance responders		Maintenance clinical remission		Maintenance sustained clinical response		No. pts entering maint.[b]	Clinical response [c]	% of responders in clinical remission	Clinical remission	
Non-biological failure			%	r	%	r	%	r	%	r	N	r	%	r	
ACT1 [49]	PBO	121	37.2%	45	19.8%	24	16.5%	20	14.0%	17	45	17		10	Assumed to be the same proportion of responders to remitters as re-randomised non-bio failure placebo arms
ACT1 [49]	INF 5mg	121	69.4%	84	45.5%	55	34.7%	42	38.8%	47	84	47	33%	28	Assumed same proportion of Induction responders in clinical remission as adalumimab.
ACT1 [49]	INF 10mg	122	61.5%	75	44.3%	54	34.4%	42	36.9%	45	75	45	33%	25	Assumed same proportion of Induction responders in clinical remission as adalumimab.
ULTRA 2 [44,94]	PBO	145	38.6%	56	24.1%	35	12.4%	18	16.6%	24	56	24		14	Assumed to be the same proportion of responders to remitters as re-randomised non-bio failure placebo arms
ULTRA 2	ADA	150	59.3%	89	36.7%	55	22.0%	33	29.3%	44	89	44	33%	29	
Biological failure															
ULTRA 2	PBO	101	28.7%	29	9.9%	10	3%	3	5.9%	6	29	6		3	Assumed to be the same proportion of responders to remitters as re-randomised bio failure placebo arms
ULTRA 2	ADA	98	36.7%	36	20.4%	20	10.2%	10	15.3%	15	36	15	25.7%	9	

ADA: adalimumab; INF: infliximab; PBO: placebo
[a] Number randomised
[b] Number of patients entering maintenance = induction responders
[c] Clinical response = sustained clinical response

Green cells indicate data taken direct from published trials, orange cells are calculations

Appendix 8 Data included in ERG maintenance-only NMA scenario

Source	Trial	Therapy	N	Clinical response	Clinical remission
Biologic failure					
CS Table 18, CS Figures 19 & 20	UNIFI	Ustekinumab 90mg q12w	70	39	16
	UNIFI	Ustekinumab 90mg q12w	91	59	36
	UNIFI	Placebo	88	34	15
Dubinsky, 2017 [102]	OCTAVE	Tofacitinib 5mg	83	37	20
	OCTAVE	Tofacitinib 10mg	93	55	34
	OCTAVE	Placebo	89	13	10
Feagan, 2017[106]	GEMINI	Vedolizumab	83	37	30
	GEMINI	Placebo	38	6	2
Imputed (see Appendix 8)	ULTRA 2	Adalimumab	36	15	9
	ULTRA 2	Placebo	29	6	3
Non biologic failure					
Company Submission B Table 18, Figures 19 & 20	UNIFI	Ustekinumab 90mg q12w	102	78	50
	UNIFI	Ustekinumab 90mg q12w	85	66	41
	UNIFI	Placebo	87	44	27
Dubinsky, 2017 [102]	OCTAVE	Tofacitinib 5mg	115	65	48
	OCTAVE	Tofacitinib 10mg	104	67	46
	OCTAVE	Placebo	109	27	12
Sandborn, 2014 [48]	PURSUIT-M	Golimumab 50mg	151	71	50
	PURSUIT-M	Golimumab 100mg	151	75	51
	PURSUIT-M	Placebo	154	48	34
Hibi, 2017[47]	PURSUIT-J	Golimumab 100mg	32	18	16
	PURSUIT-J	Placebo	31	6	2
Feagan, 2017[106]	GEMINI	Vedolizumab	145	88	68
	GEMINI	Placebo	79	21	15
Imputed (see Appendix 8)	ACT1	Infliximab 5mg	84	47	28
	ACT1	Infliximab 10mg	75	45	25
	ACT1	Placebo	45	17	10
Imputed (see Appendix 8)	ULTRA 2	Adalimumab	89	44	29
	ULTRA 2	Placebo	56	24	14

Appendix 9 Summary of key issues for cost-effectiveness

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
Modelled decision problem				
Population	The modelled patient population is described in CS section B.3.2.1.	The model population is appropriate for the scope, the anticipated marketing authorisation and UNIFI trial population.		
	Results are reported for two subgroups: • Biologic failure • Non biologic failure	We agree with the decision to present results for the subgroups only and not for the whole ITT population. The subgroups are defined by biologic failure, rather than biologic exposure as requested in the scope, but this is unlikely to affect the results.		
	Baseline characteristics for the two modelled cohorts are based on the UNIFI trial (CS Table 34).	Baseline demographics in the model are broadly reflective of the ustekinumab and comparator trial populations and similar to patients starting biologic treatment for UC in the UK. There were variations in mean age, body weight and the proportion of men between trials, but we confirm that the model is not sensitive to these parameters.		
Intervention & comparators	The CS states that the model includes all comparators in the NICE scope for both subgroups (CS B.3.2.3), although infliximab and golimumab are not included for the biologic failure subgroup.	The model includes all scope comparators except infliximab and golimumab in the biologic failure subgroup. This omission is unavoidable because RCTs for these drugs excluded people with previous biologic treatment.		
	The model includes biosimilar versions of infliximab and adalimumab, with the same assumed	The inclusion of available biosimilars is appropriate. We anticipate increasing use of biosimilars, but presentation of results		

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	clinical effects and safety profile as the original licensed brands but at lower cost.	for the original biologics as well is useful for comparison.		
Assumptions about treatment				
Extended induction for delayed response	The model allows an extended induction period for people who have not responded by the end of standard induction, as per SmPC recommendations (CS Table 36). Scenario: no extended induction.	The model appropriately reflects recommended induction regimens, including extended induction for delayed response. The 'no extended induction' scenario illustrates the effect of possible variations in clinical practice.		
	In the base case, the loss of response rate in maintenance is assumed to be the same for delayed responders as for initial responders. Scenario: loss of response rates for delayed responders estimated from trial data.	Maintenance efficacy may well differ for initial and delayed responders, but evidence is sparse, so the company's base case assumption of equal loss of response rates for initial and late responders is reasonable.		
Maintenance dose escalation	The model includes recommended maintenance treatment, including escalated regimen.	The model appropriately reflects recommended maintenance regimens, including escalation to higher dose or more frequent treatment when indicated.		
	The company assume that 30% of patients on maintenance have recommended escalated regimens. Scenarios: 10% and 50% (CS B.3.2.3).	The assumption that 30% of patients on maintenance have the escalated regimen is reasonable, with exploration of uncertainty through scenario analysis.		
	The higher (10mg/kg) dose of infliximab is excluded from the	Clinical advice to the ERG is that dose adjustment for infliximab is common in practice. This suggests that the same dose	**ERG base case**: Dose escalation for infliximab as for other	MEDIUM

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	economic analyses, because it is not recommended in the SmPC.	escalation assumptions should be made for infliximab as for other comparators.	treatments (30% high costs with pooled effects).	
	The dose escalation percentage is used in the model to adjust the cost of maintenance therapy and, for the biologic-failure subgroup only, also its effectiveness. The company pools effectiveness rates for the standard and escalated regimens in the non-biologic failure subgroup, arguing that there is not evidence of an exposure-response relationship in this subgroup.	The ERG view is that evidence supporting dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup is weak. We think that the same dose-pooling approach should be used in both subgroups. We prefer pooled estimates, because of high uncertainty over the exposure-response relationships, but use scenario analysis around the company's base case to illustrate the impact of pooling.	**ERG base case:** pooled maintenance regimens for both subgroups **Scenario CS model**: Unpooled regimens for both subgroups **Scenario CS model**: Standard regimens for both subgroups	HIGH MEDIUM
Constant loss of response (no waning)	The risk of loss of response is assumed to be constant over time – both during the trial follow-up period (approximately one year) and subsequently (until loss of response or death). This is justified by the company based on precedent in TA547 and the lack of data to estimate changes in risk over time. A scenario analysis was presented to illustrate the impact of a declining loss of response rate (-25% after 2 years).	In the absence of interim response/ remission data for the trials or longer-term follow-up it is difficult to predict how the absolute or relative risks of loss of response change over time. We therefore agree with the assumption of a constant risk over time.		

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
Treatment continuation (no stopping rule)	CS analyses assume that responders to induction continue maintenance therapy until loss of response or death. The model includes a stopping rule option but this is not used. The model option allows discontinuation at a defined time, with subsequent (constant) loss of response based on either: i) trial data for responders to active induction re-randomised to placebo (UST, GOL, VED and TOF only); or ii) the same rate as for CT (trial data for responders to placebo induction, PBO-PBO).	Given uncertainty over routine use of a 'stopping rule' for biologics in UC, we think it is appropriate to assume continued treatment until loss of response in the base case. We use the 'stopping rule' option in the model to illustrate the impact of discontinuation at one-year, but note uncertainty over this scenario. It is not clear if the assumed post-discontinuation loss of response rates are accurate or whether the scenario reflects trial of discontinuation in practice: which is usually restricted to patients with remission, with re-initiation of treatment after relapse.	**Scenario CS model**: one-year stopping rule, with subsequent loss of response based on trial data: i) PBO-PBO for all treatments; ii) active induction re-randomised to PBO (UST, GOL, VED & TOF only)	MEDIUM
Treatment sequencing	In the base case, after discontinuation of the initial treatment all patients are assumed to continue on conventional treatment until surgery or death. The model has the flexibility to allow one line of subsequent treatment. The company presents a scenario analysis, with vedolizumab as the second line treatment for all other treatments and adalimumab after vedolizumab.	Many patients who might be considered for ustekinumab would not have exhausted all other treatment options. Sequential use of therapies is common in practice, but variable, and cost-effectiveness is potentially sensitive to the choice of subsequent treatment.		
Model structure and framework				
Model type	Hybrid model with decision tree to reflect induction outcomes and a	The overall model structure is appropriate, consistent with previous TA models and		

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	Markov model for maintenance, subsequent standard care and surgery (CS Figure 37 and 38).	accurately implemented. The only major exception is the omission of response and remission health states after failure of the initial treatment (see below).		
Cycle length	The duration of the induction phase varies from 8 to 16 weeks, according to the recommended lengths of standard and extended induction for delayed response (see CS Table 36). The Markov section of the model uses a 2-week cycle, to allow induction periods of different length (CS B.3.2.2.2).	The 2-week Markov cycle is short (e.g. 8 weeks was used in TA547). This will cause some underestimation of costs if symptom recurrence is not always detected and treatment discontinued within 2 weeks. Experts have advised the ERG that clinics provide fast access on request, but this may not be consistent at all times throughout the NHS. However, delays in treatment discontinuation are unlikely to have a significant impact on costs.		
Half cycle correction	A half cycle correction was applied by using the mean number of patients in each health state at the beginning and end of each cycle to calculate costs and QALYs (CS B.3.2.2.2)	Consistent with methods guidance.		
Time horizon	50 years (patients enter the model at 41 years of age)	Consistent with a lifetime horizon and previous appraisals.		
Response and remission after failure of initial treatment	The model assumes that after failure of the initial treatment, patients switch to conventional treatment alone and continue with Active UC until they have surgery or die (CS B.3.2.2.2). The company argue that	The omission of response and remission health states after failure of the initial treatment option is a major limitation. This implies that all patients follow a chronic active or progressive form of disease, which is inconsistent with previous NICE	**ERG base case**: add response and remission health states after the switch to CT.	HIGH

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	the impact of introducing response and remission health states after failure of initial treatment would be negligible, as it would affect all treatments in a similar manner (Clarification Response B1).	appraisals and unrealistic. For face validity, the model should reflect long-term patterns of disease. This is also necessary for accurate estimation of the downstream benefits of inducing and retaining initial response.		
Surgical treatment pathway	The model includes surgery as an option for patients with active UC after failure of initial therapy. Two phases of surgery are modelled, each lasting for six months to allow for staged procedures. If the first phase is successful, patients stay in remission until death. However, some patients have chronic complications after surgery, including pouch failure which may require a second phase of surgery for revision. The model assumes that all patients achieve remission after revision surgery. (CS B.3.2.2.3)	In previous TAs, surgery was modelled as a one-off event. However, the current model better reflects the usual process of staged procedures: subtotal colectomy with ileostomy followed by either IPAA (pouch) surgery or permanent ileostomy (phase 1); and subsequent revision surgery if needed due to pouch failure (phase 2). The assumption of remission after revision surgery is a reasonable simplification.		
Mortality	Mortality rates are assumed to be the same as for the general population, except for a small mortality risk associated with surgery.	This approach is consistent with previous TAs and the ERG consider it a reasonable simplification.		
Clinical parameters				
Response & remission rates	Standard induction: NMA response and remission rates at the end of standard induction (CS Table 40).	ERG replication of the company's induction NMAs found some discrepancies (see section 3.3.6.1 above). We would prefer	**ERG base case**: ERG replication of FE	MEDIUM

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	Fixed effects model in the base case and random effects in a scenario (Scenario 1).	the random effects model, due to heterogeneity. However, this gives very wide credible intervals. We therefore use the fixed effects model for our base case and test the random effects model in a scenario.	induction NMA (Table 25 and Table 27) **ERG scenario:** Induction NMA RE (Table 26)	MEDIUM
	Maintenance phase loss of response estimates from direct trial data in the base case (CS Table 43). Maintenance NMA scenario based on company 1-year NMA, conditional on response.	The ERG has strong concerns about use of absolute response rates from individual arms of RCTs, as in the company's base case. We therefore prefer the company's maintenance NMA scenario over their base case. Due to potential heterogeneity, we prefer the random effects approach.	**ERG base case**: Company 1-year NMA conditional on response, RE ERG replication (Table 30 & Table 31)	HIGH
		The ERG alternative maintenance NMA followed methods applied in the TA547 appraisal (see 3.1.7.5.5). We conducted a scenario analysis with this 'no carry over' NMA for consistency with TA547 and to explore uncertainty associated with the assumption of carry over.	**ERG scenario:** ERG maintenance only NMA ('no carry over'), RE (Table 32 & Table 33)	HIGH
	Direct trial data is used to estimate response and remission rates at the end of extended induction period for people who did not respond during standard induction (CS Table41). Direct trial data is also used to estimate loss of response rates for delayed responders (CS Table 44).	There is high uncertainty over the direct trial estimates of response and remission for extended induction and loss of response rates for delayed responders. The company's scenario excluding extended induction tests the impact of assumptions about delayed response.		

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
Adverse events	Serious infections were the only adverse events included in the model. This is consistent with previous NICE UC appraisals. Rates of serious infections in the model are based on a multinational registry for systemic treatment of psoriasis: the PSOLAR study[32], which included 7,300 patients treated with ustekinumab, infliximab or adalimumab over a total of 13,349 person years (mean follow up 22 months). Risks with vedolizumab, tofacitinib and CT are assumed to be the same as for ustekinumab; and those with golimumab and the infliximab biosimilar to be the same risk as infliximab. Scenario: same rate of serious infections (0.83%) for all treatments (Scenario 11).	Overall the rates of serious infections used in the model appear reasonable. Despite uncertainties over use of the PSOLAR data and assumptions, this is still the best available source of evidence and the model is not sensitive to plausible changes in serious infection rates.		
Incidence of surgery and complications	Misra et al. (2016)[77] was used as the source for the initial incidence of surgery (0.47% per year). This was a large UK-based study, used in TA547. Chronic complication rates within 6 months of first surgery (33.5%) were based on the 2013 national clinical	We agree with the use of UK estimates for the incidence of first surgery and rates of early and late complications. The first two of these estimates were also used in TA547. A different source was used for late complications in TA547 (Ferrante et al. 2008), although the model is not sensitive to this difference. The company's assumption that the incidence of revision		

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	audit for inpatient care for adults with UC[78] and the rate for late chronic complications (3.25% per year) was based on Segal et al. (2018).[79] Despite its small sample size (39 patients), this was the only UK study. The company assumes that the probability of a second phase of surgery for revision of pouch failure is the same as for the initial surgery.	surgery for patients with chronic complications is the same as that for initial surgery is arbitrary, but this only affects a small proportion of the cohort and the model is not sensitive to this assumption. Use of the same set of parameters to characterise the incidence and complications of surgery for patients with and without prior biologic failure is a reasonable simplification.		
Mortality	The model uses general population all-cause mortality rates adjusted for age and gender from UK Life tables. The only excess mortality for UC was a relative risk of 1.3 for surgery from a meta-analysis by Jess et al. (2007)[83] applied during the six-month surgery health states. This approach is similar to that in TA547 and TA329, although TA342 applied excess mortality to all active UC and post-operative health states.	The company's assumptions about mortality are reasonable, with an excess risk for surgery, but otherwise the same risks as for the general population. We note that model is not sensitive to the relative risk assumed during surgery.		
Utilities				
Health state utilities	General population utility (EQ-5D-3L) by age and gender from Ara and Brazier (2010)[88]. Health state utilities from Woehl et al. (2008)[84], used to calculate multipliers with respect to	We agree with the company's decision not to use utility estimates from the UNIFI EQ-5D data: primarily because they are inconsistent with the values used in previous NICE appraisals for UC. However,	**ERG scenario**: UNIFI utilities applied to ERG base case	MEDIUM

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	remission. This was a UK EQ-5D-3L study of 180 UC patients used in TA329, TA342 and TA547. UNIFI EQ-5D-5L data (valued using the cross-walk method [89]) is used in scenario analysis. Utility multipliers for the surgery health state were taken from Arseneau et al. (2006)[85], a US TTO study for 48 UC patients undergoing ileostomy and J pouch. These were assumed to apply to both first and second stages of surgery.	the number of observations in the three severity health states is large and the analysis appears to have been well-conducted. The ERG therefore considers the scenario analysis with UNIFI utility estimates to be important.		
Disutility for serious infection	A disutility for serious infections was derived from a company model for TA329, as reported by Stevenson et al. [86] This is applied as a one-off decrement for each SI.	The QALY decrement for serious infections appears to have been overestimated as the disutility of 0.156 is not adjusted in the model for the expected duration of symptoms (assumed to be 28 days in TA329).	**ERG base case**: disutility for serious infections (0.156) applied for estimate duration of 28 days (0.012 QALY loss)	
Costs and resource use				
Drug acquisition costs	Drugs are costed according to licensed regimens, with unit costs sourced from the BNF, TA342, TA457 and MIMS. Wastage assumptions are applied for weight-based medications.	Changes to assumptions about the use and costs of CT are unlikely to be influential in the model because of their low cost and similar impact on cost-effectiveness of comparators. Nevertheless, for face validity we update the assumptions about use of conventional	**ERG base case**: CT drug usage as per RCP 2016 audit (TA547). **ERG base case**: include concurrent	

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	Costs of CT are estimated as a treatment mix of 6 drugs. The weights of each of the CT treatment taken from NICE TA342. We note that these usage assumptions were updated in TA547, using results from the 2016 RCP audit of biologic treatment for IBD.[67] Costs for concurrent conventional treatment drugs were not included alongside biologics or JAK inhibitors.	therapy drugs as a comparator and concurrent with other treatments as per TA547.	use of conventional drugs alongside other comparators as per RCP 2016 audit (TA547).	
Administration costs	Administration costs for intravenous drugs were included, with a cost of an outpatient visit based on 2017/18 NHS Reference Costs. No administration cost was included for self-injection treatment.	Currently distribution and patient education for self-administration is organised and paid for by the drug companies, so no cost to the NHS. If this changed it would add to NHS cost of self-administered drugs (?), but likely to be modest.		
Other health care costs	Health state resource use: Mostly based on Tsai et al. 2008, similar to TA543. Hospitalisation rates for the pre-surgery health states were obtained from Sandborn et al. 2016 and adjusted by the proportion of non-surgery related hospitalisations, to derive the inpatient care without colectomy rates. Cost of surgery are based on - Buchanan	Estimates of health state, surgery and adverse event costs are reasonably consistent with previous UC appraisals.		
Adverse event costs	The cost of a serious infection was estimated as a weighted average of	This is reasonable.		

Issues	*Features of the company model*	*ERG comments*	*ERG analysis*	*Priority*
	HRG costs for five types of infection: sepsis, pneumonia, urinary tract infection, respiratory infection and bronchitis (NHS reference costs 2016/17).			

Appendix 10 Comparison of the company's cost effectiveness results when SoC/CT results are pulled from Sheet!Markov_UK in the company base case model

Table 62 Comparison of the ICERs for ustekinumab vs CT: non-biologic failure

Scenario	Description	Company (Results from Markov_SOC sheet)	ERG (Results from Markov_UK sheet)	Difference
Company base case		***£23,446***	***£23,450***	***£4***
Scenario 1: Induction NMA	NMA random effect model	£23,446	£23,451	£5
Scenario 2: Maintenance NMA	Alternative efficacy source for the maintenance phase	£24,575	£24,581	£6
Scenario 3: Non-constant loss of response	Max Tx to apply linear loss of response: 2; after max tx loss of response reduced by 25%	£23,053	£23,056	£3
Scenario 4: Utility values from UNIFI trial	Utilities for active UC, remission, response without remission	£78,091	£78,227	£136
Scenario 5: Utility values from Swinburn et al 2012 [87]	Utilities for 1st surgery, post-1st/2nd surgery remission, post-1st surgery complications	£23,363	£23,369	£6
Scenario 6: Subsequent treatment	Upon loss of response, a second treatment is initiated for each comparator (except CT)	£27,785	£27,817	£32
Scenario 7: Dose escalation set to 10%	Dose escalation is set to 10% for all treatment	£21,701	£21,705	£4
Scenario 8: Dose escalation set to 50%	Dose escalation is set to 50% for all treatment	£25,191	£25,195	£4
Scenario 9: Delayed responder loss of response	Delayed responder efficacy is taken from individual trials rather than the assumption that efficacy is the same as early responders	£23,297	£23,302	£5
Scenario 10: Exclude delayed responders	Delayed responders are removed from the analysis	£21,870	£21,876	£6
Scenario 11: Serious infection	All treatments have the same rate of serious infection as ustekinumab (0.83%)	£23,446	£23,450	£4

Source: CS Table 69

Table 63 Comparison of the ICERs for ustekinumab vs CT: biologic failure

Scenario	Description	Company (Results from Markov_SOC sheet)	ERG (Results from Markov_UK sheet)	Difference
Company base case		***£26,205***	***£26,213***	***£8***
Scenario 1: Induction NMA	NMA random effect model	£26,334	£26,342	£8
Scenario 2: Maintenance NMA	Alternative efficacy source for the maintenance phase	£28,018	£28,028	£10
Scenario 3: Non-constant loss of response	Max Tx to apply linear loss of response: 2; after max tx loss of response reduced by 25%	£25,711	£25,718	£7
Scenario 4: Utility values from UNIFI trial	Utilities for active UC, remission, response without remission	£86,723	£87,035	£312
Scenario 5: Utility values from Swinburn et al 2012 [87]	Utilities for 1st surgery, post-1st/2nd surgery remission, post-1st surgery complications	£26,106	£26,116	£10
Scenario 6: Dose escalation set to 10%	Dose escalation is set to 10% for all treatment	£24,733	£24,741	£8
Scenario 7: Dose escalation set to 50%	Dose escalation is set to 50% for all treatment	£27,705	£27,712	£7
Scenario 8: Delayed responder loss of response	Delayed responder efficacy is taken from individual trials rather than the assumption that efficacy is the same as early responders	£25,880	£25,890	£10
Scenario 9: Exclude delayed responders	Delayed responders are removed from the analysis	£23,525	£23,537	£12
Scenario 10: Serious infection	All treatments have the same rate of serious infection as ustekinumab (0.83%)	£26,205	£26,213	£8

Source: CS Table 70

Appendix 11 Additional scenarios conducted by the ERG in the company's base case model (ERG replication)

Table 64 Additional ERG scenarios conducted on the company's base case model (ERG replication), Non-biologic failure sub group (ustekinumab vs comparators), company's proposed CMU arrangement price for ustekinumab; list prices for comparators

Scenario	Infliximab	Infliximab biosimilar	Golimumab	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
Company Base Case (ERG replication)	**£14,710**	**£16,606**	**£12,025**	**£18,047**	**£19,146**	**£1,762**	**£13,465**	**£23,450**
Scenario 1: Unpooled dose regimen (higher regimen)	£12,524	£14,998	£9,576	£17,174	£18,522	Dominant	£9,215	£23,761
Scenario 2: Standard regimen (lower regimen)	£16,881	£18,274	£14,594	£19,126	£19,980	£6,490	£16,560	£23,334
Scenario 3: 1-yr stopping rule with subsequent loss of response based on SoC data	Dominant	Dominant	Dominant	Dominant	£2,283	Dominant	Dominant	£13,726
Scenario 4: 1-yr stopping rule with subsequent loss of response based on active induction re-randomised to placebo	Dominant	Dominant	Dominant	Dominant	£695	Dominant	Dominant	£10,470
Scenario 5: Utility for subsequent surgery health state: 0.55 (assuming a 10% decline from the baseline estimate of 0.614)	£14,709	£16,606	£12,025	£18,047	£19,146	£1,762	£13,465	£23,450

Table 65 Additional ERG scenarios conducted on the company's base case model (ERG replication), Biologic failure sub group (ustekinumab vs comparators), company's proposed CMU arrangement price for ustekinumab; list prices for comparators

Scenario	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
Company Base Case (ERG replication)	**£18,210**	**£19,670**	**Dominant**	**£5,394**	**£26,213**
Scenario 1: Unpooled dose regimen (higher regimen)	£18,210	£19,670	Dominant	£5,394	£26,213
Scenario 2: Standard regimen (lower regimen)	£19,099	£20,656	Dominant	£5,486	£27,479
Scenario 3: 1-yr stopping rule with subsequent loss of response based on SoC data	£1,606	£3,972	Dominant	Dominant	£16,377
Scenario 4: 1-yr stopping rule with subsequent loss of response based on active induction re-randomised to placebo	£1,324	£3,587	Dominant	Dominant	£15,590
Scenario 5: Utility for subsequent surgery health state: 0.55 (assuming a 10% decline from the baseline estimate of 0.614)	£18,210	£19,670	Dominant	£5,394	£26,212