Vedolizumab for the treatment of adults with moderately to severely active Crohn's disease: A Single Technology Appraisal

Produced by School of Health and Related Research (ScHARR), The University of Sheffield

Authors Rachid Rafia, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Alison Scope, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Sue Harnan, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

John W Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Anthea Sutton, Information Resources Group Manager/Senior Information Specialist, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Kath Dickinson, Information Specialist, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Miles Parkes, Consultant Gastroenterologist, Addenbrooke's Hospital, Cambridge

John Mayberry, Consultant Physician and Honorary Professor, University Hospitals of Leicester NHS Trust

Alan Lobo, Consultant Gastroenterologist and Honorary Professor of Gastroenterology, Sheffield Teaching Hospitals NHS Foundation Trust, University of Sheffield

Correspondence to Rachid Rafia, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Date completed October 2014

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 13/128/01.

Declared competing interests of the authors

None

Acknowledgements

We would like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Gill Rooney, Programme Administrator, ScHARR, for her help in preparing and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Rafia R., Scope A., Harnan S., Stevens J.W., Stevenson M., Sutton A., Dickinson K., Parkes M., Mayberry J., Lobo A. Vedolizumab for the treatment of adults with moderately to severely active Crohn's disease: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2014.

Contributions of authors

Alison Scope and Sue Harnan summarised and critiqued the clinical effectiveness data reported by the company. John W Stevens critiqued the statistical analyses undertaken by the company. Anthea Sutton and Kath Dickinson undertook the literature searches run by the ERG and critiqued the searches by the company. Rachid Rafia and Matt Stevenson critiqued the health economic analysis submitted by the company. Professor John Mayberry, Professor Alan Lobo and Dr. Miles Parkes provided clinical advice to the ERG throughout the project. All authors were involved in drafting and commenting on the final document.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

anti-TNF-α	tumour necrosis factor-alpha antagonist
ASAs / 5-ASAs	Aminosalicylates
MP / 6-MP	6-mercaptopurine
AE	Adverse event
CD	Crohn's disease
c.i.	Confidence interval
CEAC	Cost-effectiveness acceptability curve
CRP	C-reactive protein
CS	Company submission
CSF	Corticosteroid-free
CSR	Clinical Study Report
ECG	Electrocardiogram
EMA	European Medicines Agency
EOW	Every other week
EQ-5D	Euroqol 5-Dimensions
ERG	Evidence Review Group
EW	Every week
FDA	Food and Drug Administration
HE	Health economics
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
IBD	Inflammatory Bowel Disease
IBDQ	Inflammatory Bowel Disease Questionnaire
ICER	Incremental cost-effectiveness ratio
IPAA	Ileal pouch-anal anastomosis
ITT	Intention-to-treat
LOCF	Last observation carried forward
i.v.	Intravenous
MeSH	Medical subject heading
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
ONS	Office for National Statistics
PAS	Patient Access Scheme

PbR	Payment by Results
PML	Progressive multifocal leukoencephalopathy
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSS	Personal Social Services
Q4W	Every 4 weeks
Q8W	Every 8 weeks
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
SAE	Serious adverse event
s.c.	Subcutaneous
SD	Standard deviation
SE	Standard error
SF-36	Short Form 36
SG	Standard gamble
STA	Single Technology Appraisal
SmPC	Summary of product characteristics
TB	Tuberculosis
TNF	Tumour necrosis factor
TNF-α	Tumour necrosis factor–alpha
UC	Ulcerative colitis
UK	United Kingdom
US	United States
VAS	Visual analogue scale
WHO	World Health Organisation
Wk(s)	Week(s)

1. SUMMARY

1.1 Critique of the decision problem in the company's submission

The population considered by the company in this assessment (adult patients with moderately to severely active Crohn's disease in whom the disease has responded inadequately to, or is no longer responding to, either conventional therapy or an anti-TNF-α, or who are intolerant to either of them) matches that defined in the final NICE scope. The intervention considered in the company submission (CS), vedolizumab, also matches the final NICE scope. According to its current marketing authorisation, the recommended dose regimen of vedolizumab is 300mg administered by intravenous (i.v.) infusion at zero, two and six weeks and every eight weeks thereafter. It should be noted that the treatment regimen used in the company's model differs from the licensing and the treatment regimen described by the company in their decision problem (Section 1.10 of the CS). The final NICE scope defines comparators to be established clinical management without vedolizumab, which may include antibiotics, drug treatment with conventional corticosteroids alone or in combination with azathioprine, mercaptopurine or methotrexate; aminosalicylates; budesonide alone or in combination with azathioprine, mercaptopurine or methotrexate and tumour necrosis factor-alpha antagonist (anti-TNF-α). The CS includes data on remission and response rates but did not include data on relapse rates. Data on surgery are not included in the CS but were provided following a request by the Evidence Review Group (ERG). No equity issues were highlighted in the CS.

1.2 Summary of clinical effectiveness evidence submitted by the company

The CS includes a systematic review and network meta-analysis (NMA) of the clinical effectiveness literature. The GEMINI II and GEMINI III trials form the main supporting evidence for the intervention. Both trials were Phase III, multicentre (GEMINI II 39 countries; GEMINI III 19 countries), randomised, double-blind, placebo-controlled trials designed to evaluate the efficacy and safety of vedolizumab. The GEMINI II trial assessed vedolizumab as an induction treatment (dosing at weeks 0 and 2 with assessment at week 6) and maintenance treatment (weeks 6 to 52), and included patients who were naïve to anti-TNF-α, and patients who had an inadequate response to, loss of response to, or intolerance to immunomodulators or anti-TNF-α. The GEMINI III trial was designed to evaluate the efficacy and safety of vedolizumab as an induction treatment with dosing at weeks 0, 2 and 6 and assessment at weeks 6 and 10. The primary analysis in the GEMINI III trial focussed on people for whom an anti-TNF-α has failed (i.e., an inadequate response to, loss of response to, or intolerance of $\geq$1 anti-TNF-α). A secondary analysis evaluated an overall population which also included patients who were naïve to anti-TNF-α, and pre-specified exploratory analyses examined the group naive to anti-TNF-α. In general, all efficacy analyses in the GEMINI II and III trials were conducted according to the intention-to-treat (ITT) principle whereby patients who withdrew prematurely were considered as treatment failures.

During the 6 week induction phase of the GEMINI II trial 368 patients were randomised (3:2 ratio) to receive 300mg vedolizumab i.v. or placebo (as saline) at Weeks 0 and 2 (Cohort 1). In order to fulfil sample size requirements for the maintenance study, an additional 748 patients were enrolled in an open-label group (Cohort 2), which also received 300mg vedolizumab i.v. During the maintenance phase, patients from both cohorts (Cohort 1 and Cohort 2) who had a clinical response (defined as $\geq$ 70 point decrease in the CDAI score) to vedolizumab at week 6 were randomised (1:1:1 ratio) to double-blind treatment with vedolizumab 300mg i.v. every 8 weeks (with placebo administered every other visit to preserve blinding), vedolizumab 300mg i.v. every 4 weeks or placebo every 4 weeks for up to 52 weeks. According to the CS, randomisation was stratified by three factors: (1) cohort; (2) concomitant use or non-use of glucocorticoids; and (3) concomitant use or non-use of immunosuppressive agents or prior use or non-use of anti-TNF-α or both. Patients in the induction study who did not have a clinical response at week 6 continued to receive their assigned study drug (vedolizumab or placebo) every 4 weeks and were followed through until week 52 separately from the maintenance study.

In the induction phase of GEMINI II, 9% [105/1115] of the total population prematurely discontinued from the study. In contrast, a larger proportion of patients discontinued during the maintenance phase (48% [242/461]). The main reasons for discontinuation in the vedolizumab and placebo groups were lack of efficacy or adverse events (AEs).

In the induction phase of the GEMINI II trial, patients treated with vedolizumab, had significantly higher rates of clinical remission (defined as CDAI $\leq$150) at week 6 compared with placebo (14.5%versus (vs) 6.8%) (the first primary outcome). The treatment difference from placebo was 7.8% (95% CI 1.2, 14.3; p = 0.0206). There was no significant difference between the vedolizumab and placebo groups for the second primary outcome which analysed the number of patients achieving an enhanced clinical response (defined a reduction of 100 points or more in the CDAI score) at week 6 (*p*-value not reported).

In the maintenance phase of the GEMINI II trial, patients treated with vedolizumab every 8 weeks (Q8W) and every 4 weeks (Q4W), had significantly higher rates of clinical remission at week 52 (defined as CDAI score of $\leq$ 150 points) compared with placebo. The treatment difference from placebo was 17.4% (95% CI 7.3, 27.5; p = 0.0007) and 14.7% (95% CI 4.6, 24.7; p = 0.0042) respectively. Patients receiving vedolizumab every 4 or 8 weeks were significantly more likely to achieve enhanced clinical response defined as a reduction of 100 points or more in the CDAI score and have a corticosteroid-free remission at week 52 compared with patients receiving placebo.

Sub-group analyses demonstrated that clinical remission rates were greater for patients treated with vedolizumab than those treated with placebo, regardless of prior exposure to anti-TNF-α. Similar improvements with vedolizumab versus placebo were found in the enhanced clinical response (defined as a reduction of 100 points or more in the CDAI score), and corticosteroid-free clinical remissions at week 52 in all sub-groups. Although a higher number of patients achieved clinical remission with vedolizumab in the immunomodulator and corticosteroid failure sub-groups than in the anti-TNF-α failure subgroup, the treatment differences between placebo and vedolizumab were generally similar among all sub-groups.

During the 10 week induction trial of the GEMINI III trial, 416 patients were enrolled. 315 patients had a previous inadequate response to, loss of response to, or intolerance of, one or more anti-TNF-α and 101 patients were naïve to anti-TNF-α. Patients were randomly assigned to receive intravenous vedolizumab (300mg) or placebo (as saline) at week 0, week 2, and week 6, with three stratification factors: (1) the presence or absence of previous anti-TNF-α failure; (2) concomitant use or non-use of glucocorticoids; and (3) by concomitant use or non-use of immunosuppressive agents.

In GEMINI III, 7% (28/416) of the total population prematurely discontinued from the study. The reasons for discontinuation in the vedolizumab and placebo groups were not reported. Discontinuation due to AEs occurred in 2% (4/209) of placebo patients and 4% (8/207) of vedolizumab-treated patients.

There was no statistically significant difference between vedolizumab and placebo in the primary endpoint of the proportion of patients achieving clinical remission at week 6 (CDAI score ≤150 points) in the anti-TNF-α failure population, therefore, statistical evaluation of the secondary endpoints is considered exploratory by the company.

Secondary efficacy endpoints included clinical remission at week 10, enhanced clinical response (defined as a reduction of 100 points or more in the CDAI score) at weeks 6 and 10 and sustained remission (defined as CDAI score ≤150 points at both Week 6 and Week 10) in the anti-TNF-α failure population, and clinical remission and enhanced clinical response at week 6 and 10 and sustained remission in the overall population. Compared with placebo, vedolizumab was associated with a higher number of patients achieving clinical remission at week 10 and an enhanced clinical response (defined as a reduction of 100 points or more in the CDAI score) at week 6 and 10 in the anti-TNF-α failure population. In the overall population, vedolizumab-treated patients had higher rates of clinical remission, and enhanced clinical response (defined as a reduction of 100 points or more in the CDAI score) at weeks 6 and 10 and sustained remission compared with placebo-treated patients.

The frequency of AEs was similar between the vedolizumab and placebo groups in both GEMINI II and GEMINI III. In GEMINI II serious adverse events (SAEs) occurred more frequently in the vedolizumab groups (24.4%) than in the placebo group (15.3%), in GEMINI III these rates were 8% in the vedolizumab group compared with 6% in the placebo group. Four patients receiving vedolizumab and one patient receiving placebo died during the GEMINI II trial. No patients died during GEMINI III. The rates of infusion-related reactions in the induction and maintenance phases were similar across the vedolizumab (5.5%) and placebo groups (3.0%) in GEMINI II. It was reported that no serious infusion-related reactions occurred in GEMINI III. No cases of anaphylaxis or progressive multifocal leukoencephalopathy (PML) were observed in either GEMINI II or GEMINI III.

Supplementary safety evidence from an ongoing GEMINI Long Term Safety study and two separate pooled safety analyses were also provided by the company. In general, the overall safety profile of vedolizumab appeared to be similar between patients with ulcerative colitis (UC) and Crohn's disease (CD) with slightly higher rates of AEs in the CD patients. As of June 2013, no cases of PML were reported in any of the >2,700 patients treated with vedolizumab, including approximately 900 patients with ≥24 months exposure. A total of 26 vedolizumab-treated patients in the integrated safety population had been diagnosed with malignancy, of which 18 met SAE criteria. Of these, skin cancers (n=5) and colon cancer (n=4) were most common. Tuberculosis was reported in a total of 4 patients (3 with CD, 1 with UC), and 13 deaths occurred across all controlled and uncontrolled studies in UC (n=4) and CD (n=9).

In the absence of any direct head-to-head randomised controlled trials (RCTs) comparing vedolizumab to other relevant biologic therapies for the treatment of moderate to severe CD, the company conducted an NMA. The NMA, as reported in the CS, compared vedolizumab, adalimumab, infliximab and placebo for the outcomes of: clinical response; enhanced clinical response; clinical remission; and discontinuation due to AEs; using data from the trials: GEMINI II; GEMINI III; CLASSIC I; Targan et al(1997); NCT00105300; NCT00445939; EXTEND; ACCENT I; CLASSIC II; NCT00445432; and CHARM. The size of the network for each outcome varied depending on the availability of the data in each study.

The company undertook separate NMAs of the anti-TNF-α naïve, anti-TNF-α experienced/failure subgroups and the entire (mixed) population. Induction phase and maintenance phase data were synthesised separately. Both Bayesian fixed and random effects models were used but only the fixed effects model results were presented. According to the CS all outcome measures were modelled using a logistic model.

In the induction phase for anti-TNF-α naïve patients, for clinical response (drop in CDAI ≥70) all treatments were statistically significantly effective versus placebo. Infliximab is statistically significantly better than vedolizumab. Vedolizumab has a lower odds ratio than adalimumab but the pairwise comparison between the two was not statistically significant. For enhanced clinical response (drop in CDAI ≥100), there were no data for infliximab. Adalimumab 40/20mg (not a licensed UK dose) and 80/40mg dose (licensed in UK as "normal" dose) were not significantly different to placebo, but adalimumab 160/80mg (licensed in UK as "accelerated" dose) and vedolizumab were. There was no significant difference in pairwise comparison between adalimumab and vedolizumab. For clinical remission, all treatments except adalimumab 40/20mg (not a UK dose) were statistically better than placebo. In pairwise comparisons, infliximab was statistically significantly better than vedolizumab at 10 or 6 weeks; vedolizumab had a better OR versus placebo than adalimumab 80/40, but worse OR versus placebo than adalimumab 160/80mg, but nether comparison was statistically significant. For discontinuations due to AEs, adalimumab 160/80mg dose was significantly better (lower) than vedolizumab, and there were no data available for infliximab.

In the maintenance phase for the anti-TNF-α naïve patients, vedolizumab every 4 weeks was only statistically different to placebo for the outcome clinical remission. Vedolizumab every 8 weeks was statistically different from placebo for both clinical response and clinical remission. Infliximab was statistically different from placebo in all three outcomes (clinical remission, clinical response, discontinuation due to AE's). The statistical significance of the difference in clinical response between vedolizumab and infliximab was not reported for the standard dose (5mg) of infliximab licenced in the UK, but infliximab 10mg was statistically significantly better than vedolizumab every 4 weeks; the clinical response OR for infliximab 5mg versus placebo was better than that for both vedolizumab every 4 weeks and every 8 weeks (both licenced in UK). The difference between vedolizumab and infliximab for the outcome clinical remission was not statistically significant. There was a high OR for discontinuation due to AE's compared to placebo for infliximab; vedolizumab was significantly better than infliximab for discontinuations due to AEs.

In the induction phase for anti-TNF-α experienced/failure network, both adalimumab and vedolizumab (infliximab was not included in this network) were significantly better than placebo, except for vedolizumab at 6 weeks for the outcome clinical remission. Vedolizumab and adalimumab were not statistically significantly different to one another in most analyses; the OR for adalimumab versus placebo was better that that for vedolizumab in most analyses, and statistically significantly superior at 6 weeks for clinical remission.

A network for anti-TNF- α failure subgroups was not possible for maintenance. Only GEMINI II reports this data.

For the "entire population" induction studies networks, headline results for the clinical response outcome include: all UK licensed treatments were significantly better than placebo; infliximab was statistically significantly better than vedolizumab (OR 5.5 (95% CrI 1.5 to 25); and the difference between adalimumab and vedolizumab was not statistically significant, with 95% CrI all crossing the line of no effect.

For the "entire population" maintenance network, headline results for the clinical response outcome include: all treatments except vedolizumab every 4 weeks were significantly better than placebo; and both adalimumab and infliximab were significantly better than vedolizumab.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The systematic review process followed by the company was adequate, although not reported fully in the CS. Further detail of the systematic review and NMA were provided in a separate document "Takeda Data on File". Despite minor limitations in the company's search strategy, the Evidence Review Group (ERG) is confident that all relevant studies of vedolizumab were included in the CS. The specified inclusion and exclusion criteria appear generally appropriate, if lacking in detail, and reflect the information given in the decision problem. The validity assessment tool used to appraise the included studies, as suggested by NICE Specification for company/sponsor submission of evidence template, was based on the quality assessment criteria for RCTs and was considered appropriate by the ERG.

The efficacy and safety of vedolizumab was positively demonstrated in GEMINI II. Owing to the high discontinuation rates in the maintenance phase of the GEMINI II trial, estimates of treatment effects (including magnitude) may be confounded; though the imputation of missing patients as failures should limit the impact of attrition on estimates of efficacy to underestimation of treatment effects, attrition may be more problematic for safety outcomes and lead to underestimates of adverse events. The trials assess response in the induction phase earlier than would be done in the UK, at six weeks. As such, the population entering the maintenance phase in GEMINI II is not fully representative of the UK spectrum, as patients who take longer to respond are excluded. This could conceivably lead to an overestimation of maintenance treatment effect, if these patients are also less likely to maintain a response when in remission. In addition, the trial of maintenance therapy was not of sufficient size or duration to estimate the risk of uncommon AEs.

The primary endpoint was not achieved in GEMINI III; therefore, statistical evaluation of the secondary endpoints is acknowledged as exploratory by the company.

Furthermore, the ERG noted a number of prognostic factors which clinical advisor to the ERG suggested may have response modulating effect in the trial populations. These included levels of faecal calprotectin which indicates active inflammation (levels were relatively high in both GEMINI II and GEMINI III), and the fact that patients with stricturing disease were excluded.

The ERG considered that the results presented in the company's NMA may have underestimated the uncertainty in treatment effects since fixed effects models were used. The networks included in the CS were of varying quality and relevance. The ERG had several observations relating to these. In summary, for both induction and maintenance networks, the anti-TNF-α naïve network was thought to be theoretically the most generalizable to UK patients in whom the disease has responded inadequately to, or is no longer responding to conventional therapy and who have not previously received an anti-TNF-α . The network presented for the induction phase which includes the Targan study was thought to be valid, and the exclusion of Targan unnecessary. The network presented for the maintenance phase was more problematic, and the ERG felt that all three analyses (two with CLASSIC II and one without CLASSIC II) should be interpreted together, but with caution. The "entire population" networks were thought to be difficult to interpret, as study populations were too heterogeneous in terms of potentially important treatment modifying effects. The anti-TNF-α failure network may have overestimated efficacy for adalimumab as primary anti-TNF-α failure patients were excluded from the adalimumab study but not the vedolizumab studies. Several studies across the evidence base excluded patients with strictures, meaning generalisation to this population is problematic, and most did not report the proportion of patients with fistulising disease, so it is unclear whether all studies were representative of UK populations in this respect. No studies included patients with CDAI>450, meaning generalisation to severe patients (if defined as CDAI 450 to 600) is problematic. Uncertainty remains around how the comparator "usual care" provided in studies compares to UK practice. No analysis for serious adverse events was provided for the anti-TNF-α naïve networks.

Additionally, for the induction networks, there were limitations with the induction schedule used in the RCTs, with fewer doses than recommended being provided, and/or assessments taking place earlier than would be done in UK practice or than stated in the licence.

Furthermore, maintenance networks were subject to potential bias from the recruitment of patients on the basis of assessment at earlier time points that would commonly be done in the UK. This means patients who take longer to respond are not represented in these trials, which may affect estimates of efficacy and/or limit generalisation to the full UK population who will be treated: the ERG do not know if these missing patients would have a differential response to treatment.

The main uncertainties in the clinical evidence relate to the duration of treatment and generalizability of the evidence to the UK population, as well as the comparability of treatments in terms of serious AEs.

1.4 Summary of cost effectiveness submitted evidence by the company

The company submitted a model-based health economic analysis as part of their submission to NICE. The analysis was undertaken from the perspective of the NHS over a 10-year time horizon. The company's analysis is presented for three populations: (1) the mixed-ITT population, which is comprised of patients who have previously received anti-TNF-α therapy and those who are anti-TNF-α naïve; (2) patients who are anti-TNF-α naïve only and; (3) patients who have previously failed anti-TNF-α therapy only. Within all three analyses, comparators include conventional non-biologic therapies (a combination of 5-ASAs, immunomodulators and corticosteroids). Other anti-TNF-α agents (infliximab, adalimumab) are included only in the analysis of the anti-TNF-α naïve subgroup; these are excluded from the analyses of the mixed-ITT and anti-TNF-α failure subgroups. All analyses include price reductions to reflect the proposed Patient Access Scheme for vedolizumab.

Within the anti-TNF-α failure subgroup, the company's model estimates the ICER for vedolizumab against conventional non-biologic therapy to be £62,903 per QALY gained within the mixed ITT population in patients with moderate to severe disease. The ICER for patients with moderate and severe CD at baseline were £21,064 and £77,382 per QALY gained respectively in the mixed ITT population.

Within the anti-TNF-α naïve subgroup, the CS estimates that vedolizumab dominates infliximab and the ICER for vedolizumab against adalimumab is £2.602 per QALY gained. However, following a request for clarification, the company reports the ICER for vedolizumab versus adalimumab to be £758,344 and infliximab versus vedolizumab to be £26,580. Based on a fully incremental analysis (constructed by the ERG), vedolizumab is subject to extended dominance. No ICER is calculated in the model for the subgroup of patients with moderate and severe disease at baseline.

Within the anti-TNF-α failure population, the company's model estimates that the ICER for vedolizumab against conventional non-biological therapy is £98,452 per QALY gained. The ICER for patients with moderate and severe CD at baseline were reported to be £55,201 and £134,330 per QALY gained respectively in this population.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The ERG critically appraised the company's health economic analysis and the model upon which this analysis is based. The ERG identified a number of concerns regarding the model structure and parameterisation of the company's model. Notably, a key concern is the derivation of the transition matrices following induction treatment. The ERG was unable to replicate the approach used by the company; and therefore cannot amend the transition matrices. This is a concern as the transition matrices are a key input parameter and are conditional on the model structure and on other input parameters. The ERG also expressed concerns that non-responders at the induction phase on conventional non-biologic treatment are assumed to remain with moderate to severe CD (and are not able to improve) and only discontinuation due to AEs is considered for biologic treatments but not discontinuation due to lack of efficacy. Similarly, the ERG expressed some concerns with efficacy data that are used, notably the comparability of data for the different biologics at the maintenance phase, and efficacy data used for conventional non-biologic treatment. The combination of all these issues leads to discrepancies between the model prediction and observed data from the GEMINI II trial.

1.6 ERG commentary on the robustness of evidence submitted by the company

1.6.1 Strengths

The company's methods for performing the clinical effectiveness systematic review were considered by the ERG to be largely appropriate. The ERG is satisfied that all relevant studies of vedolizumab (published and unpublished) were included in the CS.

1.6.2 Weaknesses and areas of uncertainty

The duration of treatment of vedolizumab in the GEMINI1 trial was 52 weeks, followed by enrolment in the ongoing GEMINI LTS study. As a result, the long-term efficacy and safety of vedolizumab is unknown. It was also noted that very few of the study sites in GEMINI II and GEMINI III were UK-based.

The ERG considered that the results of the NMA may underestimate the uncertainty in treatment effects since fixed effects models were used. There were also problems with the generalizability of findings to patients with strictures, patients with severe disease (CDAI >450) and to maintenance in patients who take longer to respond to induction therapy. Any generalisations to UK practice should be done with due consideration for the limitations of the evidence base.

The health economic model submitted by the company is subject to a number of issues which limit the credibility of the company's results. These include (a) potential omission of key aspects of the

condition such as the relapsing-remitting nature of CD, (b) simplifying and debatable assumptions regarding surgery, (c) the difficultly associated with parameterising the company's chosen structure notably the derivation of the transition matrices, and (d) debatable key structural assumptions such as assuming the same induction duration, end of scheduled maintenance at one year irrespective of achievement of remission, omission of discontinuation due to lack of efficacy and the assumptions that non-responders at the induction phase on conventional non-biologic treatment remain with moderate to severe CD (and are not able to improve). The ERG compared the model prediction with data from the GEMINI II trial and showed discrepancies between the model prediction and trial data.

The ERG is unclear whether the ICER would improve or deteriorate following amendment of the identified structural issues.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

For the sake of transparency and completeness, the ERG conducted additional scenarios analyses. The number of scenarios was limited given challenges arising from making changes to the model structure: in isolation, these had little impact on the ICER.

However, as indicated, the ERG expressed concern regarding the model structure and is not able to provide a robust ICER for vedolizumab. The ERG is unclear whether the ICER would improve or deteriorate following amendment of the identified structural issues.

Based on the company's model, vedolizumab does not appear to have an ICER below £30,000 per QALY gained in all analyses presented by the company, with the exception of patients with moderate disease at baseline for the mixed ITT population (£21,064 per QALY gained). However, the ERG is unable to confirm results from this analysis due to discrepancies in the data used and the lack of transparency regarding the derivation of model parameters. Furthermore, this analysis is compared with conventional non-biologic therapy alone and no indication of the ICER for vedolizumab compared with adalimumab or with infliximab is reported.

2 BACKGROUND

This report provides a review of the evidence submitted by the company[1] in support of vedolizumab for the treatment of adults with moderate to severe Crohn's disease (CD). It considers both the original submission received on the 2nd September 2014[1] and a subsequent response to clarification questions supplied by Takeda in batches between 7th and 17th October 2014.[2]

2.1 Critique of company's description of underlying health problem

The company[1] provided a reasonable description of the underlying health problem, which is briefly summarised in this section. The company submission (CS)[1] describes the underlying health problem as 'moderately to severely active CD', and as one of two major illnesses comprising inflammatory bowel diseases (IBD). The CS[1] describes CD as characterised by chronic relapsing inflammation that mainly affects the gastrointestinal tract and is often accompanied by abdominal pain, fever, malaise, anorexia, diarrhoea, weight loss, and clinical signs of bowel obstruction or diarrhoea with passage of blood or mucus, or both.[3,4] The CS[1] also outlines (see CS[1] pg. 39) that CD may lead to intestinal obstruction due to strictures, fistulae (often perianal), or abscesses.[4]

The description[1] includes details on how diagnosis and assessment of CD is performed (see CS[1] Section 2.1). The CS[1] states that diagnosis of CD is complex and must integrate patient history, physical symptoms, and evidence from imaging and laboratory studies.[3] Disease activity, in combination with phenotypic and endoscopic features, allows stratification of patients and selection of appropriate therapeutic strategies.[3] The Harvey Bradshaw Index is used internationally to assess disease activity in daily clinical practice and the CD Activity Index (CDAI) is the gold standard for classifying disease activity in clinical trials. The submission also explained that a CDAI score ≤ 150 indicates clinical remission and a CDAI score > 450 indicates severe disease (Yoshida et al.[5]).

The CS[1] provides prevalence estimates of CD in the UK from the NICE TA 187.[6] This was reported as approximately 50-100 per 100,000 people and that in total, it affects approximately 60,000 people in the UK.

Clinical advisors to the ERG considered the description of the underlying health problem,[1] including diagnosis and assessment to be largely appropriate and relevant to the decision problem. Clinical advisors commented that the evidence of the societal burden of CD appeared to be overly restricted to evidence from the US, with no reference to UK or European evidence.

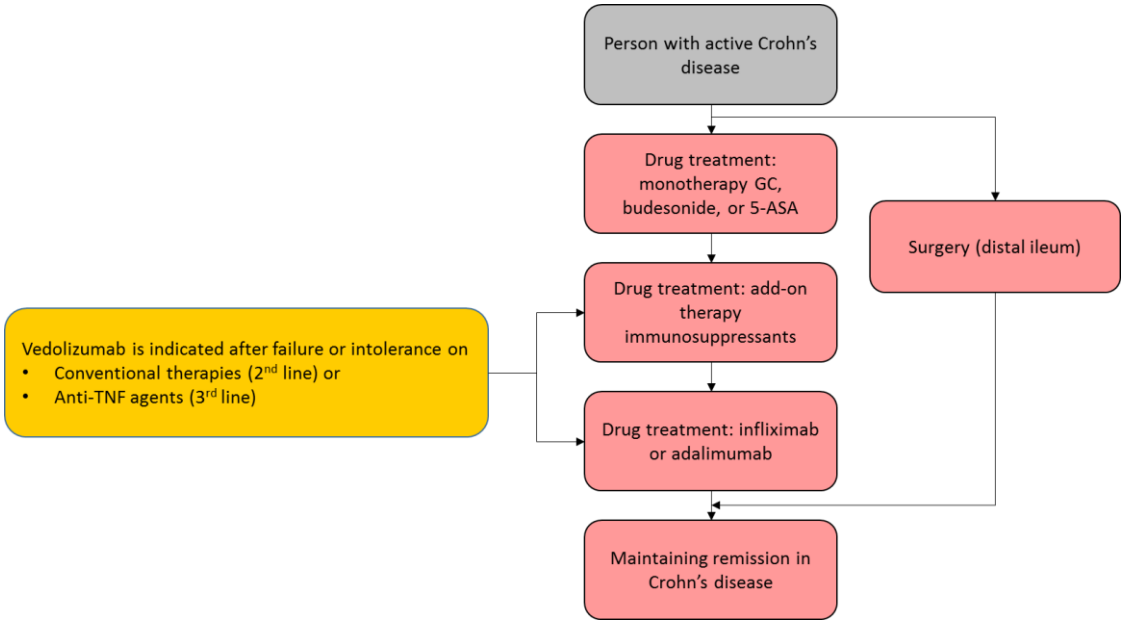
2.2 Critique of company's overview of current service provision

The company[1] states that the aim of drug treatment is to induce and maintain remission, with the optimal outcome of maintaining corticosteroid-free-remission, reducing CD complications and the need for hospitalisations and surgery. The company[1] describes (see CS[1] pg. 45) UK practice as utilising a '*standard step-up approach*' to the treatment of CD. The CS[1] refers to and summarises the CD clinical guidelines,[7] and The British Society for Gastroenterology guidelines (BSG) for the treatment of CD.[4] The management of CD with reference to the guidelines is described in the CS[1] as involving the following steps:

- the initial use of monotherapy with a conventional glucocortiscosteroid to induce remission in people with a first presentation or a single inflammatory exacerbation of CD in a 12-month period, or that budesonide and 5-ASAs can be considered in those who cannot tolerate, or a conventional glucocorticosteroid is contraindicated.
- azathioprine or mercaptopurine can be added to a conventional glucocorticosteroid or budesonide to induce remission of CD if there are 2 or more inflammatory exacerbations in a 12-month period, or the glucocorticosteroid dose cannot be tapered.
- consider adding methotrexate for those who cannot tolerate azathioprine or mercaptopurine, or in whom thiopurine methyltransferase activity is deficient.
- infliximab and adalimumab are recommended as treatment options for adults with severe active CD whose disease has not responded to conventional therapy, or who are intolerant of or have contraindications to conventional therapy. Infliximab or adalimumab should be given as a planned course of treatment until treatment failure (including the need for surgery), or until 12 months after the start of treatment, whichever is shorter.
- for maintenance of remission azathioprine or mercaptopurine as monotherapy should be offered when previously used with a conventional glucocorticosteroid or budesonide to induce remission.
- consider methotrexate to maintain remission only in people who need methotrexate to induce remission, or have tried but did not tolerate azathioprine or mercaptopurine for maintenance or these drugs are contraindicated.
- treatment with infliximab or adalimumab should only be continued if there is clear evidence of ongoing active disease as determined by clinical symptoms, biologic markers, and investigation, including endoscopy if necessary. People whose disease relapses after treatment is stopped should have the option to start treatment again.
- consider azathioprine or mercaptopurine to maintain remission after surgery in people with adverse prognostic factors otherwise consider 5-ASA treatment to maintain remission after surgery. Surgery is usually considered the final option although if CD is limited to the distal ileum surgery may be considered as an alternative early in the course of the disease.

The company[1] proposes, in line with its licence, that vedolizumab will fit in the clinical pathway as an option following failure/intolerance on conventional therapies (second-line) or tumour necrosis factor-alpha antagonist [anti-TNF-α] (third-line).

Figure 1 Proposed positioning of vedolizumab in current NICE clinical guidelines treatment path for adults with CD (reproduced from Figure 4.5.1 in CS[1] pg. 49)



The ERG and their clinical advisors agree with the broad description of management of CD, and the clinical advisors to the ERG felt that vedolizumab would more likely fit into clinical practice as a third-line treatment, after failure on existing anti-TNF-α agents and/or for people in whom ileostomy is the last option.

As described by the company,[1] two anti-TNF-α agents are currently licensed in the UK for the treatment of moderate to severe CD. These are infliximab and adalimumab. Both are recommended by NICE for use in severe CD.[7] Discussion with clinical experts indicated that these are also used in patients with moderate disease refractory to other therapies in clinical practice.

3. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

A summary of the decision problem as outlined in the final scope issued by NICE[8] and addressed in the company submission[1] is presented in Table 1.

Table 1 Decision problem as outlined in the final scope issued by NICE and addressed in the company's submission

	Decision problem outlined in final scope issued by NICE[8]	**Decision problem addressed in the CS[1]**
Population	Adults with moderately to severely active Crohn's disease in whom the disease has responded inadequately to, or is no longer responding to, either conventional therapy or an anti-TNF-α, or who are intolerant to either of them .	Adult patients with moderately to severely active Crohn's disease in whom the disease has responded inadequately to, or is no longer responding to, either conventional therapy or an anti-TNF-α, or who are intolerant to either of them (i.e. matches the population in final NICE scope)
Intervention	Vedolizumab	Vedolizumab
Comparator(s)	• Conventional treatment strategies without vedolizumab (including antibiotics, drug treatment with conventional corticosteroids alone or in combination with azathioprine, mercaptopurine or methotrexate; aminosalicylates; budesonide alone or in combination with azathioprine, mercaptopurine or methotrexate) • anti-TNF-α (infliximab and adalimumab)	• Conventional therapy, as defined in the GEMINI II and III study including concomitant use of glucocorticoids, immunosuppressive agents and mesalamine. • anti-TNF-α licensed for the treatment of Crohn's disease in the UK (infliximab and adalimumab)
Outcomes	disease activity surgery adverse effects of treatment health related quality of life.	The CS includes data on the remission and response rates but did not include data on the relapse rates. Data are on surgery are not included.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	The submission includes a model-based cost-utility analysis of vedolizumab compared against infliximab, adalimumab and conventional non-biologic therapies. The analysis was undertaken over a 10-year time horizon from the perspective of the NHS. A Patient Access Scheme (PAS) is included for vedolizumab.

	Biosimilars are not expected to be in established NHS practice at the time of appraisal and are not included as comparators	
Subgroups to be considered	If evidence allows following subgroups will be considered: • People who have not previously received an anti-TNF-α • People for whom an anti-TNF-α has failed • People for whom anti-TNF-α are not suitable because of intolerance or contraindication.	The company present analyses for 1. anti-TNF-α naïve population 2. anti-TNF-Failure population (people for whom an anti-TNF-α has failed) 3. mixed population (includes both anti-TNF-α naïve and anti-TNF-Failure subgroups)

3.1 Population

Vedolizumab has a therapeutic indication for the treatment of adult patients with moderately to severely active CD who have had an inadequate response with, lost response to, or were intolerant to conventional therapy including anti-TNF-α.[9,10]

The population described in the final NICE scope[8] was adults with moderately to severely active CD in whom the disease has responded inadequately to, or is no longer responding to, either conventional therapy or an anti-TNF-α, or who are intolerant to either of them.

The company does not provide a clear statement of the population included in the submission, but states that *"the patient population considered within this appraisal is in line with the final scope population"* (see CS[1] pg. 63).

The patient population included in the submission reflects patients included in both the GEMINI II and GEMINI III trials.[11,12] The company states that the eligibility criteria of the GEMINI II and III trials are identical (see CS[1] pg. 83). The ERG noted some differences in the inclusion criteria between the GEMINI II[11] and III[12] studies (see Section 4.2).

Demographic, baseline disease characteristics and medication history of patients in the GEMINI II (see CS[1] Table 6.3.4.1 pg. 84) and III (see CS[1] Table 6.3.4.2 pg. 86). In the GEMINI II trial,[11] patients had an overall mean age of 36.1 (standard deviation [SD] =12.1) years, were predominantly white (89.2%) as a cohort, and 46.6% were male with a mean body weight of 69.8 kg (SD=19.4). Mean duration of disease was 9.0 (SD=7.8) years and patients had a mean CDAI score of 324 (SD=69). The site of the disease was in the ileum only, colon only or both in 16.2%, 28.3% and 55.4% of patients respectively. Concomitant medications for CD included glucocorticoids only (34.2%),

immunosuppressant only (16.2%), glucocorticoids and immunosuppressant (17.0%) and no glucocorticoids or immunosuppressant (32.6%). 61.8% had received prior anti-TNF-α treatment with 57.8% having experienced ≥1 failure of an anti-TNF-α therapy.

In the GEMINI III trial[12] (ITT population), patients had an overall mean age of 37.9 (SD=12.66) years, were predominantly white (90%) and female (57%) as a cohort, with a mean body weight of 70.4 kg (SD=18.50). Mean duration of disease was 10.3 (SD=8.37) years and patients had a mean CDAI score of 307.7 (SD=54.38). The site of the disease was in the ileum only, colon only or both in 15%, 24% and 61% of patients respectively. 44% had a history of surgery and 36% a history of fistulising disease. Concomitant medications for CD at baseline included corticosteroid only (35%), immunomodulators only (16%), both (18%) or no (31%) corticosteroid and immunomodulators. 75% had received prior anti-TNF-α treatment.

In the economic section, the company presents results for three patient populations:

- a mixed population representing the intention to treat (ITT) population of the GEMINI trials (hereafter referred as the ITT mixed population),[11,12] which includes both patients who have never received an anti-TNF-α (referred as anti-TNF-α naïve) and patients who have previously been exposed to an anti-TNF-α (referred as anti-TNF-α failure),
- anti-TNF-α naïve,
- and anti-TNF-α failure, which includes both primary failure (no initial response to anti-TNF-α agents) and secondary failure (loss of response after initially responding to anti-TNF-α agents).

The ERG considered the GEMINI populations[11,11,12] included in the CS to reflect broadly the population and subgroups described in the final NICE scope.[8] However, it is unclear whether the proportions of anti-TNF-α failure patients were representative of UK norms. It should be noted that both studies had very different proportions (approximately 47% in GEMINI III and 76% in GEMINI II) of anti-TNF-α failure patients. It should be noted that the faecal calprotectin in the GEMINI trials was deemed to be high indicating that patients may had had significant active inflammation. This point may be important as anti-TNF-α treatments are thought to mediate their response through targeting inflammatory pathways and may therefore mediate a proportionately greater response in patients with severe inflammation.

As the CS[1] notes, the CDAI score is not routinely used in clinical practice and its limitations widely acknowledged, but is the standard used in clinical trials. The range chosen appears to be consistent with other trials of moderate to severe disease, though it should also be noted that patient at the higher end of the CDAI spectrum were excluded (CDAI >450). This upper range is variably described as

"severe", "very severe" or even "extremely severe" elsewhere in the literature.[13] The current NICE clinical guidelines[7] defines "severe active Crohn's" as "very poor general health and one or more symptoms such as weight loss, fever, severe abdominal pain and usually frequent (3–4 or more) diarrhoeal stools daily. People with severe active Crohn's disease may or may not develop new fistulae or have extra-intestinal manifestations of the disease". The guidelines[7] also state that *"this clinical definition normally, but not exclusively, corresponds to a Crohn's Disease Activity Index (CDAI) score of 300 or more"*. As such, it is likely that patients with the most severe disease have not been included in the GEMINI trials, and generalisation to this population may not be possible.

3.2 Intervention

The intervention described in the CS[1] matches the intervention described in the final scope issued by NICE.[8]

Vedolizumab (brand name Entyvio®) is a humanized monoclonal antibody that binds exclusively to the α4β7 integrin on gut-homing T helper lymphocytes and selectively inhibits adhesion of these cells to mucosal addressing cell adhesion molecule-1 (MAdCAM-1) and fibronectin, but not vascular cell adhesion molecule-1 (VCAM-1).[1] The gut-selective mechanism of action of vedolizumab is described in the CS[1] as being novel, with the potential to reduce adverse effects beyond the gut seen with current anti-TNF-α inhibitors (see CS[1] pg. 58).

Vedolizumab has a therapeutic indication for the treatment of adult patients with moderately to severely active Crohn's disease who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or an anti-TNF-α.[9,10]

Vedolizumab is available as a powder for concentrate for solution for infusion. Each pack contains one vial containing 300mg of vedolizumab. Based on correspondence between the company and NICE (21st August 2014), the basic NHS list price of vedolizumab is £2,050 per 300mg vial, although at the time of writing the product was not listed on the British National Formulary (BNF). The company's model includes a lower drug acquisition cost to reflect the agreed Patient Access Scheme (PAS) for vedolizumab; the price used in the model is [redacted] per 300mg vial. The agreed PAS takes the form of a simple price discount (a reduction of [redacted] of the NHS list price) for the NHS.

In adherence with the licensing of the drug,[9,10] the company[1] states the treatment regimen for vedolizumab to be the following (see CS[1] Table 1.10.1 pg. 35):

"…300 mg administered by intravenous infusion at 0, 2 and 6 weeks and then every 8 weeks thereafter. Patients with Crohn's disease, who have not shown a response may benefit from a dose of Vedolizumab at Week 10. Continue therapy every 8 weeks from Week 14 in responding patients.

Therapy for patients with Crohn's disease should not be continued if no evidence of therapeutic benefit is observed by Week 14."

It should be noted that the treatment regimen used in the company's model[55] differs from the licensing and the treatment regimen described by the company[1] in their decision problem in Section 1.10 of the CS[1] (See CS[1] pg. 35) . Further details are provided in Section 5.2.4.

In their description of the decision problem (see CS[1] pg.35), the company[1] asserts that patient will usually be treated until relapse, intolerance or discontinuation due to side effects. This differs from the company's model;[55] where reasons for discontinuation are (a) lack of primary response to induction, (b) end of scheduled maintenance (assumed to be approximately one year), (c) discontinuation due to AEs, (d) surgery and (e) death. Discontinuation following relapse (lack of efficacy) is not included. See Section 5.2.3 for further details.

In line with the licensing, the company[1] (see CS[1] Table 1.10.1 pg.36) adds that *"if therapy is interrupted and there is a need to restart treatment with Vedolizumab, dosing at every 4 weeks may be considered".* The company[1] states that in the trials, efficacy was still evident upon vedolizumab re-treatment with no apparent increase in infusion-related reactions or other adverse events.

The company[1] states (see CS[1] pg. 38) that vedolizumab will be added-on to existing therapies in clinical practice. It should be noted that in the licensing, the use of vedolizumab in conjunction with other biologics is not recommended.

In response to a request for clarification (see clarification response[2] question A5), the company states that : *"vedolizumab is a hospital-based product, typically expected to be administered in an outpatient setting by a specialist healthcare professionals experienced in the diagnosis and treatment of Crohn's disease. It is an IV product which requires reconstitution and dilution prior to administration over a 30 minute infusion. According to the SPC, patients should be monitored during and after infusion. For the first two infusions, they should also be observed for approximately two hours following completion of the infusion for signs and symptoms of acute hypersensitivity reactions. For all subsequent infusions, patients should be observed for approximately one hour following completion of the infusion".*

The company states (see CS pg. 36) that *"Vedolizumab is contraindicated in patients with active tuberculosis (TB). Before starting treatment with Vedolizumab, patients must be screened for TB according to the local practice….vedolizumab treatment should not be initiated in patients with active, severe infections until the infections are controlled, and physicians should consider*

withholding treatment in patients who develop a severe infection while on chronic treatment with Vedolizumab". It should be noted that the licensing contraindicates vedolizumab in patients with active severe infections such as tuberculosis, sepsis, cytomegalovirus, listeriosis, and opportunistic infections such as Progressive Multifocal Leukoencephalopathy (PML) or patients with hypersensitivity to the active substance or to any of the excipients.

Clinical experts to the ERG indicated that for the anti-TNF-α agents currently in use in the UK, in addition to screening for TB, screening must also be undertaken for HIV, Hepatitis B and C and for heart conditions, and that treatment may be problematic in those receiving a flu vaccine. Clarification was requested from the company[2] regarding whether screening is also required for people taking vedolizumab for viruses such as HIV, Hepatitis B and C and cardiac conditions; the company[1] believes that this was not necessary for vedolizumab (see clarification response[2] question A3).

In adherence with the licensing,[9,10] the company states that patients should be monitored closely for infections before, during and after treatment, monitor for emerging neurological signs/symptoms and monitor for signs and symptoms of acute hypersensitivity reactions with respect to administration (infusion-related reactions). Patients receiving vedolizumab should also be monitored for PML and new onset or worsening of neurological signs and symptoms.

Finally, it should be noted that the licensing[9,10] mentions that no vedolizumab clinical trial data are available for patients previously treated with natalizumab or rituximab and that caution should be exercised when considering the use of vedolizumab in these patients.

3.3 Comparators

The final NICE scope[8] describes appropriate comparators to be: established clinical management without vedolizumab, which may include antibiotics, drug treatment with conventional corticosteroids alone or in combination with azathioprine, mercaptopurine or methotrexate; aminosalicylates; budesonide alone or in combination with azathioprine, mercaptopurine or methotrexate and anti-TNF-α agents (infliximab, adalimumab).

The CS[1] states that included comparators were conventional therapy (as defined in the GEMINI II[11] and III[12] studies and used in UK clinical practice based on the UK IBD audit[14]) and TNF-α inhibitors licensed for the treatment of CD disease in the UK (infliximab, adalimumab).

The main comparator (used in all three populations in the company's model:[15] mixed ITT, anti-TNF-α naïve and anti-TNF-α failure subgroup) was described by the company as being standard care, consisting of 5-ASAs, corticosteroids and immunosuppressants, reflecting baseline treatments in CD

in the GEMINI trials.[11,12] Patients in the GEMINI trials[11,12] received vedolizumab or placebo alongside conventional treatments as background therapies. In the company's model,[15] a comparison is presented against anti-TNF-α agents (adalimumab, infliximab) for the anti-TNF-α naïve subgroup only (they are excluded from both the analyses of the mixed ITT and anti-TNF-α failure subgroups).

3.4 Outcomes

The final NICE scope[8] specified outcomes for consideration as follows:

- disease activity
- Surgery
- adverse effects of treatment
- health-related quality of life

The CS[1] states that the outcomes considered were in line with those specified in the final NICE scope.[8] The company presented data on response (defined as a reduction in CDAI score of 70 points or more), enhanced clinical response (drop in CDAI score of 100 points or more), remission (CDAI score ≤ 150), adverse events and discontinuation due to adverse events and health-related quality of life. Data on relapse rates were not presented in the CS.[1] Furthermore, data on surgery were not presented in the CS, but surgery was included as an outcome in the health economic model. In response to a request for clarification (see clarification response[2] question B51), the company presented data on the number of patients who underwent bowel surgery randomised to the maintenance phase of the GEMINI II trial.[11] However, this outcome was not considered in the network meta-analysis.

3.5 Other relevant factors

No equity issues were highlighted in the CS.[1] Discussion with one of our clinical advisor indicated that ethnic minority patients' access to biologics is much reduced when compared to white British (though we have found no empirical evidence to support this view). The reasons for this are unclear, but could be associated with perceived safety, ingredients of the drugs, relationship with health professionals and communication barriers. Another clinical advisor noted that creation of a stoma or receipt of any surgery may be problematic for some people form particular cultures and backgrounds. As such, a treatment that could delay or reduce the risk of such procedures could be important in terms of equity.

4. CLINICAL EFFECTIVENESS

This section presents a review of evidence relating to the clinical effectiveness of vedolizumab in adult patients with moderately to severely active CD. Section 4.1 presents a critique of the company's conduct of the systematic review[1] and Section 4.2 provides a summary of the clinical effectiveness results (efficacy and safety) and critique of included vedolizumab trials. Section 4.3 and 4.4 provide critiques and summaries of the trials included in the network meta-analysis, and the methods and results of the network meta-analyses (NMA) included within the CS.[1] Finally, Section 4.5 provides the conclusions of the clinical effectiveness section.

4.1 Critique of the methods of review(s)

One systematic review was included in the CS,[1] and was described in an accompanying 498 page document[16] submitted alongside the CS[1] (referred as the Takeda on file document[16]). The CS[1] included a brief summary of the methods and findings from this review. Given the time constraints, the ERG only looked at the information provided within the CS[1], and referred back to the Takeda on file document[16] when details were lacking from the CS[1]. The review was commissioned by the company. The methods are critiqued in the following sections.

A systematic search to identify existing reviews on which to base the network meta-analysis was not performed, though searches to identify reviews as sources of additional trials were. As data relating to vedolizumab trials had only recently been published, this seems logical, though theoretically an existing review could have been updated.

4.1.1 Searches

Main searches

The search strategy was newly developed for the purposes of the STA, and was not based on any previous published search strategies.

The original searches were conducted on 9 April 2013, followed by an update search on 12 February 2014. In both instances, the following databases were searched:

- MEDLINE/Medline (R) In-Process (via PubMed)
- Embase (via Elsevier)
- The Cochrane Library (CDSR, DARE, CENTRAL, HTA, NHSEED) (via Wiley)

No date limit was applied to the original searches. The update searches were limited to material published from 1 April 2013 onwards. In all instances, no language restrictions were applied.

Additional searches (Internet)

Searches of the following websites were performed in addition to the main database searches:

- ClinicalTrials.gov (for the original review and update searches)
- World Health Organization's International Clinical Trials Registry Platform Search Portal (http://apps.who.int/trialsearch/) (for the update searches only)

Despite its inclusion in the protocol, the United European Gastroenterology website was not searched due to technical problems with the website.

The report does not mention any restrictions being applied to these additional searches, and the dates on which the searches were performed are not provided within the CS.[1]

Reference tracking

The reference lists of identified systematic reviews and meta-analyses were checked for additional relevant publications.

Search critique

Overall the search strategies for clinical effectiveness are deemed by the ERG to be appropriate, although the reporting of the search strategies and the exact methods used to search for clinical data are not ideal. A complete list of issues is provided below:

1. The search filter used to gather RCT and non-RCT evidence is unfamiliar to the ERG and has not been referenced. After further clarification was sought,[2] it appears that the filters have been constructed based on the company's previous experiences of conducting reviews and recommendations from the project team (including Takeda and RTI Health Solutions). The search filters used include some published strings from the well-known InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Resource. This is deemed by the ERG to be acceptable practice, as it demonstrates a systematic approach to searching the evidence base.
2. The above mentioned search filter was also used in Cochrane, which is not standard practice since the different elements of The Cochrane Library effectively act as filters. However, this will not have impacted negatively on the results and their relevance to the search topic.
3. For the free-text elements of the search, word variations are given a new line in the search strategy, rather than being incorporated as part of a single, truncated free-text search. This is a minor point and will not have affected the number of results retrieved.

4. The systematic review element of the strategy is excluded from the main results in the final line – it would appear that this evidence was incorporated as part of the review, but the reporting of the searches makes this initially unclear.
5. Reference tracking (i.e. checking of reference lists for additional relevant publications) was performed on the five most up-to-date and robust systematic reviews and meta-analyses, as identified during screening for each review. Whilst this is deemed by the ERG to be good practice, neither the references for the five systematic reviews/meta-analyses or the results of the reference tracking are provided in the report, and so it is not possible to determine the quality of this element of the searches.
6. Overall, the way in which the searches are discussed and reported within the report is not always clear, and it is difficult and time-consuming to ascertain exactly how the searches were conducted. Referencing of search filters and more detailed explanation of how the searches were conducted would have been beneficial to the ERG.

The ERG believes that the issues identified above did not impact on the overall quality of the searches, which were deemed to be sufficient.

4.1.2 Inclusion criteria

The methods used to select relevant primary literature were mostly of a good standard. Two reviewers independently screened titles, and a third reviewer was consulted in cases of uncertainty.

Study selection was split over two stages, with separate inclusion/exclusion criteria. Criteria for level 1 screening are presented in Table 2, and criteria for level 2 screening are presented in Table 3. It is not entirely clear why this was necessary and makes the ERG's job of interpreting them more difficult. The reasons why each study was excluded from the review at full-text stage were missing from the CS[1] and the Takeda data on file document,[16] making it difficult to audit the selection process.

In addition, there are several points of lack of clarity and detail within the PICOS framework that reduce the quality of the review.

- Population

The population is defined as "*Patients with Crohn's disease (both biologic treatment-naïve and biologic treatment-experienced)*", whereas it appears that only studies in moderate to severely active patients have been included in the review and network meta-analysis. The includable population should have been more clearly defined. The ERG requested clarification[2] of the inclusion criteria (see clarification response[2] question A49), and the company agreed this was an omission, but assured the ERG that only studies in moderate to severe Crohn's patients were included.

- Intervention
  - Doses were not defined in the inclusion criteria.
  - Surgery was listed as an includable intervention in the CS[1], but this is not listed in the NICE scope.[8] However, no further data for these studies is in fact presented.
- Comparators
  - No comparator inclusion/exclusion details are provided, only interventions. If all comparators were eligible for inclusion, this should have been stated.
- Study design
  - Prospective studies with more than 1 treatment arm were includable, but none appeared in the list of included studies. Reasons for this were requested by the ERG in the clarification letter (see clarification response[2] question QA9) and the company indicated that two studies had been identified but excluded from the NMA. The ERG accepts this was for valid reasons.

Table 2 Table of inclusion and exclusion criteria used for level 1 screening in the CS[1] (reproduced from Table 6.2.1.1 pg.67 CS[1])

Criteria	Included	Excluded
Study design	• Randomised, controlled, prospective clinical trials • Non-randomised, controlled clinical trials • Long-term follow-up studies (e.g., open-label follow-up of randomised clinical trials) • Prospective observational studies (e.g., Phase 4 studies) • Systematic reviews and meta-analyses[a]	• Single-arm clinical trials • Preclinical studies • Phase 1 studies • Pilot studies • Prognostic studies • Retrospective studies • Case reports • Commentaries and letters (publication type) • Consensus reports • Non-systematic reviews
Population	Patients with CD (both treatment-naïve and treatment-experienced)	Patients who do not have CD
Intervention	Biologics search: [b] • Vedolizumab • Certolizumab (Cimzia) • Natalizumab (Tysabri, Antegren) • Infliximab (Remicade) • Adalimumab (Humira) Additional search: • Surgery (of any type)	Studies that do not investigate one of the biologics of interest in at least one of the arms
Outcomes	None	None: the studies were not excluded on the basis of outcomes at the level 1 screening process

[a] Systematic reviews and meta-analyses were used for identification of primary studies.

[b] We have extracted and presented information on biologics relevant for this appraisal, i.e., vedolizumab, infliximab, and adalimumab only. Natalizumab and certolizumab have not been approved for use in CD in the UK.

Table 3 Table of inclusion and exclusion criteria used for level 2 screening in the CS[1] (reproduced from table 6.2.1.2 pg.68 CS[1])

Criteria	Included	Excluded
Study design	• Randomised, double-blind clinical trials • Randomised, open-label clinical trials • Randomised, open-label follow-up studies • Prospective studies with more than 1 treatment arm	Same as the criteria for level 1 (Table 2), with the addition of systematic reviews and meta-analyses:
Population	Patients with CD (both biologic treatment-naïve and biologic treatment-experienced)	Patients who do not have CD
Intervention	Same as the criteria for level 1 (Table 2)	Same as the criteria for level 1 (Table 2)
Outcomes[a]	• Clinical response • Sustained clinical response • Durable clinical response • Durable clinical remission • Mucosal healing All of the above with timing and definition • Safety outcomes (AEs, SAEs, specific AEs of interest) • Quality-of-life outcomes, including IBDQ • Surgery • Hospitalizations • Change in CDAI from baseline • Mean CDAI at baseline and each subsequent visit • Amended search for studies of surgery: o Any clinical outcomes as noted above o Any surgical outcomes, including complications	None For IBD articles, exclude if IBD results not broken down into CD and ulcerative colitis (UC)

AE = adverse event; CDAI = Crohn's Disease Activity Index; IBD = inflammatory bowel disease; IBDQ = Inflammatory Bowel Disease Questionnaire; SAE = serious adverse event

[a] Outcomes to be included were finalized following review of the clinical study reports. As definitions of response, remission, and mucosal healing, along with the timings of outcome measurement, may differ between studies, heterogeneity of reporting was considered during data extraction

The ERG identified one study from an existing systematic review[17] which had not been included in the systematic review reported in the CS,[1] which should have been according to the level 2 inclusion and exclusion criteria listed on pg.68 of the CS[1]and reproduced in Tables 2 and 3:

- a maintenance trial[18] extension of Targan et al.[19] (which was an induction trial using infliximab). The maintenance dose in this trial was 10mg/kg, whereas the licensed dose in the UK is 5mg/kg. However, according to the inclusion criteria it should have been included (no definition of dose was provided in the selection criteria).

Whilst this study should technically have been included, according to the level 2 inclusion criteria reported in the CS[1], its relevance to the decision problem is marginal and its exclusion probably appropriate. This suggests that the problem with the review is likely to be documentation rather than execution, that is the level 2 inclusion criteria appear to be incomplete in the CS[1].

4.1.3 Critique of data extraction

The extraction of data has been performed in a transparent manner; with the exception that it is unclear whether data-checking of any form was conducted. The Takeda data on file document[16] states (see Takeda data on file[16] pg. 15) that ***"****data were extracted by one researcher and quality-checked by an independent reviewer….*" From this description it is not clear whether the intended meaning is that a) all data were checked, b) a sample were checked, or in fact c) that no data were checked, as appendix E (from the Takeda data on file[16]) relates to quality assessment of studies included, not data checking of extracted data. In the case of b) it is unclear what would have been done had a high rate of errors been identified. As such, there is the potential that data extraction errors have not been minimised through high quality methods. Indeed, during the course of the assessment, the ERG identified that data had been missed from one study,[20] which is described in section 4.3.

4.1.4 Quality assessment

The quality assessment appears to follow recommendations given in the NICE Specification for company/sponsor submission of evidence template,[21] as the Takeda data on file document[16] lists the appropriate quality assessment items as suggested by NICE as a minimum. However, no narrative synthesis of these assessments is given, and no attempt has been made to integrate the quality assessment into the reporting of the findings or into the reporting of the network meta-analysis. Although quality has been assessed, the overall impact of the quality of the included studies on the results is unclear. Whilst the Takeda data on file[16] statement quoted above appears to relate to data extraction in general, a similar sentence is used in the CS[1] in relation to quality assessment. Therefore, it would appear that checking of quality assessment of studies (at least) was conducted.

4.1.5 Evidence synthesis

Neither the Takeda data on file document[16] nor the CS[1] describes which analyses and analysis methods were pre-planned. As such, there is theoretically a high risk that bias may have been introduced through ad-hoc analyses and methodologies.

The company undertook a narrative synthesis of the evidence for vedolizumab; however, no explicit details were provided in the CS [1] on how this approach was undertaken. Ideally, a narrative synthesis approach should be justified, rigorous (i.e. describe results without being selective or emphasising some findings over others) and transparent to reduce potential bias. The ERG note that not all outcomes detailed in the CSR for each trial[22,23], such as time to treatment failure were presented in the CS[1], therefore critique of these data were not possible. Despite the lack of transparency regarding the methods adopted, the ERG acknowledges that the narrative synthesis approach undertaken by the company was acceptable for the two main trials.

An NMA was used to perform indirect comparisons of vedolizumab, adalimumab, infliximab and placebo. In the CS[1] the company presented NMA for the outcomes of clinical response (drop in CDAI ≥70), enhanced clinical response (drop in CDAI ≥100), clinical remission (defined as a CDAI score of ≤150 points) and discontinuations due to adverse events (AEs). Networks for the sub-populations, anti-TNF-α naïve and anti-TNF-α failure, and in the induction and maintenance phases separately were presented where data allowed. The ERG will focus their critique on the outcomes presented in the CS[1], although it should be noted that many more outcomes were presented in the Takeda data on file document.[16] The ERG considers the company's outcome selection to be relevant and appropriate.

For the statistical analysis (see CS[1] Section 6.7), the company undertook separate NMAs for the anti-TNF-α naïve and anti-TNF-α experienced/failure subgroups and the ITT population. Induction phase data and maintenance phase data were synthesised separately. Clinical response and remission were modelled separately using a logistic model. The company suggests that Bayesian and Frequentist fixed and random effects models were conducted; although not all models are reported within the CS.[1] The models are reported in Section 6.7 of the CS.[1]

4.2 Critique of trials of the technology of interest, their analysis and interpretation

4.2.1 Studies included in/excluded from the submission

The company[1] presented a systematic review of the clinical effectiveness and safety of vedolizumab for the treatment of moderately to severely active CD in adults who were naïve to anti-TNF-α and those who are intolerant of, or whose disease has an inadequate response or loss of response to conventional therapy or anti-TNF-α. The systematic review aimed to assess the best available evidence to evaluate the efficacy and safety of all biologics in patients with moderate to severe CD to inform a NMA. A review of vedolizumab only was not performed. The CS[1] included a description of a separate search for surgery, although this is not relevant to this appraisal as it is not listed as a comparator. The CS[1] documents that an initial search was undertaken in April 2013, with update searches performed on February 12th 2014 and limited to publications from April 1st 2013. These searches had a global remit to assess vedolizumab against certolizumab and natalizumab in countries where they are licensed for use, as these biologics are not licensed for use in the UK they are not relevant to this assessment and were therefore excluded at sifting stage. The company's Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (formerly QUOROM) flow diagram[1] relating to the literature searches does not conform exactly to the PRISMA statement flow diagram (http://www.prisma-statement.org/statement.htm). Despite minor discrepancies, the flow diagram (see CS[1] pg. 71) represents the identification and selection of relevant biologic therapies for the treatment of CD (i.e. for the systematic review of vedolizumab and for the systematic review/ potential NMAs incorporating infliximab and adalimumab indicated for the treatment of moderate to severe CD using indirect comparisons) and appears to be an adequate record of the literature searching and screening process. For clarity, a separate PRISMA flow diagram for each of the reviews would have been beneficial as it would aid the transparency of the identification and selection process for each of the reviews. A separate flow diagram relating to a review of surgery is presented in the CS[1] (page 72) but is not presented here as it is outside the scope of the evaluation.

The PRISMA flow diagram indicates of a total of 1,648 potentially relevant records were identified, 1,491 were excluded at title/abstract sift (level 1) and 81 articles were excluded at full paper (level 2). Subsequently, according to the CS[1]18 studies were included in the NMA, of which 10 were relevant to this appraisal. However, this includes all biologics and citations from both the original and update searches. These 81 exclusions are separated into categories indicating broad reasons for exclusion for the original and update searches respectively. These categories included study design (original search n=15; updated search n=34), population (original search n=4; updated search n=0), intervention (original search n=5, updated search n=6), and outcomes (original search n=12, updated search n=5). Excluded studies relating to the systematic review of vedolizumab are not documented in the CS[1]. However, reasons for excluding studies from the network meta-analysis are provided in Takeda data

on file[16] (see Takeda data on file[16] Table 11 on pg. 33) and therefore excluded studies which relate to vedolizumab can be identified from this information.

The ERG identified from the Takeda data on file[16] (see Takeda data on file[16] Table 11 on pg. 33) that one study of vedolizumab was excluded from the systematic review. Feagan et al.[24] was a Phase 2 randomised, double-blind, placebo controlled study of vedolizumab. According to the Takeda data on file[16] it was excluded as vedolizumab was administered at low dose and at various dosing regimens based on weight. Patients were randomised to receive vedolizumab 2.0 mg/kg (n = 65), vedolizumab 0.5 mg/kg (n = 62), or placebo (n = 58) by intravenous infusion on days 1 and 29. As the license indication[9,10] is based on a fixed dosing schedule (300mg at zero, two and six weeks and then every eight weeks thereafter), the ERG agrees that this study was appropriately excluded from the company's review.

For the systematic review and NMA of other biologic therapies, seventeen further potential citations were excluded. This includes studies on certolizumab and natalizumab which although included in the NMA were not part of the decision problem. In five studies (6 citations) the patient population was considered not of interest (Hyams et al.,[25] Veeremans et al.,[26] Present et al.,[27] Van Assche et al.,[28] Sands et al.,[29] Regueiro et al.[30]). In four studies there was no placebo arm (Mazzouli et al.,[31] Lichtenstein et al.,[32] Colombel et al.,[33] Bhatia et al.[34]). In three studies the drug combination was not of interest (Duan et al.,[35] Lemann et al.,[36] D'Haens et al.,[37]). One maintenance study only included 26 week data (Schreiber et al.[38]), one study only included preliminary analyses (Sands et al.[39]), a further study did not include a suitable time point for analysis (Panaccione et al.[40]). The ERG agrees that the design and context of these studies were not suitable for inclusion in the NMA. One final study in which patients were re-randomised into maintenance phase based on remission status (Sandborn et al.[41]) was reported in Takeda data on file[16] (see Takeda data on file[16] Table 11 on pg. 33) as excluded from the NMA for this reason, although it was in fact included in a secondary analysis.

Main evidence for vedolizumab: GEMINI II[11] and GEMINI III trials[12]

The CS[1] included two Phase III, multicentre, randomised, double-blind, placebo-controlled trials designed to evaluate the efficacy and safety of vedolizumab as induction therapy (dosing at weeks 0 and 2 with assessment at week 6 in GEMINI II,[11] and dosing at weeks 0, 2 and 6 with assessment at weeks 6 and 10 in GEMINI III[12]) and maintenance therapy (weeks 6 to 52 in GEMINI II[11] only) in adults with moderately to severely active CD who had an inadequate response to, loss of response to, or intolerance to immunomodulators or anti-TNF-α. It is noteworthy that although the studies were designed against placebo, conventional therapies (5-ASAs, corticosteroids, immunomodulators, antibiotics, probiotics, and antidiarrheal) were concomitantly administered to patients in both treatment arms. However, as noted in the European Public Assessment Report (EPAR),[10] the lack of an anti-TNF-α compound comparator arm represents a limitation of the studies.

The GEMINI II trial[11] included patients who were naïve to anti-TNF-α, and patients who had an inadequate response to, loss of response to, or intolerance to immunomodulators or anti-TNF-α.

The primary analysis in the GEMINI III trial[12] focussed on people for whom an anti-TNF-α has failed (i.e., an inadequate response to, loss of response to, or intolerance of ≥1 anti-TNF-α). A secondary analysis evaluated an overall population which included patients who were naïve to anti-TNF-α, and pre-specified exploratory analyses examined the group naïve to anti-TNF-α.

GEMINI II Trial[11]

An overview of the induction and maintenance phases in the GEMINI II trial[11] is provided in Figure 2. Although the study was designed to compare vedolizumab with placebo, conventional therapies (5-ASAs, corticosteroids, immunomodulators, antibiotics, probiotics, and antidiarrheal) were concomitantly administered to patients in both treatment arms.

The GEMINI II trial[11] was conducted at 285 medical centres in 39 countries from 2008 to 2012. Of the 285 sites, enrolment at 9 sites was discontinued because of concerns about the ability to fully comply with good clinical practice. At 6 of these sites, enrolment was later resumed. Enrolment was also permanently discontinued at country level in India due to concern for patient safety affecting 8 sites. This arose as serious adverse events (SAEs) led to 2 deaths at sites in India. The cause of death attributed by the principle investigators at each site, were study-related 'septic shock', and study-related 'sepsis' (further details are provided in the supplementary appendix to Sandborn et al.[11]).

Patients eligible for inclusion in GEMINI II[11] were required to be aged between 18 to 80 years with moderate to severely active CD as determined by having;

- CD for ≥3 months,

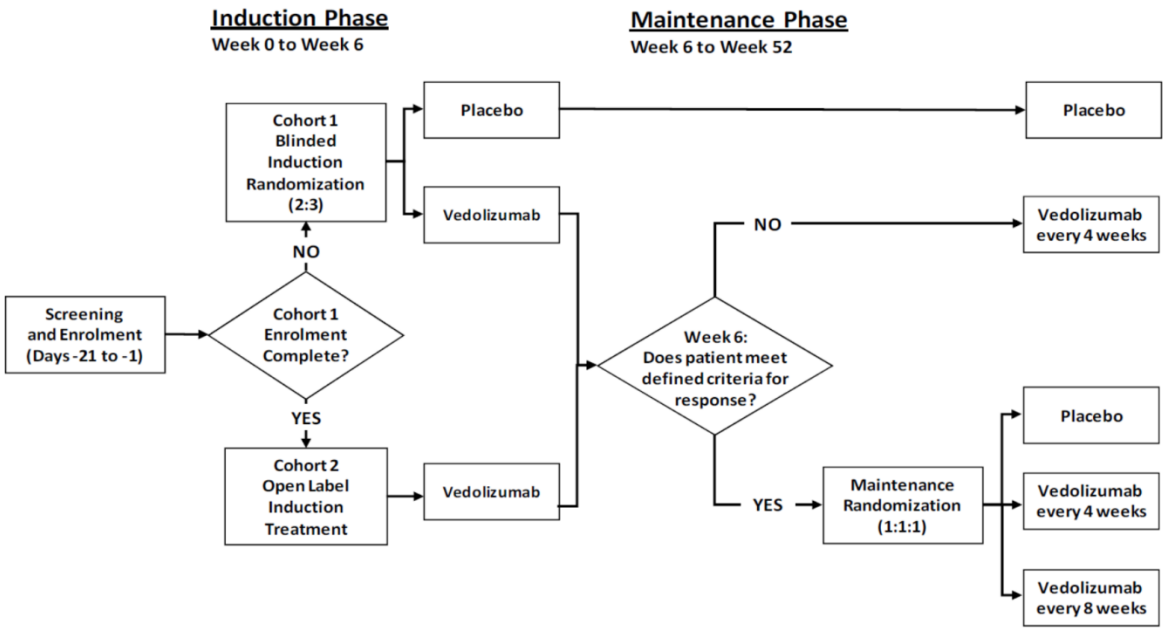
- a CDAI score of 220 to 450,
- the presence of one of the following:
  - a serum C-reactive protein (CRP) >2.87 mg/L during the screening period,
  - colonoscopic findings showing ≥3 large ulcers or ≥10 aphthous ulcers,
  - or faecal calprotectin concentrations ≥250 mcg/g of stool
- plus evidence of ulcers on computed tomography or magnetic resonance enterography, small bowel radiography, or capsule endoscopy,
- and has demonstrated, over the previous 5-year period, an inadequate response to, loss of response to, or intolerance of at least 1 of the following:
  - immunomodulators (including oral azathioprine, or methotrexate);
  - anti-TNF-α (including infliximab, adalimumab, or certolizumab pegol),
  - or for patients outside of the US, corticosteroids.

The key exclusion criteria [22] were;

- severe gastrointestinal symptoms requiring surgical treatment and patients with extensive surgeries (including abdominal abscess, extensive colonic resection, subtotal or total colectomy, history of > 3 small bowel resections or diagnosis of short bowel syndrome, ileostomy, colostomy, or known fixed symptomatic stenosis of the intestine),
- evidence of or treatment for *C. difficile* infection or other intestinal pathogen within 28 days prior to enrolment,
- history or evidence of adenomatous colonic polyps that have not been removed,
- history or evidence of colonic mucosal dysplasia,
- infectious diseases such as chronic hepatitis B or C infection, active or latent TB,
- or laboratory abnormalities during the screening period.

Figure 2 shows how patients were randomised to the induction and maintenance phase of GEMINI II.

Figure 2 Overview of the induction and maintenance phase in the GEMINI II trial[11] (reproduced from figure 6.3.2.1 in CS[1] pg. 77)



The GEMINI II trial[11] assessed the efficacy of vedolizumab compared with placebo in both an induction phase and in a maintenance phase. Patients were therefore randomised at two different time points within the trial.

In the induction study, 368 patients were randomised in a 3:2 ratio to receive intravenous (i.v.) vedolizumab (300mg) or placebo (as saline) at week 0 and week 2 (Cohort 1), with two stratification factors: (1) concomitant use of glucocorticoids and (2) by concomitant use of immunosuppressive agents or prior use of anti-TNF-α or both. The proportion of patients with prior anti-TNF-α exposure was limited to 50% to ensure that the efficacy of vedolizumab could be evaluated in patients who are naïve to anti-TNF-α. In order to fulfil sample size requirements for the maintenance study, an additional 748 patients were enrolled in an open-label group (Cohort 2), which received the same active induction regimen (vedolizumab 300mg i.v. at week 0 and 2) given in the blinded study (Cohort 1).

The two primary endpoints in the induction trial phase were enhanced clinical response at week 6 (defined as $\geq$100-point decrease in CDAI score), and clinical remission at week 6 (defined as a CDAI score of $\leq$150 points). The secondary end point was the mean change in C-reactive protein levels from baseline to week 6.

In the maintenance study, patients from both cohorts (Cohort 1 and Cohort 2) who had a clinical response (defined as $\geq$70 point decrease in the CDAI score) to vedolizumab at week 6 (n=461) were

randomised in a 1:1:1 ratio to double-blind treatment with vedolizumab (300mg administered intravenously) every 8 weeks (with placebo administered every other visit to preserve blinding), vedolizumab every 4 weeks or placebo every 4 weeks for up to 52 weeks. Randomisation was stratified by three factors: (1) cohort, (2) concomitant use of glucocorticoids, and (3) concomitant use of immunosuppressive agents or prior use of anti-TNF-α. Patients in the induction study; in both vedolizumab cohorts not having clinical responses at week 6 continued to receive vedolizumab every 4 weeks and were followed through to week 52. Patients who received placebo in the induction phase continued to receive placebo and followed in a similar fashion irrespective of response at week 6.

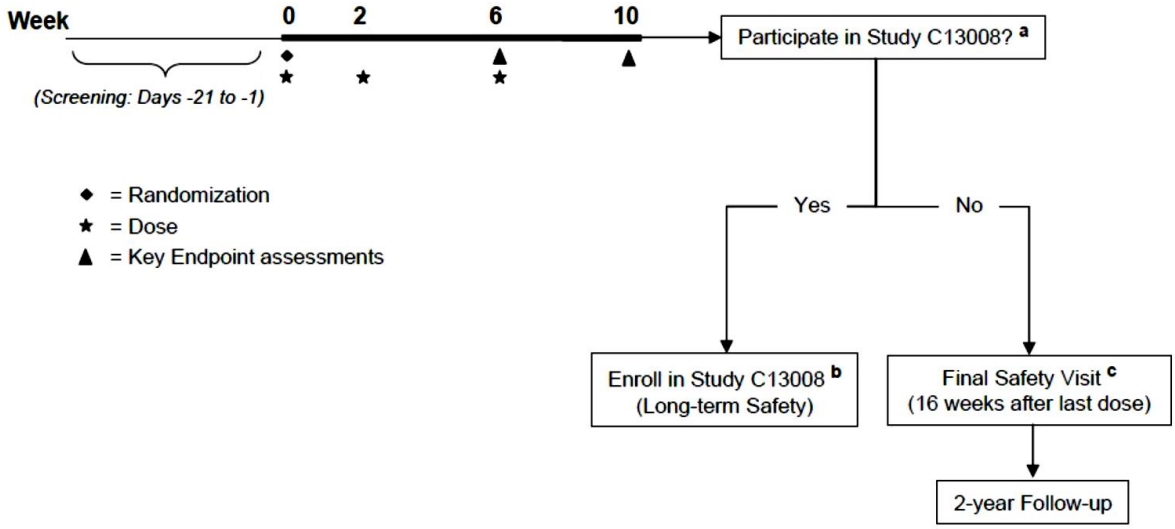
The primary endpoint for the maintenance trial phase was clinical remission at week 52. Secondary outcome measures included enhanced clinical response (defined as a 100 point reduction or more from baseline in CDAI score) at 52 weeks, glucocorticoid-free remission at week 52 in patients receiving glucocorticoids at baseline, and durable clinical remission (defined as clinical remission at ≥ 80% of study visits, including the final visit). The proportion of patients meeting these end points was analysed.

A summary of the study design and population characteristics is provided in Table 4.

GEMINI III Trial[12]

An overview of the GEMINI III trial[12] which only included an induction phase is provided in Figure 3.

Figure 3 Overview of the GEMINI III (induction only) trial[12](reproduced from Figure 6.3.2.2 in CS[1] pg. 78)



a After completing the Week 10 assessments, patients were eligible to enroll in Study C13008 if study drug was well tolerated and no surgical intervention for CD occurred or was required.

b Eligible patients were to have enrolled in Study C13008 within 5 weeks after their final dose in this study.

c Patients who were not eligible for or declined enrollment in Study C13008 were to return for an on-study Final Safety visit (Week 22, or 16 weeks after last dose) and complete the 2-year follow-up.

The GEMINI III trial[12] was conducted at 107 medical centres in 19 countries from 2010 to 2012. Patients eligible for inclusion in GEMINI III were required to be aged between 18 and 80 years with moderate to severely active CD as determined by having;

- CD for ≥3 months,
- a CDAI score of 220 to 400,
- the presence of one of the following:
    - a serum CRP >2.87 mg/L during the screening period,
    - colonoscopic findings showing ≥3 large ulcers or ≥10 aphthous ulcers,
    - or faecal calprotectin concentrations ≥250 mcg/g of stool plus evidence of ulcers on computed tomography or magnetic resonance enterography, small bowel radiography, or capsule endoscopy,
- and has demonstrated, over the previous 5-year period, an inadequate response to, loss of response to, or intolerance of at least 1 of the following:
    - immunomodulators (including oral azathioprine or methotrexate);
    - anti-TNF-α (including infliximab, adalimumab, or certolizumab pegol),
    - or for patients outside of the US, corticosteroids.

The key exclusion criteria were;

- severe gastrointestinal symptoms requiring surgical treatment and patients with extensive surgeries (including abdominal abscess, extensive colonic resection, subtotal or total colectomy,
- history of > 3 small bowel resections or diagnosis of short bowel syndrome, ileostomy, colostomy, or known fixed symptomatic stenosis of the intestine),
- evidence of, or treatment for, *C. difficile* infection or other intestinal pathogen within 28 days prior to enrolment,
- history or evidence of adenomatous colonic polyps that have not been removed,
- history or evidence of colonic mucosal dysplasia,
- infectious diseases such as chronic hepatitis B or C infection, active or latent TB,
- and laboratory abnormalities during the screening period.

A summary of the study design and population characteristics is provided in Table 4.

Four hundred and sixteen patients were enrolled. 315 patients had a previous inadequate response to, loss of response to, or intolerance of, one or more anti-TNF-α, and 101 patients were naïve to anti-TNF-α. Patients were randomly assigned to receive i.v. vedolizumab (300mg) or placebo (as saline) at week 0, week 2, and week 6, with three stratification factors: (1) the presence or absence of previous anti-TNF-α failure, (2) concomitant use or non-use of glucocorticoids and (3) by concomitant use or non-use of immunosuppressive agents.

The primary endpoint in the GEMINI III trial[12] focussed on people for whom an anti-TNF-α has failed (i.e., an inadequate response to, loss of primary response to, loss of secondary response to, or intolerance of ≥1 anti-TNF-α) (pre-specified to be 75% of the recruited population), and was the proportion of patients in clinical remission (CDAI score ≤150 points) at week 6. A secondary analysis evaluated an overall population which included patients who were naïve to anti-TNF-α, and pre-specified exploratory analyses examined the group naïve to anti-TNF-α. Secondary endpoints were the proportion of patients in the overall study population (including the additional 25% anti-TNF-α naïve) in remission at week 6; the overall and failure population in remission at week 10; the overall and failure population with remission at both week 6 and 10 (sustained clinical remission); and failure population with an enhanced clinical response (defined as a 100 point reduction or more from baseline in CDAI score) at week 6.

Table 4 Characteristics of included studies (see CS[1] pg. 79-82)

Study	Location (sites)	Design	Population	Interventions (n=randomised)	Comparator	Primary outcome measures	Duration
GEMINI II[11] CSR13007[22]	285 medical centres in 39 countries	Phase III randomised, double-blind, placebo-controlled, induction and maintenance trial	Patients aged 18 to 80 years with moderate to severe active CD (defined as; CD for ≥3 months CDAI score 220-450), inadequate response to, loss of response to, or intolerance of at least 1 of conventional therapy or anti-TNF-α .	Induction phase Vedolizumab (IV) 300 mg week 0 and 2 Cohort 1 (n=220), Cohort 2 (n=747) Maintenance phase Vedolizumab (IV) 300mg every 8 weeks (n=154) , every 4 weeks (n=154)	Induction phase Placebo (IV) at week 0 and 2 (n = 148) Maintenance phase Placebo (IV) every 4 weeks (n=153)	Induction Phase Clinical remission at week 6 (CDAI score of ≤150) Clinical response at week 6 (≥100 decrease in CDAI score) Maintenance Phase Clinical remission at week 52	Induction phase 6 weeks Maintenance phase 52 weeks
GEMINI III[12] CSR 13011[23]	107 medical centres in 19 countries	Phase III randomised, double-blind, placebo-controlled, induction trial	Patients aged 18 to 80 years with moderate to severe active CD (defined as; CD for ≥3 months CDAI scores 220-400.	Vedolizumab (IV) 300 mg at 0, 2, and 6 weeks (n = 416)	Placebo (IV) at 0, 2 and 6 weeks (n = 207)	Clinical remission at week 6 (CDAI score of ≤150) in patients with prior anti-TNF-α failure.	10 weeks

Despite the CS[1] stating that the eligibility criteria of the GEMINI II[11] and GEMINI III[12] trials were identical the ERG did note some key differences between the two trials. The CDAI cut-off used in GEMINI II[11] was 450, yet a cut-off of 400 was used in GEMINI III.[12]

The ERG also noted that in the listing of exclusion criteria in the GEMINI III CSR,[23] additional to those listed in the GEMINI II CSR,[22] included that minor surgical procedures to treat complications of CD (e.g., fistulotomy) are acceptable, and that patients should be excluded from GEMINI III if laboratory abnormalities during the screening period relating to Albumin 2.0 g/dL were identified.

In both GEMINI II and GEMINI III, various exclusion criteria around stenosis are reported: patients with fixed stenosis, small bowel stenosis with prestonic dilation and patients with intestinal stricture are excluded. The clinical advisors to the ERG noted that if patients with stricturing disease were excluded this may limit the generalizability of the findings to only those with inflammatory disease.

- *Ongoing studies of vedolizumab (CS page 33)*

As reported in the CS[1] (see CS[1] pg. 33), there do not appear to be any relevant ongoing studies that will be completed in the next 12 months. For completeness, a brief summary of ongoing relevant vedolizumab studies (identified by the ERG via clinicaltrials.gov and WHO ICTRP on 30th September 2014) which are planned for completion in the next 5 years is presented in Table 5. (The interim results for GEMINI LTS NCT00790933 (C13008)[42] will be summarised and critiqued in section 4.2.4.2).

Table 5 List of ongoing studies as identified by the ERG in searches of ClinicalTrials.gov and WHO ICTRP

Ongoing/ planned Study	Design	Objective	Duration and planned recruitment	Expected start date and end date
GEMINI LTS NCT00790933 (C13008)[42] Sponsor: Millennium Pharmaceuticals, Inc.	Interventional, Phase III, open-label, single arm, multicentre study	To determine the long-term safety of vedolizumab in patients with UC and CD. Eligible patients included those who had previously been treated in Study C13004 (Phase II long-term follow-up), Study C13006 (GEMINI I), Study C13007 (GEMINI II), or Study 13011 (GEMINI III). Primary objectives are to determine AEs, SAEs, results of standard laboratory tests and electrocardiograms (ECG), time to major IBD-related events (hospitalisations, surgeries or procedures), and improvements in quality of life.	Duration up to a maximum of 7 years Estimated enrolment of 2,200 patients	Start date: May 2009 Expected end date: August 2016 Interim safety results provided by company up to July 2012
Phase III Study of **MLN0002** (300 mg) in Treatment of CD NCT02038920 Sponsor: Takeda	Phase III, multicentre, randomised, double-blinded, placebo-controlled, parallel-group study	To examine the efficacy, safety, and pharmacokinetics of **MLN0002** (**Vedolizumab**) in induction and maintenance therapy in Japanese patients with moderately or severely active CD.	Duration up to 4 years.	March 2014 April 2018 (final data collection date for primary outcome measure)

4.2.2 Details of relevant studies not included in the submission

The ERG and its clinical advisors were satisfied that all relevant vedolizumab studies were included in the CS.[1]

4.2.3 Summary and critique of company's analysis of validity assessment

The validity assessment tool used to appraise the GEMINI II[11] and III[12] trials in the CS[1] (see CS[1] pg. 100) is based on the quality assessment criteria suggested by NICE.[21] The completed validity assessment tool for the GEMINI II[11] and III[12] trials, as reported in the CS[1] is reproduced (with minor changes in the table headings made by the ERG) in Table 6.

Table 6 Company's quality assessment results for included RCTs (reproduced from CS[1] pg. 100)

Quality assessment criteria	**Trial**	
	GEMINI II[11]	**GEMINI III[12]**
Was randomisation carried out appropriately?	Yes	Yes
Was the concealment of treatment allocation adequate?	Yes	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes
Were there any unexpected imbalances in drop-outs between groups?	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No
Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes. All patients who prematurely discontinued for any reason were to be considered as not achieving remission for the primary efficacy analysis.	Yes. All patients who prematurely discontinued for any reason were to be considered as not achieving remission for the primary efficacy analysis.

The CS[1] (see CS[1] pg. 79-80) states that randomisation was computer generated centrally for both GEMINI II[11] and III.[12] Participants and investigators were blinded to treatment allocation (double-blind) in both trials. It was not specified if imputation of missing data was undertaken. The ERG acknowledges that adequate methods of randomisation, allocation concealment and blinding were used in the conduct of GEMINI II[11] and III.[12] The quality assessment was not incorporated into the discussion of the results in the CS.[1]

4.2.4 Summary and critique of results

This section presents the results (as reported by the company[1]) from the GEMINI II[11] and III trials,[12] which forms the pivotal evidence in the CS[1] for the efficacy and safety of vedolizumab in the induction and maintenance treatment of patients with moderate to severe active CD. Additional information, not reported in the CS,[1] was provided by the company in their response to the clarification questions[2] raised by the ERG, and in a supplementary document – Takeda data on file document.[16] Where applicable, data have been re-tabulated by the ERG to ensure clarity.

GEMINI II Trial[11]

In the GEMINI II trial,[11] at induction phase, patients were predominantly white (89.2%) with a mean age of 36.1 years. The mean body weight was 69.8kg and 46.6% were male. The mean duration of disease was 9 years, patients had a mean CDAI score of 324, and the mean faecal calprotectin score was 1,254. Concomitant medications for CD included glucocorticoids only (34.2%), immunosuppressant only (16.2%), glucocorticoids and immunosuppressant (17%) and neither glucocorticoids nor immunosuppressant (32.6%). 61.8% of patients had received prior anti-TNF-α treatment. The CS[1] (page 84), suggests that no relevant differences in baseline demographic or clinical characteristics were observed between the treatment groups (*p*-values were not provided). In the US, patients were required to have failed either an immunomodulator (6-MP or azathioprine) or an anti-TNF-α agent, whilst outside of the US, failing corticosteroids alone was sufficient for study entry. It is unclear to the ERG how the different criteria might have impacted on the study results.

All study withdrawals were adequately described in the CSR[22] and all patients were accounted for, this included 9% (105/1115) of the total population in the induction phase who prematurely discontinued from the study (vedolizumab Cohort 1, 10% [21/220], placebo Cohort 1, 7% [11/148], and vedolizumab Cohort 2, 10% [73/747]). The primary reason for discontinuation in the induction phase was due to adverse events 5% (7/148) in the placebo arm, and 3% (33/968) in the combined vedolizumab arm, followed by lack of efficacy 1% (1/148) in the placebo arm, and 3% (31/968) in the combined vedolizumab arm. In general, the validity of a study may be threatened if attrition is more than 20%.[43] As such, the ERG acknowledges that attrition bias should be considered low in the induction phase of the GEMINI II trial.[11] The maintenance phase ITT population only includes vedolizumab patients who had a clinical response at week 6. At the start of the maintenance phase, these patients were randomised to one of two vedolizumab dosing regimens (300 mg every 4 weeks or every 8 weeks) or placebo. During the maintenance phase, of the ITT population, 58% (89/153) discontinued in the placebo arm, 53% (81/154), and 47% (72/154) discontinued in the vedolizumab Q8W and Q4W arms respectively. The main reason for discontinuation

in the maintenance phase was due to lack of efficacy, 42% (64/153) in the placebo arm, and 38% (58/154) and 31% (48/154) in the vedolizumab every 8 weeks (Q8W) and the vedolizumab every 4 weeks (Q4W) arms respectively. As noted earlier it has been argued that loss to follow-up of 20% or greater means that the validity of the study may be threatened.[43] The ERG acknowledges that in a study of this length, whereby patients are continued on placebo for an extended period of time, greater discontinuations may be expected. However, attrition rates as these levels have the potential to impact on the maintenance study results, posing a serious threat to external validity. As withdrawals were counted as treatment failures for the efficacy outcomes, the ERG believes that the estimates of efficacy are problematic more in terms of generalizability rather than estimation of the treatment effect within the trial. However, the ERG believes that the loss of patients may be problematic for the assessment of adverse events.

GEMINI III Trial[12]

In the GEMINI III trial,[12] most patients were white (90%). The mean age was 37.9 years, mean body weight was 70.4kg and 43% were male. Other baseline characteristics were reported only for each treatment group (vedolizumab vs. placebo). Median duration of disease was 8.4 years in the vedolizumab group and 8 years in the placebo group. Patients in the vedolizumab group had a mean CDAI score of 301.3, and 313.9 in the placebo group. Median faecal calprotectin score was 1148.1 in the vedolizumab group, and 1426.5 in the placebo group. Concomitant medications for CD included corticosteroid use (53% in the vedolizumab group and 52% in the placebo group), immunosuppressant use (34% in the vedolizumab group and 33% in the placebo group), and 5-ASA use (33% in the vedolizumab group and 29% in the placebo group). In each group 76% of patients had had a prior anti-TNF-α failure. The CS[1] (see CS[1] pg. 84), suggests that most baseline demographics were similar between the treatment groups with the exception of the vedolizumab-treated patients who had a slightly higher baseline CDAI compared to the placebo group (313.9 vs 301.3, p=0.015), and more placebo-treated patients (51%) were <35 years of age compared to vedolizumab-treated patients (42%) (*p*-values were not provided).[12,23] In the US, patients were required to have failed either an immunomodulator (6-MP or azathioprine) or an anti-TNF-α agent, whilst outside of the US, failing corticosteroids alone was sufficient for study entry. It is unclear to the ERG how these different criteria might have impacted on the study results.

All study withdrawals were adequately described and all patients were accounted for; this included 7% (28/416) of the total population who prematurely discontinued from the study (vedolizumab anti-TNF-α failures, n=7; vedolizumab anti-TNF-α naïve, n=6; placebo anti-TNF-α failures, n =12; placebo anti-TNF-α naïve, n = 3). The primary reason for discontinuation was not provided in the CS[1], Takeda data on file[16] or the CSR[23]. Discontinuation due to AEs was reported in 2% (4/209) of placebo patients and in 4%

(8/207) the vedolizumab-treated patients. As the attrition rate is less than 20% it is considered not to threaten the validity of a trial, as such, the ERG acknowledges that attrition bias should be low in the GEMINI III trial.[12]

GEMINI II[11] & GEMINI III[12]

Clinical advisors to the ERG expressed the view that the concomitant conventional therapy used in the GEMINI trials[11,11,12] may not reflect those used in UK clinical practice in all cases. The company, in response to clarification (see clarification response[2] question B29) appears to agree and states that *'the use of conventional therapy within the GEMINI II and GEMINI III trials was protocol driven and the trial was international and may not represent treatment patterns in England and Wales…'*. It is unclear to the ERG how the potential lack of generalizability of conventional therapy might have impacted the study results.

It should be noted that the faecal calprotectin in the GEMINI trials was deemed to be high; indicating that patients may had had significant active inflammation. It is unclear to the ERG how this affects generalizability of the findings across the GEMINI trials, and also how it may impact on the generalizability of the findings to the UK patient population.

In the CSR for GEMINI II[22] several amendments to the inclusion criteria are detailed. One amendment relates to the CDAI cut-off used to include and exclude patients. The CSR detailed that this was amended from 220-480 (210-490 for the per protocol population) down to 220-450. However, as detailed in the CSR the range of scores for the included patients were 93 – 584. Although no such amendments were detailed in the CSR for GEMINI III [23] again the range of baseline CDAI scores were 166-564. Although the ERG note the proportion of patients scoring above 450 or below 220 is likely to be very small and thus should not affect the results of the trial, it is unclear why the range of scores does not represent the inclusion criteria.

4.2.4.1 Efficacy

GEMINI II Trial[11]

- *Induction phase of GEMINI II*

As reported in the CS[1], and presented in Table 7 for the primary outcome, patients treated with vedolizumab, had significantly higher rates of clinical remission (defined as CDAI ≤150) at week 6 compared with placebo (14.5% vs. 6.8%). The treatment difference from placebo was 7.8% (95% CI 1.2, 14.3; p = 0.0206). There was no significant difference between the vedolizumab and placebo groups for the second primary outcome which analysed the number of patients achieving enhanced clinical response (defined as a 100-point reduction from baseline in CDAI score) at week 6 (p-value not reported).

The secondary endpoint relating to changes from baseline in CRP at week 6 was not significantly different between treatment groups (p-value not reported). Other key endpoints reported in the CS[1] were clinical remission and enhanced clinical response by week 10 and 14 in induction non-responders. Of patients who had not achieved clinical remission to vedolizumab by week 6 (n=86 from cohort 1; n=265 from cohort 2; total=351), 6.8% (24 patients) achieved clinical remission at week 10 (an additional 4 weeks of treatment/1 additional infusion), and 10.5% (37 patients) achieved clinical remission at week 14 (an additional 8 weeks of treatment/2 additional infusions).

Table 7 Clinical remission and enhanced clinical response at week 6 – ITT population (reproduced from Table 6.5.3.1 in CS[1] pg. 103)

	Clinical remission[a]		**Enhanced clinical response[b]**	
	Placebo n=148	**Vedolizumab n=220**	**Placebo n=148**	**Vedolizumab n=220**
Number (%) achieving endpoint 95% CI	10 (6.8) (2.7, 10.8)	32 (14.5) (9.9, 19.2)	38 (25.7) (18.6, 32.7)	69 (31.4) (25.2, 37.5)
Difference from placebo[c] 95% CI for difference from placebo P-value for difference from placebo[d]		7.8 (1.2, 14.3) 0.0206		5.7 (-3.6, 15.0) 0.2322
Relative risk[e] 95% CI for relative risk		2.1 (1.1, 4.2)		1.2 (0.9, 1.7)

Abbreviations: CI = confidence interval

a Clinical remission is defined as CDAI score ≤ 150 points.

b Enhanced clinical response is defined as a 100-point reduction from baseline in CDAI score.

c Difference and 95% CI: adjusted percent vedolizumab - adjusted percent placebo and its 95% CI.

d P-value is based on the CMH chi-square test, with stratification according to: 1) concomitant use of oral corticosteroids (yes/no); 2) previous exposure to anti-TNF-α and/or concomitant immunomodulatory use (yes/no).

e Adjusted Relative Risk and its 95% CI.

Sub-group analyses

The subgroup analyses reported in the CS[1] showed a trend in people who have not previously received an anti-TNF-α and people for whom an anti-TNF-α has failed, with a greater proportion of vedolizumab-treated patients achieving clinical remission at Week 6 (treatment difference 8.2% and 6.2% respectively). The treatment benefit of vedolizumab over placebo was maintained in patients with prior corticosteroid failure for the endpoint of clinical remission at week 6. A trend favouring vedolizumab was observed in patients with prior immunomodulatory failure.[44] In general, analyses of clinical remission in sub-groups of patients according to baseline concomitant corticosteroid or immunomodulator use showed trends that were supportive of the primary efficacy analysis population as a whole.[44]

Patient-Reported Outcomes

The CS[1] and CSR[22] report health-related quality of life (HRQoL) assessments using the Inflammatory Bowel Disease Questionnaire (IBDQ) total score, Medical Outcomes Study 36-item Short Form (SF-36) mental and physical component scores, Euroqol 5-Dimensions (EQ-5D) questionnaire and EQ-5D Visual Analogue Score (VAS).

The results showed that patients receiving induction therapy with vedolizumab reported higher scores on all IBDQ domain scales and the total score compared with the placebo group at week 6. Although the 95% CIs for differences from baseline to Week 6 included zero for most scales, except for Bowel Function, the increases in all IBDQ domain scale scores and IBDQ total score were considered to be clinically meaningful improvements, according to the definition used in the CS[1]. Higher scores were observed for vedolizumab patients on the SF-36 physical and mental component summary scores and all SF-36 scales except for the physical functioning scale compared to the placebo group at week 6. Additionally, for the Role-physical, Bodily Pain and Social Functioning scales, the 95% CI of differences from baseline to week 6 excluded zero. Patients receiving vedolizumab also had greater improvements in HRQoL as measured by EQ-5D and EQ-5D VAS scores compared to placebo at week 6; however, the 95% CIs in the difference of scores between the two groups included 0. The decrease in the EQ-5D score was reported as clinically meaningful in both groups according to the definition used in the CS[1] (Table 8). A significant higher improvement in IBDQ score was seen for anti-TNF-α naïve subgroups compared to anti-TNF--α failure subgroups (Table 9).

Table 8 Overall observed changes in HRQL from baseline to week 6 in GEMINI II (reproduced from Table 6.5.3.2 in CS[1] pg. 106)

	Placebo	Vedolizumab
IBDQ Total Score[a]	**n=146**	**n=212**
Adjusted mean (SE) change from baseline (95% CI)[b]	16.5 (2.75) (11.1 to 21.9)	23.1 (2.28) (18.6 to 27.6)
Difference in adjusted change from baseline vs placebo, mean (SE) (95% CI)[c]	---	6.5 (3.58) (–0.5 to 13.6)
SF-36 Physical Component Summary[a]	**n=144**	**n=211**
Adjusted mean (SE) change from baseline (95% CI)[b]	2.4 (0.56) (1.3 to 3.6)	3.5 (0.47) (2.6 to 4.4)
Difference in adjusted change from baseline vs placebo, mean (95% CI)[c]	---	1.0 (0.73) (–0.4 to 2.5)
SF-36 Mental Component Summary[a]	**n=144**	**n=211**
Adjusted mean change from baseline (95% CI)[b]	2.4 (0.86) (0.8 to 4.1)	4.6 (0.71) (3.2 to 6.0)
Difference in adjusted change from baseline vs placebo, mean (SE) (95% CI)[c]	---	2.2 (1.11) (0.0 to 4.4)
EQ-5D Score[a]	**n=146**	**n=211**
Adjusted mean change from baseline (95% CI)[b]	–0.3 (–0.5 to –0.0)	–0.5 (–0.7 to –0.3)
Difference in adjusted change from baseline vs placebo, mean (95% CI)[c]		–0.2 (–0.5 to 0.1)
EQ-5D VAS Score[a]	**n=146**	**n=208**
Adjusted mean (SE) change from baseline (95% CI)[b]	5.4 (1.65) (2.2 to 8.7)	6.9 (1.38) (4.2 to 9.6)
Difference in adjusted change from baseline vs placebo, mean (95% CI)[c]	---	1.5 (2.15) (–2.8 to 5.7)

Abbreviations: CI=confidence interval; EQ=EuroQol; HRQL=health-related quality of life; IBDQ=Inflammatory Bowel Disease Questionnaire; SF-36=Short Form-36; VAS=visual analog scale.

a Higher IBDQ, SF-36, and EQ-5D VAS scores indicate improvements in HRQL; lower EQ-5D scores indicate improvements in HRQL.

b Mean changes were adjusted within the ANCOVA model with factors for treatment and baseline measurement.

c Difference = adjusted mean change for vedolizumab – adjusted mean change for placebo.

Table 9 Observed changes in HRQL in Anti-TNF-α naïve and TNF-α -Failure from baseline to week 6 in GEMINI II (reproduced from Table 6.5.3.3 in CS[1] pg. 108)

	PRIOR Anti-TNF-α – Failure		**No PRIOR Failure**	
	Placebo	**Vedolizumab**	**Placebo**	**Vedolizumab**
IBDQ Total Score[a]	n=69	n=104	n=77	n=108
Adjusted Mean (SE) change from baseline (95% CI)[b]	13.0 (3.65) (5.8, 20.2)	15.3 (2.97) (9.4, 21.2)	19.6 (3.94) (11.8, 27.4)	30.6 (3.33) (24.1, 37.2)
Difference in adjusted change from baseline vs placebo, Mean (SE) (95% CI)[c]		2.3 (4.72) (-7.0, 11.6)		**11.0 (5.18)* (0.8, 21.3)**
Physical Component Summary	n= 67	n= 103	n= 77	n= 108
Adjusted Mean (SE) change from baseline (95% CI)[b]	1.6 (0.83) (-0.1, 3.2)	3.0 (0.67) (1.7, 4.3)	3.1 (0.76) (1.6, 4.6)	3.9 (0.64) (2.7, 5.2)
Difference in adjusted change from baseline vs placebo, Mean (SE) (95% CI)c		1.4 (1.07) (-0.7, 3.5)		0.8 (1.00) (-1.2, 2.8)
Mental Component Summary	n= 67	n= 103	n= 77	n= 108
Adjusted Mean (SE) change from baseline (95% CI)[b]	1.2 (1.22) (-1.2, 3.6)	2.4 (0.98) (0.4, 4.3)	3.6 (1.19) (1.3, 6.0)	6.7 (1.00) (4.7, 8.7)
Difference in adjusted change from baseline vs placebo, Mean (SE) (95% CI)[c]		1.2 (1.57) (-1.9, 4.3)		3.1 (1.56) (0.0, 6.2)
EQ-5D VAS Score	n= 69	n= 100	n= 77	n= 108
Adjusted Mean (SE) change from baseline (95% CI)[b]	1.7 (2.48) (-3.2, 6.6)	2.7 (2.06) (-1.3, 6.8)	8.4 (2.06) (4.3, 12.4)	11.0 (1.74) (7.6, 14.4)
Difference in adjusted change from baseline vs placebo, Mean (SE) (95% CI)[c]		1.0 (3.22) (-5.3, 7.4)		2.6 (2.71) (-2.7, 8.0)

Abbreviations: SE = Standard Error; CI=confidence interval; EQ=EuroQol; HRQL=health-related quality of life; IBDQ=Inflammatory Bowel Disease Questionnaire; SF-36=Short Form-36; VAS=visual analog scale.

a Higher IBDQ, SF-36, and EQ-5D VAS scores indicate improvements in HRQL; lower EQ-5D scores indicate improvements in HRQL.

b Mean changes were adjusted for individual baseline measurements.

c Difference = adjusted mean change for vedolizumab – adjusted mean change for placebo.

*** denotes statistically significant results. (*p*-value cut-off not reported in CS)**

- *Maintenance phase of GEMINI II*

The Maintenance Study ITT Population includes vedolizumab-treated patients who had a clinical response at week 6 (defined as ≥70-point decrease in CDAI score); at the start of the maintenance phase, these patients were randomised to 1 of 2 vedolizumab i.v. dosing regimens (300 mg Q4W or Q8W, n=154 each) or placebo (n=153). The data presented here is for the intention to treat (ITT) population.[11,22]

As presented in Table 10 patients treated with vedolizumab every 8 weeks (Q8W) and every 4 weeks (Q4W), had significantly higher rates of clinical remission at week 52 (defined as CDAI score of ≤150 points) compared with placebo. The treatment difference from placebo was 17.4% (95% CI 7.3, 27.5; p = 0.0007) and 14.7% (95% CI 4.6, 24.7; p = 0.0042) respectively.

Table 10 Clinical remission at week 52 – ITT population – GEMINI II (reproduced from Table 6.5.3.4 in CS[1] pg. 109)

	Clinical remission[a]		
	Placebo **n=153**	**Vedolizumab Q8W** **n=154**	**Vedolizumab Q4W** **n=154**
Number (%) achieving endpoint 95% CI	33 (21.6) (15.1, 28.1)	60 (39.0) (31.3, 46.7)	56 (36.4) (28.8, 44.0)
Difference from placebo[b] 95% CI for difference from placebo P-value for difference from placebo[c]		17.4 (7.3, 27.5) 0.0007	14.7 (4.6, 24.7) 0.0042
Relative risk[d] 95% CI for relative risk		1.8 (1.3, 2.6)	1.7 (1.2, 2.4)

Abbreviations: CI = confidence interval

a Clinical remission is defined as CDAI score ≤ 150 points.

b Difference and 95% CI: adjusted percent vedolizumab - adjusted percent placebo and its 95% CI.

c P-value is based on the CMH chi-square test, with stratification according to: 1) concomitant use of oral corticosteroids (yes/no); 2) previous exposure to anti-TNF-α and/or concomitant immunomodulatory use (yes/no).

d Adjusted Relative Risk and its 95% CI.

The secondary endpoints demonstrated that patients receiving vedolizumab every 4 or 8 weeks were significantly more likely to achieve enhanced clinical response (defined as a ≥100-point reduction in CDAI score from baseline) and have a corticosteroid free remission at week 52 compared with patients

receiving placebo (See Table 11). In contrast, the number of patients with durable clinical remission (defined as CDAI score ≤ 150 points at ≥ 80% of study visits including final visit) did not differ significantly between the study groups. The company state that this was due to baseline differences at re-randomisation.[11]

Sub-group analyses

The sub-group analyses reported in the CS[1] demonstrate that clinical remission rates were greater for patients treated with vedolizumab than those who were treated with placebo, regardless of prior exposure to anti-TNF-α (Table 12) Similar improvements with vedolizumab versus placebo were found for enhanced clinical response, and corticosteroid-free clinical remissions at week 52 in all sub-groups. Although a higher number of patients achieved clinical remission with vedolizumab in the immunomodulator and corticosteroid failure sub-groups than the anti-TNF-α failure subgroup, the treatment differences between placebo and vedolizumab were generally similar among all sub-groups.[44]

Table 11 Enhanced clinical response, corticosteroid free remission, and durable clinical remission at week 52 – ITT population (reproduced from Table 6.5.3.6 in CS[1] pg. 111)

	Enhanced clinical response[a]			Corticosteroid-free Clinical Remission[b]			Durable Clinical Remission[c]		
	Placebo n=153	VDZ Q8W n=154	VDZ Q4W n=154	Placebo n=82	VDZ Q8W n=82	VDZ Q4W n=80	Placebo n=153	VDZ Q8W n=154	VDZ Q4W n=154
Number (%) 95% CI	46 (30.1) (22.8, 37.3)	67 (43.5) (35.7, 51.3)	70 (45.5) (37.6, 53.3)	13 (15.9) (7.9, 23.8)	26 (31.7) (21.6, 41.8)	23 (28.8) (18.8, 38.7)	22 (14.4) (8.8, 19.9)	33 (21.4) (14.9, 27.9)	25 (16.2) (10.4, 22.1)
Difference from placebo[d] 95% CI P-value[e]		13.4 (2.8, 24.0) 0.0132	15.3 (4.6, 26.0) 0.0053		15.9 (3.0, 28.7) 0.0154	12.9 (0.3, 25.5) 0.0450		7.2 (-1.5, 16.0) 0.1036	2.0 (-6.3, 10.2) 0.6413
Relative risk[f] 95% CI		1.4 (1.1, 1.9)	1.5 (1.1, 2.0)		2.0 (1.1, 3.6)	1.8 (1.0, 3.3)		1.5 (0.9, 2.4)	1.1 (0.7, 1.9)

Abbreviations: CI = confidence interval VDZ, Vedolizumab

a Enhanced clinical response is defined as a ≥100-point reduction in CDAI score from baseline.

b Corticosteroid-free clinical remission is defined as patients using oral corticosteroids at baseline who had discontinued corticosteroids and were in clinical remission at week 52.

C Durable clinical remission is defined as CDAI score ≤ 150 points at ≥ 80% of study visits including final visit (week 52).

d Difference and 95% CI: adjusted percent vedolizumab - adjusted percent placebo and its 95% CI.

e P-value is based on the CMH chi-square test, with stratification according to: 1) concomitant use of oral corticosteroids (yes/no); 2) previous exposure to anti-TNF-α and/or concomitant immunomodulatory use (yes/no).

f Adjusted Relative Risk and its 95% CI.

Table 12 Results at week 52 by prior anti-TNF-α status (reproduced from Table 6.5.3.6 in CS[1] pg. 112)

Study Endpoint	Vedolizumab Every 8 Wks (n=82)	Vedolizumab Every 4 Wks (n=77)	Placebo (n=78)	Between Group Difference (95% CI)	
				Every 8 Wks vs Placebo	Every 4 Wks vs Placebo
Patients With Prior anti-TNF-α Failure[a]					
Clinical Remission (%)	28.0	27.3	12.8	15.2 (3.0 to 27.5)	14.5 (2.0 to 26.9)
CDAI-100 Response (%)	29.3	37.7	20.5	8.8 (-4.6 to 22.1)	17.1 (3.1 to 31.2)
Patients Without anti-TNF-α Exposure[b]					
	Vedolizumab Every 8 Wks (n=66)	**Vedolizumab Every 4 Wks (n=71)**	**Placebo (n=71)**	**Between Group Difference (95% CI)**	
				Every 8 Wks vs Placebo	**Every 4 Wks vs Placebo**
Clinical Remission (%)	51.5	46.5	26.8	24.8 (8.9 to 40.6)	19.7 (4.2 to 35.2)
CDAI-100 Response (%)	60.6	53.5	38.0	22.6 (6.3 to 38.9)	15.5 (-0.7 to 31.7)

CDAI=Crohn's Disease Activity Index; CI=confidence interval; TNF=tumour necrosis factor; Wks=weeks

a Treatment failure (inadequate response, loss of response, or intolerance) defined as follows: inadequate response to anti-TNF-α =persistently active disease despite induction treatment with specified agents; loss of response to anti-TNF-α =recurrence of symptoms during maintenance dosing following prior clinical benefit; intolerance=occurrence of treatment-related protocol-defined toxicities.

b Patients without prior exposure to anti-TNF-α therapy (i.e., anti-TNF-α -naïve patients)

Patient-Reported Outcomes

The CS[1] and CSR[22] report HRQoL assessments using the IBDQ total score, SF-36 mental and physical component scores, EQ-5D and EQ-5D VAS.

Maintenance therapy with vedolizumab either every 4 weeks or every 8 weeks resulted in higher scores on all IBDQ domain scales and higher IBDQ total score from baseline to week 52 compared with placebo, with the mid-point increases considered clinically meaningful according to the definition used in

the CS.[1] There were no major differences between vedolizumab and placebo in the improvements in the total IBDQ scores at week 30.

There were no major differences between vedolizumab and placebo in the SF-36 assessments at week 30, but at week 52, both vedolizumab regimens resulted in higher scores on all SF-36 scales and the physical and mental component summary scores compared with the placebo group. For vedolizumab every 8 week group, the 95% CI of the differences from baseline to week 52 excluded zero for all scales, except the mental component summary score and the Mental Health scale. For vedolizumab every 4 weeks, the 95% CI of the differences from baseline to week 52 excluded zero for the Role-Emotional, General Health, Bodily pain, Physical functioning scales and the physical component summary score.[22]

Both vedolizumab maintenance treatment regimens resulted in greater improvements in the EQ-5D score and EQ-5D VAS score from baseline to week 52 compared with placebo, with the improvements in all groups considered clinically meaningful according to the definition used in the CS.[1] From baseline to week 30, the 95% CIs for the differences in the EQ-5D scores and EQ-5D VAS scores between vedolizumab and placebo included zero.[22]

A higher proportion of vedolizumab-treated patients compared with placebo patients had clinically meaningful improvements in some HRQL endpoints at week (Table 13).

Table 13 Proportion of patients with clinically meaningful improvement difference from baseline compared with placebo at week 52 (GEMINI II) (reproduced from Table 6.5.3.10 in CS[1] pg. 117).

	Placebo	**Vedolizumab Q8W**	**Vedolizumab Q4W**
IBDQ Total Score	n=82	n=79	n=92
Number (%) Achieving Clinically Meaningful Improvement	54(65.9)	59(74.7)	73(79.3)
95% CI	(55.6 to 76.1)	(65.1 to 84.3)	(71. to 87.6)
Difference from Placebo	---	8.8	**13.5***
95% CI for Difference from Placebo	---	(-5.2, 22.9)	(0.3, 26.7)
P-value for Difference from Placebo	---	0.2222	0.0460
SF 36 Physical Component Summary	n=82	n=79	n=91
Number (%)Achieving Clinically Meaningful Improvement	46(56.1)	57(72.2)	56 (61.5)
95% CI	(45.4 to 66.8)	(62.3 to 82.0)	(51.5 to 71.5)
Difference from Placebo	---	**16.1***	5.4
95% CI for Difference from Placebo	---	(1.5 to 30.7)	(-9.2 to 20.1)
P-value for Difference from Placebo	---	0.0345	0.4689
SF-36 Mental Component Summary	n=82	n=79	n=91
Number (%)Achieving Clinically Meaningful Improvement	44(53.7)	52(65.8)	55(60.4)
95% CI	(42.9 to 64.5)	(55.4 to 76.3)	(50.4 to 70.5)
Difference from Placebo	---	12.2	6.8
95% CI for Difference from Placebo	---	(-2.9 to 27.2)	(-8.0 to 21.5)
P-value for Difference from Placebo	---	0.1169	0.3694
EQ-5D VAS Score	n=81	n=79	n=89
Number (%)Achieving Clinically Meaningful Improvement	53(65.4)	62(78.5)	71(79.8)
95% CI	(55.1 to 75.8)	(69.4 to 87.5)	(71.4 to 88.1)
Difference from Placebo	---	13.0	**14.3***
95% CI for Difference from Placebo	---	(-0.7 to 26.8)	(1.0 to 27.6)
P-value for Difference from Placebo	---	0.0673	0.0361

Abbreviations: SE = Standard Error; CI=confidence interval; EQ=EuroQol; HRQL=health-related quality of life; IBDQ=Inflammatory Bowel Disease Questionnaire; SF-36=Short Form-36; VAS=visual analog scale.

a Higher IBDQ, SF-36, and EQ-5D VAS scores indicate improvements in HRQL; lower EQ-5D scores indicate improvements in HRQL.

b Mean changes were adjusted for individual baseline measurements.

c Difference = adjusted mean change for vedolizumab – adjusted mean change for placebo.

* denotes statistically significant results.

An increase of ≥ 16 points in the IBDQ Total score, ≥ 5 in IBDQ Bowel Function domain scores, ≥ 6 in IBDQ Emotional Function domain scores, or ≥ 2.5 in IBDQ Systemic and Social Function domain scores, represents clinically meaningful improvements in HRQL for patients.

An increase of ≥ 5 points in the Physical Component Scale, the Mental Component Scale, and SF-36 subscales represents a clinically meaningful improvement in HRQL for patients.

A decrease of ≥ 0.3 points in the EQ-5D score represents a clinically meaningful improvement in HRQL for patients. An increase of ≥ 7 points in the EQ-5D VAS score represents a clinically meaningful improvement in HRQL for patients.

GEMINI III Trial[12]

There was no statistically significant difference between vedolizumab and placebo in the primary endpoint of the proportion of patients achieving clinical remission at week 6 (CDAI score ≤150 points) in the anti-TNF-α failure population (Table 14); therefore, statistical evaluation of the secondary endpoints is acknowledge as exploratory by the company.[12] Nominal *p* values, relative risks, and 95% CIs are presented for descriptive purposes to fully characterize the effect of vedolizumab induction treatment in this population.

Secondary efficacy endpoints included: clinical remission at week 10; enhanced clinical response (defined as a reduction of 100 points or more in the CDAI) at weeks 6 and 10; and sustained remission (defined as CDAI score ≤150 points at both Week 6 and Week 10) in the anti-TNF-α failure population and clinical remission and enhanced clinical response at week 6 and 10, and sustained remission in the overall population. As shown in Table 15 compared to placebo, vedolizumab was associated with a higher number of patients achieving clinical remission at week 10 and an enhanced clinical response at week 6 and 10 in the anti-TNF-α failure population. The company[1] asserts that these results suggest that a potential treatment benefit for vedolizumab in the anti-TNF-α failure population may be achieved beyond the 6-week period used to evaluate the primary endpoint in this study.[23] In the overall population, vedolizumab-treated patients had higher rates of clinical remission, and enhanced clinical response at weeks 6 and 10 and sustained remission compared with placebo-treated patients (Table15). As these are exploratory analyses the ERG note that the findings should be interpreted with caution.

Table 14 Efficacy outcomes in anti-TNF-α **failure population in GEMINI III – ITT population (reproduced from Table 6.5.3.11 in CS[1] pg. 119)**

	Clinical remission[a]				**Enhanced clinical response[b]**				**Sustained Remission[c]**	
	week 6		**week 10**		**week 6**		**week 10**			
	Placebo n=157	**VDZ n=158**	**Placebo n=157**	**VDZ n=158**	**Placebo n=157**	**VDZ n=158**	**Placebo n=157**	**VDZ n=158**	**Placebo n=157**	**VDZ n=158**
Number (%) 95% CI	19 (12.1) (7.0, 17.2)	24 (15.2) (9.6, 20.8)	9 (12.1) (7.0, 17.2)	42 (26.6) (19.7, 33.5)	35 (22.3) (15.8, 28.8)	62 (39.2) (31.6, 46.9)	39 (24.8) (18.1, 31.6)	74 (46.8) (39.1, 54.6)	13 (8.3) (4.0, 12.6)	19 (12.0) (7.0, 17.1)
Difference from placebo[d] 95% CI P-value[e]		3.0 (-4.5, 10.5) 0.433		14.4 (5.7, 23.1) 0.0012		16.9 (6.7, 27.1) n/a		22 (11.4, 32.6) n/a		3.7 (-2.9, 10.3) 0.2755
Relative risk[f] 95% CI		1.2 (0.7, 2.2)		2.2 (1.3, 3.6)		1.8 (1.2, 2.5)		1.9 (1.4, 2.6)		1.4 (0.7, 2.8)

Abbreviations: CI = confidence interval; n/a, not available; TNF=tumour necrosis factor; VDZ, Vedolizumab

a Clinical remission is defined as CDAI score ≤ 150 points.

b Sustained remission is defined as CDAI score ≤ 150 points at both week 6 and week 10

c Enhanced clinical response is defined as a ≥100-point reduction in CDAI score from baseline.

d Difference and 95% CI: adjusted percent vedolizumab - adjusted percent placebo and its 95% CI.

e P-value is based on the CMH chi-square test, with stratification according to: 1) concomitant use of oral corticosteroids (yes/no); 2) previous exposure to anti-TNF-α and/or concomitant immunomodulatory use (yes/no).

f Adjusted Relative Risk and its 95% CI.

Table 15 Efficacy outcomes in overall population in GEMINI III – ITT population (reproduced from Table 6.5.3.12 in CS[1] pg. 120)

	Clinical remission[a]				Enhanced clinical response[b]				Sustained Remission[c]	
	week 6		week 10		week 6		week 10			
	Placebo n=207	VDZ n=209	Placebo n=207	VDZ n=209	Placebo n=207	VDZ n=209	Placebo n=207	VDZ n=209	Placebo n=207	VDZ n=209
Number (%) 95% CI	25 (12.1) (7.6, 16.5)	40 (19.1) (13.8, 24.5)	27 (13.0) (8.5, 17.6)	60 (28.7) (22.6, 34.8)	47 (22.7) (17.0, 28.4)	82 (39.2) (32.6, 45.9)	50 (24.2) (18.3, 30.0)	100 (47.8) (41.1, 54.6)	17 (8.2) (4.5, 12.0)	32 (15.3) (10.4, 20.2)
Difference from placebo[d] 95% CI P-value[e]		6.9 (0.1, 13.8) 0.0478		15.5 (7.8, 23.3) < 0.0001		16.4 (7.7, 25.2) n/a		23.7 (14.5, 32.9) n/a		7.0 (0.9, 13.1) 0.0249
Relative risk[f] 95% CI		1.6 (1.0, 2.5)		2.2 (1.4, 3.3)		1.7 (1.3, 2.3)		2.0 (1.5, 2.6)		1.9 (1.1, 3.2)

Abbreviations: CI = confidence interval; n/a not available; TNF=tumour necrosis factor; VDZ, Vedolizumab

a Clinical remission is defined as CDAI score ≤ 150 points.

b Sustained remission is defined as CDAI score ≤ 150 points at both week 6 and week 10

c Enhanced clinical response is defined as a ≥100-point reduction in CDAI score from baseline.

d Difference and 95% CI: adjusted percent vedolizumab - adjusted percent placebo and its 95% CI.

e P-value is based on the CMH chi-square test, with stratification according to: 1) concomitant use of oral corticosteroids (yes/no); 2) previous exposure to anti-TNF-α and/or concomitant immunomodulatory use (yes/no).

f Adjusted Relative Risk and its 95% CI.

Sub-group analyses

In the anti-TNF-α naïve population proportions of patients were greater with vedolizumab than with placebo for the following outcomes: clinical remission at week 6 (vedolizumab, 31.4%; placebo, 12.0%; *p*=0.012; relative risk, 2.6 [95% CI: 1.1, 6.2]); remission at week 10 (vedolizumab, 35.3%; placebo, 16.0%; P=0.025; relative risk, 2.2 [95% CI: 1.1, 4.6]); remission at both weeks 6 and 10 (vedolizumab, 25.5%; placebo, 8.0%; *p*=0.018; relative risk, 3.2 [95% CI: 1.1, 9.1]); enhanced clinical response (defined as a ≥100-point reduction in CDAI score from baseline) at week 6 (vedolizumab, 39.2%; placebo, 24.0%; *p*=0.088; relative risk, 1.6 [95% CI: 0.9, 2.9]); and enhanced clinical response at week 10 (vedolizumab, 51.0%; placebo, 22.0%; *p*=0.002; relative risk, 2.3 [95% CI: 1.3, 4.2]).

Patient-Reported Outcomes

The CS[1] and CSR[23] report HRQoL assessments using the IBDQ total score, SF-36 mental and physical component scores, EQ-5D and EQ-5D VAS. These assessments were completed during screening (and prior to dosing) at weeks 6 and 10 (or early termination visit). The results showed that patients receiving induction therapy with vedolizumab in both the anti-TNF-α failure subgroup and the overall population achieved greater improvements in the IBDQ total score and on all the IBDQ domain scales at week 6 and week 10 compared with patients receiving placebo. The improvements in HRQL in the vedolizumab groups were considered to be clinically meaningful improvements, as defined by the study authors.[23] For both the anti-TNF-α failure sub-population and the overall population, although the vedolizumab treatment groups achieved greater increases in the week 6 and week 10 SF-36 physical and mental component summary scores compared with the placebo group, the 95% CIs for the treatment differences included 0 except for the week 10 SF-36 mental component summary score.[23] For the anti-TNF-α failure sub-population that received vedolizumab treatment, the decreases in the EQ-5D scores and the increases in the EQ-5D VAS scores were considered clinically meaningful improvements in HRQL at both week 6 and week 10. The 95% CIs for the differences between vedolizumab and placebo in the EQ-5D scores included 0 at week 6 but not at week 10, demonstrating improvements in HRQL for vedolizumab over placebo. Compared to patients receiving placebo, patients receiving vedolizumab demonstrated greater improvements on the EQ-5D VAS scores at both week 6 and week 10. Similar results were seen in the overall study population.[23]

4.2.4.2 Safety and tolerability

This section provides the main safety evidence for the use of vedolizumab in patients with moderate to severe CD available from the GEMINI II[11] and GEMINI III[12] trials. The company[1] also provided supplementary supporting evidence on the safety of vedolizumab from three further sources. These were GEMINI LTS (C13008):[45] interim results from an ongoing Phase III, single-arm, open-label study where the objective is to determine the long-term safety and efficacy of vedolizumab in patients with ulcerative colitis (UC) and CD; GEMINI I (UC) and GEMINI II (CD) pooled safety analysis,[46] and results from an integrated safety analysis of six vedolizumab randomised placebo-controlled in IBD (UC and CD). This analysis includes data from the GEMINI LTS plus from patients enrolled in randomised studies who did not enrol into the open-label extension.[45] The CS[1] (see CS[1] pg. 156) confirms that no separate search was undertaken for safety, as safety was a secondary outcome of the GEMINI II and III trials.

GEMINI II[11]

The rates of discontinuation for all ITT participants in the induction phase of the GEMINI II trial[11] were 9% (105/1115), with no notable difference between the combined vedolizumab and placebo-treated groups. In the ITT population, discontinuation due to AEs was reported in 5% (148) of placebo treated patients and in 3% (33/968) of the vedolizumab-treated patients. The ERG notes that the low numbers of discontinuation during this phase is likely to be due to the short duration (6-weeks) of the induction phase. During the maintenance phase, of the ITT population, 58% (89/153) discontinued in the placebo arm, 53% (81/154), and 47% (72/154) discontinued in the vedolizumab Q8W and Q4W arms respectively. Discontinuation due to AEs was reported in 10% (15/153) of placebo patients and in 8% (12/154) of the vedolizumab Q8W patients, and in 6% (9/154) of the vedolizumab Q4W patients. There were no notable differences between the vedolizumab and placebo groups in terms of discontinuation due to AEs during the maintenance phase.

The safety population in the GEMINI II trial[11] included all enrolled patients, including both Cohort 1 and Cohort 2. The placebo safety group (n=301) includes patients who received placebo in Cohort 1 (n=148) and patients who responded to vedolizumab in the induction phase (up to week 6) and were randomised to placebo in the maintenance phase (up to week 52). The vedolizumab safety group (n=814) includes patients from Cohort 1 and Cohort 2 who responded to vedolizumab in the induction phase (up to week 6) and were randomised to vedolizumab (Q4W, n=154 or Q8W, n=154) in the maintenance phase (up to week 52) and patients from Cohort 1 and Cohort 2 who received but did not respond to vedolizumab in the induction phase (up to week 6) and received vedolizumab (Q4W, n=506) in the maintenance phase (up to week 52).

The overall incidence of AEs was similar across treatment groups in both the induction and maintenance studies.[11,22] At least one AE was reported in 59% of patients receiving placebo, 56% of patients receiving double-blind vedolizumab, and 57% of patients receiving open-label vedolizumab in the induction study; and 84% of patients receiving placebo, 88% of patients receiving vedolizumab every 8 weeks, and 84% of patients receiving vedolizumab every 4 weeks in the maintenance study. Table 16 provides the most common AEs reported in at least 5% of vedolizumab-treated patients.[11]

Serious AEs occurred more frequently in the vedolizumab groups (24.4%) than in the placebo group (15.3%).[11] In the maintenance study, one case each of latent TB, carcinoid tumours in the appendix, squamous-cell carcinoma, and basal-cell skin carcinoma were diagnosed in the vedolizumab groups, and a borderline ovarian tumour developed in one placebo patient. Five deaths occurred during the study period. Four patients receiving vedolizumab died compared to one in the placebo group. Causes of death in the vedolizumab group were CD with sepsis, intentional overdose of prescription medication, myocarditis, and septic shock. The death in the placebo group was caused by bronchopneumonia. One patient discontinued the study because of a serious infusion reaction (presumably in the vedolizumab group, but this is not clear), and no cases of anaphylaxis were reported. The company stated that the rates of infections and serious infections (5.5% vs 3.0%) were higher in the vedolizumab group compared to the placebo group, although it was not reported if this difference was statistically significant or not. No cases of PML were identified. The ERG notes that an EMA risk management plan exists for vedolizumab.

Table 16 Treatment-emergent adverse events in the overall safety population in the GEMINI II trial (reproduced from Table 6.9.2.1 in CS[1] pg. 159)

Event, n (%)	Placebo[a] (n=301)	Vedolizumab[b] (n=814)
Any AEs	246 (82)	706 (87)
Serious AEs	46 (15.3)	199 (24.4)
Serious infection	9 (3.0)	45 (5.5)
Any cancer	1 (0.3)	4 (0.5)
Adverse events occurring in >5% of Vedolizumab patients, categorized by preferred term		
CD exacerbation	65 (21.6)	164 (20.1)
Arthralgia	40 (13.3)	110 (13.5)
Pyrexia	40 (13.3)	103 (12.7)
Nasopharyngitis	24 (8.0)	100 (12.3)
Headache	47 (15.6)	97 (11.9)
Nausea	30 (10.0)	90 (11.1)
Abdominal pain	39 (13.0)	79 (9.7)
Upper respiratory tract infection	17 (5.6)	54 (6.6)
Fatigue	14 (4.7)	53 (6.5)
Vomiting	23 (7.6)	49 (6.0)
Back pain	12 (4.0)	38 (4.7)

a The placebo group includes patients who did not receive maintenance therapy with vedolizumab (i.e., those who were randomly assigned to placebo during the induction phase plus those who had had a response to Vedolizumab induction therapy and were randomly assigned to placebo for the maintenance trial). † A serious infection was defined as a SAE of infection according to the classification for adverse event reporting in Medical Dictionary for Regulatory Activities (MedDRA).

b The vedolizumab group includes patients who received maintenance therapy with vedolizumab (i.e., those who had had a response to Vedolizumab induction therapy and were randomly assigned to receive vedolizumab every 8 weeks or every 4 weeks as maintenance therapy plus those who did not have a response to vedolizumab induction therapy and continued to receive vedolizumab every 4 weeks during the maintenance trial);

c A serious infection was defined as a serious adverse event of infection according to the classification for adverse event reporting in MedDRA.

d The cancer in the placebo group was a borderline ovarian carcinoma, which is defined as a subset of epithelial ovarian tumours that are considered to be of low malignant potential. The cancers in the vedolizumab group included one case each of basal-cell skin carcinoma, breast cancer, carcinoid tumour in the appendix, and squamous-cell carcinoma of the skin.

GEMINI III[12]

The rates of discontinuation for all ITT participants in the induction trial (GEMINI III) were 7% (28/416) with no notable difference between the combined vedolizumab and placebo groups. In the ITT population, discontinuation due to AEs was reported in 2% (4/209) of placebo patients and in 4% (8/207) of the vedolizumab-treated patients. The ERG notes that the low numbers of discontinuation during this phase is likely to be due to the short duration (6-weeks) of the induction phase.

The overall safety population was defined as all patients who received any amount of study drug.[12,23] The incidence of AEs was similar between the treatments,[12,23] with treatment-emergent AEs reported in 56% and 60% of the vedolizumab and placebo patients, respectively. The most common AEs in the vedolizumab group are reported in Table 17. Among these events, the vedolizumab group had higher incidences of nausea (6% vs. 2%), upper respiratory tract infection (4% vs. 2%), vomiting (4% vs. 2%), fatigue (3% vs. < 1%), and urinary tract infection (3% vs. 0%) compared with the placebo group, whereas the placebo group had higher incidences of CD (10% vs. 3%) and pyrexia (6% vs. 3%) compared with the vedolizumab group.

Serious AEs were reported in 6% of patients receiving placebo and 8% of patients receiving vedolizumab. A breakdown of the serious AEs was not provided in the CS.[1], although it was reported that serious infection AEs occurred in 2 patients in the vedolizumab group and no patients in the placebo group . No cases of PML were reported. No deaths occurred and no serious infusion-related or anaphylactic reactions were reported.[12,23]

Table 17 Treatment-emergent adverse events in the overall safety population in the GEMINI III trial (reproduced from Table 6.9.2.2 in CS[1] pg. 161)

Event, n (%)	**Placebo n=207**	**Vedolizumab n=209**
Any AEs	124	117 (56)
Drug-related AEs	34 (16)	34 (16)
Discontinued because of AEs	8 (4)	4 (2)
Serious AEs	16 (8)	13 (6)
Serious infection	0	2 (<1)
Drug-related SAEs	1 (<1)	1 (<1)
Discontinued because of SAEs	5 (2)	4 (2)
Adverse events occurring in >1% of Vedolizumab patients, categorized by preferred term		
Nausea	5 (2)	12 (6)
Headache	15 (7)	11 (5)
Arthralgia	9 (4)	10 (5)
Nasopharyngitis	8 (4)	9 (4)
Abdominal pain	6 (3)	9 (4)
Upper respiratory tract infection	5 (2)	9 (4)
Vomiting	5 (2)	9 (4)
Pyrexia	13 (6)	7 (3)
Crohn's disease	21 (10)	6 (3)
Fatigue	2 (< 1)	6 (3)
Urinary tract infection	0	6 (3)
Dizziness	4 (2)	5 (2)
Anaemia	1 (< 1)	5 (2)
Aphthous stomatitis	3 (1)	4 (2)
Musculoskeletal pain	0	4 (2)
Diarrhoea	4 (2)	3 (1)
Back pain	3 (1)	3 (1)
Insomnia	3 (1)	3 (1)
Oedema peripheral	2 (< 1)	3 (1)
Oropharyngeal pain	2 (< 1)	3 (1)
Asthenia	1 (< 1)	3 (1)
Decreased appetite	1 (< 1)	3 (1)
Erythema nodosum	1 (< 1)	3 (1)
Hypertension	1 (< 1)	3 (1)
Hypoaesthesia	1 (< 1)	3 (1)
Muscular weakness	1 (< 1)	3 (1)
Dyspepsia	0	3 (1)
Gastroenteritis	0	3 (1)

Supplementary safety evidence

GEMINI LTS (C130008)[45] (see CS[1] pg. 161-163 CS)

The GEMINI LTS[45] is a Phase III, open-label, multicentre, long-term safety study which is ongoing and evaluating vedolizumab in patients with UC and CD. The objective of this study is to collect and characterise important clinical safety events resulting from chronic vedolizumab administration. The primary outcome measures are safety parameters: AEs; serious AEs; results of standard laboratory tests and ECGs; time to major IBD-related events (i.e., hospitalisations, surgeries, or procedures); and improvements in quality of life.

Limited interim results (as of July 2012) are presented in the CS[1] (see CS[1] pg. 161-163) and are summarised in Table 18. The mean age was 41.3 years (standard deviation [SD] 13.30) for patients with UC and 37.7 years (SD 12.52) for those with CD. Vedolizumab exposure was ≥6, ≥12, and ≥24 months for 1534, 1149, and 502 patients, respectively. The safety profile of vedolizumab in this study was similar to that observed in the prior 12-month Phase III trials. Drug-related AEs were similar between CD and UC patients with the most common AEs being headache 6%, nasopharyngitis 4%, nausea 4%, arthralgia 4%, upper respiratory infection 3%, and fatigue 3%. SAEs occurred in <1% of patients, both overall and by indication (UC or CD), except for anal abscess and abdominal pain, which occurred in 2% of CD patients but at a rate less than 1% in UC patients. No cases of systemic candidiasis, disseminated herpes zoster, cytomegalovirus hepatitis or encephalitis, pneumocystis pneumonia or PML were reported. AEs that most commonly led to discontinuation were gastrointestinal, with exacerbations of UC and CD most commonly reported (5% each). Malignancies were observed in <1% of patients (two cases of colon cancer and two malignant melanomas). A breakdown of serious infection and infusion-related reactions was not provided in the CS.[1]

Table 18 GEMINI LTS - interim safety results (as of July 2012) (reproduced from Table 6.9.2.3 in CS[1] pg. 163)

AE category, n (%)	UC Patients (n=704)	CD Patients (n=1118)
Drug-related AE	258 (37%)	447 (40%)
AE leading to discontinuation	61 (9%)	108 (10%)
SAE	127 (18%)	285 (25%)
Serious infection	30 (4%)	74 (7%)
Drug related	15 (2%)	51 (5%)
Leading to discontinuation	23 (3%)	65 (6%)
Death	3 (<1)[a]	3 (<1) [b]

AE, adverse event; CD, Crohn's disease; SAE, serious adverse event; UC, ulcerative colitis

a Respiratory failure, acute stroke, pulmonary embolism

b Septicaemia, traumatic intracranial haemorrhage, suicide

Pooled safety analyses

The company undertook two separate pooled safety analyses[47,11]. The first was of two Phase III, randomised, placebo-controlled, double-blind studies in adults with moderately to severely active UC (GEMINI 1)[47] or CD (GEMINI II)[11] despite previous anti-TNF-α and/or other therapy. The results of this analysis found that patients receiving vedolizumab (300mg vedolizumab i.v. every 4 weeks) had higher rates of overall adverse events and serious adverse events (including gastrointestinal disorders and infections) compared with placebo; however, the overall incidence of adverse events, adjusted for patient-years, was higher for the placebo groups than the vedolizumab group. Further details are provided on pages 163-166 of the CS.[1]

A second pooled safety analysis of six studies included two Phase II trials[48,49], three Phase III trials (GEMINI I, GEMINI II, GEMINI III)[11,12,47] and one open-label long-term safety study (GEMINI LTS).[45] In general, as noted in the CS[1] (see CS[1] pg. 164), the baseline characteristics of the safety population were comparable between studies, with the mean age ranging from 36 to 40 years, approximately 70% of patients with disease activity of >3 years and anti-TNF-α failure ranging from 41% to 75%. The results of this analysis found that the safety profile of vedolizumab was similar between UC (n=1107) and CD patients (n=1723) with the most common adverse events being nasopharyngitis (combined UC and CD group: 18.1%, [511/2830]), headache (combined UC and CD group: 16.1%, [457/2830]) and arthralgia (combined UC and CD group: 15.5% [439/2830]). Further details are provided on pages 165-166 of the CS.[1]

It is reported in the CS[1] that SAEs were low with vedolizumab treatment with the most common being exacerbation of CD, exacerbation of UC, abdominal pain anal abscess. As of June 2013, no cases of PML were reported in any of the >2,700 patients treated with vedolizumab, including approximately 900 patients with ≥24 months exposure. All patients entering vedolizumab studies were pre-screened for TB. Across the integrated safety population, TB was reported in a total of 4 patients (3 with CD, 1 with UC), with all cases occurring within the first 18 months of vedolizumab treatment. No extra pulmonary manifestations or dissemination were reported. However, the absence of long term safety data has been pointed out as a concern in the EMA assessment report[9]. As a result, the risk of PML is being monitored in the post-approval safety studies. A total of 26 vedolizumab treated patients had been diagnosed with malignancy, of which 18 met SAE criteria. Of these skin cancers (n=5) and colon cancer (n=4) were most common.

Limited information on deaths was provided in the CS.[1] As noted in the FDA briefing document [50], a total of 13 deaths (as of June 2013) occurred across all controlled and uncontrolled studies in UC and CD: GEMINI I[47] (UC patient, n=1 [vedolizumab cohort 2 group]), GEMINI II[11] (CD patients, n=5 [1 in placebo group and 4 in vedolizumab group]) and 7 in the GEMINI LTS study[45] (UC patients, n=3; CD patients, n=4).

The EMA assessment report[10] documents 9 post-study deaths occurring up to March 2013 in the vedolizumab clinical program. This includes 2 in GEMINI 1,[47] 1 in GEMINI II[11] and 5 in GEMINI LTS[45] and one in a Phase II study.[42] Of these 9 deaths, sepsis was reported in a total of 3 subjects, malignancies occurred in 2 of the deaths (both UC patients with colon cancer) and the remaining 4 deaths were, cardiorespiratory arrest, multi-organ failure, cardiac arrest and pulmonary embolism. The EMA assessment report [9,10] concluded that none of the post-study deaths could be ascribed with any reasonable degree of certainty to vedolizumab.

4.3 Critique of trials identified and included in the network meta-analysis, and of the networks constructed

In the absence of any direct head-to-head RCTs comparing vedolizumab and infliximab or adalimumab for the treatment of moderate to severe CD, the company conducted a NMA. This is an extension of the conventional pairwise meta-analysis, and allows the combination of direct and indirect evidence from RCTs. This approach allows simultaneous comparisons of multiple treatments from trials comparing different sets of treatments (providing there is a connected network) and ensures that the estimates produced between the pairwise comparators are not discrepant. It is typically performed in a Bayesian manner to allow for all sources of uncertainty and to allow probabilistic statements to be made about population parameters.

The company conducted a systematic review to retrieve published RCTs which had assessed the efficacy and safety of biologic therapies prescribed for the treatment of CD. The methods for the systematic review are described in Section 4.1. As reported in the CS[1] (see CS[1] pg. 64) the review and NMA had a global remit and therefore included biologic therapies not licenced in the UK for CD (certolizumab and natalizumab). The company states that data presented in the CS[1] would not include these two drugs, however some of these data do remain in some of the networks; the CS[1] states (pg. 127) that *'the inclusion of appropriate evidence for treatments in the network not licenced in the UK is not expected to affect the integrity of the analyses.'* The ERG agrees that the inclusion of these data should not be considered problematic. It should be noted that some of those trials will have contributed to the baseline placebo response in the economic model. The data relating to these studies has not been presented in the CS[1] but study and patient characteristics are included in the Appendix of this report.

The company's systematic review[1] identified 18 RCTs that compared vedolizumab, infliximab, adalimumab, certolizumab or natalizumab with placebo in adult patients with moderate to severe CD, of which the company state that 10 are relevant to this appraisal (see CS[1] PRISMA diagram pg. 71). Due to variability in terms of patients recruited, outcomes and subgroups reported, not all trials contributed to all analyses.

This critique of the NMA networks considers relevance to UK practice, the decision problem and clinical heterogeneity within the networks. The critique of the trials identified and included in the NMA includes a consideration of the study design and methodological quality (risk of bias) of the included studies together with an assessment of the heterogeneity and relevance of studies in terms of clinical characteristics and UK practice, using the PICOS framework.

Clinical relevance of networks

The CS[1] presents a matrix of networks (see CS[1] pg. 137), which were presented in Appendix 5 of the CS. This matrix is reproduced in Table 19. As can be seen from this table, networks were constructed for both the induction and maintenance phases, for four main outcomes (clinical response, enhanced clinical response, clinical remission and discontinuation due to AEs), in three main populations (Anti-TNF-α-naïve, anti-TNF-α experienced/failure and the entire population). Clinical response, enhanced clinical response and remission were modelled separately using a logistic model. The models are reported in Section 6.7 of the CS.[1]

Table 19 Summary of data available for the analyses that are presented in Appendix 5 (reproduced from Table 6.7.4.1 in the CS[1])

Study Population (Study Phase)	Clinical Response (drop in CDAI ≥70)	Enhanced Clinical Response (drop in CDAI ≥100)	Clinical Remission	Discontinuation due to AEs
Anti-TNF-α Naïve (Induction)	√*	√*	√*	√*
Anti-TNF-α Naïve (Maintenance)	√*		√*	√*
anti-TNF-α Experienced/Failure (Induction)	√*	√*	√*	√*
Entire population (Induction)	√	√	√	√
Entire population (Maintenance)	√	√	√	√

* Diagrammatic representation of network provided

The CS[1] presented four main network diagrams, reproduced here as Figures 4 to 7. The networks were for:

- anti-TNF-α-Naïve, induction:
  - clinical response (drop in CDAI ≥70)
  - clinical remission (CDAI <150)
- anti-TNF-α-Naïve, induction:
  - enhanced clinical response (drop in CDAI ≥100)
  - discontinuation due to AEs
- anti-TNF-α-Naïve, maintenance:
  - clinical response (drop in CDAI ≥70)
  - clinical remission (CDAI <150)
  - discontinuation due to AEs
- anti-TNF-α failure experienced/failure, induction:
  - clinical response (drop in CDAI ≥70)
  - enhanced clinical response (drop in CDAI ≥100)
  - clinical remission (CDAI <150)
  - discontinuation due to AEs

The networks for the entire population were not presented diagrammatically in the CS,[1] but were presented in the Takeda data on file document,[16] along with several other networks and sensitivity analyses, including networks for serious adverse events. The ERG has not critiqued these networks in detail.

Figure 4 Network diagram of the interventions compared for the outcomes of clinical remission and clinical response (drop in CDAI ≥ 70) in the anti-TNF-α **–Naïve sub-population in induction treatment (reproduced from Figure 6.7.3.1 in CS[1] pg.135)**

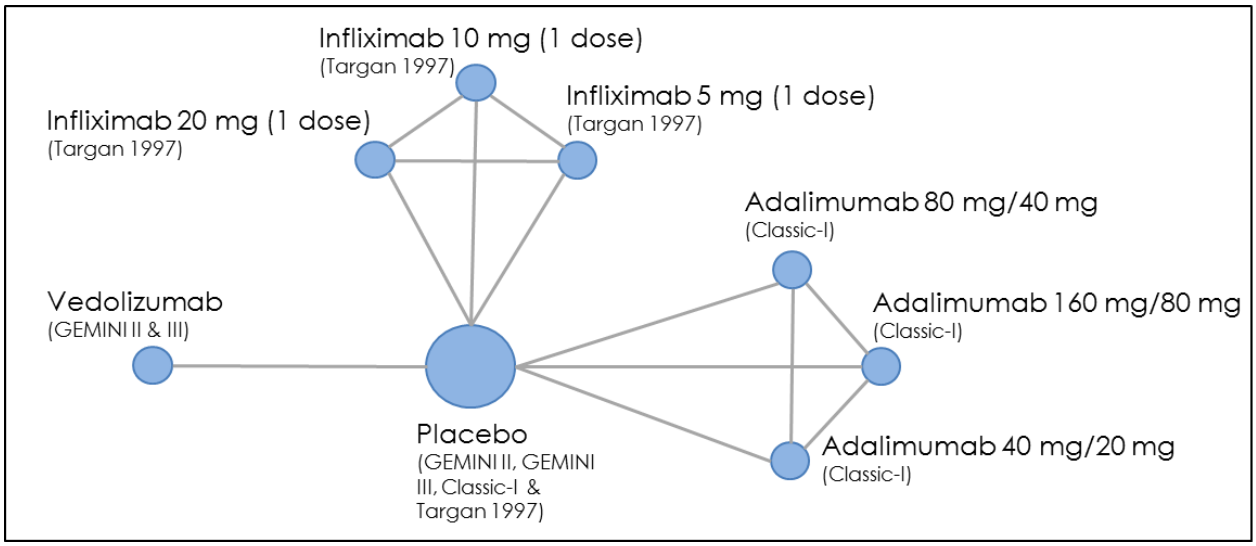


Figure 5 Network diagram of the interventions compared for the outcomes of enhanced clinical response (drop in CDAI ≥ 100) and discontinuation due to AE in the anti-TNF-α –Naïve sub-population in induction treatment (reproduced from Figure 6.7.3.2 in CS[1] pg.135)

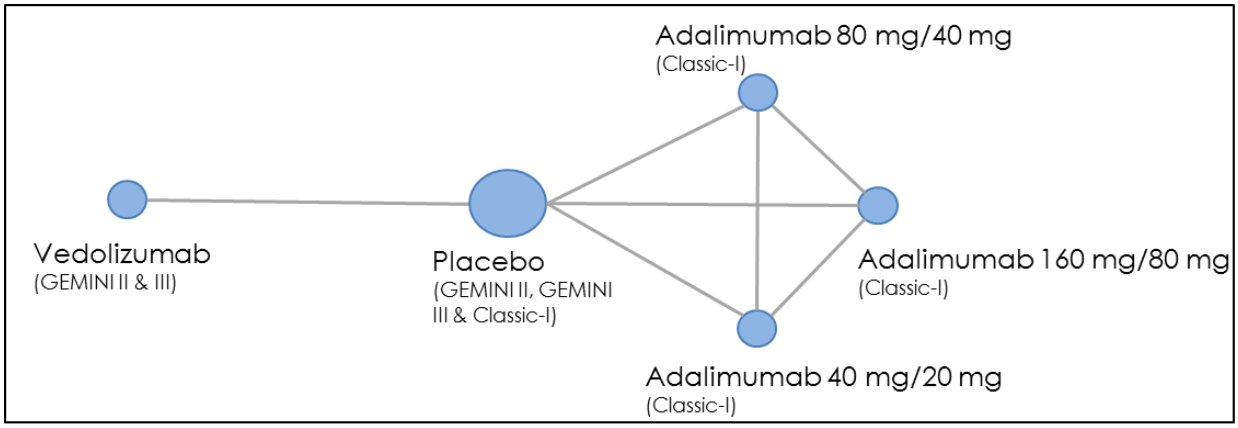


Figure 6 Network diagram of the interventions compared for the outcomes of clinical remission, clinical response (drop in CDAI ≥ 70) and discontinuation due to AE in the anti-TNF-α –Naïve sub-population in maintenance treatment (reproduced from Figure 6.7.3.3 in CS[1] pg.136)

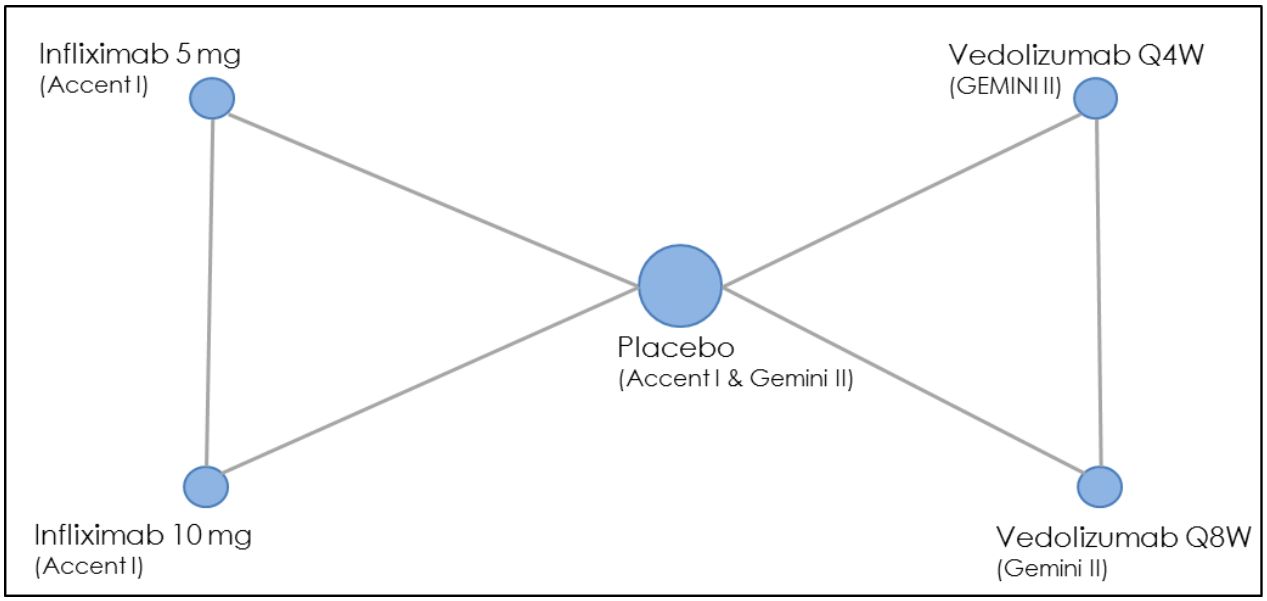
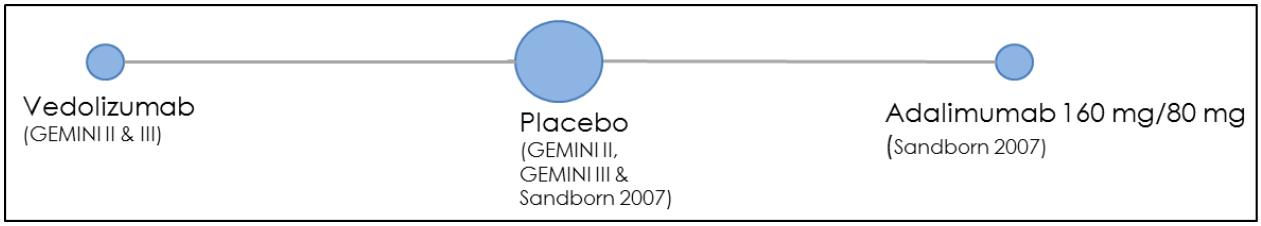


Figure 7 Network diagram of the interventions compared for the outcomes of clinical response (drop in CDAI ≥ 70), enhanced clinical response (drop in CDAI ≥ 100), clinical remission and discontinuation due to AEs in the anti-TNF-α **–Experienced/Failure sub-population in induction treatment (reproduced from Figure 6.7.3.4 in CS[1] pg.136)**



The ERG has considered the relevance of the listed networks to the scope issued by NICE.[8] Within the scope,[8] the population defined is:

"*Adults with moderately to severely active Crohn's disease in whom the disease has responded inadequately to, or is no longer responding to, either conventional therapy or a TNF-α antagonist, or who are intolerant to either of them.*"

Within this population patients are at different stages in the UK treatment pathway. There are (categories not mutually exclusive):

- people in whom the disease has responded inadequately to, or is no longer responding to conventional therapy and who have not previously received an anti-TNF-α (anti-TNF-α naïve);
- people in whom the disease has not responded to an anti-TNF-α at the induction phase (primary non-responders);
- people in whom the disease has responded to an anti-TNF-α at the induction phase but whom the disease has not responded to the maintenance phase (secondary non-responders);
- people who are intolerant to or contraindicated against anti-TNF-α therapy;
- people in whom the disease has not responded to or who have lost response to both anti-TNF-α available in the UK (adalimumab and infliximab).

Clinical advice to the ERG suggests that people in whom prior treatment with anti-TNF-α failed are less likely to respond to a different anti-TNF-α than other patients. This is also stated in the CS[1] (see CS[1] Section 2.6).[1] This potentially causes difficulties in the interpretation of efficacy results in the mixed population.

With respect to the treatment pathway in the UK, those patients coming directly from conventional therapy are probably best represented by the anti-TNF-α naïve subgroup. This is the primary NMA analysis presented in the CS.[1] Those patients who have already tried one or more anti-TNF-α treatments and whose disease did not respond or lost response are best represented by the anti-TNF-α failure subgroup, which is a network meta-analysis presented in the CS.[1] Limitations to these analyses are discussed in later sections.

The "entire population" analysis presented within the CS[1] is problematic in terms of interpretation of the results. This analysis included studies which did not select patients on the basis of anti-TNF-α failure, but also some that either included or excluded patients on this basis,[19,20,51-53] or included a pre-specified proportion who were/were not (as in the case of GEMINI II and GEMINI III).[11,12] As such, the study populations within the NMA for the "entire population" are not clinically homogeneous and the results may not represent a clinically meaningful population. The presentation of this analysis as a secondary analysis is appropriate, however, as it was not known a priori whether failure to previously available anti-TNF-α treatments would confer a higher risk of failure to vedolizumab, as vedolizumab has a different mode of action to adalimumab and infliximab, and may conceivably not be subject to the same issue. In this case, evidence from several studies would have been rejected unnecessarily. On this point, it is worth noting that the analyses provided in GEMINI trials[11,12] show a different response level in naïve versus failure patients, which supports the hypothesis that those who fail treatment with a previous anti-TNF-α are less responsive to vedolizumab.

Quality assessment of studies included in the NMA

The CS[1] provides a table assessing the quality of studies that were included in the network and which assessed UK-licenced treatments (see CS[1] Appendix 5). The ERG has reproduced this table with some additions and comments given in footnotes (see Table 20). As can been seen from this table, risk of bias was generally low. Whilst no studies were of a quality low enough to imply the need for sensitivity analyses for quality reasons, some studies did not score low risk for all items.

- two studies did not describe the methods of randomisation.[20,41]
- there is some doubt around which studies conducted ITT analyses. Hanauer et al.[51] and Watanabe et al.[20] were scored as having done so in the CS[1] quality assessment, but the ERG could not verify this from the primary publications[20,51] (see footnote in Table 20). Targan et al.[19] was scored as not having performed an ITT analysis, though the description within the primary publication[19] appeared to the ERG to describe such an analysis (see Table 20 footnote), though this study failed to impute missing data for one missing participant. All four other studies appeared to conduct ITT analysis, and all imputed missing data as failures.[11,12,52,53]

Table 20 Quality assessment of studies included in the NMA, adapted from Appendix 5 of the CS[1]

Study	Induction or maintenance?	Networks included in?	Was randomization appropriate?	Was the concealment of treatment allocation adequate?	Were the groups similar at the outset of the study in terms of prognostic factors?	Were the care providers, participants, and outcome assessors blind to treatment allocation?	Did the analysis include an intention-to-treat analysis?	Were there any unexpected imbalances in drop-outs between groups?	Is there any evidence to suggest that the authors measured more outcomes than they reported?
Induction									
GEMINI-II[11]	Induction	Naïve; failure; all	Yes	Yes	Yes	Yes	Yes	No	No
GEMINI-III[12]	Induction	Naïve; failure; all	Yes	Yes	Yes	Yes	Yes	No	No
CLASSIC I (Hanauer et al.[51])	Induction	Naïve	Yes	Yes	Yes	Yes	Yes*	Yes	No
Targan et al., 1997[19]	Induction	Naïve; all	Yes	Yes	Yes	Yes	Not clear**	Not clear	No
NCT00105300 Sandborn et al.[52]	Induction	Failure	Yes	Yes	Yes	Yes	Yes	No	No
NCT00445939 (Watanabe et al.[20])	Induction	Naïve***; all	Not clear	Yes	Yes	Yes	Yes****	No	No
EXTEND (Rutgeerts et al., 2012)[53]	Induction	All	Yes	Yes	No	Yes	Yes	No	No
Maintenance									
GEMINI-II[11]	Maintenance	Naïve; all	Yes	Yes	No	Yes	Yes	No	No
ACCENT I (Hanauer et al.[54])	Maintenance	Naïve; all	Yes	Yes	Not clear	Yes	Yes	No	No
CLASSIC II (Sandborn et al.[41])	Maintenance	All	Not clear	Yes	No	Yes	Yes	No	No
NCT00445432 (Watanabe et al.[20])	Maintenance	All	Not clear	Yes	Yes	Yes	Yes	No	No

CHARM (Colombel et al.[55])	Maintenance	All	Yes	Yes	Not clear	Yes	Not clear	No	No
EXTEND (Rutgeerts et al.[53])	Maintenance	All	Yes	Yes	No	Yes	Yes	No	No

* ERG were not able to verify this score from the primary publication: though all patients were included in the efficacy analyses it was not clear if they were assessed according to their treatment assignment
** study states "The original study protocol did not specify the use of intention-to-treat analysis… all patients were analyzed according to the treatment to which they were assigned" In addition, patients not receiving study drug were excluded from analysis. This analysis seems to be in accordance with broad definitions of ITT analysis.
*** anti-TNF-α naïve patient data not identified in CS, or included in MTC.
**** study states "All patients enrolled in the induction trial who were randomly assigned to 1 of the 3 treatment groups and received at least 1 dose of study drug constituted the full analysis set (FAS) of the induction trial and were included in the primary and secondary efficacy analyses, and in the safety analyses for the induction trial." But it was not clear if they were assessed according to their treatment assignment.

Induction studies

This section considers the specifics of included induction studies both as a whole body of evidence, and where relevant, in the context of the networks they have been included in.

Study and patient characteristics

Study and patient characteristics for induction studies are presented in Tables 21 and 22. All studies aimed to recruit patients with moderate to severe CD, and definitions of this were broadly similar. As with the GEMINI II and III trials,[11,12] the definition usually extended to CDAI 450, which means that the upper range of severe patients (usually defined as up to a CDAI score of 600) was missing. There were some variations in exclusion criteria amongst studies; clinical advice to the ERG suggested that the patient characteristics most likely to have a response-modulating effect were:

- the proportion of patients with fistulising disease, as these patients may respond differently to anti-TNF-α treatment than non-fistulising patients
- the proportion of patients with stricturing disease, as these patients may respond differently to anti-TNF-α treatment than patients without stricturing disease
- the proportion of patients who have not responded or lost response (whilst on treatment) to previous anti-TNF-α treatment (referred to as failures in the CS[1]), as these patients are less likely to respond to other anti-TNF-α treatments
- The severity of disease in the patients, as assessed by:
    - CDAI
    - Faecal calprotectin
    - CRP

Stricturing disease; all studies: As can be seen from Table 21, GEMINI II,[11] GEMINI III,[12] CLASSIC I[51] and Targan et al. 1997[19] all excluded at least some patients with strictures. Three adalimumab trials[20,52,53] did not list these as exclusion criteria and are assumed to have included these patients. These differences in patient spectrum may affect the estimates of efficacy; however the extent of this is unclear.

Fistulas; all studies: It was not possible to assess whether the proportion of active fistulas in the studies were comparable, as data were not reported in four studies.[12,19,20,53] Of the three remaining studies, GEMINI II[11] had the highest proportion (17.3% in the intervention arm, 15.5% in the placebo arm), but differences were small (other study arms were 8, 5, 13, 16, 15 and 13%). Clinical advice to the ERG suggests that these proportions were within normal UK ranges.

Table 21 Induction studies: key patient characteristics

Study	Population	Treatment	% TNF naïve	% TNF failure	Mean baseline CDAI (Intervention: control)	Fistulising	Stricturing
GEMINI II CSR13007[22]	Moderate to severe: Failed at least one CT (IMs & anti-TNF, unclear if CS)	Vedolizumab	50; 51	47.7; 47.3	324.6; 327.3	History of fistulising disease: 37% Draining fistulae at baseline: Intervention 17.3%; placebo 15.5%	Intestinal stricture patients excluded
GEMINI III CSR 13011[23]	Moderate to severe: Inadequate response, loss of response, intolerance of 1 or more IM or anti-TNF	Vedolizumab	24.2-24.4	76; 76	297.4; 311.4	History of fistulising disease: 36%	Intestinal stricture patients excluded
CLASSIC I Hanauer et al.[51]	Moderate to severe: Anti-TNF-naive	Adalimumab	100	0	295; 301	Enterocutaneous or perianal fistula at screening and baseline (% in placebo, 160mg, 80mg and 40mg groups): 8, 5, 13, 16	Excluded symptomatic strictures
Targan et al.[19]	Moderate to severe (CDAI 220 to 400): excluded those exposed to humanized anti-TNF-α (Certolizumab; Natalizumab)	Infliximab	NR	NR	288 to 318	NR	Excluded those with symptomatic stenosis or ileal strictures
NCT00105300 Sandborn et al.[52]	Moderate to severe: all intolerant or lost response to infliximab. Excluded primary non-responders	Adalimumab	0%	Lost response: 52; 48 Intolerant: 57; 60	313	Abdominal or perianal fistula at baseline: 15;13	

Study	Population	Treatment	% TNF naïve	% TNF failure	Mean baseline CDAI (Intervention: control)	Fistulising	Stricturing
				Lost response and intolerant: 13; 12			
NCT00445939 Watanabe et al.[20]	Moderate to severe: Prior exposure to anti-TNF-α other than adalimumab allowed, but excluded primary non-responders	Adalimumab	42%	Primary non-responders: 0%	303.3 (SD65)	NR	
EXTEND Rutgeerts et al.[53]	Moderate to severe: Prior exposure to anti-TNF-α other than adalimumab or natalizumab allowed, but excluded primary non-responders.	Adalimumab	48%	Primary non-responders: 0%	320 (SD 70)	NR	

IMS, immunomodulators; CS, corticosteroids; NR, not reported; SD, standard deviation.

Anti-TNF-α failures; all studies: The potential impact of the proportion of anti-TNF-α failure patients is mitigated by the decision of the company to analyse data for anti-TNF-α naïve patients as a separate network, and making this the primary analysis. Furthermore, the ERG believe that in relation to the treatment pathway, anti-TNF-α naïve subgroups are likely to be the best (potentially exact) match to populations presenting post-conventional therapy failure who have not yet received first-line anti-TNF-α treatment. This results in the exclusion of three trials from the primary network,[20,52,53] all of which are adalimumab trials, leaving only one adalimumab trial in the network.[51] For the "entire population" network where all studies were included regardless of prior anti-TNF-α experience or failure, the proportion of anti-TNF-α failure patients ranges from 0%[52] to 100%,[51,56] which potentially could impact on estimates of efficacy.

Severity; all studies: Studies were similar in terms of baseline CDAI (means ranging from 262 to 327). However, data relating to faecal calprotectin and CRP levels were not presented in the CS,[1] and have not been assessed by the ERG due to time restraints.

In addition there are some specific comments relating to specific networks:

Anti-TNF-α naïve network; missing data: Watanabe et al.[20] is excluded from the primary analysis (i.e. the analysis of anti-TNF-α naïve subgroups) on the basis of no data for the anti-TNF-α naïve subgroup being reported (see CS[1] Table 6.7.3.3 in the listed under "*Trials included in the MTC but excluded from primary analysis for only presenting mixed anti-TNF-α exposure data*"). However the ERG identified data in Watanabe et al.[20] for the induction phase for the anti-TNF-α naïve subgroup that appears to have been missed by the CS,[1] as it neither appears in the Takeda data on file document,[16] nor in the data extraction spreadsheet embedded in that document. These missing data appear to have been eligible for inclusion in the anti-TNF-α naïve population network. The odds ratio of response (reviewer calculated) for treatment with 80/40mg regime of adalimumab was 2.72 (95% CI 0.51 to 14.49) in Watanabe et al.[20] versus 2.47 (95% CI 1.28 to 4.78) in Hanauer et al.,[51] the only other adalimumab study included in the NMA. As a smaller study (n=57 versus n=149) with less precise results, its impact is likely to have been relatively small. It should also be noted that this particular study had other limitations which may or may not have impacted on its estimates of efficacy: from a risk of bias perspective, randomisation was not well described and it scored poorly for this item; the analysis of patients by anti-TNF-α exposure was a post-hoc analysis; the population was Japanese, and ethnicity may impact on responsiveness; and the study was judged by the company in the "entire population" analysis to be an outlier in investigations of heterogeneity for having a small placebo response, and was removed in a sensitivity analysis. It is unclear whether it would also have been judged heterogeneous in an analysis of the anti-TNF-α naïve subgroup, had the data been identified and included in this analysis.

Regardless of these potential issues, the ERG believes that this study should have been included in the NMA of the anti-TNF-α naïve subgroup as it met all inclusion criteria, and (at most) removed in a sensitivity analysis for concerns about its quality and heterogeneity.

Anti-TNF-α failure/experienced network; population: The failure/experienced network included three studies, namely GEMINI II, GEMINI III and Sandborn 2007,[11,12,41] which reported vedolizumab, vedolizumab and adalimumab results respectively. As noted in the CS,[1] Sandborn et al., 2007 included those who were intolerant or lost response (secondary non-responders), whilst GEMINI II[11] and GEMINI III[12] included primary non-responders as well as secondary non-responders and those who were intolerant. As such, the populations in the GEMINI trials are potentially likely to be less responsive to treatment, and in comparison to Sandborn et al.,[52] may produce underestimates of efficacy. This should be borne in mind when interpreting the results of this network.

"Entire population" network; population: This network was not formally presented in the CS,[1] and has already been discussed above in the paragraph entitled "anti-TNF-α failures: all studies". As previously stated, the study populations within the NMA for the "entire population" are not clinically homogeneous and the results may not represent a clinically meaningful population.

Interventions

All studies: Not all studies administered full induction periods as stated in the relevant SmPC, as they assessed outcomes before the end of the recommended induction period. Data relating to induction periods are listed in Table 22, and discussed in more detail in the "outcomes" section below. In brief and of particular relevance to the treatment under assessment, GEMINI II[11] only administered two vedolizumab doses during the induction phase, whereas the licence calls for three doses.[9,10]

However, within the time period of assessment all studies used a dosing regimen in accordance with UK licensing except one.[19] Targan et al.[19] was included in the anti-TNF-α naïve network and used only a single dose of infliximab and assessed response at 4 weeks, where a second dose at week 2 should have been administered to conform to UK licensing.[57] This is likely to underestimate the efficacy outcomes for infliximab in this network, but may also underestimate adverse events. This is the only available induction trial with a placebo arm for infliximab.

Adalimumab has two recommended doses,[58] the 160/80mg (accelerated) dose and the 80/40mg (normal) dose. The former is recommended for use when a rapid response is required, and only with consideration of the impact of increased adverse events. Two trials were multi-arm trials[20,51] and used both doses in different arms, whilst two trials only used the accelerated dose. It is unclear to what

extent the 160/80 mg dose is used in practice in the UK. One trial of adalimumab, which used only the accelerated dose was included in the failure/experienced network.[52] This may impact on generalizability of these results to UK practice, though it may also be a clinically relevant dose for this failure population. Both accelerated dose trials were included in the "entire population" analysis.[52,53]

Comparators

All studies used usual care with placebo as the comparator. Details of what "usual care" comprised in each study was not presented in the CS.[1] It is possible that usual care has changed over time, which may contribute to the low placebo rate seen in earlier trials, for example the infliximab trial which had a low placebo rate and was conducted more than 15 years ago.[19] A meta-regression (by date) was not attempted to explore this.

Outcomes

Outcomes of clinical response (drop in CDAI≥70), enhanced clinical response (drop in CDAI≥100) and clinical remission (CDAI≤150) were remarkably well standardised, though not all studies reported all outcomes. One study used a slightly more stringent definition of clinical response, where the fall in CDAI had to be independent of a change in concomitant medications.[19] This is likely to result in a bias towards underestimation rather than overestimation of treatment effect; the study in question was of infliximab,[19] which was shown to be superior to other treatments in all NMA where it was included. The most comprehensive networks were available for clinical response and clinical remission.

The time of assessment of outcomes is more problematic (this section is also relevant to the critique of the interventions delivered above). Discussion with clinical experts indicated that in practice, response is typically assessed between 10 to 14 weeks, but response may be assessed sooner in accordance with the licensing of the drugs. Specifically, in relation to the evidence included in the review:

- For adalimumab, the SmPC[58] states that response should be assessed at 4 weeks, though it is noted that some patients who have not responded by week 4 may show a response by week 12 if they continued on maintenance therapy. As such, both week 4 and week 12 could be considered acceptable assessment points in adalimumab trials.
    - CLASSIC I,[51] NCT00445939[20] and NCT00105300[41] assessed patients at the 4 week time point. Patients who may have responded by week 12 are therefore missing from these assessments. It is unclear how this impacts on generalizability to UK practice, and whether exclusion of these patients would have affected estimates of efficacy.
        - This affects the anti-TNF-α naïve network, the failure/experienced network and the "entire population" network.

- EXTEND[53] assessed response at 12 weeks and 52 weeks. The CS classed this study as an induction trial. However, it should be noted that this study is described by its authors as a maintenance trial: all patients received the accelerated dose before being randomised to placebo or ongoing therapy at week 4. As such, it provides some information about induction response at 12 weeks, but the placebo arm also received treatment during the first 4 weeks, which may impact on the comparative efficacy.
    - This affects the "entire population" network only.

- For infliximab, the SmPC[57] states that assessment should be conducted at week 6, and treatment stopped after two doses if no response is observed. As such, the induction period is unambiguously 6 weeks long and comprises two doses (week 0 and 2).
    - Targan et al.[19] delivered fewer doses that recommended in the licence, and assessed response at the wrong time point (4 weeks), the likely effects of which are underestimation of treatment effect.
        - This affects the anti-TNF-α naïve and the "entire population" networks
- For vedolizumab, the induction period is not clearly stated, but maintenance therapy starts in patients from week 14, implying this is the induction period. Patients not responding at week 10 can be given another dose at this time point, and reassessed at week 14. As such, week 10 and 14 are considered acceptable assessment time points in vedolizumab trials.
    - GEMINI II[11] assessed patients at week 6 rather than week 14, and patients received 2 doses instead of the 3 recommended dose in the SmPC.[9,10]
    - GEMINI III[12] assessed patients at week 6 and 10 rather than week 14, and administered correct dosages up those points
        - This affects all networks.

In addition, the company have chosen to provide an analysis of "discontinuations due to adverse events". It is not clear what clinical relevance this outcome has. An analysis of serious adverse events would have been more appropriate and informative.

Study design

Most studies followed a standard RCT design without a run-in period. Four studies included both an induction phase and a subsequent maintenance phase,[11,20,51,53] which were assessed separately. In theory, this should not affect estimates of efficacy, however, as already described, Rutgeerts et al.[53] included a run-in period where all patients were treated with adalimumab from week 0 to week 2, and patients were randomised at week 4, stratified for response to adalimumab. It is unclear how the doses received by the placebo group may impact on estimates of efficacy at 12 weeks. It should be noted that this study was not included in the primary analysis for the anti-TNF-α naïve subgroup so does not impact on this network, but is included in the secondary analysis including all patients.

Table 22 Induction studies: treatment regimens and outcome analyses available

Study	Analysis methods	Intervention/Comparator (n=randomised)	Population	Outcome Time point (week)			Comparison to UK licence
				CR	ECR	CRem	
GEMINI II CSR13007[22]	ITT Missing data counted as failures	Vedolizumab 300 mg week 0 and 2 (n = 220) Placebo (n = 148)	Mixed: Naïve: Experienced: Failure:	6 6 NR 6	6 (P) 6 NR 6	6 (P) 6 NR 6	Licenced for response to occur up to 14 weeks and should include an additional dose at week 6 and optional dose at week 10
GEMINI III CSR 13011[23]	ITT Missing data counted as failures	Vedolizumab 300 mg at 0, 2, and 6 weeks (n = 209) Placebo (n = 207)	Mixed: Naïve: Experienced: Failure:	6* 6*, 10* NR 6*, 10*	6, 10 6, 10 NR 6, 10	6, 10 6, 10 NR 6 (P), 10	Licensed for response to occur in up to 14 weeks and can include an optional dose at week 10
CLASSIC I Hanauer et al.[51]	NR if ITT Missing data counted as failures	Adalimumab (week 0/week 2) 160mg/80mg (n = 76)** 80mg/40mb (n = 75) 40mg/20 mg (n = 74) Placebo (n = 74)	Mixed: Naïve: Experienced: Failure:	NR 1, 2, 4 NR NR	NR 1, 2, 4 NR NR	NR 1, 2, 4 (P) NR NR	Week 4 is end of induction period, but licenced for a response to occur in up to 12 weeks 40mg/20mg not UK dose
Targan et al.[19]	ITT*** No imputation (1 pt missing at 4 weeks)	Infliximab, single administration 20 mg/kg (n = 28) 10mg/kg (n = 28) 5 mg/kg, (n = 27) Placebo n = 25	Naïve to humanized anti-TNF	2,4 (P)	NR	2,4	Licenced for second dose at week 2 and assessment at week 4. Only 5mg/kg is UK dose
NCT00105300 Sandborn et al.[52]	ITT Missing data counted as failures	Adalimumab 160 mg at week 0 and 80 mg at week 2 (n = 159) Placebo n = 166	Failure (Intolerant or lost response to infliximab)	4	4	1, 2, 4 (P)	Week 4 is end of induction period, but licenced for a response to occur in up to 12 weeks Dose is the accelerate regime**
NCT00445939	NR if ITT Missing data	Adalimumab at week 0/week 2 (n) 160 mg/80mg (n = 33)	Mixed (excluded primary non-	2, 4 2,4	2, 4 2,4	2, 4 (P)	Week 4 is end of induction period, but licenced for a response

Study	Analysis methods	Intervention/Comparator (n=randomised)	Population	Outcome Time point (week)			Comparison to UK licence
				CR	ECR	CRem	
Watanabe et al.[20]	counted as failures	80mg/40mg (n = 34) Placebo n = 32	responders): Naïve:			2,4	to occur in up to 12 weeks Both are licensed doses**
EXTEND Rutgeerts et al.[53]	ITT Missing data counted as failures	Adalimumab 160mg/80mg at weeks 0 and 2. Then 40 mg every other week from week 4 to 52 n = 64 Comparator: As treatment arm weeks 0 and 2. Then placebo every other week. N=65	Mixed (excluded primary non-responders):	NR	NR	12	Week 4 is end of induction period, but licenced for a response to occur in up to 12 weeks Accelerated regime, then normal maintenance**

CR, clinical response (drop in CDAI ≥70); ECR, enhanced clinical response (drop in CDAI≥100); CRem, clinical remission (CDAI score ≤150); P, primary analysis; NR, not reported

* Not listed in CS or CSR, listed in Takeda data on file
** accelerated regime should only be used where a rapid response is required, and with considerations of increased adverse events.
*** study states "The original study protocol did not specify the use of intention-to-treat analysis… all patients were analyzed according to the treatment to which they were assigned" In addition, patients not receiving study drug were excluded from analysis. This analysis seems to be in accordance with broad definitions of ITT analysis.

Maintenance studies

Study and patient characteristics

Study and patient characteristics for maintenance studies are presented in Tables 23 and 24 respectively. As with induction studies, all aimed to recruit moderate to severe patients at baseline (before induction therapy), with broadly similar definitions, but the upper range of severe patients were missing from all studies, as none reported recruiting patients with CDAI>450.

With reference to the patient characteristics believed by the clinical advisors to the ERG to be most likely to have a response-modifying effect (fistulising disease; stricturing disease; anti-TNF-α failures; disease severity), general observations are as follows:

- fistulas at induction baseline are largely comparable across groups, with proportions ranging from 11% to 16% in all studies.
- symptomatic strictures were excluded from one study in the primary NMA analysis (GEMINI II)[11] which may bias results possibly towards an overestimate of efficacy compared with populations where strictures are included, however, it is not clear to what extent. All other studies did not state that patients with strictures were excluded, apart from CLASSIC II (adalimumab) where these data were only included as a sensitivity analysis.[41]
- disease severity at induction baseline was similar across studies (range of means CDAI 295 to 325). Severity at randomisation to maintenance phase was only reported for one study.[41]

There are also some specific observations relating to each network:

Anti-TNF-α naïve NMA, maintenance: The primary NMA analysis included only anti-TNF-α naïve patients, which the ERG agrees is appropriate. This left a network of only two studies[11,54] covering vedolizumab and infliximab.

- This resulted in the exclusion of three studies[20,53,55] which had anti-TNF-α failure subgroups of 0%, unreported and 0% respectively. Though Watanabe et al.[20] reported data for anti-TNF-α naïve patients in the induction phase of their trial, and these data were missed by the CS,[1] they did not report data for this group for the maintenance phase. As such, the network does not provide an estimate of efficacy for adalimumab.

Table 23 Maintenance studies: key study and patient characteristics

Study	Population at recruitment • Mean age (years) • % Male	Treatment	Induction phase: time and dose	Eligibility for maintenance phase: assessment time and criteria	% TNF naïve at baseline	% TNF failure at baseline	Mean baseline CDAI	Fistulising disease	Strictures
GEMINI II[11] CSR13007[22]	Moderate to severe: Failed at least one CT (IMs & anti-TNF, unclear if CS) • 34.9-38.6 • 44-53	Vedolizumab	week 0: 300mb i.v. week 2: 300mg i.v.	week 6: Clinical response	Induction: 46; 46; 43 Maintenance: NR	Induction: 47.7; 47.3 Maintenance: NR	Induction: 317; 325.5;325.2	History of fistulising disease: 37% Draining fistulae at baseline: 14.8%	Intestinal stricture excluded
ACCENT I Hanauer et al.[54]	Moderate to severe (CDAI 220-400). CT failure NR • 35-37 • 38-39	Infliximab	week 0: 5mg/kg i.v.	week 2: Clinical response AND 25% reduction in CDAI	100	0	Induction: 297 (260–342) week 2 responders: 299 (264–342)	NR	NR
CHARM Colombel et al.[55]	Moderate to severe (CDAI 220-450). CT failure NR ▪ 36.7, 4-week responders; 37.1 all patients	Adalimumab	week 0: 80mg SC week 2: 40mg SC	week 4: Primary outcome is for clinical responders only, though all patients were randomised, stratified by	Induction: 50.4 week 4 responders: 52.3	NR	Induction: 313.9 week 4 responders: 316.6	Enterocutaneous or perianal fistula at both screening and baseline Induction: 15.2% week 4	NR

Study	Population at recruitment • Mean age (years) • % Male	Treatment	Induction phase: time and dose	Eligibility for maintenance phase: assessment time and criteria	% TNF naïve at baseline	% TNF failure at baseline	Mean baseline CDAI	Fistulising disease	Strictures
	▪ % male: 37.7, 4-week responders; 38.2 all patients			responder status and previous exposure to anti-TNF.				responders: 12.8%	
EXTEND Rutgeerts et al.[53] [[p. 1102]]	Moderate to severe CD (with mucosal ulceration) ▪ 37.1-37.2 ▪ 37-38	Adalimumab	week 0: 160 mg week 2: 80 mg	week 4: Clinical response	48.1	0%: Primary non-responders excluded	319.9	≥1 Draining cutaneous fistulas (all perianal) 12.4%	NR
Watanabe et al.[20] NCT00445432	Moderate to severe CD (primary anti-TNF-α non responders were excluded) ▪ 30.8-31.6 ▪ 60-64	Adalimumab	week 0: 160 or 80mg week 2: 80 or 40mg	week 4: Clinical response*	Induction: 42 Maintenance: 46*	0%: Primary non-responders excluded	Induction: 303.3 (SD 65.2) Maintenance: 311.1 (SD 64.9)*	NR	NR
CLASSIC II Sandborn et al.[41] [[p. 1232]]	Moderately to severely active CD who were naïve to anti-TNF-α therapy at time of	Adalimumab	Classic I trial: week 0: 160, 80 or 40mg week 2:	Clinical remission at week 0 (week 4 of CLASSIC I) and week 4	100	0	Induction: 295; 301 Maintenance: 106; 88; 107	Enterocutaneous or perianal fistula at screening and baseline (% in placebo,	Excluded symptomatic strictures

Study	Population at recruitment • Mean age (years) • % Male	Treatment	Induction phase: time and dose	Eligibility for maintenance phase: assessment time and criteria	% TNF naïve at baseline	% TNF failure at baseline	Mean baseline CDAI	Fistulising disease	Strictures
	induction (CLASSIC I) • 34-38 years • 33-50		80, 40 or 20mg Classic II: week 0/4: 40 mg week 2/6: 40 mg	(i.e. Pts had a total of 4 doses over 6 weeks before selection at week4/ 8)				160mg, 80mg and 40mg groups): 8, 5, 13, 16	

CT, conventional therapy; IMs, immunomodulators; CS, corticosteroids; i.v. intravenous; NR, not reported; SC, subcutaneous; SD, standard deviation.

* Watanabe et al.[20] randomised induction-phase responders from both the placebo arm and the vedolizumab arms to maintenance treatment. However, the primary analysis for the maintenance phase was conducted only on those who had received vedolizumab in the induction phase. Demographic data relates to all randomised patients in the maintenance phase.

- Both studies[11,54] in the company's anti-TNF-α naïve network recruited patients who had shown a clinical response (drop in CDAI ≥70 or drop plus 25% reduction) in an induction phase. According to the respective licences, infliximab response should be assessed at six weeks (after two doses) and vedolizumab at 14 weeks (after three doses). Both studies therefore assess response earlier than would be done in the UK, at two and six weeks respectively. As such, the population entering the maintenance phase is not fully representative of the UK spectrum as patients who take longer to respond are excluded. This could conceivably lead to an overestimation of maintenance treatment effect, if these patients are also less likely to maintain a response when in remission. This is currently unknown.
- Hanauer et al.[54] only selected patients who have also demonstrated a 25% reduction in CDAI from baseline, as well as a drop in CDAI ≥70. This could equally lead to an overestimation of efficacy. Hanauer et al.[54] was a trial of infliximab.
- In a sensitivity analysis, CLASSIC II was added to the network.[41] This study had been removed from the network because it selected only patients in clinical remission (CDAI ≤150) at both 4 and 8 weeks from the initiation of therapy (adalimumab) to enter the maintenance trial. The CS[1] states that it was excluded "*since patients that met the remission criteria are likely to have experienced a much bigger drop in CDAI compared to those patients that were only classified as responders*." The same could be argued for Hanauer et al.[54] It is therefore unclear why the CS[1] considered only CLASSIC II[41] to be at risk of bias but not Hanauer et al.[54] The CLASSIC II criteria are arguably more stringent, but both impose a risk of bias and are a source of heterogeneity.
    - However, a better explanation and exploration of this issue is provided in Takeda data on file,[16] and is discussed in section 4.4.2.
- The ERG is not convinced by the argument regarding the exclusion of CLASSIC II given in the CS.[1] As such, both available networks (with and without CLASSIC II) should be given consideration.
    - A better approach may have been to use a random effects analysis to formally consider heterogeneity
    - It may also have been valid to consider that no network was possible due to clinical heterogeneity in patient spectrums and outcome assessment.

Anti-TNF-α failure NMA, maintenance: no network was possible for this population as Sandborn et al.[52] was an induction-only trial, and no other trials recruiting or reporting a failure population (of any definition) were identified.

Entire population NMA, maintenance: This network was not reported in full in the CS,[1] but was reported in the Takeda data on file document.[16] It included four studies.[11,20,54,55]

- These studies recruited a mixed population with no exclusions on the basis of anti-TNF failure,[55] a mixed population with intentional stratification to 50% naïve,[11] a mixed population with primary non-responders excluded[20] and an anti-TNF-α naïve study.[54] As such, the study populations in this network are likely to be heterogeneous in terms of proportion of anti-TNF-α failure patients (proportions not reported in most studies), which may affect the comparability of estimates of efficacy.
- All four studies recruited patients who had shown a clinical response in an induction phase. All assessed response earlier (6, 2, 4 and 4 weeks respectively) than would be done in UK clinical practice or is recommended in the licensing of the drugs.[9,10,57,58] It is possible these patients may be more responsive to treatment, which could result in a bias towards overestimation of treatment effect compared to the UK population who would enter the maintenance phase of treatment. Whether this is likely to affect results is unclear.

Interventions

All studies reporting maintenance phases used doses in line with UK licensing in at least one treatment arm. All studies included tapering of corticosteroids. The initiation of tapering varied by a few weeks between studies (between 6 and 12 weeks), which the ERG do not feel is problematic given the trials were of 52 weeks' duration.

Comparator

All studies used a placebo comparator in addition to usual care. Details of what "usual care" comprised in each study was not presented in the CS[1] and as such it is not clear if usual care was similar to UK practice. Usual care may have changed over time, but no meta-regression (by date) was attempted to explore this.

Outcomes

All studies reported data for at least two of the three main outcomes relating to clinical efficacy. In ACCENT I[54] (infliximab) the definition of clinical response was stricter than in other trials, as patients also had to have a 25% decrease in CDAI from baseline. It is likely that fewer patients would have met these criteria than the sole criteria of a fall in CDAI of 70 points or more, which may affect estimates of efficacy in comparison to studies that used a fall in CDAI of 70 points or more only. It is not clear whether the outcome definition would, in this case, have a differential effect on placebo versus treated patients, and what effect this might have on comparative efficacy. It should be noted, however, that patients who entered the trial were selected on the basis of having a drop in CDAI ≥70 as well as a 25% change from baseline, and therefore the outcome is assessing maintenance of that

state of health. As such, it may be less problematic, unless we believe that treatments have a differential effect on the maintenance of a drop in CDAI ≥70 versus CDAI ≥70 plus a 25% change from baseline. This study was included in the anti-TNF-α naïve population and the entire population networks.

In addition, the company have chosen to provide an analysis of "discontinuations due to adverse events". It is not clear what clinical relevance this outcome has. An analysis of serious adverse events would have been more appropriate and informative.

All studies reported their outcomes at or around 1 year from initiation of treatment. There are no data to inform performance of the treatments beyond this timeframe from RCT trials.

Study design

All studies were RCTs. Four were part of two stage trials, where patients were randomised to treatment or placebo during an induction phase study, and then responders were re-randomised for the maintenance phase as essentially a separate RCT.[11,20,41,53] The two other studies had only one phase, but had a run in period where all patients received an induction dose at baseline and then randomised only the responders.[54,55] One study[20] selected and randomised induction-phase responders from both the placebo arm and the vedolizumab arms to maintenance treatment. However, the primary analysis for the maintenance phase was conducted only on those who had received vedolizumab in the induction phase. The ERG does not think any of these differences in study design are likely to markedly affect the results. The effects of re-randomisation on patient spectrums and the success of randomisation in producing balanced groups were, however, often not well documented.

Table 24 Maintenance studies: treatment regimens and outcome analyses available

Study	Analysis methods	Interventions (n=randomised)	Protocol (CS tapered? Crossover?)	UK dosing regimen?	Comparator	Population (at induction baseline)	Outcome Time point (week)		
							Clinical response	Enhanced clinical response	Clinical remission
GEMINI II CSR13007[22]	ITT Missing data counted as failures – to check	Vedolizumab (IV) 300 mg every 4 weeks (n = 154) Vedolizumab (IV) 300 mg every 8 weeks (n = 154)	Tapered at week 6	Every 8 weeks is a UK dose Every 4 weeks should be used in those whose response decreases	Placebo (induction responders randomized) (n = 153) Placebo* (randomized before induction) (n = 148)	Mixed: Naïve: Experienced: Failure:	52 52 NR 52	52 52 NR 52	52 (P) 52 NR 52
ACCENT I Hanauer et al. 2002 [54]	ITT Missing data or switch to retreatment counted as failure	Grp II: Infliximab (IV) 5 mg/kg at weeks 2 and 6 and then every 8 weeks to 46 weeks (n = 113) Grp III: Infliximab (IV) 5 mg/kg at weeks 2 and 6, then 10mg every 8 weeks to 46 weeks (n = 112)	Tapered at week 6 Pts losing response eligible to switch to episodic retreatment (counted as failures)	Grp II – UK dose Grp III – higher dose than UK	Placebo (n = 110)	Naïve:	30 (P), 54**		30, 54
CHARM	ITT NR	Adalimumab (SC) 40	Tapered at	Yes, both	Placebo	Mixed:	26, 56	26, 56	26, 56(P)

Study	Analysis methods	Interventions (n=randomised)	Protocol (CS tapered? Crossover?)	UK dosing regimen?	Comparator	Population (at induction baseline)	Outcome Time point (week)		
							Clinical response	Enhanced clinical response	Clinical remission
Colombel et al.[55]	Missing data or switch to open label counted as failure	mg weekly randomized responders (n = 157) Adalimumab (SC) 40 mg every other week Randomized responders (n = 172)	week 8 Pts could switch to open label arm if lose response (counted as failure)	are UK doses	Randomized responders (n = 170)				
EXTEND Rutgeerts et al.[53]	ITT Missing data or switch to open label counted as failure	Adalimumab (IV) 40 mg every other week from week 4 to 52 n = 6	Tapered at week 12 Pts could switch to open label if lose response (counted as failure)	Yes	Placebo (n = 65)	Mixed (excluded primary non-responders)	52 (P)	52	52
NCT00445432 Watanabe et al.[20]	NR if ITT Missing data counted as failures	Adalimumab (IV) 40 mg every other week from week 4 to 52 (n = 25)	Tapered at week 8 Pts could switch to open label if lose response (counted as	Yes	Placebo (n = 25)	Mixed (excluded primary non-responders)	52 (P) (other data available from graph)	52 (other data available from graph)	52 (other data available from graph)

Study	Analysis methods	Interventions (n=randomised)	Protocol (CS tapered? Crossover?)	UK dosing regimen?	Comparator	Population (at induction baseline)	Outcome Time point (week)		
							Clinical response	Enhanced clinical response	Clinical remission
			failure)						
CLASSIC II Sandborn et al.[41]	ITT, LOCF	Adalimumab (SC) 40 mg every week (n = 18) Adalimumab (SC) 40 mg every other week (n = 19)	Tapered at week 8 Pts could switch to open label if lose response (counted as failure)	Yes	Placebo (n = 18)	Naive	24, 56 (P) (other data available from graph)	24, 56 (other data available from graph)	24, 56 (other data available from graph)

NR, not reported; Grp, group; Pts, patients; SC, subcutaneous; IV, intravenous.

4.4 Critique of the indirect comparison and/or network meta-analysis

This section is split into two parts, one of which critiques the methods used in the network meta-analysis, and one of which summarises the results of the networks and their limitations.

4.4.1 Critique of the methods used in the network meta-analysis

The company undertook separate NMAs of the anti-TNF-α naïve, anti-TNF-α experienced/failure subgroups and the entire mixed-ITT population. Induction phase and maintenance phase data were synthesised separately.

The statistical account of the analyses within the CS[1] is often unclear and several statistical claims are made that suggest a misunderstanding of issues or that are simply wrong.

The following bullets describe limitations of the analyses conducted by the company[1] or statements believed by the ERG to be incorrect.

- Definition of a closed loop

There appears to be a misunderstanding in the CS[1] about what is meant by a closed loop, which seems to be confused with replication of studies. Inconsistency in a network meta-analysis refers to a difference between the direct and indirect estimates of treatment effect. It was claimed that a consistency check was performed on closed loops formed by GEMINI II[11] and GEMINI III.[12] However, these studies do not form a closed loop and there cannot be inconsistency only heterogeneity between their study-specific estimates of the population treatment effect.

In addition, the company[1] say that, *"Frequentist random effects MTCs cannot run unless there is at least one closed loop.",* which is not true.

- Frequentist versus Bayesian methods

The company[1] used results from the NMA using frequentist fixed effect models in the economic model. The ERG believes that a Bayesian approach is preferable because the Bayesian approach:

(a) allows the ability to incorporate external evidence in addition to the sample data. This includes the ability to incorporate reasonable prior beliefs about the between-study standard deviation and adjustments for study quality,

(b) the ability to model data exactly. Frequentist methods provide approximations to Bayesian methods and are only asymptotically correct. and

(c) the ability to make probabilistic statements about parameters: An objective of the evidence synthesis is to characterise uncertainty about true values in an economic model and this is done using the joint posterior distribution. The true underlying joint distribution will not

follow any standard multivariate parametric distribution and will only be asymptotically multivariate normal. It should be noted that the company used marginal univariate normal distributions to characterise uncertainty about inputs to the economic model which not only approximates the true distribution but ignores correlation between parameters,

Furthermore, the CS[1] compare results from frequentist and Bayesian fixed effects models a as way of justifying the robustness of the fixed effect results. The ERG believes this comparison to be of very limited value because both models are treating the treatment effects observed in each study as being estimates of the same population treatment effect with no additional sources of uncertainty; and therefore results are expected to be similar whether a Bayesian or frequentist approach is used. The comparison that might be of some interest would be between frequentist and Bayesian random effects models; a frequentist random effect model treats the between-study standard deviation as if it was known and equal to the estimated value, whereas a Bayesian random effects model treats the between-study standard deviation as unknown with its own posterior distribution. Essentially, a Bayesian random effects model is a compromise between a fixed effect model and a frequentist random effects model. We would expect these to be different but a Bayesian approach is preferred given the context.

- Fixed versus random effects

It is stated that when there is relatively little sample data (i.e. studies) with which to estimate parameters in a random effects model and that frequentist methods will tend to underestimate the between-study standard deviation, whereas Bayesian models will tend to overestimate the between-study standard deviation, although this will only be true if the prior distribution does not represent reasonable prior beliefs. The results presented in this submission are from a fixed effect model which assumes that the between-study standard deviation is zero with probability one. The authors claim that they performed a Bayesian random effects analysis, although no results are presented in the CS. They also claim that the results were similar irrespective of the choice of prior distribution, which is unlikely given the limited evidence with which to estimate the between-study standard deviation.
Bayesian methods are commonly implemented using reference (or so-called non-informative) prior distributions. Reference prior distributions are not intended to represent reasonable prior beliefs but are usually acceptable when there is sufficient sample data to dominate the prior distribution. When there are limited sample data (i.e. studies) some thought is required to incorporate reasonable beliefs about the distribution of treatment effects in the population.

Results based on a fixed effect model are likely to underestimate the uncertainty in the treatment effects because they assert that the between-study standard deviation is known to be zero. The authors should have considered using a random effects model with a weakly informative prior distribution for

the between-study standard deviation, τ; for example, such that, $\tau \sim HN(0, 0.32^2)$, where HN represents a half-normal distribution.

- Between Study Heterogeneity

The CS[1] claims that heterogeneity is a problem in network meta-analyses. In fact, it is not a problem in network meta-analyses any more than it is a problem in standard pairwise meta-analyses. The important issues to consider when doing the analysis are: 1) the use of an appropriate model for the data that allows for heterogeneity between studies (i.e. a random treatment effects model), and 2) there is some attempt to deal with heterogeneity either by using meta-regression to explain it or by presenting results as the predictive distribution of the treatment effect(s) in a new study.

It was claimed that most of the NMAs conducted did not show much degree of heterogeneity, although the estimates of the between-study standard deviations and their 95% credible intervals were not provided and it is not clear from the CS that random effects models were implemented.

- Meta-regression

Meta-regression is a technique that is used to explain variation in treatment effects between studies and the following covariates were apparently assessed by the company:

- proportion of anti-TNF–naïve patients
- proportion of males
- mean age
- baseline CDAI
- week (primary endpoint) (for induction)

It is unclear how meta-regression could be performed using five potential treatment effect modifiers for the following reasons:

- the networks were said to contain relatively few studies given the number of parameters to be estimated
- the main inferences were based on a fixed effect model in which each study is assumed to be estimating the same treatment effect
- with multiple treatments it is possible that treatment effect modifiers:
  - operate identically for each treatment
  - operate differently for each treatment
  - operate differently but in a related way for each treatment

Somewhat confusingly, in the clarification response it was stated, "The networks contained too few studies to perform this type of analysis."

- Bayesian NMA

The Markov chains were run with only every 50th iteration being retained. This implies very high autocorrelation between successive samples of the Markov chain. Whist this is not necessarily a problem providing that sufficient samples are taken with which to estimate parameters, it is questionable whether a burn-in of 20,000 iterations would be sufficient in this situation and there is no information reported in the CS[1] on the Markov chain error with which to assess the accuracy of estimates.

- Placebo response rates

Differences in placebo response rates *per se* are not a problem. Variation in placebo response rates between studies is expected and this is dealt with in the statistical analysis by estimating treatment effects within studies.

Variation in the placebo response rates was assessed on the absolute scale. However, an aspect that is important is whether the treatment effect depends on the baseline response rate; given that treatment effects are being estimated on the log odds scale, the question is whether the log odds ratio is related to the true baseline log odds. If this is the case then it would be important to identify placebo arm covariates that explain the difference in the baseline log odds. If it is believed that placebo response rate have improved over time then year of publication could be assessed in a meta-regression.

It is not clear what is meant by the statement, *"Meta-regression techniques were used to fit the response to treatment"*. A simple random effects model of the placebo arms from each study would have sufficed.

- Repeated measures random effects MTC

The report claims that time was treated as a random effect, whereas it should have been treated as a fixed effect because time cannot be random.

A limitation of the Dakin model reported in the CS[1], which is used to analyse repeated measures data, was that it did not account for correlation and presumably this is also true of any repeated measures analyses in this submission.

- Goodness-of-fit

There is no assessment provided of goodness of fit. An absolute assessment of the goodness-of-fit can be performed by calculating the residual deviance for each observation and the total residual deviance. It is not clear whether the models provide a reasonable representation of the data.

4.4.2 Summary of the results of the network meta analyses and their limitations

The CS[1] presents three main tables of results – reproduced here for convenience. The tables were for:

- anti-TNF-α naïve network: Induction (Table 25)
- anti-TNF-α naïve network Maintenance (Table 26)
- anti-TNF-α experience/failure network: induction (Table 27)

A narrative summary was also provided for each of the above, and in addition for:

- entire population network: Induction studies
- entire population network: Maintenance studies

Anti-TNF-α naïve network: Induction

The results of the network are given in Table 25. In the discussion of the results for this network, the CS[1] notes that infliximab had significantly better clinical response than vedolizumab, but that this was based on one study (Targan et al).[19] The CS[1] goes on to argue that this study has major limitations that amount to a good rationale to exclude the study. The three reasons given were (see CS[1] pg. 140):

- *"A nonstandard dose was used*
- *There was a low placebo rate meaning the active treatment (infliximab) was more likely to demonstrate a significant effect*
- *Population sizes were small (fewer than 30 patients in each arm)"*

The ERG does not share the same degree of concern about this study for the following reasons:

- the standard dose was lower than should be used, and would likely underestimate treatment effects
    - for efficacy outcomes, infliximab was shown to be superior to vedolizumab, so an underestimation in efficacy would not alter the conclusions regarding the relative efficacy.
    - for safety outcomes, this may be a concern in terms of assessing adverse events as these may also be underestimated. However, the only analysis presented for adverse events (discontinuations due to adverse events) did not include data for infliximab, so this will not have affected the network
- the low placebo rate could have been dealt with using more appropriate statistical analysis methods
- small sample sizes could have been dealt with using more appropriate statistical analysis methods

As such, the ERG does not place much emphasis on the sensitivity analysis where Targan et al.[19] is removed.

In summary, the NMA results for clinical response (drop in CDAI ≥70) suggest to the ERG:

- all treatments are statistically significantly effective versus placebo.
- infliximab is statistically significantly better than vedolizumab
    - removal of Targan et al[19] from the network results in no data for infliximab and gives a result of no statistically significant difference between adalimumab and vedolizumab.
- vedolizumab versus placebo has a lower odds ratio than adalimumab versus placebo: odds ratio (OR) (95% credible interval (95% CrI)) for vedolizumab vs placebo week 10: 1.9 (1.2 to 3.1); adalimumab 80/40mg versus placebo (week4): 2.5 (1.3 to 4.9). A statistical comparison between these results was not presented in the CS,[1] but pairwise comparisons were provided in Takeda data on file.[16] These show the difference is not statistically significant.

For enhanced clinical response (drop in CDAI ≥100):

- no data for infliximab
- adalimumab 40/20mg (not a UK dose) and 80/40mg dose (normal dose) not significantly different to placebo, but adalimumab 160/80mg (accelerated dose) and vedolizumab are.
- no significant difference between adalimumab and vedolizumab, but OR versus placebo for UK doses of adalimumab were the same or higher than vedolizumab

For clinical remission (CDAI ≤ 150):

- all treatments except adalimumab 40/20mg were statistically better than placebo
- infliximab (OR versus placebo (95% CrI): 25.0 (4.1 to 451.0) is statistically significantly better than vedolizumab at 10 or 6 weeks (OR versus placebo (95% CrI): 2.7 (1.4 to 5.4) and 2.9 (1.5 to 6.0) respectively)
- vedolizumab has a better OR versus placebo (OR (95% CrI): 2.7 (1.4 to 5.4)) than adalimumab 80/40mg (OR (95% CrI): 2.3 (1.0 to 5.9), but worse than adalimumab 160/80mg (OR versus placebo (95% CrI): 4.1 (1.8 to 10.0). Statistical significance between adalimumab and vedolizumab was not reported
- removal of Targan et al[19] from the network results in no data for infliximab, and no statistically significant difference between adalimumab and vedolizumab; the OR versus placebo was higher for vedolizumab (OR 3.0 (95%CrI 1.6 to 6.2) than adalimumab 80/40mg (OR 2.4 (95% CrI 1.0 to 5.8), but lower for vedolizumab than adalimumab 160/80mg (OR 4.1 (95% CrI 1.8 to 10.0).

For discontinuations due to AEs

- adalimumab 160/80mg dose was significantly better (lower rates of discontinuations) than vedolizumab
- no data was available for infliximab

Anti-TNF-α naïve network: Maintenance

This network excluded CLASSIC II,[41] and so only comprises two studies and only provides a comparison between infliximab and vedolizumab. The results are given in Table 26, reproduced from the CS.[1] The ERG also discusses the results of the sensitivity analysis where CLASSIC II was included, as the ERG believes that this analysis had potential to be informative (see section 4.3).

The main findings of the NMA are listed on page 143 of the CS[1]. The ERG's interpretation of the findings is as follows:

- infliximab was statistically different to placebo in all three outcomes (clinical remission, clinical response, discontinuation due to AEs)
    - There was a high OR for discontinuation due to AEs compared with placebo.
- vedolizumab every 4 weeks was only statistically different to placebo for the clinical remission outcome
- vedolizumab every 8 weeks was statistically different to placebo for both clinical response and clinical remission
- the statistical significance of the difference in clinical response between vedolizumab and infliximab was not reported for the dose (5mg) of infliximab licenced in the UK. Infliximab 5mg dose OR versus placebo was better than both vedolizumab every 4 weeks and every 8 weeks (both licenced in UK). Pairwise comparisons were reported in Takeda data on file[16] but due to time constraints, the ERG were not able to assess this data.
- the difference between vedolizumab and infliximab for the outcome clinical remission was not statistically significant. The OR versus placebo were however different: (OR (95% CrI) for vedolizumab every 4 weeks, 4.2 (1.2 to 4.9); vedolizumab every 8 weeks, 2.9 (1.4 to 6.1); infliximab 5mg/kg, 2.5 (1.3 to 5.2); infliximab 10mg/kg 4.0 (2.1 to 8.1)
- vedolizumab was significantly better than infliximab for discontinuations due to AEs

The inclusion of CLASSIC II,[41] referred to on page 129 of the CS[1] and found on page 2 of appendix N (see Takeda data on file[16] pg. N-2) presented some difficulties. CLASSIC II recruited only patients who had gone into clinical remission (CDAI ≤150) after induction treatment. The outcome "clinical response" (drop in CDAI ≥70) may have been affected by this, in that clinical response is assessed by comparison with the induction baseline, and therefore the placebo arm is likely to have a high response rate (as all of them would have achieved a CDAI ≤150 already, which in nearly all cases will also be a fall in CDAI≥70, as there are only 70 points between CDAI 220 (the lower end of the

recruitment selection criteria) and CDAI 150). As such, there could be a large response in the placebo arm, assuming at least some patients maintain some or all of their improvement on conventional (placebo) therapy, which would disadvantage adalimumab when analysed as an odds ratio.
To avoid this problem, the Takeda data on file instead takes the following approach:

"Assume that the relative difference for responders at end of induction to number of responders at 12 months is equivalent to [the] number in remission at end of induction to [the] number in remission at 12 months, and compare the remission data from CLASSIC II with the response data for other studies. Although this network contains data for two different endpoints, as far we are able to reason there is no obvious source of bias in favour of or against adalimumab." (see Takeda data on file[16] pg. N-2)

It is unclear why the analysis was not done comparing remission data from all trials, though this may have disadvantaged the estimates of efficacy where remission was not used as a selection criterion (i.e. disadvantaged infliximab and vedolizumab); reassessment of GEMINI II[11] by the company may have been possible.

The results were not tabulated in the CS[1] but were presented in appendix N of the Takeda data on file document.[16] There were two Bayesian analyses: a fixed effects analysis comparing clinical response data from CLASSIC II[41] with clinical response data from GEMINI II[11] and ACCENT I;[54] and a fixed effects analysis comparing remission data from CLASSIC II[41] with response data from GEMINI II[11] and ACCENT I;[54] The pairwise ORs for these analyses are presented in Figure 8 and Figure 9, reproduced from Takeda data on file.[16]

- In the first analysis (response data from CLASSIC II),
    - Only infliximab 5mg and vedolizumab every 8 weeks were statistically significantly better than placebo.
    - Ranks were difficult to interpret (see Takeda data on file[16] pg. N-8), but seem to suggest infliximab 10mg (non-UK dose) and 5mg are most likely to be ranked 1st and 2nd place.
- In the second analysis (remission data from CLASSIC II), of treatments relevant to the UK,
    - all treatments except vedolizumab 4 weekly were statistically significantly better than placebo
    - vedolizumab every 8 weeks had nearly a 50% probability of being ranked 5th most effective treatment, vedolizumab every 4 weeks had >60% probability of being ranked 6th most effective treatment, with placebo having >90% probability of being ranked 7th (out of 7 treatment arms).

In the analysis without CLASSIC II, vedolizumab every 8 weeks is statistically significantly better than placebo, with an OR of 1.8 (95% CrI 1.1 to 3.0), versus an OR of 2.6 (95% CrI 1.3 to 5.6) in the first analysis with CLASSIC II, and OR of 2.6 (95% CrI 1.3 to 5.2) in the second analysis. In addition, the following observations can be made:

- All analyses report a statistically significant difference for vedolizumab 300mg every 8 weeks versus placebo, but not for vedolizumab 300mg every 4 weeks.
    - The relative efficacy of vedolizumab and adalimumab is uncertain
    - It is likely that vedolizumab is less effective than infliximab, regardless of which analysis is preferred.

Table 25 Summary of NMA induction anti-TNF-α -Naïve sub-population (odds ratio vs. placebo [95% CrI]) – Reproduced from Table 6.7.6.1 in CS[1]

		Comparator							
Outcome Measured		**Vedolizumab 300 mg**	**Adalimumab 80/40**	**Adalimumab 160/80**	**Adalimumab 40/20**	**Infliximab 5**	**Infliximab 10a**	**Infliximab 20a**	**Conclusion**
Clinical response (drop in CDAI ≥ 70)	Week 6 for vedolizumab	1.8* (1.1, 3.0)	2.5* (1.3, 4.8)	2.6* (1.3, 4.8)	2.0* (1.1, 4.0)	25.0* (6.2, 128.0)	5.3* (1.5, 23.0)	9.8* (2.6, 41.0)	infliximab significantly better than vedolizumab
	Week 6 for vedolizumab (Targan et al., 1997 removed)	1.8* (1.1, 3.0)	2.5* (1.3, 5.0)	2.5* (1.3, 5.0)	2.1* (1.1, 3.9)	NA	NA	NA	vedolizumab not significantly different from adalimumab
	Week 10 for vedolizumab	1.9* (1.2, 3.1)	2.5* (1.3, 4.9)	2.5* (1.4, 4.9)	2.1* (1.1, 4.0)	25.0* (6.3, 118.0)	5.3* (1.5, 22.0)	9.7* (2.6, 42.0)	infliximab significantly better than vedolizumab
Enhanced clinical response (drop in CDAI ≥ 100)	Week 6 for vedolizumab	1.9* (1.1, 3.1)	1.9 (0.9, 4.0)	2.9* (1.4, 5.9)	1.5 (0.7, 3.1)	NA	NA	NA	vedolizumab not significantly different from adalimumab
	Week 10 for vedolizumab	2.3* (1.4, 3.8)	1.9 (0.9, 4.0)	2.9* (1.4, 5.9)	1.5 (0.7, 3.0)	NA	NA	NA	vedolizumab not significantly different from adalimumab
Clinical remission	Week 6 for vedolizumab	2.9* (1.5, 6.0)	2.3* (1.0, 6.2)	4.1* (1.8, 10.0)	1.5 (0.6, 4.0)	26.0* (4.0, 425.0)	8.4* (1.3, 148.0)	8.7* (1.4, 160.0)	infliximab significantly better than vedolizumab
	Week 6 for vedolizumab (Targan et al., 1997 removed)	3.0* (1.6, 6.2)	2.4* (1.0, 5.8)	4.1* (1.9, 10.0)	1.6 (0.6, 4.2)	NA	NA	NA	vedolizumab not significantly different from adalimumab
	Week 10 for vedolizumab	2.7* (1.4, 5.4)	2.3* (1.0, 5.9)	4.1* (1.8, 10.0)	1.5 (0.6, 4.1)	25.0* (4.1, 451.0)	8.7* (1.4, 156.0)	8.8* (1.4, 180.0)	infliximab significantly better than vedolizumab
Discontinuation due to AEs		1.4 (0.3, 7.4)	0.4 (0.0, 5.6)	0.0* (0.0, 0.7)	0.5 (0.0, 5.9)	NA	NA	NA	adalimumab 160 mg/80 mg significantly better than vedolizumab

AE = adverse event; CDAI = Crohn's Disease Activity Index; CrI=credible interval; NA = not applicable. * = significant vs. placebo. [a] = non-standard dose, should not be included in comparisons

Table 26 Summary of NMA maintenance anti-TNF-α -Naïve sub-population (odds ratio vs. placebo [95% CrI]) – reproduced from Table 6.7.6.2 in CS[1]

Outcome Measured	Comparator				Conclusion
	Vedolizumab Q4W	Vedolizumab Q8W	Infliximab 5 mg/kg	Infliximab 10 mg/kg	
Clinical response (drop in CDAI ≥ 70)	1.8 (0.9, 3.5)	2.6* (1.3, 5.0)	3.4* (1.9, 6.5)	5.0* (2.6, 9.4)	infliximab 10 mg significantly better than vedolizumab Q4W
Clinical remission	2.4* (1.2, 4.9)	2.9* (1.4, 6.1)	2.5* (1.3, 5.2)	4.0* (2.1, 8.1)	vedolizumab not significantly different
Discontinuation due to AEs	0.8 (0.3, 2.7)	0.5 (0.1, 1.8)	6.6* (2.8, 20.0)	3.4* (1.3, 10.0)	vedolizumab significantly better than infliximab

AE = adverse event; anti-TNF-α = tumor necrosis factor antagonist; CDAI = Crohn's Disease Activity Index; CrI=credible interval; Q4W = every 4 weeks; Q8W = every 8 weeks.

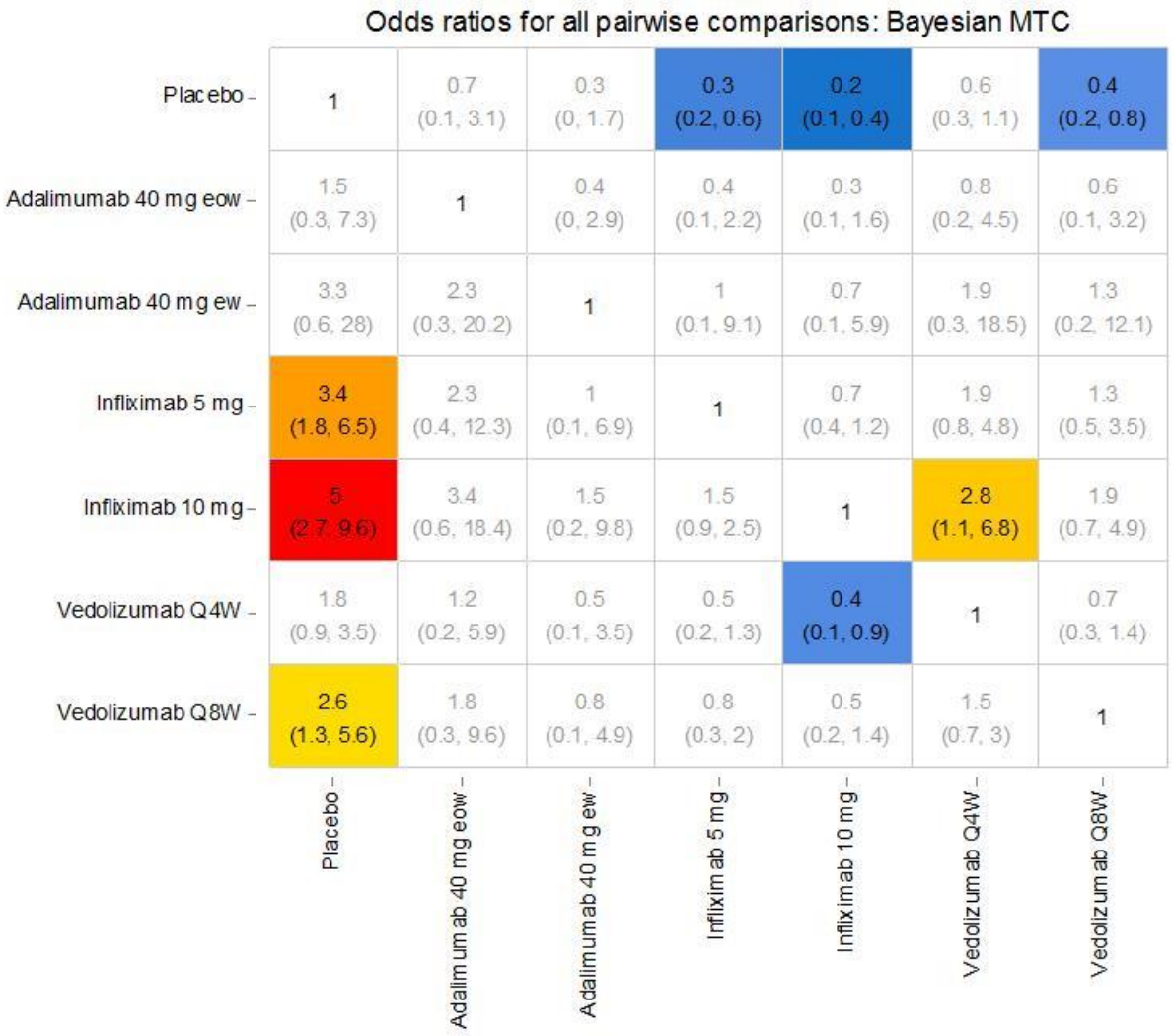
* = significant vs. placebo.

Table 27 Summary of NMA induction anti-TNF-α -Experienced/Failure sub-population (odds ratio vs. placebo [95% CrI]) – reproduced from Table 6.7.6.3 in CS[1]

Outcome Measured		Comparator		Conclusion
		Vedolizumab 300 mg	Adalimumab 160 mg/80mg	
Clinical response (drop in CDAI ≥ 70)	Week 6 for vedolizumab	1.9* (1.3, 2.8)	2.1* (1.4, 3.3)	vedolizumab not significantly different from adalimumab
	Week 10 for vedolizumab	1.9* (1.3, 2.8)	2.1* (1.4, 3.3)	vedolizumab not significantly different from adalimumab
Enhanced clinical response (drop in CDAI ≥ 100)	Week 6 for vedolizumab	1.7* (1.2, 2.6)	1.9* (1.2, 3.1)	vedolizumab not significantly different from adalimumab
	Week 10 for vedolizumab	2.0* (1.3, 3.0)	1.9* (1.2, 3.1)	vedolizumab not significantly different from adalimumab
Clinical remission	Week 6 for vedolizumab	1.4 (0.8, 2.6)	3.6* (1.8,7.1)	adalimumab significant benefit over vedolizumab
	Week 10 for vedolizumab	2.5* (1.5, 4.3)	3.5* (1.8, 7.4)	vedolizumab not significantly different from adalimumab
Discontinuation due to AEs		0.4* (0.1, 0.9)	0.5 (0.1, 2.4)	vedolizumab not significantly different from adalimumab

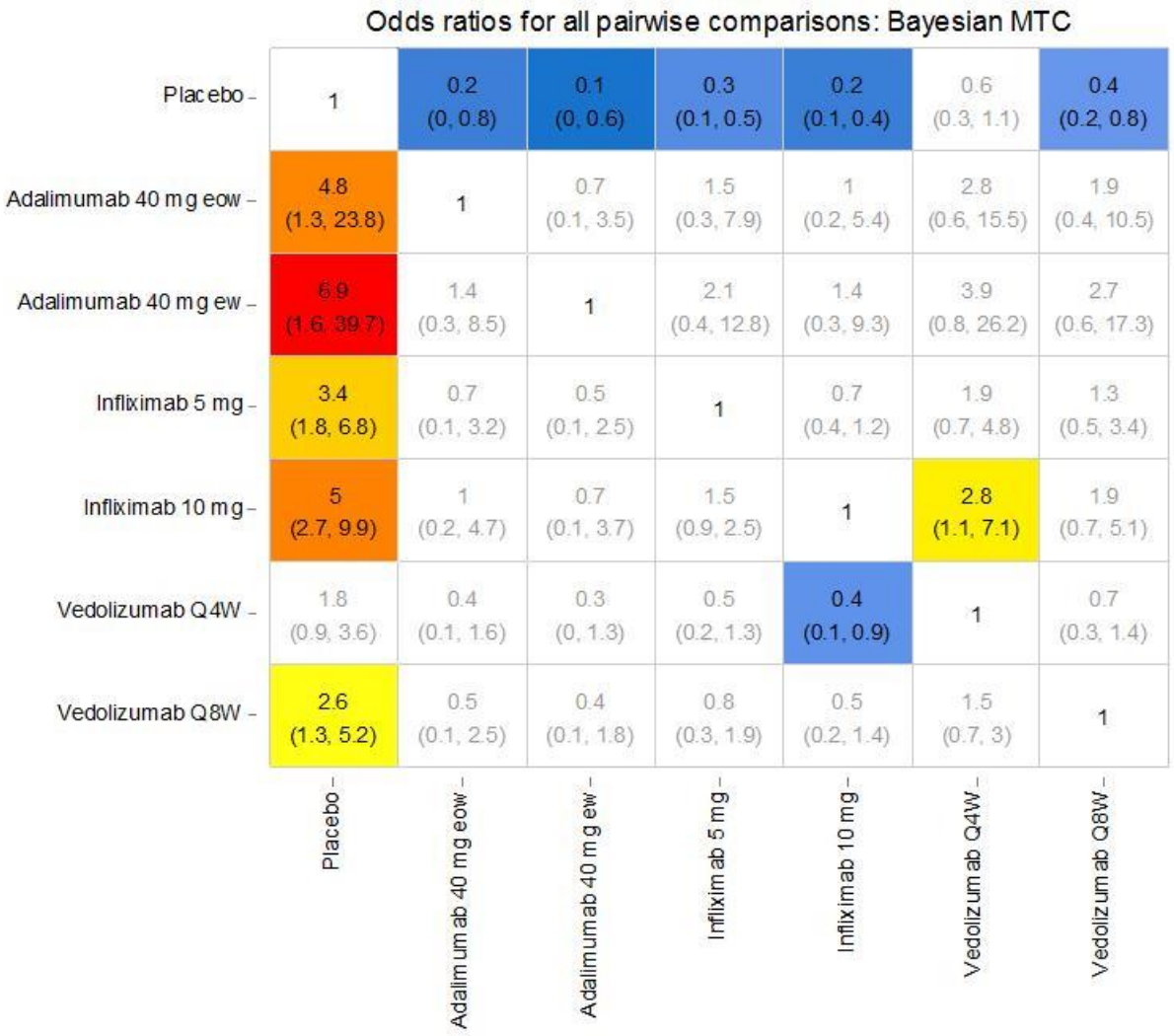
AE = adverse event; anti-TNF-α = tumor necrosis factor antagonist; CDAI = Crohn's Disease Activity Index; CrI=credible interval; Q4W = every 4 weeks; Q8W = every 8 weeks.* = significant vs. placebo.

Figure 8 **All Pairwise Odds Ratios From MTC For anti-TNF-α –Naïve Maintenance Patients Sustained Response Including Response Data From CLASSIC II (Reproduced from Figure N-3 from Takeda data on file[16])**



anti-TNF-α = tumor necrosis factor antagonist; eow = every other week; ew = every week; MTC = mixed treatment comparison; Q4W = every 4 weeks; Q8W = every 8 weeks.

Figure 9 **All Pairwise Odds Ratios From MTC for anti-TNF-α –Naïve Maintenance Patients Sustained Response, Including Remission Data From CLASSIC II (Reproduced from Figure N-9 from Takeda data on file[16])**



Anti-TNF-α experienced/failure network: Induction (no network for maintenance)

This network comprised three trials, the two GEMINI trials of vedolizumab [11,12] and one of adalimumab.[52] The table of results provided in the CS[1] is reproduced here as Table 27. The ERG notes the following about the results:

- The patient spectrum in the adalimumab trial[52] was more likely to be responsive to anti-TNF-α treatments as it excluded primary non-responders. The trial also used the accelerated dose of adalimumab, though this may be appropriate to UK practice as all patients in the trial had experienced treatment failure and a rapid response is likely to be desirable.
    - On balance, the ERG feels that this network is likely to overestimate adalimumab treatment effects.
- Both treatments were statistically significantly different to placebo, except:
    - vedolizumab at 6 weeks for clinical remission
    - adalimumab for discontinuation due to AEs
- There was insufficient evidence to conclude whether adalimumab was different to placebo for the outcome "discontinuation due to AEs"
- There was insufficient evidence to conclude whether vedolizumab and adalimumab were statistically significantly different to one another in most cases; the OR versus placebo was better for adalimumab in most analyses and statistically significantly superior at 6 weeks for clinical remission.

"Entire population" network: Induction

This analysis was not presented in full in the CS[1] but rather summarised as a number of bullet points, which can be found on page 147 of the CS (not reproduced here).[1]

A total of ten analyses over 83 pages were presented in Takeda data on file[16]for this analysis. Due to time constraints the ERG were not able to fully assess them. In addition, as discussed in section 4.3, the ERG did not feel this analysis would produce results that would be easy to interpret as study populations are not clinically homogeneous and the results may not represent a clinically meaningful population.

Instead, the ERG has selected the entire population analysis for clinical response (drop in CDAI ≥70) at week 10 as the most relevant to the decision problem for the following reasons:

- infliximab is included in the analysis; infliximab data were not available for enhanced clinical response

- The 10 week time point is closer to the 10 to 14 week time point indicted by the clinical advisors to the ERG as being usual in UK clinical practice.

Figure 10 provides the pairwise comparisons from this network. The results are very similar to previous networks in terms of the relative efficacy of vedolizumab, infliximab and adalimumab, namely that:

- All UK licenced treatments were significantly better than placebo
- infliximab was statistically significantly better than vedolizumab (OR 5.5 (95% CrI 1.5 to 25)
- there was insufficient evidence to conclude that there was a statistically significant difference between adalimumab and vedolizumab, with 95% CRI all crossing the line of no effect.
-

The ERG would also like to draw attention to the serious adverse event NMA referred to on page 147 of the CS,[1] in a bullet point:

"*For the entire population, analysis of SAEs was carried out no significant differences were found*." (pg. 147 of the CS)[1]

"Entire population" network: Maintenance

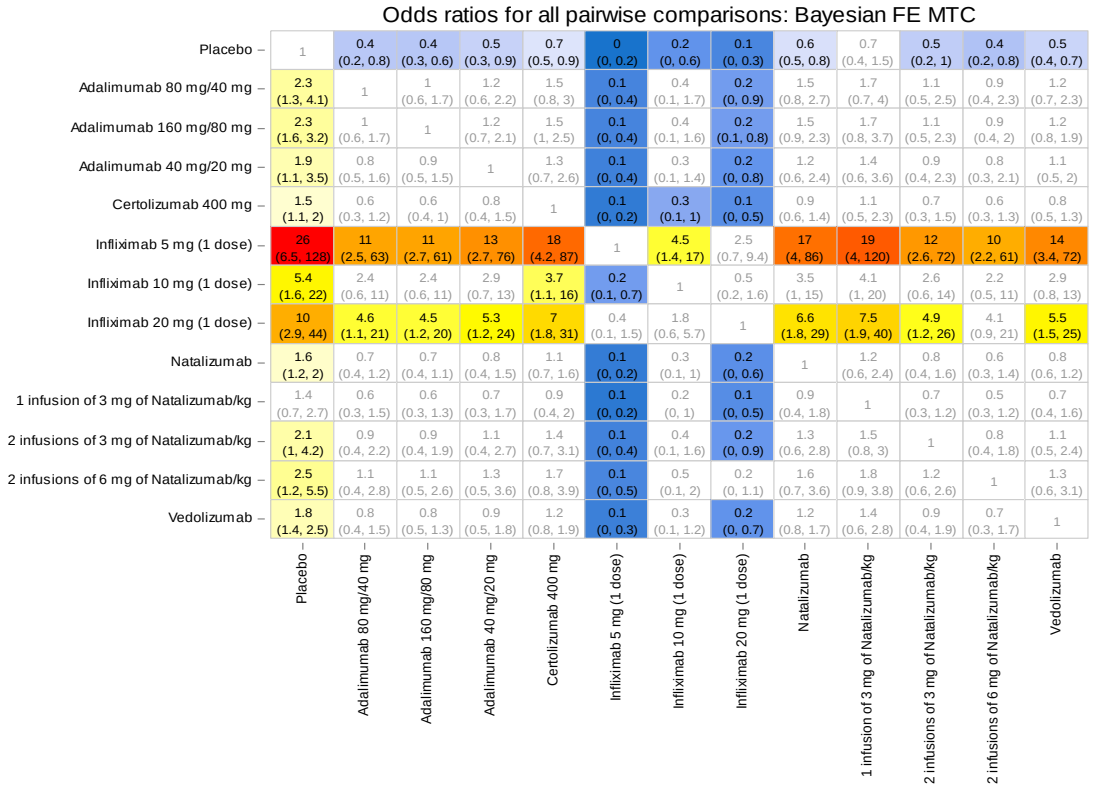
This analysis was not presented in full in the CS[1] but rather summarised as a number of bullet points, which can be found on pages 148-149 of the CS (not reproduced here).[1]

As above, the ERG has selected the entire population analysis for clinical response (drop in CDAI ≥70) as an example which includes infliximab.

Figure 11 provides the pairwise comparisons for this network. The following observations can be made:

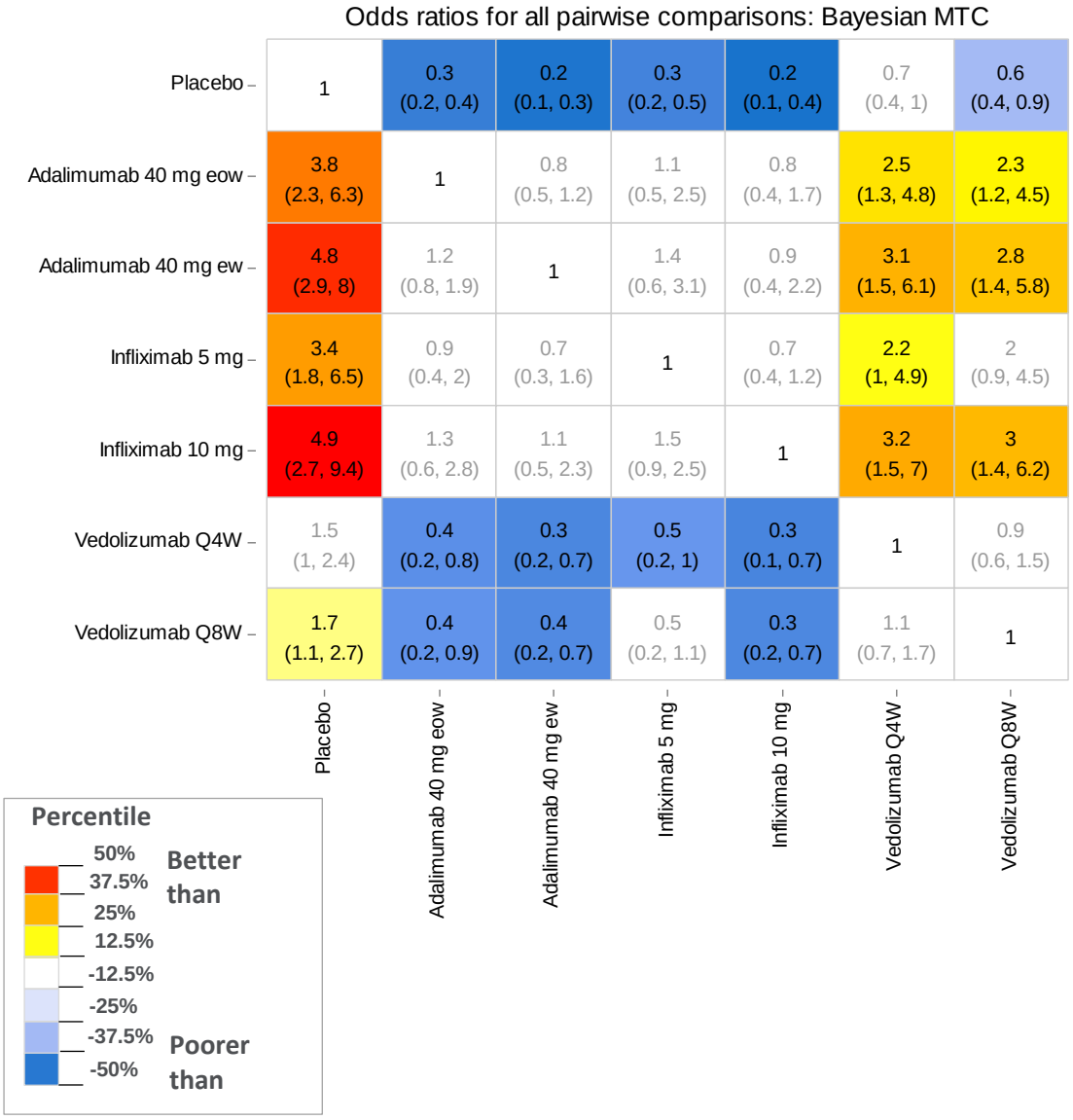
- all treatments except vedolizumab every 4 weeks were significantly better than placebo
- both adalimumab and infliximab were significantly better than vedolizumab

Figure 10 All Pairwise Odds Ratios From MTC All Patients Induction Week 10 Clinical Response (CDAI ≥ 70) – Reproduced from Figure H-17 from Takeda data on file[16]



CDAI = Crohn's Disease Activity Index; FE = fixed effects; MTC = mixed treatment comparison.

Figure 11 Pairwise Odds Ratios From MTC All Patients Maintenance Durable Response (CDAI ≥ 70) – Reproduced from Figure H-86 from Takeda data on file.[16] All



CDAI = Crohn's Disease Activity Index; MTC = mixed treatment comparison.

4.5 Additional work on clinical effectiveness undertaken by the ERG

As the company undertook a comprehensive systematic review (no major limitations were noted) of vedolizumab of treatment of adults with moderate to severe active CD, no additional work, apart from some minor data extractions to complete study and patient characteristic tables, was undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

4.6.1 Completeness of the CS with regard to relevant clinical studies and relevant data within those studies

The clinical evidence in the CS[1] is based on a systematic review of the clinical effectiveness and safety of vedolizumab for the treatment of adults with moderately to severely active CD. The ERG is satisfied that all relevant (published and unpublished) studies of vedolizumab were included in the CS.[1]

The same is true for the network meta-analysis, with the exception of data for the induction period having been missed in one trial.[20] The ERG believe this is a data extraction error, rather than a problem with identification of relevant studies.

4.6.2 Interpretation of treatment effects reported in the CS in relation to relevant population, interventions, comparator and outcomes

A key issue that may limit the robustness of the efficacy and safety data for vedolizumab reported in the CS[1] relates to the high attrition rates in the maintenance phase of the GEMINIII[11] trial. High rates of discontinuation were observed across all treatment groups. 58% (89/153) discontinued in the placebo arm, 53% (81/154), and 47% (72/154) discontinued in the vedolizumab Q8W and Q4W arms respectively. The ERG believe that attrition rates at these levels have the potential to impact on the maintenance study results, posing a serious threat to external validity.

Furthermore, whilst GEMINI II[11] achieved his primary endpoint, the primary endpoint was not achieved in GEMINI III;[12] therefore, statistical evaluation of the secondary endpoints is acknowledge as exploratory by the company.

Table 28 summarises the ERG's interpretation of the treatment effects given in the NMAs. In the induction NMAs there were a number of observations about the relevance of the populations. These included:

- the anti-TNF-α naïve population was thought to be most generalizable to UK in whom the disease has responded inadequately to, or is no longer responding to conventional therapy and who have not previously received an anti-TNF-α
- the "entire population" analysis mixed populations with differential proportions of characteristics that are thought to be treatment-modifying, namely proportion of anti-TNF-α failure populations, making the results of this analysis difficult to generalise to any particular population, and difficult to interpret as a whole
- the anti-TNF-α failure/experienced network may have overestimated efficacy for adalimumab as primary anti-TNF-α failure patients were excluded from the adalimumab study but not the vedolizumab studies
- several studies across the evidence base excluded patients with strictures, meaning generalisation to this population is problematic
- some studies did not report the proportion of patients with fistulising disease, so it is unclear whether all studies were representative of UK populations in this respect
- no studies included patients with CDAI>450, meaning generalisation to the upper range of severe patients (defined as CDAI 450 to 600) is uncertain

For the maintenance NMAs the following observations about the population were made:

- the anti-TNF-α naïve population was thought to be most generalizable to UK patients in whom the disease has responded inadequately to, or is no longer responding to conventional therapy and who have not previously received an anti-TNF-α
- the "entire population" analysis mixed populations with differential proportions of characteristics that are thought to be treatment-modifying, namely proportion of anti-TNF-α failure populations, making the results of this analysis difficult to generalise to any particular population, and difficult to interpret as a whole
- no studies included patients with CDAI>450, meaning generalisation to the upper range of severe patients (defined as CDAI 450 to 600) is problematic
- patients with strictures were excluded from GEMINI II[11] only which may confer an advantage to estimates of efficacy for vedolizumab and cause problems with generalisation of efficacy results to those with strictures
- patients were selected to enter the maintenance phase in both trials included in the anti-TNF-α naïve maintenance network on the basis of assessment at earlier time points than would commonly be done in the UK.[9,10,57,58] This means patients who take longer to respond are not represented in these trials, which may affect estimates of efficacy and/or limit generalisation to the population of patients who take longer to respond: the ERG do not know if these patients would have a differential response to treatment.

- CLASSIC II recruited patients who achieved remission, which may mean results of this adalimumab trial may not be generalizable to those who continue maintenance treatment on the basis of clinical response.

The ERG also noted the following about the interventions used:

- induction periods were not always in line with UK licencing and clinical practice, meaning not all studies delivered a full induction dose
- studies which used the adalimumab accelerated dose did not overtly attempt to recruit patients who would receive this dose according to UK licencing
- maintenance doses were usually in line with UK licencing

Comparators used in the network were not assessed by the CS[1] and it is unclear how similar usual care was to UK practice, and whether usual care may have changed over time.

Given the above, the following conclusions can be drawn (Table 28):

- clinical significance of results is unclear
- generalisation to patients with strictures is uncertain
- generalisation to severe patients (CDAI >450) is uncertain
- generalisation of maintenance studies to UK practice should be done with awareness that those who take longer to respond to induction therapy were not included in these trials
- anti-TNF-α naïve population, induction:
    - if the Targan et al.[19] study is included in the network, infliximab appears to be significantly better than vedolizumab for clinical response and clinical remission
    - regardless of the inclusion of Targan et al.,[19] there is insufficient evidence to conclude there is a difference between vedolizumab and adalimumab on all other efficacy outcomes
    - adalimumab appears to result in fewer discontinuations due to AEs than vedolizumab
- anti-TNF-α naïve population, maintenance:
    - none of the presented analyses were without considerable limitations
    - across the three analyses presented, vedolizumab every 4 weeks appears significantly worse than infliximab 10mg for clinical remission. However, other pairwise comparisons between treatments adalimumab, vedolizumab and infliximab are not statistically significant
    - for discontinuations due to AEs, vedolizumab appears significantly better than adalimumab, though this should be interpreted with reference to the numbers who discontinued for each treatment in the induction period.

- in agreement with the CS,[1] the ERG did not feel the anti-TNF-α failure/experienced network would give a robust assessment of comparative treatment effects due to differences in patient populations
- in agreement with the CS[1] the ERG did not feel that the "entire population" analysis was of great relevance to this assessment for either induction or maintenance periods.
- no analysis for serious adverse events was provided for the anti-TNF-α naïve network.

Table 28 Summary of the ERGs interpretation of the treatment effects reported from relevant NMAs

Network	ERG Comments	ERG conclusion
Induction phase		
All induction networks	• Number of studies excluded patients with strictures • No studies recruited CDAI>450 • Unclear if comparators are comparable between trials and to UK practice • Time of outcome assessment means some responders are missed for all treatments • Not all studies delivered full induction periods • infliximab dose was less than UK licence • No analysis of serious adverse events was presented	• Clinical relevance of differences uncertain • Generalisation to patients with strictures uncertain • Generalisation to severe patients (CDAI >450) uncertain • Generalisation to UK practice should only be done with due consideration of the limitations of the evidence base • Uncertainty about serious adverse events
Anti-TNF-alpha naive	• Best match to patients presenting post-conventional therapy failure • 10 week preferred to 6 weeks data • Data from Watanabe et al.[20] missing	**Most relevant analysis at wk 10** **CR, CRem (Wk 10 & 6):** infliximab is statistically significantly better than vedolizumab **CR, ECR (Wk 10& 6):** no data for infliximab; insufficient evidence to conclude there is a difference between vedolizumab and adalimumab – **Discontinuation due to AEs:** adalimumab accelerated dose significantly better than vedolizumab
Anti-TNF--α naïve, Targan et al.[19] removed	• Complete removal of Targan et al.[19] not considered appropriate by the ERG • Better statistical analysis possible	**Assessment not robust** **CR, CRem (Wk 6):** Insufficient evidence to conclude there is a difference between vedolizumab and adalimumab
Anti-TNF-α failure/experienced	• Patient populations were not comparable between trials, which may bias estimates in favour of adalimumab	**Most relevant analysis at Wk 10:** Insufficient evidence to conclude there is a difference between vedolizumab and adalimumab **Wk 6**: as wk 10 except for CRem at wk 6: adalimumab significantly better than vedolizumab **Discontinuation due to AEs:** insufficient evidence to conclude there is a

Network	ERG Comments	ERG conclusion
		difference between vedolizumab and adalimumab
Entire population	• Mixes studies with very different populations in terms of anti-TNF-α failure proportions • Only network to report analysis for serious adverse events	**Assessment difficult to interpret in context of UK population** **CR (WK10):** infliximab statistically significantly better than vedolizumab, insufficient evidence to conclude there is a difference between vedolizumab and adalimumab **Serious adverse events:** no significant differences between treatments
Maintenance phase		
All maintenance networks	• One study excluded patients with strictures • No studies recruited CDAI>450 • unclear if comparators are comparable between trials and to UK practice • Re-randomisation of patients after initial randomisation may affect patient spectrums	• Generalisation to patients with strictures uncertain • Generalisation to severe patients (CDAI >450) uncertain • Generalisation to UK practice should only be done with due consideration of the limitations of the evidence base
Anti-TNF-alpha naïve (without CLASSIC II)[41]	• Best match to patients presenting post-conventional therapy failure • Recruitment criteria (response) differ between the two trials: response assessed early in both studies and response defined differently in infliximab study • Definition of outcome CR differs between the two trials • No data for adalimumab	**CR:** infliximab 10 mg significantly better than vedolizumab Q4W **CRem:** Insufficient evidence to conclude there is a difference between vedolizumab and infliximab **Discontinuation due to AEs:** vedolizumab significantly better than infliximab
Anti-TNF-alpha naïve (with CLASSIC II)[41]	• Best match to patients presenting post-conventional therapy failure • Recruitment criteria differ between all three trials • To address problems with recruitment criteria and outcome definitions, two analyses were presented, with different results • Patients were assessed for response to induction therapy, for inclusion in the trial, at a time point probably earlier than would be done in UK practice	**Evidence base presents difficulties, and neither analysis is without limitations** Regardless of which of the two analyses including CLASSIC II[41] is preferred, it is likely that vedolizumab is less effective than infliximab. No significant difference between adalimumab and vedolizumab.
Anti-TNF-α failure	NA	NA

Network	ERG Comments	ERG conclusion
Entire population	• Mixes studies with very different populations in terms of anti-TNF-α failure proportions • Definition of CR outcome differed for one trial	**Assessment difficult to interpret in context of UK population** Both adalimumab and infliximab are significantly better than vedolizumab

CR, clinical response, drop in CDAI ≥70; ECR, enhanced clinical response, drop in CDAI ≥100; CRem, clinical remission, CDAI ≤150; AEs, adverse events; NA, not applicable; Wk, week

4.6.3 Uncertainties surrounding the reliability of the clinical effectiveness

The main uncertainties in the clinical evidence primarily relate to duration of treatment and generalizability to the UK population. Further details are provided below.

Duration of treatment

The duration of treatment of vedolizumab in the GEMINI II trial was 52 weeks, followed by enrolment in the ongoing GEMINI LTS study. As a result, the long-term efficacy and safety of vedolizumab is unknown and the optimum duration of therapy remains unclear. There are no data on strategies for withdrawal of the drug in those on maintenance therapies or with respect to how to predict instances in which this can be successfully achieved. The SmPC for vedolizumab[9,10] recommends monitoring and reporting of any suspected adverse reactions after authorisation especially for new onset or worsening of neurological signs and symptoms.

Generalizability to the population of England and Wales

In GEMINI II,[11] at induction phase, patients were predominantly white (89.2%) with a mean age of 36.1 years. The mean body weight was 69.8kg and 46.6% were male. The mean duration of disease was 9 years, patients had a mean CDAI score of 324, and the mean faecal calprotectin score was 1,254. In GEMINI IIIl,[12] most patients were white (90%). The mean age was 37.9 years, mean body weight was 70.4kg and 43% were male. Median duration of disease was 8.4 years in the vedolizumab group and 8 years in the placebo group. Patients in the vedolizumab group had a mean CDAI score of 301.3, and 313.9 in the placebo group. Median faecal calprotectin score was 1148.1 in the vedolizumab group, and 1426.5 in the placebo group. It should be noted that the faecal calprotectin in the GEMINI trials was deemed to be high, indicating that patients may had had significant active inflammation. Although information on the number of UK-based study sites was not available, it appears that very few were used and very few UK patients included in either GEMINI II[11] or GEMINI III.[12] In comparison, a large number of study sites were US-based. In the US, patients were required to have failed either an immunomodulator (6-MP or azathioprine) or an anti-TNF-α agent, whilst outside of the US, failing corticosteroids alone was sufficient for study entry. It is unclear to the ERG how the different criteria might have impacted on the study results. The trials also assess response in the induction phase earlier than would be done in the UK, at six weeks. As such, the population entering the maintenance phase in GEMINI II is not fully representative of the UK spectrum, as patients who take longer to respond are excluded. This could conceivably lead to an overestimation of maintenance treatment effect, if these patients are also less likely to maintain a response when in remission.

Clinical advisors to the ERG expressed the view that the concomitant conventional therapy used in the GEMINI trials may not reflect those used in UK clinical practice. The company, in response to question B29 of the clarification questions,[2] appear to agree, *'The use of conventional therapy within the GEMINI II and GEMINI III trials was protocol driven and the trial was international and may not represent treatment patterns in England and Wales…'*. It is unclear to the ERG how the potential lack of generalizability of conventional therapy might have impacted the study results. As such there is some uncertainty regarding the generalizability of the evidence to the clinical population of England and Wales.

Furthermore, the safety and efficacy of vedolizumab has not been established in children aged below 17 years, in pregnant women, in women of childbearing potential, lactating mothers, patients with renal or hepatic impairment, or in concomitant use with biologic immunosuppressants.[9,10]

In the NMA, the ERG considered that the results presented may underestimate the uncertainty in treatment effects since fixed effects models were used, and there is clear evidence of heterogeneity among the trials included in the NMAs.

5. COST-EFFECTIVENESS

Section 5.1 provides a brief summary and critique of the review of cost-effectiveness evidence included in the CS.[1] A summary of the economic evidence submitted by the company in support of this Single Technology Appraisal (STA) is provided in Section 5.2. Additional works undertaken by the evidence review group (ERG) are presented in Section 5.3.

The CS[1] includes a review of published cost-effectiveness evidence for the treatment of CD and a description of, and results from, a *de novo* cost-utility model evaluating the cost-effectiveness of vedolizumab for the treatment of CD in adults with moderate to severe disease. In addition to the economic evidence provided in the CS,[1] the company submitted a Microsoft© Excel-based economic model[15] (referred as the company's model).

5.1 ERG comment on the company's review of cost-effectiveness evidence

Brief description of the company's economic review included in the CS[1]

The CS[1] includes a systematic review of evidence relating to the cost-effectiveness of vedolizumab and other treatment for patients with CD. A systematic literature review was initially performed by the company in April 2013 and updated in March 2014. Search terms for databases included combinations of free text and MeSH headings incorporating terms related to the disease, interventions, comparators and study type. The searches also included terms relating to specific aspects of health economic evaluations (e.g. costs and utilities). The CS[1] states that searches were restricted to studies published after 2002 as prior to that date, biologic drugs used in the treatment of CD had not been approved for use in the UK, and resource use and cost studies would be out of date.

The company's search strategy was comprised of searches of the following databases:

- MEDLINE
- MEDLINE In-Process
- EMBASE
- Econlit
- The Cochrane Library

The company's electronic database searches were supplemented with a search of the following:

- NICE website
- Cost effectiveness analysis registry
- International Society for Pharmacoeconomics and Outcomes Research: Research Digest, at http://www.ispor.org/research_study_digest/research_index.asp
- European Crohn's and Colitis Organisation, at https://www.ecco-ibd.eu/
- Digestive Disease Week
- United European Gastroenterology Week
- American College of Gastroenterology.

Bibliographic reference lists of included studies and systematic reviews were also screened for relevant publications.

The company's selection of studies for inclusion in the review was guided by inclusion and exclusion criteria (described in the CS[1] on pg. 185). Non-UK economic evaluations were excluded from the review. Studies were screened over two stages: titles and abstracts were reviewed by one researcher and 5% were checked by a second researcher to ensure that the inclusion criteria had been applied correctly. The full texts of studies included during the first level screening were then obtained and independently reviewed by two researchers.

Five full UK economic evaluations[59-63] were included in the company's systematic review (see Table 7.1.2.1 in CS[1] pg. 189-193). Dretzke et al.[63] and Bodger et al.[60] evaluated the cost-effectiveness of infliximab and adalimumab against standard care (Table 29). Loftus et al.[61] assessed the cost-effectiveness of adalimumab against non-biologic therapy; Lindsay et al.[62] evaluated the cost-effectiveness of infliximab against non-biologic therapy. Finally, Clark et al.[59] evaluated the cost-effectiveness of infliximab against placebo. All studies were conducted from the perspective of the national health service (NHS) and personal social services (PSS). Time horizons varied between 1-year to a lifetime. A quality assessment of the included studies is presented in the CS[1] in Table 7.1.3.1 and Table 7.1.3.2 (see CS[1] pg. 195 – 203).

Results are presented in Table 29. infliximab and adalimumab appear to have an ICER below £30,000 per QALY gained in Bodger et al,.[60] Loftus et al.[61] and Lindsay et al.[62] The ICER is above £30,000 per QALY gained in Dretzke et al.[63] and Clark et al.[59] However, it is difficult to interpret and compare results from the different studies due to differences in time horizon (1 year vs. lifetime), population included (fistulising CD, luminal…) or decision problem (induction, maintenance, episodic treatment…).

From the information that was provided regarding the cost-effectiveness searches, it would appear that this element of the review was conducted appropriately and to a sufficiently high standard. However, the absence of certain information makes it difficult to provide a full and thorough critique.

The ERG is largely satisfied with the company's systematic review of economic evidence[1] but notes the following;

- the economic analysis which informed the recent guideline for the management of CD[7] has not been included within the company's systematic review;[1] the omission of this study is not justified.
- the company restricted searches to studies published after 2002 as prior to that date, biologic drugs used in the treatment of CD had not been approved for use in the UK; studies evaluating conventional non-biologic therapy may have been published prior 2002 and may provide useful information
- finally, non-UK analyses were excluded; these may have provided useful information.

Table 29 Summary of studies included in the company's cost-effectiveness review[1]

Study	Dretzke et al.[63]	Bodger et al.[60]	Loftus et al.[61]	Lindsay et al.[62]	Clark et al.[59]
Analysis type	Cost-utility analysis	Cost-utility analysis	Cost-utility analysis	Cost-utility analysis	Cost-utility analysis
Population	Patients with 'moderate-to-severe' CD	Adult patients with moderate to severely active CD.	Patients with moderate-to-severe CD	Patients suffering from active luminal or fistulising CD.	CD patients chronic active or fistulising disease
Economic comparisons included	• infliximab • adalimumab • conventional non-biologic treatment	• infliximab • adalimumab • conventional non-biologic treatment	• adalimumab • conventional non-biologic treatment	• infliximab • conventional non-biologic treatment	• infliximab • placebo
Perspective	NHS perspective	NHS perspective	NHS perspective	NHS perspective	NHS perspective
Time horizon	1 years	Lifetime	Lifetime	5 years	Lifetime
Key results	**Infliximab, severe:** - SC dominated by IFX induction. - ICER for maintenance versus induction: £5.03M **Infliximab, moderate:** - ICER IFX vs SC: £94,321. - ICER for maintenance versus induction: £13.09M **adalimumab, severe:** - SC dominated by ADA induction. - ICER for maintenance versus induction: £4.98M **adalimumab, Moderate:** - SC dominated by ADA induction. - ICER for maintenance versus induction: £13.9M	**ICER against standard care** Infliximab 5 mg/kg, 1 year: £19,050 Infliximab 5 mg/kg, 2 years: £21,300 Adalimumab 80 mg, 1 year: £7,190 Adalimumab 80 mg, 2 years: £10,310	**Patients with moderate-to-severe CD:** £ 33,731 **Patients with severe CD:** £16,064	**Fistulizing CD:** £29,752 **Severe, active luminal CD:** £26,128	**Infliximab compared with placebo (5 mg/kg):** 93,244 (single dose) 62,016 (episodic) **Infliximab compared with placebo (all doses):** 135,333 (single dose) 72,261 (episodic)
CD = crohn's disease; ADA = adalimumab; IFX = infliximab; SC = standard care					

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

Consistency of the CS[1] against the requirements set out in the NICE Reference Case[64] for the base case analysis is provided in Section 5.2.1. A description and critique of the model structure is provided in Section 5.2.2. The populations, interventions and comparators, perspective, time horizon and discounting are described in Sections 5.2.3, 5.2.4 and 5.2.5 respectively. Input parameters used for treatment effectiveness, health-related quality of life (HRQoL), resources and costs are described in Section 5.2.6, 5.2.7 and 5.2.8 respectively. Data used for the subgroup analyses are summarised in Section 5.2.9. Base case results included within the CS[1] are presented in Section 5.2.10 with results from the sensitivity and scenario analyses presented in Section 5.2.11.

The CS[1] includes a health economic model[15] constructed in Microsoft Excel which compares vedolizumab versus conventional non-biologic therapy (a combination of 5-ASAs, immunomodulators and corticosteroids) in a mixed population and subgroup of patient who are anti-TNF-α naïve and anti-TNF-α failure from the perspective of the UK NHS. Anti-TNF-α agents used in the UK (adalimumab and infliximab) are only evaluated in the anti-TNF-α naïve subgroup.

It should be noted in the CS[1] that the description of the model structure, input parameters and results are on some occasions brief, with scant detail and on some occasions is inaccurate. The description of the economic evaluation submitted by the company provided hereafter is typically based on information provided within the CS[1] when this is consistent with the company's model.[15] When there is a discrepancy between the values reported in the CS[1] and company's model,[15] the values used in the latter are reported and highlighted in this report.

Finally, it should be noted that an updated Excel-based model was submitted by the company following the clarification process.[2] The main amendments relate to (a) the functionality to assess outcomes separately for patients with moderate and severe disease at baseline (b) correction of errors and (c) the updating of costs.

5.2.1 Adherence to the NICE reference case

Table 30 summarises the ERG's appraisal of the company's economic evaluation[1,15,15] against the requirements set out in the NICE Reference Case for the base case analysis.[64]

Table 30 Adherence of the company's economic analysis[1,15,15] to the NICE Reference Case[64]

Attribute	Reference case	Does the *de novo* economic evaluation match the reference case?
Decision problem	The scope developed by NICE	The scope of the company's health economic analysis is in adherence with that developed by NICE.[8]
Comparator(s)	Alternative therapy routinely used in the NHS	Conventional non-biologic treatment are considered in all three population analyses considered (mixed-ITT, anti-TNF-α naïve and anti-TNF-α failure subgroups) Other biologic agents (infliximab and adalimumab) are evaluated only for the anti-TNF-α naïve subgroup. Other biologic agents are not considered within the mixed-ITT or anti-TNF-α failure subgroups. Comparators included in the company's health economic analysis are broadly in adherence with the list of comparators set out in the NICE final scope.[8] It should be noted that; • further anti-TNF-α agents may be used in patients after failure of prior anti-TNF-α therapy (although the ERG recognises that the effectiveness in this population is uncertain), • the mixed-ITT population includes anti-TNF-α naïve patients and therefore anti-TNF-α agents may be a relevant comparator (although the ERG is unsure of the relevance of analyses conducted within this population)
Perspective costs	NHS and Personal Social Services	A NHS perspective is considered. Costs borne by PSS are excluded from the company's economic analysis; the company states that these are expected to be minimal (see Table 7.2.6.1 in CS[1] pg. 213).
Perspective benefits	All health effects on individuals	Yes.
Form of economic evaluation	Cost-utility analysis	The company undertook a cost-utility analysis

Time horizon	Sufficient to capture differences in costs and outcomes	A 10-year time horizon is used in the company's base case analysis. A lifetime horizon is considered in a sensitivity analysis.
Synthesis of evidence on outcomes	Systematic review	Main efficacy parameters are taken (when possible) from a network meta-analysis (NMA) of the effects of biologic and conventional non-biologic treatment based on a systematic review of the published literature. Transition probabilities (relating to surgery) are drawn from published sources.
Outcome measure	Quality adjusted life years	Health outcomes are valued using quality-adjusted life years (QALYs), derived from patients with CD using the EQ-5D questionnaire.
Health states for QALY	Described using a standardised and validated instrument	
Benefit valuation	Time-trade off or standard gamble	
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	
Discount rate	An annual rate of 3.5% on both costs and health effects	Costs and benefits are discounted at 3.5%
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	No additional equity weighting is applied to estimated QALY gains.
Sensitivity analysis	Probabilistic sensitivity analysis	A probabilistic sensitivity analysis (PSA) is conducted; although ICERs are not reported and the ERG has concerns regarding the PSA conducted (arbitrary distributions assumed).

5.2.2 Model structure

Brief description of the model structure and the logic of the company's model[15]

The description of the model structure and logic provided within the CS[1] are incomplete and brief. To aid understanding of the model structure and the validity of the key structural assumptions, the description of the model structure/logics is based on information provided within the CS,[1] when possible and accurate and the ERG's understanding of the company's model[15]when necessary.

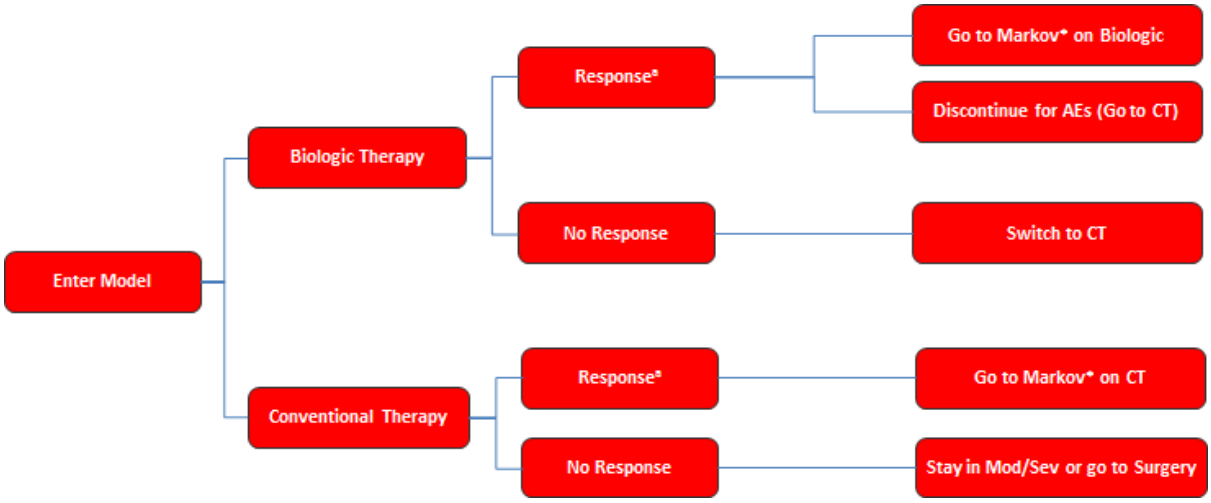
The company's model[15] structure is based on the structure published by Bodger et al.[60] The company (see CS[1] pg. 206-207) states that other models[63] did not include partial response and consequently the structure from Bodger et al.[60] was deemed to be more appropriate to capture the treatment effect. A reference to a previous NICE Decision Support Unit (DSU) assessment in CD[65] recognising the importance of partial response was provided (see clarification response[2] question B6).

The company's model[15] adopts a hybrid approach whereby a decision tree is used to evaluate outcomes at the end of the initial induction therapy (during which all patients receive initial treatment to induce response – assumed to be 6 weeks for all biologic and non-biologic therapy) followed by a Markov structure (8-week cycle) to evaluate subsequent outcomes.

The company's diagrammatic representations of the model structure (see CS[1] Figure 7.2.2.1 in pg. 207 and Figure 7.2.2.2 in pg. 210) for induction and maintenance treatment are presented in Figures 12 and 13, respectively. It should be noted that the description of model states within the company's diagrams does not directly reflect the actual health states included in the Markov component of their model[15] as (a) it does not account for patients with moderate to severe CD who are responders and those who are non-responders and (b) does not account for whether patients are receiving biologic or conventional non-biologic therapy (patients switch from the biologic Markov structure to the non-biologic Markov structure following discontinuation from biologic therapy).

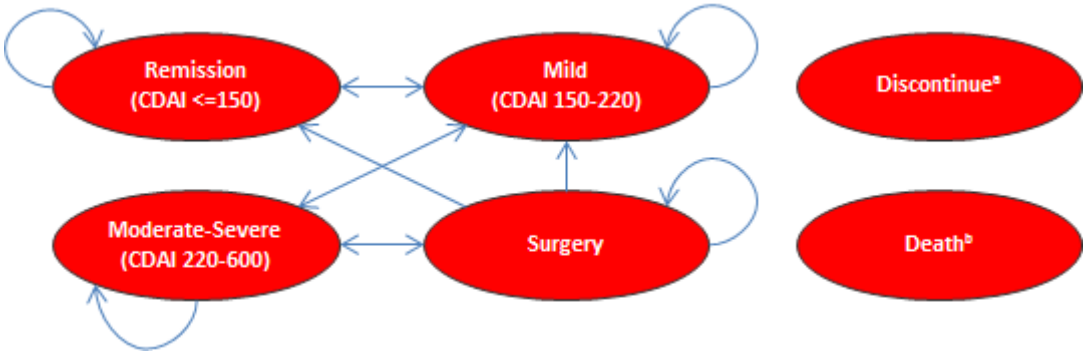
The general model structure is the same for patients commencing biologic and conventional non-biologic treatment. The company's model[15] includes a total of 12 mutually exclusive health states, separated into two identical Markov paths on (a) whether patients are currently receiving biologic treatment (referred as 'Markov on biologics') or (b) conventional non-biologic treatment (referred as 'Markov on CT').

Figure 12 Decision-tree for induction treatment (reproduced from Figure 7.2.1.1 in CS[1] pg. 207)



[a] Response is defined as a drop in CDAI of 70 points or more; * The Markov structures; AE = adverse event; CDAI = Crohn's Disease Activity Index; CT = conventional therapy.

Figure 13 Markov model schematics for CD maintenance phase and beyond (reproduced from Figure 7.2.1.2 in CS[1] pg. 210)



[a] Reasons for discontinuation include lack of response and adverse events. Discontinuation due to adverse events is applicable only to responders on biologic treatments, because non-responders on biologics switch to conventional therapy and continue receiving such until the end of the model's time horizon.

[b] Patients may transition to death from any health state during any cycle.

Each Markov structure ('Markov on biologics' and 'Markov on CT') is composed of six mutually exclusive health states, namely: remission CD; responder with mild CD; responder with moderate to severe CD; non-responder (assumed to have moderate to severe CD); surgery; and death.

Patients initiating conventional non-biologic treatment enter the model in the 'Markov on CT' portion of the model in the non-responder state; patients commencing biologic treatment enter the model in the 'Markov on biologics' portion the model in the non-responder state.

Irrespective of the treatment initiated, response is assessed at week 6 (end of induction treatment - first model cycle), defined as a drop in the CDAI score of 70 points or more.

At the end of induction therapy (referred as induction phase), patients commencing conventional non-biologic therapy are redistributed across the health states of the 'Markov on CT' portion of the model according to the induction therapy vector (see Section 5.2.6) for patients treated with conventional non-biologic therapy (referred as 'initial induction vector on CT'). Patients commencing biologic treatment are redistributed across the health states of the 'Markov on biologics' portion of the model according to initial induction therapy vectors (see Section 5.2.6) for patients treated with biologic therapy (hereafter referred as 'initial induction vector on biologics'). It should be noted that the 'initial induction vector on biologics' is different for each biologic (see Section 5.2.6).

Table 31 summarises the key structural assumptions following induction therapy.

Table 31 Key structural assumptions

In patients commencing conventional non-biologic treatment	In patients commencing biologic treatment
• patients remain in the conventional non-biologic portion of the model for the remainder of the model time horizon (i.e. they cannot subsequently receive biologic treatment) irrespective of their response to the induction phase.	• patients initiating biologic treatment receive conventional non-biologic therapy following discontinuation due to either AEs, surgery or end of schedule treatment. Retreatment using the same of different biologic therapy is not allowed.
• non-responders are assumed to have moderate to severe CD	
• patients are treated with conventional non-biologic therapy following surgery	
• patients who do not achieve the 'required' level of response at week 6 to the induction phase remain in the non-responder moderate to severe CD health state (and continue treatment with conventional non-biologic therapy) unless surgery or death.	• patients who do not achieve the 'required' level of response at week 6 to the induction phase discontinue biologic treatment and subsequently receive conventional non-biologic therapy. These patients are redistributed across the health state of the 'Markov on CT' portion of the model according to the 'initial induction vector on CT'.
• patients who achieve the 'required' level of response to the induction phase enter a maintenance phase (and continue treatment with conventional non-biologic therapy). These patients are able to transition between any health states of the 'Markov on CT' portion model according to an 8-week transition matrix.	• Patients who achieve the 'required' level of response at week 6 to the induction phase enter a maintenance phase (and continue to receive the same biologic treatment as maintenance therapy) irrespective of their CDAI score. During the maintenance phase, patients can transition between any health states of the 'Markov on biologics' portion model according to an 8-week treatment-specific transition matrix.
	• during the maintenance phase, patients remain on biologic treatment, provided (a) they do not experience an adverse event sufficient to warrant discontinuation, (b) they have not received biologic treatment for more than 1-year (end of scheduled maintenance), (c) they do not undergo surgery and (d) do not die.

	• at approximately 1-year, a forced treatment switch is applied to all patients receiving biologic treatment (end of scheduled maintenance); any patients who are currently receiving biologic therapy at this point are assumed to discontinue and subsequently receive conventional non-biologic treatment, irrespective of their current health states.
	• patients in the remission or mild CD health states at the time of discontinuation (due to AEs or forced switch at approximately one year) are treated with conventional non-biologic therapy and enter the 'Markov on CT' portion of the model. These patients are assumed to follow the transition matrix for the maintenance phase of patients treated with conventional non-biologic therapy according to their previous health state (before discontinuation from biologic treatment).
	• in contrast, patients with moderate to severe disease at the time of discontinuation enter the 'Markov on CT' portion of the model but are redistributed across the health state according to the 'initial induction vector on CT' (i.e. a proportion of patients is assumed to respond subsequently to conventional non-biologic therapy – same effectiveness as for patients initially treated with conventional non-biologic treatment)

In addition to the CDAI health states, the company's model[15] includes a surgery health state (see CS[1] pg. 209), defined as a mix of procedures (including panproctocolectomy with ileostomy or anal pouch formation, extended right hemicolectomy, drainage procedures, sigmoid colectomy, and ileal resection). Patients can only enter the surgery health state from the responder in moderate to severe CD and non-responder (assumed to have moderate to severe CD) health state. Following surgery, a proportion of patients may remain in the surgery health state and are assumed to undergo further surgery (see Section 5.2.6). The remaining patients are redistributed across the CDAI health states of

the 'Markov on CT' portion of the model according to a set of transition probabilities taken from Bodger et al.[60]

Finally, patients may transition to death (which is an absorbing state) from any health state during any cycle. The risk of mortality is applied as an age-specific baseline other-cause mortality rate, with state-specific relative risks to reflect an excess risk of death due to CD (see Section 5.2.6).

ERG's comments on the company's model structure

The choice of model structure is justified by the company as this was used in a previous economic evaluation by Bodger et al.[60] and due to the importance of capturing partial response in addition to remission, as recognised in a previous DSU assessment in CD[65] (see CS[1] pg. 206-207 and response to clarification[2] question B6). The ERG is largely satisfied with the justification provided by the company on the choice of model structure.

However, whilst the chosen model structure, adapted from Bodger et al.[60] may include partial response; the ERG expresses the following concerns: (a) potential omission of key aspects of the condition, (b) simplifying and debatable assumptions regarding surgery, (c) the difficultly associated with parameterising the company's chosen structure, and (d) debatable key structural assumptions. These issues are discussed in turn below.

The company's model[15] captures two key aspects of the condition: changes in disease severity (measured by the CDAI score) and the risk of surgery. The model ignores a key aspect of the condition in that CD is relapsing (exacerbation) and remitting (some patients may improve spontaneously). In the company's model,[15] the company assumes that patients who do not respond to conventional non-biologic therapy at week 6 remain in the non-responder state (and are assumed to have moderate to severe CD) for the remainder of the model until death or surgery; this is overly pessimistic. It should be noted that within the Bodger et al.[60] structure, about 15% of non-responders are able to improve (go to partial response or full response) every 8 weeks. To a lesser extent, as stated by the company (see CS[1] pg. 48) the aim of treatment is to induce and maintain remission and to maintain corticosteroid-free remission; the latter aspect is not captured within the company's model.[15]

In accordance with Bodger et al.,[60] surgery is modelled as a single health state representing a mix of procedures. The ERG believes this to be overly simplistic given that the type of subsequent surgery is likely to be conditional on the previous surgery received. Ideally, patients undergoing resection (removal of inflamed area of the intestine) should be distinguished from patients undergoing

ileostomy (disconnection of the small intestine from the colon and re-routed through a stoma). However, the ERG recognises the possible lack of data to distinguish resection from ileostomy and believes that the impact on results would be minimal given the lack of evidence suggesting that the type of surgery is conditional on the treatment administered.

A particular concern with the chosen model structure is the difficulty in parameterising required variables. Given the short time constraint for this STA, the absence of the electronic version of the Bodger's mathematical model,[60] and limited details included within the publication,[60] the ERG was unable to conduct a full assessment of the economic evaluation upon which the company's model is based.[15] However, the ERG notes that the model published by Bodger et al.,[60] which is similar to the company's model,[15] relies on a series of adjustments and assumptions in an attempt to replicate the results from the pivotal trials. Whilst the ERG recognises the need to calibrate model inputs on occasions, it is unclear from the Bodger publication[60] what the model predictions are calibrated against and how the transition probabilities were derived. The company's model[15] also uses a calibration approach to estimate the transition probabilities during the maintenance phase; however, the calibration relies on a series of constraints which are not adequately justified by evidence (see Section 5.2.6).

The following differences between the Bodger's model structure[60] and the company's model[15] should be noted: (a) the company's model[15] attempts to combine data from different trials whilst Bodger et al.[60] appear to use data from a single trial for each treatment, (b) the two models appear to calibrate model inputs to different outcomes (although it is unclear from Bodger et al.[60] what the model is fitted to), (c) the company's model[15] distinguishes patients with moderate to severe CD with and without response and (d) the company's model[15] assumes that patients with no response remain in this health state for the remainder of the time horizon.

The model structure also relies on a series of debatable structural assumptions. It should be noted that the derivation of transition probabilities (See Section 5.2.6) are conditional on these structural assumptions.

- Non-responders are assumed to have moderate to severe disease (see clarification response[2] question B12). This is inappropriate, as a non-responder may have mild disease (defined as CDAI between 150-220). For instance, a patient with a CDAI score of 250 at baseline with a drop in CDAI of 50 would be classified as a non-responder, but at the end of the induction phase will be in the mild health state (CDAI 150 – 220).
- No distinction is made between responders with moderate to severe CD and non-responders (except for continuation on biologic treatment following induction). The ERG believes that outcomes

(HRQoL, management and the probability of surgery) are likely to differ between responders and non-responders.

- The same induction phase duration is assumed for all therapy, leading to inconsistencies.
- Response is defined as a drop of 70 points or more in the CDAI score in the base case. A scenario analysis is conducted in which response corresponds to a drop of 100 points or more in the CDAI score (enhanced clinical response). This response criterion was chosen (see clarification response[2] question B10) to reflect the definition of response used in the GEMINI studies[11,12] and other trials.[19,51,66] Whilst the ERG recognises that this was the response definition used in the trials, it should be noted that it is unclear how such a criterion relates to clinical practice as the CDAI is not used.
- All patients who are still receiving anti-TNF-α therapy at approximately 1-year are assumed to discontinue (end of scheduled maintenance) and subsequently receive non-biologic treatment, irrespective of whether they are currently responding to treatment. A scenario analysis is conducted assuming a 3-year maximum treatment duration. There is uncertainty with respect to the long-term efficacy of biologic therapy as the randomised phases of trials of these therapies adopted a maximum follow-up of 54 weeks. Furthermore, the wording of the marketing authorisations for the biologics does not stipulate if or when responding patients should discontinue therapy.[9,10,57,58] In response to a request for clarification (see clarification response[2] question B8), the company states that *"in the absence of a stopping rule in clinical guidelines, it is uncertain what the average duration of treatment would be with vedolizumab, adalimumab and infliximab for the NHS…. A no stopping rule was not considered because based upon informal discussions with clinical experts, lifetime treatment with a biologic is unlikely".* The ERG is partly satisfied with the justification provided by the company. It should be noted that NICE recommendation for infliximab and adalimumab[6] suggests that *"specialists should discuss the risks and benefits of continued treatment with patients and consider a trial withdrawal from treatment for all patients who are in stable clinical remission. People who continue treatment with infliximab or adalimumab should have their disease reassessed at least every 12 months to determine whether ongoing treatment is still clinically appropriate. People whose disease relapses after treatment is stopped should have the option to start treatment again".* Therefore a discontinuation rule for patients in remission may be appropriate, but not for patients who are not in stable clinical remission. The ERG recognises that this is an area of uncertainty, but believes that the discontinuation rule assumed by the company is inappropriate.
- Following withdrawal from biologic therapy, patients previously in the remission or mild CD health states receive conventional non-biologic therapy and follow the transition matrix for the maintenance phase of patients treated with conventional non-biologic therapy according to their previous health state (before discontinuation from biologic treatment). This assumption is overly optimistic as relapse following withdrawal from biologic treatment is a recognised effect according to our clinical experts.[67] In response to a request for clarification (see clarification response[2] question

B13), the company confirmed that *"following biologic cessation, patients may transition from remission (or mild disease) to mild or moderate severe. Transition from response or remission back to moderate/severe disease can be considered relapse and is included in the model."* The ERG disagrees with the justification provided by the company, as following biologic discontinuation/withdrawal, the disease is likely to deteriorate, go back to baseline or worsen.

- The efficacy for patients commencing conventional non-biologic therapy is applied to patients who have previously discontinued biologic therapy; this assumes that response to non-biologic treatment is independent of previous biologic use. The efficacy of conventional non-biologic therapy is likely to be different following previous biologic use.
- During the maintenance phase, patients may discontinue due to AEs, the forced 1-year treatment stopping rule (end of scheduled maintenance), surgery or death. Therefore responders to the induction phase (primary response) remain on treatment even if they lose response (secondary failure). The ERG believes that discontinuation due to lack of efficacy should be included. Data from the GEMINI II trial[22] indicates that amongst patients who discontinued treatment randomised to the vedolizumab every 8-week (Q8W) arm (n=81), the most common reason for discontinuation was lack of efficacy (n=58), followed by AEs (n=12), withdrawal of consent (n=6), lost to follow-up (n=3) and protocol violation (n=2). Similarly, In GEMINI II,[22] the probability of having disease worsening (defined as ≥ 100-point increase in CDAI score from the week 6 value on 2 consecutive visits and a CDAI score ≥ 220 points) and treatment failure (defined as disease worsening, need for rescue medications or surgical intervention for treatment of CD, or study drug-related AE leading to discontinuation from the study) at One year was 19% and 39% in the vedolizumab Q8W arm respectively. Consequently, the ERG believes that discontinuation due to lack of efficacy should be included in the economic model.
- The model attempts to combine efficacy data from induction and maintenance trials (typically 2 separate trials). The ERG recognises that this was necessary for infliximab and adalimumab in order to inform the NMA. However, such an approach, may lead to inconsistencies if the distribution of CDAI score at the end of induction in responders is different to the distribution of CDAI score at the beginning of the maintenance trials. For the comparison of vedolizumab against conventional non-biologic therapy, an analysis could have been conducted using data from GEMINI II.[11]
- The same model structure was used for all biologic treatments; however, there are differences in (a) the criteria for entering maintenance, (b) induction phase duration, (c) trial durations and (d) outcomes evaluated within the trials. See Section 4.3 for further details on the differences between the trials.
- Patients discontinue biologic following surgery in the first cycle (induction phase); but not after primary response.

- The same approach is also used for patients on biologic and non-biologic treatment (i.e. induction vector and derivation of a transition matrix for responders). However, data used for the maintenance phase in patients on conventional treatment are not estimated amongst the same population (i.e. responder on conventional therapy) but instead uses data from patients receiving conventional treatment following primary response to biological therapy.

5.2.3 Population

Population included in the economic model

The population entering the company's model[15] reflects the population included in the GEMINI trials[11,12] (see clarification response[2] question B48) and includes patients with moderately to severely active CD (namely CDAI score of between 220 and 450) who have had an inadequate response with, lost response to, or are intolerant to either a conventional therapy or anti-TNF-α agents.

Results are presented (see CS[1] pg. 205) for adults with moderate to severe disease (defined as CDAI score >220) for three patient groups;

- a mixed population representing the intention to treat (ITT) population of the GEMINI trials (referred as the mixed-ITT population),[11,12] which includes both people who have never received an anti-TNF-α therapy (referred as anti-TNF-α naïve) and people who have previously been exposed to an anti-TNF-α agent (referred as anti-TNF-α failure),
- the anti-TNF-α naïve subgroup,
- and the anti-TNF-α failure subgroup, which includes intolerance to anti-TNF-α agents, primary failure (no initial response to anti-TNF-α agents) and secondary failure (loss of response after initially responding to anti-TNF-α agents).

In addition to the analyses in adults with moderate to severe disease (defined as CDAI>220), the company provided subgroup analyses in patients with moderate (CDAI 220-330) and severe disease (CDAI > 330) at baseline separately. Results from these analyses are presented in the CS[1] in Section 7.7.

ERG comments on the population described in the CS[1] and included in the company's model[15]

Table 32 summarises the populations and subgroups outlined in final scope issued by NICE.[8] The ERG is satisfied that the populations and subgroups addressed by the company are in adherence with the NICE final scope for this STA.[8] In the GEMINI trials,[11,12] patients were eligible if they had no

response to or had had unacceptable side effects from one or more of the following: glucocorticoids, immunosuppressive agents (i.e., azathioprine, mercaptopurine, or methotrexate), or anti-TNF-α.

Table 32 Populations and subgroups outlined in the NICE final scope[8]

Population	Adults with moderately to severely active CD in whom the disease has responded inadequately to, or is no longer responding to, either conventional therapy or a TNF-α antagonist, or who are intolerant to either of them
Subgroups to be considered	• People who have not previously received a TNF-α antagonist • People for whom a TNF-α antagonist has failed • People for whom TNF-α antagonists are not suitable because of intolerance or contraindication.

It is unclear whether the population recruited in the GEMINI trials[11,12] is reflective of a typical clinical population, notably;

- the GEMINI trials[11,12] included patients from a large number of centres worldwide. In response to a request for clarification regarding current practice in the population recruited in the GEMINI trials[11,12] (see clarification[2] question B29), the company confirmed that "*the use of conventional therapy within the GEMINI II and GEMINI III trials was protocol driven and the trial was international and may not represent treatment patterns in England and Wales"*.
- the trial[11,12] included patients with a CDAI score between 220 to 450; therefore excluded patient at the higher end of the CDAI (very severe) spectrum (CDAI > 450).
- It should be noted that the faecal calprotectin in the GEMINI trials was deemed to be high, indicating that patients may had had significant active inflammation It should also be noted that patient at the higher end of the CDAI (very severe) spectrum were excluded (CDAI >450).

The CS[1] reports results from the mixed-ITT population, a combination of patients who have previously received anti-TNF-α agents and those who are anti-TNF-α naïve; as suggested in the NICE final scope.[8] The interpretation of results and the relevance of this population to the decision problem are open to debate. The ERG believes that patients who have previously received anti-TNF-α agents and those who are anti-TNF-α naïve are two distinct, defined patient groups, with different characteristics and propensities to respond to treatment, as demonstrated in the GEMINI trials.[11,12] The appropriate comparators as chosen by the company are also different within these two populations. It is unclear how results from the mixed-ITT population can be interpreted. The ERG advices that the subgroup of patients who have previously received anti-TNF-α agents and those who are anti-TNF-α naïve should be interpreted separately, but recognises this may be open to debate.

Finally, it should be noted that the updated economic model included the functionality to assess outcomes for patients with moderate or severe disease at baseline separately in accordance with results reported in the CS.[1] The ERG is satisfied an analysis for these subgroups may be informative, despite not being defined in the NICE final scope for this STA;[8] however the ERG expresses concerns with the conduct of these analyses (see Section 5.2.9).

5.2.4 Intervention and comparators

Intervention, comparators and treatment regimens included in the CS[1] and company's model[15]

Table 33 summarises the treatment regimens included within the company's model[15].

Table 33 Description of interventions/comparators assessed in the company's model[15]

Treatment	Induction regimen	Maintenance regimen	Administration
Vedolizumab	300mg at week 0 and 2[a]	300mg every 8 weeks[a]	i.v. infusion
Infliximab	5mg/kg at week 0 and 2[a]	5mg/kg every 8 weeks[a]	i.v. infusion
Adalimumab	80 mg at week 0 and 40 mg at week 2, 4 and 6[a]	40mg every 2 weeks[a]	self-administered s.c. injection
Conventional non-biologic treatment	Not specified – all treatment appear to be assumed to be given daily indefinitely		Not specified

[a] given with concomitant medications (conventional non-biologic therapy)
i.v. intravenous
s.c. subcutaneous

- **Interventions**

The intervention under consideration is vedolizumab (trade name Entyvio®), 300 mg powder for concentrate for solution for infusion given as an intravenous (i.v.) infusion.

The treatment regimen assumed by the company for the base case analysis is 300 mg i.v. infusion at weeks 0 and 2 with assessment at week 6 to reflect the treatment regimen used in the GEMINI II trial[11] (see CS[1] pg. 209 and pg. 216 and clarification response[2] question B27). Vedolizumab 300 mg i.v. infusion is assumed to be continued every eight weeks in responding patients only (referred as maintenance therapy).

Scenario analyses are conducted assuming treatment response is assessed at week 10 or 14 to reflect the labelling of vedolizumab (see clarification response[2] question B55).

- **Comparators**

Table 34 summarises the comparators included in the company's model[15] according to the population under evaluation. Within all three analyses (mixed-ITT, anti-TNF-α naïve and anti-TNF-α failure subgroups), conventional non-biologic therapy (a combination of 5-ASAs, immunomodulators and corticosteroids) is included as a comparator. Anti-TNF-α agents (infliximab, adalimumab) are included only in the analysis for the anti-TNF-α naïve subgroup (these agents are excluded from the analyses of the mixed-ITT and anti-TNF-α failure subgroups).

Table 34 Comparators included in the company's model[15]

Population	Comparators
mixed-ITT population	• conventional non-biologic therapy (a combination of 5-ASAs, immunomodulators and corticosteroids)
anti-TNF-α naïve subgroup	• conventional non-biologic therapy (a combination of 5-ASAs, immunomodulators and corticosteroids) • adalimumab • infliximab
anti-TNF-α failure subgroup	• conventional non-biologic therapy (a combination of 5-ASAs, immunomodulators and corticosteroids)

Infliximab is assumed to be given at a dose of 5 mg/kg i.v. at weeks 0 and 2 with assessment at week 6, followed by infliximab 5 mg/kg i.v. every eight weeks (maintenance phase) based on the license for infliximab (see clarification response[2] question B27).

Adalimumab is assumed to be given at a dose of 80 mg subcutaneous (s.c.) injection at week 0 and 40 mg s.c. at week 2, 4 and 6, followed by 40 mg s.c. every other weeks based on the license for adalimumab according to the company (see clarification response[2] question B27).

The company further adds (see CS[1] pg. 209) that the chosen treatment regimens are consistent with the regimens from the trials from which the efficacy is derived and assumes that all therapy have the same induction phase (assessment at week 6) with costs adjusted accordingly.

Conventional non-biologic therapy is a mix of therapy. Within the company's model,[15] the efficacy reflects the mix of therapy used in the GEMINI trials[11,12] and includes a combination of corticosteroids (prednisone, budesonide, methylprednisolone, prednisolone, hydrocortisone, beclometasone, dexamethasone), immunomodulators (azathioprine, methotrexate, mercaptopurine) and 5-ASAs (mesalazine, sulfasalazine, balsalazide). In contrast, costs are derived from the mix of

conventional non-biologic therapies reported by the IBD Audit Steering Group (see Table 7.2.7.1 in CS[1] pg. 214).[14] The methods of administering these therapies are not specified by the company.

ERG's comments on the treatment regimens and comparators included within the CS[1] and company's model[15]

- **Treatment regimens assumed in the company's model**[15]

The ERG expresses several concerns with the treatment regimens that are assumed in the company's model.[15]

The company assumes the same induction phase duration for all therapy (6 weeks), adjusting the cost accordingly. No rationale for this is provided in the CS.[1] In response to a request for clarification (see clarification response[2] question B9), the company states that the same induction period was assumed to simplify the model (same decision tree for the induction phase), but this assumption could be relaxed to allow for modelling a different induction period for each therapy. The ERG assessment found that assuming the same induction period does not simplify the model; in contrary, this assumption led to discrepancies in the company's model (in terms of costing, cycle length and efficacy).[15] The ERG believes that the induction duration for each biologic should be used as there are no obvious benefits for using the same induction phase duration.

Where possible, the ERG believes that the treatment regimen should reflect the drug license,[9,10,57,58] efficacy data that are used in the company's model[11,12,41,51,54,66,66] and clinical practice. However, the ERG recognises that in some occasion, the treatment regimens used in the clinical trials may not entirely reflect the labelling (Table 35) and/or clinical practice. Discussion with clinical experts indicated that in practice, response is typically assessed between 10 to 14 weeks, but response may be assessed sooner in accordance with the licensing of the drugs. In response to a request for clarification (see clarification response[2] question B27), it appears that the company based the treatment regimen for adalimumab and infliximab on the labelling of the drug[57,58] and for vedolizumab on the regimen used in GEMINI II[11] rather than the licensing.[9,10]

Table 35 summarises the treatment regimens from the labelling, efficacy data used in the company's model[15]/induction phase of the trial, the regimen used in the company's model and the ERG's preferred treatment regimen. Discrepancies for each biologics are discussed in turn.

Table 35 Comparison of the treatment regimen recommended in the labelling, used in trial, assumed in CS[1] and ERG's preferred regimen

	Labelling	Trial	Assumed in CS[1] and company's model[15]	ERG's preferred
Vedolizumab	The recommended dose regimen of Entyvio is 300 mg administered by intravenous infusion at zero, two and six weeks and then every eight weeks thereafter. Patients with Crohn's disease, who have not shown a response may benefit from a dose of Entyvio at week 10 (see section 4.4). Continue therapy every eight weeks from week 14 in responding patients. Therapy for patients with Crohn's disease should not be continued if no evidence of therapeutic benefit is observed by week 14 (see section 5.1). Some patients who have experienced a decrease in their response may benefit from an increase in dosing frequency to Entyvio 300 mg every four weeks. In patients who have responded to treatment with Entyvio, corticosteroids may be reduced and/or discontinued in accordance with standard of care.	**GEMINI II[11]** **(Induction phase only)** Randomised patients were treated with infusions at weeks 0 and 2. Patients were assessed for treatment response at week 6. **GEMINI III[12]** Randomised patients were treated with infusions at weeks 0, 2 and 6. Patients were assessed for treatment response at week 6 and 10.	**Doses:** 300mg at week 0 and 2 **Assessment:** week 6 **Maintenance:** 300 mg every 8 weeks	**Doses:** 300mg at week 0, 2 and 6 **Assessment:** week 10/14 **Maintenance:** 300 mg every 8 weeks (from week 14 for responders at week 10)

Adalimumab	The recommended Humira induction dose regimen for adult patients with moderately to severely active Crohn's disease is 80 mg at week 0 followed by 40 mg at week 2. In case there is a need for a more rapid response to therapy, the regimen 160 mg at week 0 (dose can be administered as four injections in one day or as two injections per day for two consecutive days), 80 mg at week 2, can be used with the awareness that the risk for adverse events is higher during induction. After induction treatment, the recommended dose is 40 mg every other week via subcutaneous injection. Alternatively, if a patient has stopped Humira and signs and symptoms of disease recur, Humira may be re-administered. There is little experience from re-administration after more than 8 weeks since the previous dose. During maintenance treatment, corticosteroids may be tapered in accordance with clinical practice guidelines. Some patients who experience decrease in their response may benefit from an increase in dosing frequency to 40 mg Humira every week. Some patients who have not responded by week 4 may benefit from continued maintenance therapy through week 12. Continued therapy should be carefully reconsidered in a patient not responding within this time period.	**CLASSIC-I (used in the NMA and company's model[15])[51]** Randomised patients were treated with subcutaneous induction regimens at weeks 0 and 2 (160/80 mg or 80/40 mg) Patients were assessed for treatment response at week 1, 2 and 4. **Watanabe(2012)[20]– Not used in company's model** Randomised patients were treated with subcutaneous induction regimens at weeks 0 and 2 (160/80 mg or 80/40 mg). Patients were assessed for treatment response at week 2 and 4.	**Doses:** 80 mg at week 0 40 mg at week 2, 4 and 6 (i.e. 5 doses of 40 mg) **Assessment:** week 6 **Maintenance:** 40 mg every 2 weeks	**Doses:** 80 mg at week 0 and 40 mg at week 2 (i.e. 3 doses of 40 mg) **Assessment:** week 4 **Maintenance:** 40 mg every 2 weeks

Infliximab	5 mg/kg given as an intravenous infusion followed by an additional 5 mg/kg infusion 2 weeks after the first infusion. If a patient does not respond after 2 doses, no additional treatment with infliximab should be given. Available data do not support further infliximab treatment, in patients not responding within 6 weeks of the initial infusion. In responding patients, the alternative strategies for continued treatment are: • Maintenance: Additional infusion of 5 mg/kg at 6 weeks after the initial dose, followed by infusions every 8 weeks or • Re-administration: Infusion of 5 mg/kg if signs and symptoms of the disease recur (see 'Re-administration' below and section 4.4). Although comparative data are lacking, limited data in patients who initially responded to 5 mg/kg but who lost response indicate that some patients may regain response with dose escalation (see section 5.1). Continued therapy should be carefully reconsidered in patients who show no evidence of therapeutic benefit after dose adjustment.	**ACT-1 trial (used in the health economic model)**[66] Patients received an initial infusion of infliximab at week 0. At week 2, patients were stratified by response status and randomised to 1 of 3 treatment strategy groups **Targan et al (1997)[19] – used in the company's NMA** Randomised patients were treated with infusions at weeks 0. Patients were assessed for treatment response at week 2 and 4.	**Doses:** 5mg/kg at week 0 and 2 **Assessment:** week 6 **Maintenance:** 5mg/kg every 8 weeks	**Doses:** 5mg/kg at week 0 and 2 **Assessment:** week 6 **Maintenance:** 5mg/kg every 8 weeks

Vedolizumab

Efficacy data for the induction phase of vedolizumab are available from two pivotal trials; (a) GEMINI II[11] in which patients were treated with infusions at weeks 0 and 2 (i.e. 2 doses) and assessed for treatment response at week 6 and (b) GEMINI III[12] in which patients were treated with infusions at weeks 0, 2 and 6 (i.e. 3 doses) and assessed for treatment response at week 6 and 10.

The labelling of the drug[9,10] recommends vedolizumab to be given at week 0, 2 and 6 (i.e. 3 doses) and then every eight weeks thereafter (Table 35). The labelling[9,10] further states that patients with CD, who have not shown a response **may benefit** from a dose of vedolizumab at week 10 and that therapy should be continued every eight weeks from week 14 in responding patients.

In the base case analysis (see clarification response[2] question B27), the company assumes that patients are treated with vedolizumab at weeks 0 and 2 (i.e. 2 doses) and assessed for treatment response at week 6 based on the schedule used in GEMINI II[11] and uses pooled efficacy data from GEMINI II[11] and GEMINI III[12] at week 6. Scenario analyses are conducted assuming assessment at weeks 10 and 14 respectively (see clarification response[2] question B55).

Clarification was sought on the rationale for using assessment at week 6 rather than week 10 or 14 for the base case analysis (see clarification response[2] question B55). In response, the company states that *"the base case model uses an assessment at 6 weeks to reflect the design of the trial: the induction period was 6 weeks and patients were re-randomised at that time point…"*

The ERG questions the treatment regimen assumed in the base case analysis for vedolizumab. Notably,

- the treatment regimen in GEMINI III[12] (i.e. doses at week 0, week 2 and week 6 with assessment at week 10) is largely in adherence with the treatment regimen recommended in the labelling of the drug (i.e. doses at week 0, week 2 and week 6 with assessment at week 10, during which only non-responders **may** receive an additional dose), compared with the treatment regimen used in GEMINI II[11] which uses a non-standard schedule (doses at week 0 and week 2 with assessment at week 6).
- the treatment regimen for infliximab and adalimumab appear to be based on the labelling rather than the induction phase of the respective trials (see clarification response[2] question B27). The approach taken for vedolizumab is inconsistent.

The ERG recognises that this is open to debate, as the patient population randomised to the maintenance phase of GEMINI II[12] (after 2 doses at week 6) may be slightly different to the population who responded to treatment in GEMINI III (after 3 doses at week 10).[11] However, given

the approach taken by the company for other drugs (regimen based on license) and for vedolizumab (using pooled efficacy from GEMINI II[11] and GEMINI III[12]), the ERG believes that the base case analysis should use the treatment regimen for the induction phase from the GEMINI III trial[12] (i.e. dose at week 0, week 2 and week 6 with assessment at week 10; with responders receiving the next dose at week 14) to reflect the labelling of the drug. The ERG recognises that the labelling recommends that in some patients who do not respond at week 10, an additional dose may be given and recognises that this cannot be captured without assumptions being required.

Adalimumab

In Table 7.2.2.1 of the CS[1] (see CS[1] pg. 207), the company suggests that efficacy data for adalimumab are available from two trials (a) the CLASSIC-I trial[51] in which patients were assessed for treatment response at week 4 and the ENACT-1 trial[68] in which patients were assessed for treatment response at week 8. This is not factually correct as the ENACT-1 trial[68] assessed the efficacy of natalizumab for the treatment of CD, not adalimumab.

The company assumed that adalimumab is given at a dose of 80 mg s.c. injection at week 0 and 40 mg s.c. at week 2, 4 and 6. In response to a request for clarification (see clarification response[2] question B27) the company confirmed that the assumed treatment regimen is based on the license for adalimumab.

The ERG has two concerns with the treatment regimen and justification provided by the company in the CS[1] and during the clarification process[2];

- the company (see clarification response[2] question B27) suggests that the induction phase in the labelling for adalimumab is 8 weeks (i.e. doses at week 0, 2, 4, 6). This is not in adherence with the labelling of adalimumab which suggests the induction phase to be 4 weeks (dose at week 0 and week 2): *"patients should receive 80 mg at week 0 followed by 40 mg at week 2…. After induction treatment, the recommended dose is 40 mg every other week via subcutaneous injection…. Some patients who have not responded by week 4 may benefit from continued maintenance therapy through week 12. Continued therapy should be carefully reconsidered in a patient not responding within this time period."* However, the ERG recognises that according to the licensing, non-responder at week 4 may receive adalimumab up to week 12 at the physician's discretion.
- the company states (see CS[1] pg. 209) that *"The model is based upon induction efficacy data as reported from the clinical trials".* However, efficacy for the induction phase of adalimumab in the company's model[15] is taken from results of the NMA which uses data from the CLASSIC-I trial at week 4.[51] In this trial, randomised patients were treated with subcutaneous induction regimens at weeks 0 and 2 and assessed for response at week 1, 2, and 4. Therefore the induction phase should be 4 week in the economic model. As an aside, the Watanabe trial[20] which was excluded from the NMA

by the company (see Section 4.3), uses a similar treatment regimen (dose at week 0 and 2 with assessment at week 4) to the CLASSIC-I trial.[51]

The ERG recognises that this is open to debate, as the labelling of adalimumab is less clear with respect to the induction period. However, the ERG believes that the following treatment regimen should be used for adalimumab to reflect efficacy data used in the economic model; a dose of 80 mg s.c. injection at week 0 followed by 40 mg s.c. injection at week 2 with assessment at week 4. It should be noted that the treatment regimen assumed in the CLASSIC-I trial[51] is largely in adherence with the labelling of adalimumab.[58] The ERG recognises that the labelling[58] suggests that some patients who have not responded by week 4 **may benefit** from continued maintenance therapy through week 12. In the absence of data, this cannot be captured without assumptions being made on the effectiveness of continued therapy with adalimumab in non-responders.

To a lesser extent no analysis is presented using the accelerated schedule for adalimumab (160 mg/80mg).

Infliximab

Efficacy data for infliximab are taken from the ACCENT-1 trial[66] in the company's model.[15] Limitations for using data from the ACCENT-1 trial[66] instead of results from the NMA using the Targan study[19] are discussed in Section 4.3.

In the ACCENT-1 trial, [66] patients received an initial infusion of infliximab at week 0. At week 2, patients were stratified by response status and randomised to 1 of 3 treatment strategy groups:

- of 5 mg/kg of infliximab at weeks 2 and 6 followed by 5 mg/kg every 8 weeks (5 mg/kg scheduled strategy),
- or of 5 mg/kg of infliximab at weeks 2 and 6 followed by 10 mg/kg every 8 weeks thereafter (10 mg/kg scheduled strategy),
- infusions at weeks 2 and 6 and every 8 weeks thereafter until week 46 of placebo (episodic strategy).

In Table 7.2.2.1 in the CS[1], the company (see CS[1] pg. 208) suggests that patients in the ACCENT-1 trial[66] are assessed for treatment response at week 6, following doses at week 0 and 2. This is misleading as patients in the ACCENT-1 trial[66] received a single dose at week 0 and were assessed for response at week 2. Whilst no information is provided in the CS,[1] it appears that the company used efficacy data from patients randomised to the scheduled maintenance strategy, and therefore received a second dose at week 2. This was confirmed during the clarification process (see clarification

response[2] question B4). The company assumed response to be assessed at week 6, using the average of the week-2 and week-10 assessments from the ACCENT-1 trial[66] of patients randomised to the scheduled strategy (combined 5 and 10 mg/kg).

The treatment regimen used for infliximab for the induction phase reflects the labelling of the drug[57] and reflects the efficacy data that are used within the economic model.[66] However, the ERG notes that this is inconsistent with the assumption made for vedolizumab where the induction period was based on the GEMINI II[11] induction phase only (despite not reflecting the labelling), rather than the subgroup of patients randomised to the maintenance phase who received a 3rd dose at week 6. Whilst the ERG does not believe the following treatment regimen to be appropriate, for consistency, the treatment regimen for the induction phase of infliximab should be based on the induction phase of the ACCENT-1 trial[66] i.e. a single dose at week 0 with assessment at week 2.

- **The comparators considered**

Within all three analyses (mixed-ITT, anti-TNF-α naïve and anti-TNF-α failure subgroup), conventional non-biologic therapy (a combination of 5-ASAs, immunomodulators and corticosteroids) is included as a comparator. Anti-TNF-α agents (infliximab, adalimumab) are included only in the analysis for the anti-TNF-α naïve subgroup (these agents are excluded from the analyses of the mixed-ITT and anti-TNF-α failure subgroups).

The mixed-ITT population represents a combination of those patients who have previously received anti-TNF-α agents and those who are anti-TNF-α naïve; a proportion of these patients are clearly suitable for treatment with adalimumab or infliximab, which are not comparators in the model. It is unclear how one should interpret the results of the analysis.

The company's analysis within the anti-TNF-α failure subgroup excludes all other biologic therapy. However, the use of a second anti-TNF-α agent following the failure of a first anti-TNF-α agent may be possible particularly where loss of response has occurred due to development of antibodies to the first anti-TNF-α therapy; however, the ERG recognises the limited efficacy evidence available. To the ERG's knowledge, no data are available on the efficacy of infliximab in patients in whom the disease has responded inadequately to, or is no longer responding to an anti-TNF-α . In contrast, clinical evidence is available regarding the efficacy of adalimumab in patients in whom the disease has responded inadequately to, or is no longer responding to infliximab; these data however only reflect patients with secondary failure (i.e. failure during the maintenance phase after demonstrating a response to induction with infliximab). The population in the adalimumab clinical trial was not deemed to be comparable to the population included in the vedolizumab trial by the company[1] as the

adalimumab trial only included secondary failure patients (primary failure patients were excluded, defined as lack of response to the induction phase).

The ERG questions the exclusion of adalimumab as a comparator for the anti-TNF-α failure subgroup. The ERG notes that despite the arguments from the company, the company reported results from a NMA using the anti-TNF-failure subgroup in the vedolizumab studies (primary and secondary failure) versus the anti-TNF–failure subgroup (secondary failure) in the adalimumab study (see CS[1] pg. 128). Whilst debatable, the ERG believes that an analysis could be presented against adalimumab for the anti-TNF-α failure subgroup for completeness.

Within the company's model,[15] the efficacy for conventional non-biologic therapy reflects the mix of therapy used in the GEMINI trials[11,12] for the mixed-ITT and anti-TNF-α failure subgroup and includes a combination of corticosteroids (prednisone, budesonide, methylprednisolone, prednisolone, hydrocortisone, beclometasone, dexamethasone), immunomodulators (azathioprine, methotrexate, mercaptopurine) and 5-ASAs (mesalazine, sulfasalazine, balsalazide). In contrast, costs are derived from the treatment mix of conventional non-biologic therapy reported by the IBD Audit Steering Group (see CS[1] Table 7.2.7.1 pg. 214).[14]

Patients in the GEMINI trials[11,12] were recruited from a large number of centres worldwide, with varying clinical practice. The generalizability of the mix of treatments from the GEMINI trials to the UK population is unclear. In response to a request for clarification (see clarification response[2] question B29) the company confirmed that "*the use of conventional therapy within the GEMINI II and GEMINI III trials was protocol driven and the trial was international and may not represent treatment patterns in England and Wales"*. Similarities in the mix of therapy used in the GEMINI trials[11,12] and the IBD audit[14] are also unknown. However, some differences in the type of corticosteroids used were noted; it is unclear what the impact would be.

5.2.5 Perspective, time horizon and discounting

Perspective and discounting

The company's model[15] adopts a NHS perspective. Costs borne by the PSS are excluded from the company's economic analysis; the company states that these are expected to be minimal (see CS[1] Table 7.2.6.1 pg. 213).

All costs and health outcomes are discounted at a rate of 3.5% per annum.

The ERG considers these to be appropriate and in adherence with the NICE reference case.[64]

Time horizon

The company's base case analysis adopts a 10-year time horizon;[1] a lifetime horizon and 1 year time horizon are considered in the sensitivity analysis. In Table 7.2.6.1 in the CS[1] (see CS pg. 213) and in response to a request for clarification (see clarification response[2] question B15), the company states that *"previous models have used time horizons between 1 year and lifetime. 10 year time-horizon chosen to balance the lifetime nature of CD and 1-year clinical trial data. Other time horizons are used in scenario analyses."*

The NICE Reference Case[64] stipulates that the time horizon of the analysis should be long enough to capture all important differences in costs or outcomes between the technologies being compared. It is not clear whether all relevant differences in health gains and costs would be captured within this 10-year period. The ERG believes that a lifetime horizon is most appropriate but notes that given the short duration of the clinical trials used to inform the model (maximum 54 weeks), the extrapolation of the available data over a lifetime horizon is subject to considerable uncertainty. It should be noted that the ICERs for vedolizumab against conventional non-biologic therapy become more favourable assuming a lifetime horizon under the company's base case assumptions.

5.2.6 Treatment effectiveness and extrapolation

Key efficacy parameters used within the company's model[15] are either (a) observed or (b) derived.

This includes:

- the probabilities of response (defined as a drop in CDAI score of 70 points or more) and remission (defined as a CDAI ≤150) to the induction phase (Observed.)
- the percentages of responder to the induction phase with moderate to severe CD (Observed.)
- the initial induction vectors (derived.)
- the probabilities of response (defined as a drop in CDAI score of 70 points or more) and remission (defined as a CDAI ≤150) at the end of the maintenance phase (Observed.)
- the transition probabilities for patients entering the maintenance phase(derived.)
- the probabilities of discontinuation due to AEs (Observed.)
- the incidence of AEs (Observed.)
- the probabilities of surgery and transition from the surgery health state
- CD-related and other-cause mortality

Key efficacy parameters used within the company's model[15] and ERG's comments are summarised in turn below.

Initial induction vectors

Table 36 summarises the initial induction vectors used within the company's model.[15]

Approach used by the company to estimate the initial induction vectors, i.e. redistribution of patients into the different health states following induction therapy.

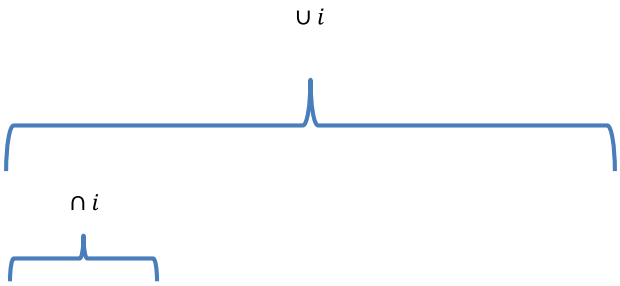
As illustrated below (Figure 14), initial induction vectors are derived from five input parameters, namely;

- the probabilities of response to the induction phase ($\cup i$)
- the probabilities of remission to the induction phase ($\cap i$)
- the proportion of responders with moderate to severe disease(ρ)
- the probabilities of surgery (φ)
- and the probabilities of death (ω).

Table 36 Initial induction vectors used within the company's model[15]

	Responders					
	Remission	**Mild**	**Moderate to severe**	**Non-responders**	**Surgery**	**Death**
Mixed-ITT population						
Conventional non-biologic therapy	9.86%	16.78%	7.16%	64.17%	2.03%	0.02%
Vedolizumab	16.78%	21.06%	10.17%	49.95%	2.03%	0.02%
anti-TNF-α naïve subgroup						
Adalimumab	29.92%	19.74%	10.77%	37.53%	2.03%	0.02%
Conventional non-biologic therapy	15.63%	15.87%	6.83%	59.64%	2.03%	0.02%
Infliximab	34.50%	17.68%	11.32%	34.47%	2.03%	0.02%
Vedolizumab	34.89%	8.67%	9.45%	44.96%	2.03%	0.02%
anti-TNF-α failure subgroup						
Conventional non-biologic therapy	10.18%	13.28%	7.52%	66.99%	2.03%	0.02%
Vedolizumab	13.08%	20.71%	10.83%	53.35%	2.03%	0.02%

Figure 14 Diagrammatic representation of the derivation on the initial induction vectors



Remission*	Mild CD*	Responder in Moderate to CD*	Non-Responder in Moderate to CD*	Surgery	Death
$\cap i$	$\cup i - [\cap i + (\cup i \times \rho)]$	$\cup i$	$1 - \cup i$	φ	ω

$\cup i = probability\ of\ response\ at\ induction\ phase$
$\cap i = probability\ of\ remission\ at\ induction\ phase$
$\rho = proportion\ of\ responders\ with\ mild\ to\ moderate\ disease$
$\varphi = probability\ of\ surgery$
$\omega = probability\ of\ death$

ERG comments on the approach used by the company to derive the initial induction vectors

The ERG notes limitations in the approach used by the company to derive the initial induction vectors. In particular, the correlation between input parameters used is ignored which may lead to inconsistencies. The company has access to patient-level GEMINI trial data[11,12] on the observed initial induction vectors for patients treated with vedolizumab and conventional non-biologic therapy; these could have been used to directly calculate the initial induction vectors and preserve the correlation between inputs (for at least the mixed-ITT and anti-TNF-α failure subgroup). The ERG recognises that assumptions may however be necessary for other biologics as the company would not have access to the data.

It should also be noted that the non-responder group is a mix of patients with mild and moderate to severe CD (i.e. patients with a drop in CDAI score of less than 70 but with a CDAI score between 150 to 220); these patients are assumed to have moderate to severe CD in the company's model[15]. This is inappropriate.

Probabilities of response ($\cup i$) and remission ($\cap i$) to the induction phase

Probabilities of response and remission to the induction phase used in the company's model[15]

Table 37 summarises the probabilities of response (drop in CDAI score of 70 points or more) and remission (CDAI ≤ 150) to the induction phase used in the company's model.[15]

Table 37 Probabilities of response and remission to the induction phase used within the company's model[15]

	Mixed-ITT		**anti-TNF-α naïve**				**anti-TNF-α failure**	
	CT	**VDZ**	**CT**	**VDZ**	**INF**	**ADA**	**CT**	**VDZ**
Response ($\cup i$)	33.80%	48.02%	38.33%	53.01%	63.50%	60.43%	30.97%	44.62%
Remission ($\cap i$)	9.86%	16.78%	15.63%	34.89%	34.50%	29.92%	10.18%	13.08%

CT = conventional non-biologic therapy; VDZ = vedolizumab; INF = infliximab; ADA = adalimumab

Within the mixed-ITT and anti-TNF-α failure subgroups, the probabilities of remission and response are taken from a pooled analysis of the GEMINI trials[11,12] (see CS[1] Table 7.3.1.4 in CS[1] pg. 221) for vedolizumab and conventional non-biologic treatment.

Within the anti-TNF-α naïve subgroup, the probabilities of remission and response are taken from the company's NMA for all therapy, except for infliximab; the company's model[15] uses the averages of the week-2 and week-10 assessments (see clarification response[2] question B4) from the ACCENT-1 trial for infliximab.[66] The company[1] argues in a footnote (see CS[1] Table 7.3.1.2 pg. 219 and Table 7.3.1.4 pg. 221) that data from the ACCENT-1 trial[66] were used as the Targan study[19] used in the NMA (a) included a very small sample size and (b) did not measure a standard dosage of infliximab.

ERG comments on the probabilities of response and remission to the induction phase used in the company's model[15]

Limited details are provided on the NMA used in the economic model. The CS[1] states that (see CS[1] pg218) *"to estimate the efficacy of each biologic treatment, we estimated odds ratios using the response and remission data from the MTC (see Section 6.7)"*, referring to results from the NMA presented in the clinical section; which uses a Bayesian framework. However, it appears from the economic model[15] and response to clarification (see clarification response[2] question A41) that the company uses results from a frequentist approach, but states in response to a separate request for clarification (see clarification[2] question A39 and A41) that results between the two approaches provided similar estimates.

The ERG questions the partial use of the NMA for the anti-TNF-α naïve subgroup. The ERG recognises that both the Targan[19] and ACCENT-1 trial[66] have some limitations. However, the ERG believes that results from the NMA (which uses the Targan trial[19]) should be used in the base case for infliximab, instead of using data from a separate single arm trial (not linked to the NMA). Contrary to the argument from the company,[1] the ERG believes that the Targan study[19] should be used as this is the only placebo-controlled trial assessing the efficacy of infliximab for the induction phase. The ERG recognises that the study recruited a small number of patients and that as such, the results need to be interpreted with caution. However, the ERG does not believe this to be a sufficient reason to dismiss this trial. Adjustment could be made to account for the small sample size. A second argument from the company is that a low placebo effect was observed in the trial. The ERG questions the validity of this argument as this was a randomised placebo controlled trial and therefore this should be reflected in the infliximab arm. However, the ERG recognises this may arise due to the small sample size, but adjustment could be made in the NMA. Finally, the company argues that the trial did not measure a standard dosage (a single dose was given with assessment at week 4). The ERG recognises that this does not reflect the licensing of the drug (dose at week 0 and week 2 with assessment at week 6); however, the treatment regimens in the model for the vedolizumab appear to be based on trial data[11] rather than the marketing authorisation.[9,10] A pessimistic assumption could be to assume that the efficacy at week 4 following a single dose is equivalent to the efficacy at week 6 following 2 doses; this is pessimistic as data from the ACCENT-1 trial[66] shows that a second dose provide more benefit.

The company uses data from the ACCENT-1 trial[66] (separated from the NMA). The ERG notes the following limitations for using the ACCENT-1 trial[66] in the economic analysis: (a) the absence of a placebo arm (b) a different definition of clinical response (defined as a reduction in CDAI ≥70 points and ≥25% from baseline) (c) the use of data in the subgroup of patients randomised to maintenance rather than the ITT induction phase. These limitations are not discussed in the CS.[1]

Following the clarification process (see clarification response[2] question B4), it appears that the company used data at week 2 in all patients randomised to infliximab (receiving a single dose a week 0) and data at week 10 in patients randomised to the scheduled strategy groups only. It should be noted that (a) Figure 4 of the Rutgeerts publication[66] suggests that clinical response at week 6 is closer to the clinical response observed at week 10 than to data at week 2 and that (b) data from the scheduled strategy group at week 10 is a combination of patients receiving a second dose of 5mg/kg or 10mg/kg at week 2.

Furthermore, as indicated in Section 4.3, the company's NMA[1] uses data from the CLASSIC-1 trial[51] only for adalimumab. However, data from the Watanabe et al. trial[20] could have been used to inform the NMA; the company excluded this trial from the primary analysis. The ERG believes that the inclusion of Watanabe et al.[20] would increase the probabilities of remission and response for adalimumab.

The company also appears to have pooled data from the placebo arms of the included trials in the NMA (see clarification[2] question B5). The ERG notes that trials pertaining to two non-licensed drug in the UK (natalizumab and cetelizumab) are included and therefore, may bias the placebo estimate in the company's model[15] if the population included were different.

Finally, little detail is provided within the CS[1] on how data from the GEMINI trials was pooled.

Percentages of responders with moderate to severe disease (Þ) and derivation of the proportion of responders with mild CD

Method used in the company's model[15] to derive the proportion of responders with mild CD and moderate to severe CD at the end of the induction phase

The company calculates the proportion of patients who respond to treatment who have moderate to severe disease based on the probabilities of response ($\cup i$) and the percentages of responders with moderate to severe CD (ρ).

The percentages of responders to induction therapy with moderate to severe disease (ρ) are summarised in Table 38 (extracted from the company's model[15]) and are taken from the pooled (see clarification response[2] question B14) percentages of responders with moderate to severe CD randomised to the vedolizumab and placebo arms of the two GEMINI trials;[11,12] the percentages are assumed to be the same for all therapies.

Table 38 Percentage of responders with moderate to severe CD (extracted from the company's model[15])

Population	No. responders[a]	No. responders with moderate to severe CD[a]	Percentage
Mixed-ITT	236	50	21.19%
anti-TNF-α naive	101	18	17.82%
anti-TNF-α failure	136	33	24.26%
[a] Taken from the company's model[15]			

The proportion of patients who respond to treatment who have mild disease is then calculated as the remaining of responders minus patients in remission and responders with moderate to severe disease.

ERG comments on the derivation of the proportion of responders with mild and moderate to severe CD

Despite clarification provided by the company (see clarification response[2] question B14), the ERG is unclear how the percentages are calculated. The company states that *"The model uses the pooled proportion of responders in moderate/severe for vedolizumab and placebo (for the conventional therapy arm of the model). The data are pooled over both the treatment arms and over both clinical trials (GEMINI II and GEMINI III). This is a conservative assumption in favour of conventional therapy. For example, the proportion of responders, treated with placebo, with moderate/severe disease was 21/85 = 24.7%. The proportion of responders, treated with vedolizumab, with moderate/severe disease was 29/151 = 19.2%".*

Table 39 summarises the number of responders at week 6 (induction phase) for patients treated with vedolizumab and placebo from both the GEMINI II[11] and GEMINI III[12] studies using information reported in the CS[1] on pg. 442 for the anti-TNF-α anti-TNF-α naïve subgroup and pg. 445 for the anti-TNF-α anti-TNF-α failure subgroup.[1]

Table 39 Number of responders at week 6 (adapted from pg. 442 and pg. 445 from the CS[1]).

	Source	Treatment arm	week	Number of patients randomised	Number of responders (drop in CDAI score of 70 points or more)
anti-TNF-α anti-TNF-α naïve	CSR13011	Placebo	6	48	18
	CSR13011	Vedolizumab	6	51	25
	CSR13007	Placebo	6	76	30
	CSR13007	Vedolizumab	6	109	61
anti-TNF-α anti-TNF-α failure	CSR13011	Placebo	6	156	50
	CSR13011	Vedolizumab	6	155	79
	CSR13007	Placebo	6	70	20
	CSR13007	Vedolizumab	6	105	37
TOTAL				**770**	**320**

In the economic model,[15] the denominators used for the number of responders (defined as a drop in CDAI score of 70 points of more) are 236, 101 and 136 for the mixed-ITT, anti-TNF-α naïve and anti-TNF-α failure subgroup respectively (Table 38).

However, using information provided in the CS[1] on pg.442 and pg.445 and presented above on Table 39, the total number of responders to the induction phase at week 6 (drop in CDAI score of 70 points of more) for the mixed-ITT population (i.e. naïve and failure patients) is 320 (134 in the anti-TNF-α naïve and 186 in the anti-TNF-α failure subgroup) pooling data from the GEMINI II and III trials and data for each treatment arm. Consequently, the ERG is unclear on how the reported percentages were derived.

Furthermore, whilst the ERG recognises the lack of data for infliximab and adalimumab; the percentages of responders with moderate to severe disease for patients receiving conventional non-biologic therapy and vedolizumab could be calculated separately from the GEMINI trials.[11,12] The ERG sought clarification (see clarification response[2] question B14) and the company states that using treatment-specific data would be expected to improve the ICER in favour of vedolizumab when compared with conventional therapy. Taking the values reported by the company in response to clarification[2] would be true, however, as indicated, the ERG is not able to confirm the data used.

Finally, the ERG is concerned with the approach used by the company to estimate the proportion of responders remaining with mild CD. The ERG believes that the current approach may lead to discrepancies when the probability of remission is high and close to the probability of response. For

instance, if data from the NMA using the Targan study are used for infliximab (as this should be the case), there are discrepancies with negative proportion of patients with mild CD. An alternative approach would be to use the proportion of responders with moderate to severe disease amongst responders not in remission.

Probabilities of surgery in patients with moderate to severe CD (φ)

- **Probabilities of surgery assumed in the company's model**[15]

A fixed proportion of patients are assumed to undergo surgery during the first induction cycle (2.03%) or every 8-week cycle (2.70%), based on Frolkis et al.[69] (reported in the company's model[15]). The risk of surgery is assumed to be constant over time.

- **ERG comments on the probabilities of surgery used in the company's model**[15]

No details or references to the Frolkis study[69] and derivation of the probability used in the company's model[15] are included within the description of the model inputs in the CS.[1]

To help assess the validity of this input, the ERG provides a brief description of this study. The Frolkis study[69] is a meta-analysis of population-based studies on the risk of surgery of patients with inflammatory bowel disease (both CD and UC) and estimated that the risk of surgery 1, 5, and 10 years after diagnosis of CD was 16.3% (95% CI, 11.4%–23.2%), 33.3% (95% CI, 26.3%–42.1%), and 46.6% (95% CI, 37.7%–57.7%), respectively.

It appears from the company's model[15] that the value at one year (16.3%) is transformed into a 6 or 8 week transition probabilities. Assuming the risk of surgery to be constant is not supported by the evidence used.[69] The ERG believes that the value used in the company's model[15] for the probability of surgery will overestimate the number of surgeries (and possibly be more favourable to vedolizumab).

In response to a request for clarification (see clarification response[2] question B51), the company reported that "w*ithin the maintenance phase of the GEMINI II trial, 3.3% (5/153) of patients randomised to placebo and 1.3% (4/308) of patients randomised to vedolizumab underwent bowel surgery".* Assuming a risk of surgery of 2.7% every 8-week would appear to be an overestimate based on data from the GEMINI II study.[2,11] Given the model structure, it is unclear what the impact would be on the ICER assuming a lower surgery rate.

Transitions between disease states during the maintenance phase

Table 40 summarises the transition matrices used during the maintenance phase (fitted).

Table 40 Fitted maintenance phase pre-surgery transition probabilities

	From state\ To state	Remission	Mild	No response
Mixed-ITT population				
Vedolizumab	Remission	99.36%	0.64%	0.00%
	Mild	4.90%	59.31%	35.79%
	No response	0.00%	6.34%	90.96%
Conventional therapy	Remission	83.28%	16.72%	0.00%
	Mild	0.00%	56.57%	43.43%
	No response	0.00%	0.00%	97.30%
Anti-TNF-α naïve subgroup				
Vedolizumab	From state\ To state	Remission	Mild	No response
	Remission	95.98%	4.02%	0.00%
	Mild	0.00%	65.44%	34.56%
	No response	0.00%	10.83%	86.47%
Conventional therapy	Remission	88.16%	11.84%	0.00%
	Mild	0.10%	60.31%	39.60%
	No response	0.00%	3.26%	94.04%
Infliximab	Remission	97.12%	2.88%	0.00%
	Mild	0.54%	71.48%	27.98%
	No response	0.00%	23.35%	73.95%
Adalimumab	Remission	99.50%	0.50%	0.00%
	Mild	1.28%	49.36%	49.36%
	No response	0.00%	0.00%	97.30%
Anti-TNF-α failure subgroup				
Vedolizumab	Remission	98.31%	1.69%	0.00%
	Mild	0.00%	50.00%	50.00%
	No response	0.00%	0.00%	97.30%
Conventional therapy	Remission	78.43%	21.57%	0.00%
	Mild	0.00%	59.80%	40.20%
	No response	0.00%	1.49%	95.80%

Description of the approach used by the company to derive the transition matrices during the maintenance phase

Transition probabilities are derived so that (a) the proportion of patients in remission at the end of the maintenance treatment (approximately at one-year) predicted by the model matches the 'expected' proportion of patients in remission at the end of the maintenance phase and (b) the proportion of patients with mild disease at the end of the maintenance phase predicted by the model matches the 'expected' percentage of responders to the induction phase with a drop of 70 points of more in the CDAI score and not in remission at the end of the maintenance phase.

The 'expected' proportion of patients in remission at the end of the maintenance phase ($\beta 1$) is calculated as follow:

$\beta 1 = \cup i \times \cap m$

Where:

$$\cup i = probability\ of\ response\ at\ induction\ phase$$

$$\cap m = probability\ of\ remission\ at\ the\ end\ of\ maintenancephase$$

The 'expected' proportion of responders to the induction phase with a drop of 70 points of more in the CDAI score and not in remission at the end of the maintenance phase ($\beta 2$) is calculated as follow:

$\beta 2 = (\cup i \times \cup m) - (\cup i \times \cap m)$

Where: $\cup i = probability\ of\ response\ at\ induction\ phase$

$$\cup m = probability\ of\ response\ at\ the\ end\ of\ maintenance\ phase$$

$$\cap m = probability\ of\ remission\ at\ the\ end\ of\ maintenancephase$$

Transitions are 'calibrated' using the Solver linear programming add-in within Microsoft Excel to minimise the sum squared error of the 'expected' and predicted estimates by manipulating seven of nine transitions probabilities (quantities x_1 to x_7 in Table 41) conditional on (a) the model structure, (b) the initial starting matrix for calibration (c) a series of arbitrary constraints defined by the company and (d) input parameters. Details of the calibration approach are included in the response to clarification (see clarification response[2] question B21).

Table 41 Cells manipulated within the calibration process

From state \ To state	Remission	Mild	Moderate to severe
Remission	x_1	x_2	Assumed to be zero
Mild	x_3	x_4	x_5
Moderate to severe	Assumed to be zero	x_6	x_7

For each biologic treatment option, the calibration process used the same initial transition matrix, as shown in Table 42. A separate transition matrix is used for conventional non-biologic treatment. The justification for using different initial matrices for different treatment is not reported within the CS.[1]

Table 42 Initial starting vectors

Biologic treatment			
From state\ To state	**Remission**	**Mild**	**Moderate to severe**
Remission	0.95	0.05	0.00
Mild	0.00	0.65	0.35
Moderate to severe	0.00	0.1	
Conventional treatment			
From state\ To state	**Remission**	**Mild**	**Moderate to severe**
Remission	0.90	0.10	0.00
Mild	0.00	0.60	0.40
Moderate to severe	0.00	0.02	

Transition probabilities are assumed to be constant and applied for the remainder of the model.

ERG's comments on the approach used by the company to derive the transition matrices during the maintenance phase

The ERG recognises that calibration method may be necessary when input parameters are not directly observable. The calibration approach adopted by the company "guesses" seven unknown parameters by fitting these to two data-points conditional on a number of assumptions regarding what these probabilities might be, as represented by constraints in the Solver routine, an assumed initial matrix for the linear program and the model structure. It should be noted that fitting seven unknown parameters to two known data-points is likely to result in over-fitting. Many possible combinations of transition probabilities could fit the two 1-year data-points on response and remission.

Whilst the ERG recognises the need for calibration on some occasions, the ERG expresses some concerns with the approach undertaken by the company.

Firstly, the ERG attempted to re-calibrate the transition matrices but was unable to replicate the approach used by the company due to lack of transparency in the economic model.[15] Transition matrices are copy-pasted.

The constraints and starting matrices (see CS[1] Section 7.3.2) are based on assumptions made by the company which do not appear to be adequately justified using evidence. Arbitrarily, a different starting matrix is used for biologic therapy and for conventional therapy. In response to a request for clarification (see clarification response[2] question B20) the company stated that *"the calibration process uses an optimization process that may provide different results based on the starting values used. In addition, there are many optimal solutions; choosing starting values that are clinically-valid will provide more clinically-valid solutions that minimize the objective function as well. Starting values were selected based on a plausibility considering the relative efficacy of biologics to conventional therapy. Specifically based on trial results, patients on conventional therapy should experience a higher probability of progressing from remission to mild disease and mild disease to moderate/severe disease."* The ERG does not believe the response from the company to be satisfactory.

The target data-points used in the fitting process relate to (a) the probability of achieving remission and (b) response (defined as a drop in CDAI score of 70 points or more) but not remission at 1-year. The company attempts to fit the proportion of patients in remission and mild health states to these target data-points. This is not correct for the second target data-point. The ERG believes that the fitting process ignores those patients who achieved response but had moderate to severe disease and therefore the target data points does not match the data point the model is fitted to, as responders to the maintenance and not in remission may be in the mild or the moderate to severe CD state.

Importantly, the derivation of these transition probabilities is dependent on structural assumptions and input parameters. Therefore, the model needs to be recalibrated if alternative assumptions were to be used (such as changes in discontinuation rates, induction phase duration, probability of surgery, effectiveness etc.). This is not automatic within the economic model and transition matrices do not appear to be recalibrated for the sensitivity/scenario analyses undertaken by the company.

Transition matrices also appear to have been copied from separate analyses and therefore it is not possible to know, without refitting all the transition matrices, whether (a) the best solution was found and (b) whether the transition matrices were manually manipulated. Notably, the transition

probabilities for vedolizumab for the anti-TNF-α failure subgroup from the mild health state to the mild and moderate to severe health state are set exactly to 50%. It is unusual for a calibration approach to provide such rounding (notably when such constraint doesn't appear to have been defined). Finally, it should be noted that an error was identified in the original model in that the model attempted to calibrate to the wrong cell (see clarification respsonse[2] question B23); despite the error is corrected in the updated version of the economic model, transition matrices are unchanged.

Finally, transition probabilities are assumed to be constant and applied for the remainder of the model. Whilst uncertain, the ERG recognises the lack of evidence after one year.

Patient-level data from the GEMINI II trial[11] are available and could have been used to estimate the transitions between remission/mild/moderate-to-severe within the maintenance phase in patients treated with conventional non-biologic therapy and vedolizumab. The ERG recognises that observed data are not available for infliximab and adalimumab for the anti-TNF-α naïve subgroup and therefore assumptions or calibration may be necessary. An assumption may be to assume the same effectiveness in the maintenance phase for all biologic treatments.

Probabilities of response ($\cup m$) and remission ($\cap m$) during the maintenance phase

Probabilities of response and remission during the maintenance phase used in the company's model[15]

Table 43 summarises the probabilities of response (drop in CDAI score of 70 points or more) and remission (CDAI≤150) to the maintenance phase used in the company's model.[15]

Table 43 Probabilities of response and remission to the maintenance phase used within the company's model[15]

	Mixed-ITT		**anti-TNF-α naïve**				**anti-TNF-α failure**	
	CT	**VDZ**	**CT**	**VDZ**	**INF**	**ADA**	**CT**	**VDZ**
Response ($\cup m$)	24.93%[a]	47.40%	39.91%	63.45%	69.44%	49.35%	26.92%	29.27%
Remission ($\cap m$)	15.61%[a]	38.96%	24.81%	49.37%	45.71%	49.35%	12.82%	28.05%

[a] taken from the company's model.[15] There were discrepancies between the company's model[15] and CS[1]

Within the mixed-ITT and anti-TNF-α failure subgroups, the probabilities of remission and response at the end of the maintenance phase are taken from the GEMINI II trial[11] (Table 7.3.1.4 in p 221 in the CS[1]) for vedolizumab and conventional non-biologic treatment.

Within the anti-TNF-α naïve subgroup, the probabilities of remission and response at the end of the maintenance phase are taken from the company's NMA[1] for all therapies except for the probability of remission for adalimumab; the probability of remission was assumed to be the same as the probabilities of response. The company argues in a footnote (see CS[1] Table 7.3.1.3 pg. 220) that it was assumed that remission was equal to response due to differences in the trial design for the adalimumab maintenance trials and therefore the odds ratio for remission is higher than the odds ratio for response for adalimumab.

ERG comments on the probabilities of response and remission during the maintenance phase used in the company's model[15]

As for the NMA for the induction phase, limited details are provided on the NMA used in the economic model for the maintenance phase. It appears from the economic model[15] and response to clarification (see clarification response[2] question A41) that the company uses results from a frequentist approach instead of results from the Bayesian approach presented in the clinical section, but states in response to a separate request for clarification (see clarification[2] question A39 and A41) that results between the frequentist and Bayesian approaches provided similar estimates.

The probabilities of remission and response for patients on conventional therapy are taken from the probabilities of patients randomised to the maintenance phase who achieved a primary response with biologic. It is unclear whether the same efficacy is expected for conventional non-biologic treatment after response to conventional or other biologic treatment.

The ERG questions the justification from the company (see CS[1] Table 7.3.1.3 pg.220) on the reason why the odds ratio for remission is higher than the odds ratio for response for adalimumab. The company argues that this is due to differences in the trial design for the adalimumab maintenance trials. The ERG notes that the NMA appears to use data from the CLASSIC II trial[41] for both response and remission. The ERG believes that the reason for this inconsistency is that response and remission are estimated as two separate outcomes; these two outcomes are correlated with each other (remission is a subset of response). This is a key structural issue with the company's approach. The ERG also notes that patients in the CLASSIC II trial[41] were re-randomised after induction based on remission not response. The implications of this are not discussed in the economic section of the CS.[1] The ERG recognises that the only data available for adalimumab for the maintenance phase were from the CLASSIC II study[41] for the anti-TNF-α naïve subgroup.

Furthermore, the ERG sought clarification on why the remission rate for adalimumab was set equal to the response rate and why this assumption was preferred to setting the response rate equal to the remission rate (see clarification response[2] question B18). In response, the company stated that "*the proportion of patients in remission was set equal to response because the analysis provided a remission percentage greater than the response percentage. This is not feasible as remission is a subset of response. The alternative assumption that the proportion in response was equal to the proportion in remission was considered less likely and was not used in the model. Whilst the model has not been re-calibrated to consider this option, it is likely that adalimumab would dominate vedolizumab, with very slightly higher QALYs and a difference in costs of about £3,500"*. The ERG recognises the uncertainty but believes that these two scenarios are equally plausible contrary to the company's view.

As for the induction phase, the company appears to have pooled data from the placebo arms of the included trials in the NMA.[1] The ERG notes that trials pertaining to two non-licensed drugs in the UK (natalizumab and certolizumab) are included and therefore, may bias the placebo estimate in the company's model[15] if the population included were different.

Probabilities of discontinuation due to AEs

Probabilities of discontinuation due to AEs used in the company's model[15]

Table 44 summarises the probabilities of discontinuation due to adverse events for biologic treatment used within the company's model.[15]

Table 44 Annual probabilities of discontinuation due to AEs assumed in the company's model[15]

		Induction	Maintenance
Mixed-ITT population	Vedolizumab	3.03%	8.89%
anti-TNF-α naïve subgroup	Adalimumab	1.33%	5.26%
	Infliximab	1.33%	5.26%
	Vedolizumab	3.07%	6.06%
anti-TNF-α failure subgroup	Vedolizumab	2.69%	8.54%

The discontinuation rates for patients on vedolizumab are taken from a pooled analysis of the GEMINI studies[11,12] and are calculated for the three populations separately (mixed-ITT, anti-TNF-α naïve and anti-TNF-α failure).

The discontinuation rates for patients on adalimumab are taken from Hanauer et al.[51] for the induction phase and Sandborn et al.[41] for the maintenance phase; the discontinuation rate for infliximab is assumed to be the same as adalimumab.

ERG's comments on the assumptions used by the company regarding discontinuation from biologics during the maintenance phase

Limited description is provided by the company on how the discontinuation rates due to AEs were calculated. This aside, the company report in the clinical section (see CS[1] Section 6.76) and in the economic model (but not used) results from a NMA for discontinuation due to AEs. It is unclear why the results from this NMA haven't been used. Furthermore, the company states that due to lack of data, infliximab discontinuation rates were assumed to be similar to adalimumab. This appears to be contradicted by evidence included in the clinical section of the CS for the NMA[1] and economic model (but not used);

- data on treatment discontinuation due to AEs appear to be available for adalimumab for the induction phase from CLASSIC I[51] and in the maintenance phase from the CLASSIC II study.[41]
- data on the treatment discontinuation due to AEs appear to be available for infliximab from ACCENT-1 trial[54] from the maintenance trial (see CS[1] Table 6.7.6.2 pg. 144); and therefore could be used within the company's model[15]

In response to a request for clarification (see clarification response[2] question B38), the company stated *"In section 6.7 of the submission, odds ratios from an MTC are provided. Discontinuation rates are presented in Table 7.3.1.6 of the submission. The discontinuation rate for Infliximab is assumed to be the same as adalimumab due to a lack of reported data."* The ERG is unable to assess the method used to derive the discontinuation rates in the company's model.[15] It should be noted that as part of a request for clarification (see clarification response[2] question B38), the company provided an analysis using the same discontinuation rates for all biologic treatment (as requested by the ERG); this analysis showed minimal impact on the ICER (or ordering of the ICER in the incremental analysis for the anti-TNF-α naïve subgroup).

Transitions from the surgery health state

Transition probabilities from the surgery health state used in the company's model[15]

Table 45 summarises the transition probabilities for patients entering the surgery health state; these are taken from Bodger et al.[60]

Table 45 Transitions (8- weekly) from the surgery health state

From / To	**Remission**	**Mild**	**Moderate to severe**	**Surgery**
Surgery	52.72%	7.71%	5.82%	33.75%

ERG's comments on the transition probabilities from the surgery health states used in the company's model[15]

It is unclear from both the CS[1] and the Bodger publication[60] how the transition probabilities for patients undergoing surgery have been calculated. The ERG notes that according to these values, approximately a third of patients undergoing surgery are assumed to undergo subsequent surgery in the next cycle (8-weekly). This appears to be high and is recognised by the company (see CS[1] pg. 225 and clarification response[2] question B53).

In response to a request for clarification (see clarification response[2] question B53) the company states that "*whilst the transition probability provided by the model Bodger et al. appears to be quite high, examination of the cohort traces suggests the use of surgery predicted by the model is reasonable (see the response to B51, above)*". The ERG is not satisfied and disagrees with this statement as the response provided to question B51[2] does not support this statement; *"Within the safety population of the GEMINI II study, 37% (111/301) of the patients randomised to placebo and 44% (355/814) of the patients that received vedolizumab at any point in the trial had undergone surgery for Crohn's disease before entering the GEMINI II study. Within the GEMINI III study, 43% (89/207) of the patients randomised to placebo and 44% (92/209) of the patients randomised to vedolizumab had undergone surgery for Crohn's disease before entering the GEMINI III. Within the maintenance phase of the GEMINI II trial, 3.3% (5/153) of patients randomised to placebo and 1.3% (4/308) of patients randomised to vedolizumab underwent bowel surgery".* It is unclear what the impact on the ICER would be correcting the transition matrix for movement between states following surgery.

CD-related and other-cause mortality (ω)

Assumptions on mortality used in the company's model[15]

Patients may transition to death from any health state (except death) during any cycle. The risk of mortality is applied as an age-specific baseline other-cause mortality rate, with state-specific relative risks to reflect an excess risk of death due to CD.

The probability of dying from other causes was modelled derived from ONS life tables.[70] A state-specific relative risk is then used to reflect the excess risk of death due to CD.

Table 46 summarises the relative risks used within the HE; taken from Lichtenstein et al.[71]

Table 46 Relative mortality risk, by health state

Health state	RR used in company's model[15]
Remission	1.00*
Mild	1.27*
Moderate-severe	2.26*
Surgery	3.22*
* values taken from the company's model[15] due to discrepancies with values reported in the CS[1]	

ERG's comments on assumptions on mortality used in the company's model[15]

No reference or details are provided in the CS[1] on the value used or the Lichtenstein study.[71] In response to a request for clarification (see clarification response[2] question B46), the company stated that *"The relative mortality risks are listed in the Lichtenstein et al. 2006 publication. Health state specific utilities were used to reflect trends seen in clinical practice, as evidenced by the variation in parameter estimates."* The ERG does not believe this to be an adequate explanation.

To help assess the validity of this input, the ERG provides a brief description of this study. The Lichtenstein study[71] is a prospective study which evaluated the risk of mortality in patients treated with infliximab and other therapy in CD. The study included 6,290 patients; of which 3,179 received infliximab (5,519 patient-years), and 3,111 received other therapy (6,123 patient-years). The mean length of follow-up evaluation was 1.9 years. The authors reported that the mortality rates were similar for infliximab and non–infliximab-treated patients (RR, 1.24; 95% CI, .73–2.10). In a multivariate logistic regression model, compared with patients in remission, the authors reported no significant differences in excess mortality in patients with mild (1.266; CI: 0.562-2.852; p = 0.57), moderate/severe (2.256; CI: 0.9-5.653; p = 0.083) and unknown (3.223; CI: 0.776-13.387; p = 0.11) disease at baseline.

It appears that the values from the multivariate logistic regression model according to severity at baseline are used within the company's model.[15] The ERG questions (a) the assumption of a

differential mortality rate in the economic model and (b) the use of the relative risk of the unknown group to represent the excess risk of mortality associated with surgery.

As mortality is conditional on the current health states in the company's model,[15] the model predicts a greater life years for patients treated with biologics compared with patients receiving non biologic therapy. The Lichtenstein study[71] suggests no statistical differences in the excess mortality rates according to disease severity at baseline. Similarly, the Lichtenstein study[71] suggests no statistical differences in mortality between infliximab and non–infliximab-treated patients. It should be noted that no increased mortality rate was observed in patients randomised to the placebo arm in the GEMINI II trial.[11]

Clarification was also sought from the company (see clarification response[2] question B22) on how the model prediction at one year compares with the trial at one year for vedolizumab. In response, the company reported the number of deaths from the GEMINI II study.[11] The ERG does not believe this to be an adequate explanation.

The ERG recognises that this open to debate. However, as indicated in Section 5.2.10, the headline cost-effectiveness results presented by the company are based on the deterministic version of the model (using point estimates of parameters) rather than the expectation of the mean. Whilst PSA was undertaken by the company, probabilistic ICERs were not presented within the CS[1] and distribution were arbitrary (see Section 5.2.11). Given that results are presented deterministically and concerns regarding the conduct of the PSA (see Section 5.2.11), the ERG believes that the same excess risk mortality should be applied to all CD health states given the lack of evidence of a differential mortality rate between treatments.

It also appears that the company used the RR for patients with unknown disease severity at baseline from the Lichtenstein study[71] to represent the excess mortality rate for the surgery health state (see clarification response[2] question B47). No rationale has been provided by the company. The ERG believes that an excess risk of death for patients undergoing surgery may be appropriate, but that the source used in the company's model[15] is inappropriate.

Incidence of adverse events

Description of the approach used by the company to estimate the incidence of AEs

Table 47 summarises the incidence of AEs used in the company's model.[15] These are used to adjust HRQoL and costs.

Adverse events included in the company's model[15] were selected based on the opinion of two clinical experts. Estimates of the incidence of adverse events were derived through a simple (unadjusted) pooling of adverse event data reported in the publications of the pivotal clinical trials of the biologics identified in the MTC.[11,12,20,22,23,33,41,51,53-55] The company calculated the incidence of AEs as number of AEs divided by the total number of patients.

Adverse event rates were assumed to be the same for all three populations.

Table 47 Adverse events incidence probabilities assumed within the company's model[15]

	Serious infection	Tuberculosis	Lymphoma	Acute hypersensitivity reactions	Skin Reactions
Vedolizumab[11,12,22,23]	1.54%	0.00%	0.00%	0.00%	0.34%
Infliximab[33,54]	4.49%	0.28%	0.00%	0.00%	0.00%
Adalimumab[20,41,51,53,55]	1.80%	0.00%	0.00%	0.00%	0.00%
Conventional Therapy[11,12,20,22,23,33,41,51,53-55]	1.89%	0.00%	0.08%	0.74%	0.16%

ERG's comments on the approach used by the company to estimate the incidence of AEs

The ERG questions the approach used by the company to estimate the incidence of AEs. Notably, the calculations from the company are simplistic and appear to be erroneous as they do not account for the trial duration. Clarification was sought from the company (see clarification response[2] question B36) on this discrepancy and the ERG asked the company to provide an amended calculation to estimate the rate of adverse events per week (to allow a fair comparison between treatments). In response, the company stated that *"This was a simplifying assumption of the model. As currently calculated adverse events contribute approximately 1% to the overall costs of care for each comparator. Weekly rates of adverse events have been calculated and can be found in the updated model in the worksheet "Weekly AE calculation". These data have not been implemented in the model as the impact on the cost-effectiveness of VEDO will be slight*". The ERG could not find the

worksheet "Weekly AE calculation" in the updated model. Furthermore, whilst the ERG recognises that AEs may have a small impact on results, the company decided to include AEs in their base case, and therefore it is unclear why these 'corrected' rates have not been implemented in the economic model. It should be noted that in the original calculation, vedolizumab had the lowest incidence of adverse events; this is no longer the case with the corrected calculation for the incidence of AEs.

It is also unclear whether all or only grade 3 or 4 AEs are included. AEs were also selected based on the opinion of 2 clinical experts. It is unclear which AEs were excluded and the basis for their exclusion. For instance, in the clinical section (see CS[1] pg. 165), the company report abdominal pain and anal abscess as serious adverse events; these are not included in the model.

The AE probabilities for conventional non-biologic therapy were calculated from rates of AEs in the placebo arms of the included trials for vedolizumab, infliximab and adalimumab. As part of the trials, placebo-treated patients received a placebo transfusion or injection. It is unclear whether the adverse events experienced by the placebo arm, notably skin reactions are due to the infusion/injection which would not happen in normal practice for patients on conventional non-biologic therapy.

Finally, it should be noted that a NMA for the incidence of serious AEs is presented in an accompanying document to the submission.[16] It is unclear why data from this NMA have not been used in the company's model.

The ERG believes that the inclusion of AEs and the impact on costs and HRQoL in the economic model is flawed. However, the ERG conducted a scenario analyses removing AEs and showed this had little impact on the ICER, despite the SA in the CS[1] showing a large impact of the incidence of AEs on the ICER (see Section 5.2.11).

5.2.7 Health related quality of life

Health states utility values

Health states utility values used in the company's model[15]

Table 48 summarises the health utility values assumed within the company's model.

Table 48 Summary of health state utility values used in the company's model[15]

Health state	**Vedolizumab Trial Data**[11,12] **(Base case)**	**Buxton et al.**[72]
Remission	0.820	0.827
Mild	0.730	0.695
Moderate-severe	0.570	0.425
Surgery	0.570*	0.425*

* assumed to be the same as for moderate to severe CD

The company obtained EQ-5D utility scores for patients in remission (CDAI<150), mild disease (CDAI 150-219) or moderate to severe disease (CDAI 220-600) based on the EQ-5D scores collected in patients from the GEMINI II[11] and GEMINI III studies.[12] The company pooled data from the GEMINI trials and estimated utility score by health state regardless of study visit or treatment received. Alternative utility values identified in the systematic review were used in scenario analyses (see CS[1] Sections 7.4.5, 7.4.6 and 7.7.9).

It should be noted that the model distinguishes patients with moderate to severe CD who respond to and not respond to treatment. No differences in utility values are assumed by response categories. The company further assumed that the utility score for non-responders equal the utility score in patients with moderate to severe CD.

For the surgery health state, the company assumed the utility value to be same as for patients with moderate to severe disease in the absence of data from the GEMINI trials[11,12] or alternative sources.

ERG's comments on approach used by the company to estimate HRQoL for the main health states

The ERG is largely satisfied with the approach used by the company to estimate utility scores for the different health states of the company's model.[15] The company obtained EQ-5D utility score which is in adherence with the NICE Reference Case.[64] It is unclear whether UK tariffs were used.

It should be noted that the same utility score is assumed for patients with moderate to severe disease who respond or not to treatment. This is unlikely to be true as it would imply that response (control of symptoms) in these patients does not improve health. Similarly, the utility score for patients with moderate to severe disease is applied to non-responders. As previously indicated, non-responders may include patients with mild disease (CDAI between 150 – 220).

In the absence of data, the company assumed that the utility value for surgery was equal to the utility value for patients with moderate to severe CD. The ERG recognises the inability of GEMINI to inform estimates of the utility of patients undergoing surgery but is unsure of the validity of the assumption made by the company given that the aim of surgery is to improve quality of life. In response to a request for clarification (see clarification response[2] question B43) the company stated that *"the value used by Bodger et al., from the study by Buxton et al., 2007, is 0.112 per 8-week cycle: a utility value of 0.728 (0.112 multiplied by 6.5 8-week periods in a year). This value was not used in the model because it appears to be inconsistent with the utilities observed in the clinical trials: in the model, a patient undergoing surgery for Crohn's disease would have almost the same utility as a patient with mild Crohn's disease (a utility value of 0.730 is used for patients with a CDAI score of 150-220). Given that a patient with surgery would have disease severe enough to warrant surgery and also have surgery in that cycle of the model, this value of 0.728 was considered to be inconsistent with values observed in the GEMINI II and GEMINI III studies. Nevertheless, using a utility value of 0.728 for surgery, the ICER for vedolizumab compared with conventional therapy is £63,199. Using the base case utility value that ICER is £62,903."*

The ERG believes that the company could use the same assumption as in Bodger et al.[60] (rather than the actual value) i.e. that patients experience 2 weeks at an equivalent state of health as non-responders, and 6 weeks at an equivalent state of health as full responders. However, the impact on the ICER is likely to be minimal, as suggested by the company.

It should be noted a slight discrepancy between the model and the value use for the mild CD state; in the model the mild CD state includes patients with a CDAI of 220 whilst HRQoL for this health state includes patients up to a CDAI score of 219.

Adjustment of utility scores to account for adverse events

The company[1] attempts to adjust the utility scores associated with the health states to account for the effects of AEs. This involves a three-step approach (1) identifying evidence on the decrement in utilities associated with the AEs of interest, (2) calculating a weighting factor for each treatment based on the incidence of AEs and the decrement in utilities and (3) adjust health states utility scores based on the estimated weighting factors.

The decrements in utilities assumed for each AE are summarised in Table 49 and are taken from the published literature.[73-77]

Table 49 Utility estimates for adverse events (reproduced from Table 7.4.9.2 in CS[1])

Adverse Event	**Disutility Estimate**	**Source**
Serious infection	-0.520	Brown et al.[73] (= 1 − 0.48)
Tuberculosis	-0.550	Porco et al.[74] (= 1 − 0.45)
Malignancy (including Lymphoma)	-0.195	Hornberger et al.[75] (= 1 − 0.805)
Acute hypersensitivity reactions	-0.110	Beusterien et al.[76]
Skin site reactions	-0.030	Beusterien et al.[77]

Decrements in utilities are multiplied by the probabilities of experiencing each adverse event per cycle to calculate a weighting factor. Table 50 summarises the weighting factors calculated by the company.[15]

Table 50 Weighting factors applied to health states utility values

	Weighting factors (Taken from the company's model[15])
Vedolizumab	99.86%
Adalimumab	99.84%
Conventional non-biologic therapy	99.81%
Infliximab	99.56%

Finally, health states utility values are adjusted according to this weighting factor.

ERG's comments on the approach used by the company to adjust utility scores to account for adverse events

The ERG has concerns regarding the approach used by the company to adjust utility weights. However, the impact on the ICER is expected to be minimal.

As indicated, the ERG has concerns regarding the approach used by the company to calculate the incidence of AEs. As the weightings factors are a function of both decrement in utilities and the incidence of AEs, the ERG expresses some reservation on the weighting factors used in the company's model.[15]

Limited details are provided within the CS[1] and in response to clarification (see clarification response[2] question B44). To help with the assessment of the validity of the values used, the ERG provides brief descriptions of the studies selected by the company;

- the disutility for serious infection was estimated using a published economic evaluation of treatment for advanced breast cancer.[73] Within this study, standard gamble (SG) methods were used to elicit utility values for a variety of health states from 180 nurses.
- the disutility for tuberculosis (TB) was estimated using a published economic evaluation of tuberculosis evaluation and treatment of newly-arrived immigrants.[74] The elicitation methods within this study are unclear; estimates appear to be based on other literature and assumptions.
- the disutility for malignancy was estimated using a published economic evaluation of rituximab plus cyclophosphamide, vincristine and prednisolone for advanced follicular lymphoma (elicited utilities via the EQ-5D questionnaire from 222 patients with lymphoma).[75]
- the disutility for acute hypersensitivity reactions was taken from a cross-sectional SG study of societal preferences for treatment outcomes in chronic lymphocytic leukaemia using members of the UK general population.[76] A disutility for grade 3/4 pyrexia was reported; this value was used directly in the company's model.
- the disutility for skin site reactions was taken from a cross-sectional SG study of societal preferences for advanced melanoma health states using members of the general public in the UK and Australia.[77] A disutility of 0.03 was reported by UK responders; this value was used directly in the company's model.

Decrements in utility for serious infection, tuberculosis and malignancy appear to have been calculated by subtracting the utility of patients experiencing that AE from a baseline value of 1. This may overestimate the disutility as it assumes that those patients who were not experiencing the event have perfect quality of life. Furthermore, it is assumed that the decrement in utility last the full duration of the AE.

5.2.8 Resources and costs

Drug acquisition costs

- **Drug acquisition costs assumed in the company's model**[15]

Table 51 summarises the drug acquisition costs included in the company's model.[15]

Table 51 Acquisition costs assumed within the company's model[15]

Product	Unit cost	Units per induction cycle	Units per maintenance cycle	Cost per induction cycle	Cost per maintenance cycle
Vedolizumab (300mg vial)	[redacted]	2	1	[redacted]	[redacted]
Infliximab (100mg vial)	£419.62	8	4	£3,356.96	£1,678.48
Adalimumab (40mg prefilled pen/syringe)	£352.14	5	4	£1,760.70	£1,408.56[a]
Conventional treatment	£3.66	Mix of various products		£52.62	£70.16*
** Assumed to be £35.08 for patients whilst receiving biologic treatment* [a] assumed to be for a 8 week period					

The basic NHS list price of vedolizumab is £2,050 per 300mg vial. The company's model includes a lower drug acquisition cost to reflect the agreed Patient Access Scheme (PAS) for vedolizumab; the price used in the model is [redacted] per 300mg vial. The agreed PAS takes the form of a simple price discount (a reduction of [redacted] of the NHS list price) for the NHS. The acquisition costs of infliximab and adalimumab are based on drug prices reported within the BNF.[78]

As indicated (see Section 5.2.4), the company assumed that for vedolizumab and infliximab, the induction phase consisted of two i.v. infusions at weeks 0 and 2 with patient assessment at week 6. Responders are subsequently treated every 8 weeks thereafter in the maintenance phase. For adalimumab, the company assumed that patients receive a loading dose of 80 mg[1] s.c. self-administered injections at week 0 and 40 mg at week 2, 4, 6 and 8.[1] During the maintenance phase, patients received 40 mg of adalimumab every other week.

It should be noted that the dose of infliximab is conditional on the body weight. Infliximab is available at a dose of 100 mg i.v. infusion. The company assumed in the base case analysis that

[1] There is a typographical error in the CS[1]. The loading dose assumed in the health economic model is 80mg rather than 160 mg as stated in p.302.

patients receive four doses of 100 mg i.v. infusion at each administration based on a mean weight of 69kg.

Acquisition costs for conventional non-biologic therapy is estimated from the mix of treatments reported by the UK IBD Audit Steering Group,[14] together with doses and unit costs derived from the BNF (2013). A cost per day is calculated.

Table 52 Doses and unit costs of conventional therapy (adapted from Table 7.5.5.3 in CS[1] pg. 304)

Treatment	Dose and Frequency	Price	% Use	Cost per day
Aminosalicylates				
Balsalazide	1.5 g twice daily, adjusted according to response (maximum: 6 g daily)	750 mg, 130-cap pack at £30.42	5%	£0.94
Mesalazine	1.2-2.4 g daily in divided doses	400 mg, 120-tab pack at £41.62	5%	£1.47
Olsalazine	500 mg twice daily	250 mg, 112-cap pack at £19.77	5%	£0.71
Sulfasalazine	500 mg 4 times daily	500 mg, 112-cap pack at £5.82	5%	£0.29
Corticosteroids				
Budesonide	3 mg 3 times daily for up to 8 weeks	3 mg net price: 100-cap pack at £75.05	6%	£2.25
Prednisolone	1 metered application (20 mg prednisolone) once or twice daily for 2 weeks	14-application canister at £48.00	19%	£0.19
Immunomodulators				
Azathioprine	1-3 mg/kg daily	25 mg net price: 28-tab pack at £6.02; 50 mg, 56-tab pack at £5.04	57%	£0.19
Mercaptopurine	Initially 2.5 mg/kg, adjusted according to response	50 mg net price: 25-tab pack at £22.54	10%	£6.95
Methotrexate	10-25 mg once weekly	2.5 mg net price: 24-tab pack at £2.39; 28-tab pack at £3.27	11%	£0.92
Total cost				£70.16

In addition, the company's model assumes that whilst patients are receiving biologic therapy, the costs associated with conventional non-biologic therapy will be half (£35.08) of those incurred by patients who are receiving conventional therapy only.

- **ERG's comments on the drug acquisition costs assumed in the company's model**[15]

The calculated drug acquisition costs are conditional on the treatment regimen assumed within the company's model.[15] As indicated in Section 5.2.4, the ERGs has some concerns with the treatment regimen assumed, notably for vedolizumab and adalimumab for the induction phase. Table 53 summarises the drug acquisition costs (induction phase) using the treatment regimens the ERG believes are correct. It should be noted that for vedolizumab, efficacy data would need to reflect the efficacy associated with 3 doses (this is not the case in the base case analysis).

Table 53 Drug acquisition costs (induction phase) according to the ERG's corrected treatment regimens

Product	Unit cost	Units per induction cycle	Induction phase duration	Cost per induction cycle	Adjusted cost (14 week)[e]	Cost per 8 weeks maintenance cycle
Vedolizumab (300mg vial)	[redacted]	3[a]	10/14 weeks[d]	[redacted]	[redacted]	[redacted]
Infliximab (100mg vial)	£419.62	8[b]	6 weeks[d]	£3,356.96[b]	£4,422.79[f]	£1,678.48
Adalimumab (40mg)	£352.14	3[c]	4 weeks[d]	£1,056.42[c]	2,120.49[g]	£1,408.56

[a] Dose at week 0, 2 and 6 with assessment at week 10; [b] Dose at week 0 and 2 with assessment at week 6; [c] Dose at week 0 and 2 with assessment at week 4; [d] licensing; [e] Estimated costs at 14 weeks, accounting for the proportion of responder to the induction phase; [f]assumed 63.50% receive a dose at week 6 based on response at week 6 for infliximab used within the company's model;[15] [g]assumed 60.43% receive a dose at week 4, 6, 8, 10 and 12 based on response at week 4 for adalimumab used within the company's model[15]

It should be noted that the induction phase duration is different for each biologics. To allow for a fair comparison of drug acquisition costs, we also present the estimated drug acquisition costs calculated over the same period (14 weeks) based on the costs for the induction phase, the proportion of responders to the induction phase and the costs for the maintenance phase for responders only. Vedolizumab appear to be more costly over this period compared with infliximab and adalimumab.

It should also be noted that the drug acquisition cost for infliximab is conditional on the patient weight. Table 54 summarises the number of vials needed per infusion according to the weight of patients. The ERG believes that using the mean weight is not appropriate and that the distribution of patients within weight band should be used instead; it is unclear whether the drug acquisition for infliximab would be affected.

Table 54 Number of vials needed according to patient's body weight for patients treated with infliximab

No. vials	Weight (max weight)
7	120 > weight ≤ 140 kg
6	100 > weight ≤ 120 kg
5	80 > weight ≤ 100 kg
4	60 > weight ≤ 80 kg
3	40 > weight ≤ 60 kg
2	20 > weight ≤ 40 kg
1	≤ 20 kg

The company arbitrarily assumed that whilst patients are receiving biologic therapy, the costs associated with conventional non-biologic therapy will be halved. This is not justified in the CS[1]. In response to a request for clarification (see clarification response[2] question B30), the company states that: *"In a scenario analysis (not in the submission but conducted for this clarification), an extreme value of 100% was used. In other words, it was assumed that patients receiving vedolizumab have the same costs of conventional therapy as patients receiving conventional therapy alone (i.e. £70.16 per cycle in the updated model)."* The company reported little changes to the ICER.

The ERG was also unclear why the cost for conventional therapy was derived from a UK audit rather than from the number and type of therapy used in the trial directly. In response to a request for clarification (see clarification response[2] question B29) the company states that: *"a detailed assessment of the use of conventional therapy alongside vedolizumab would be complex. The use of conventional therapy within the GEMINI II and GEMINI III trials was protocol driven and the trial was international and may not represent treatment patterns in England and Wales. A full analysis of the use of conventional therapy within the trial would involve assessment of frequency, dosing and duration and still would not replicate NHS treatment patterns. The model, as submitted, was intended to provide a reasonable assumption of the use of conventional therapy in real-world, NHS use"*.

Finally, the ERG is unclear how robust is the approach used by the company to estimate the drug acquisition costs for patients receiving conventional non-biologic therapy. The company used the mix of treatments reported in the IBD audit and assumed that treatment are taken daily indefinitely. This is likely to overestimate the cost for conventional non-biologic therapy. To a lesser extent, the specific products assumed are not specified by the company in either their model[15] or submission. However, the impact on the ICER is likely to be minimal.

Administration costs

- **Assumptions on administration used in the company's model**[15]

The costs associated with the administration of infusional biologics (infliximab and vedolizumab) were taken from the PbR tariff 2012/13[79] and were assumed to be £308 per administration visit. No administration costs were assumed for adalimumab or conventional non-biologic therapy.

- **ERG's comments on assumptions from the company on administration costs**

No details are provided by the company on administration costs in the CS[1] in the economic section (see Section 7 in the CS[1]). However, the ERG is satisfied with the administration cost estimate assumed by the company.

Assessment cost

The cost associated with assessment of response has not been explicitly included in the company's model.[15] Discussion with clinical experts indicated that in practice, assessment is likely to happen during monitoring visits to the gastroenterologist. The omission of the assessment cost is unlikely to impact results.

CD health state resource costs

- **Health states costs assumed in the company's model**[15]

Management costs for the different health states are taken from Bodger et al.[60] inflated to 2012 using the Pay and Price Index.[80]

Table 55 Per-cycle cost, by health state

Health states	Cost inflated from Bodger et al.[60]
Remission	£109.80
Mild	£313.38
Moderate-severe	£489.51
Surgery	£10,580.51

In addition, for patients in the surgery health state, the company included the costs of treating surgical complications. Complications were included based on expert opinion. The probabilities of surgery related complication were taken from a pooled estimated of the systematic review of the published literature[1] and are presented in Table 56. Costs are estimated from the NHS Reference Costs[81] and expert opinion.

Table 56 Probabilities and costs of surgery-related complications (Table 7.3.1.8 in CS[1])

	Proportion	Cost	Source
Wound infection	8.13%	£1,724.87	NHS Reference Costs 2011/12.[81] Assumed 4 additional hospital days and 1 outpatient visit according to expert clinical opinion
Prolonged ileus/bowel obstruction	4.52%	£1,609.39	NHS Reference Costs 2011/12.[81] Assumed 4 additional hospital days according to expert clinical opinion
Intra-abdominal abscess	1.61%	£2,011.73	NHS Reference Costs 2011/12.[81] Assumed 5 additional hospital days according to expert clinical opinion
Anastomotic leak	4.00%[a]	£2,816.43	NHS Reference Costs 2011/12.[81] Assumed 7 additional hospital days according to expert clinical opinion
[a] Taken from the company's model			

- **ERG's comments on health states costs assumed in the company's model**[15]

No details are provided by the company[1] on how the costs were estimated or which resources were included in the Bodger et al[60] analysis. Notably, the company included an additional cost for complications due to surgery. It is unclear from the Bodger study[60] whether the costs associated with complications due to surgery are already included.

Costs of managing adverse events

- **Cost of managing adverse events used in the company's model**[15]

Unit costs associated with the management of AEs associated with biologic and non-biologic treatment are taken from the NHS Reference Costs 2011-2012[81] and three previous NICE Technology Appraisals (see Table 57).[82-84]

Table 57 Unit costs associated with managing adverse events

Adverse Event	Total Cost	Source
Serious infection	£1,470	NHS Reference Costs 2011/12.[81] Average of 5 different types of serious infections: sepsis, pneumonia, urinary tract infection, respiratory infection, and bronchitis
Tuberculosis	£2,272	NHS Reference Costs 2011/12.[81] Average of non-elective short-stay and long-stay tuberculosis
Lymphoma	£14,975	NICE (2003), NICE (2012), and NICE (2011). Average of lymphoma costs from three technological appraisals for rituximab (TA65, TA243, and TA226) [82-84]
Hypersensitivity	£3,188	NHS Reference Costs 2011/12.[81] Average of non-elective short-stay and long-stay pyrexia
Injection site reactions	£1,363	NHS Reference Costs 2011/12.[81] Average of procedures associated with skin disorders

- **ERG's comments on the cost of managing adverse events used in the company's model**[15]

In response to a request for clarification (see clarification response[2] question B39), the company states that "*only serious adverse events were included in the model. By definition, these adverse events required hospitalisations".* The ERG is satisfied with the justification provided by the company and notes that the impact of AEs on the ICER is minimal. It should be noted that the company's model does not use the latest version of the NHS reference cost. In response to a request for clarification (see clarification response[2] question B33), the company states that "*2012 / 13 NHS Reference costs have been included in an update to the model*." After assessment of the model, the ERG notes that costs have not been updated for adverse events or surgery complications.

5.2.9 Moderate and severe subgroup analysis

In the CS[1] the company reports results for patients with moderate and severe disease at baseline separately. It was unclear how these analyses were conducted because the company's model[15] submitted as part of the original submission[1] did not appear to include the option to conduct an analysis for these subgroups of patients. In response to clarification (response to clarification[2] question B2) the company states that:

"The submitted results of the analysis were generated with a variation of the submitted model that included the ability to choose among baseline disease severity and experience with biologics. This version of the model was not provided with the submission in error.

The efficacy data used to populate this model were based on response and remission rates from subgroup analysis of pooled trial results from the VDZ-CT head-to-head clinical trials. Similar calibration procedures were used to define transition matrices between health states. In cycle 1, patients enter the Mild and Moderate-Severe states based on the observed progression of the moderate or severe subgroups, as seen in the analysis of these subgroups within trial data.

As the subgroups are only specified at baseline, utilities and costs are still defined on the basis of the defined health states.

The updated model includes these data points in the 'Data Store' and 'Calibration' worksheets."

- **ERG's comments**

It appears from the company's model[15] that the following inputs are changed when analysing the moderate population or the severe population:

- Probability of response and remission
- Proportion of responders with moderate to severe disease
- Transition matrices at maintenance phase
- Discontinuation due to AEs on vedolizumab

Table 58 and Table 59 summarises the number of patients in remission (CDAI ≤150), response (defined as a drop of 70 points or more in CDAI score) and total number of patients used to calculate the probabilities of response and remission for each population group in the mixed ITT and anti-TNF-α failure subgroups in the vedolizumab and placebo arm respectively. The ERG is concerned that the number of patients with moderate to severe disease regularly does not equate to the number of patients with moderate disease plus the number of patients with severe disease. For instance, for the mixed-ITT population, for the induction phase, the company report 206 responders (defined as drop in CDAI ≥ 70 points) for the moderate and severe group combined; those with moderate disease (n=64) plus those with severe disease (n=42) does not equate to the reported number of responder in the combined group (n=206).

Table 58 Total number of patients, number of responders (drop in CDAI score of 70 points or more) and remission (CDAI≤150) used in the mixed ITT anti-TNF-α failure subgroup (defined as experienced in the company's model[15]) for the vedolizumab arm

	No of responders (drop in CDAI of 70 points or more)	No of patients in remission	Total number of patients
Mixed ITT population – induction			
Moderate to severe CD	206	72	429
Moderate CD	64	27	108
Severe CD	42	11	108
Anti-TNF-α failure - induction			
Moderate to severe CD	116	34	260
Moderate CD	18	5	39
Severe CD	14	3	40
Mixed ITT population – maintenance			
Moderate to severe CD	73	60	154
Moderate CD	36	36	78
Severe CD	31	23	75
Anti-TNF-α failure - maintenance			
Moderate to severe CD	24	23	82
Moderate CD	14	14	39
Severe CD	10	9	43

Table 59 Total number of patients, number of responders (drop in CDAI score of 70 points or more) and remission (CDAI≤150) used in the mixed ITT anti-TNF-α failure subgroup (defined as experienced in the company's model[15]) for the placebo arm

	No of responders (drop in CDAI of 70 points or more)	No of patients in remission	Total number of patients
Mixed ITT population – induction			
Moderate to severe CD	120	35	355
Moderate CD	22	5	69
Severe CD	16	3	68
Anti-TNF-α failure - induction			
Moderate to severe CD	70	23	226
Moderate CD	7	2	29
Severe CD	6	0	29
Mixed ITT population – maintenance			
Moderate to severe CD	86	64	345/410[a]
Moderate CD	29	24	86
Severe CD	17	9	67
Anti-TNF-α failure - maintenance			
Moderate to severe CD	21	10	78
Moderate CD	8	7	35
Severe CD	8	3	43

[a] discrepancy in the data for the denominator – the model uses n=345 when calculating the response rate but uses n=410 when calculating the remission rate

Furthermore, the ERG has concerns regarding the validity of the calibrated transition probabilities as these appear to be pasted into the model as values and it is not possible to know without refitting them whether the best solution was found. Notably, it appears that the probability of transition from the mild health state to mild or moderate to severe health state for the mixed population with moderate disease at baseline is set to be the same.

It should also be noted that in the CS[1] analyses for adalimumab and infliximab were presented for the subgroup of patients with moderate or severe CD at baseline for the anti-TNF-α naive population. In the updated company's model submitted in response to clarification, no data appear to be available for patients treated with infliximab and adalimumab.

Due to the above reasons, the ERG is unable to confirm results from these analyses.

5.2.10 Cost effectiveness results

All analyses include price reductions to reflect the proposed PAS for vedolizumab.

It should be noted that within the CS;[1]

- for the anti-TNF-α naïve subgroup, the company arbitrarily reported outcomes obtained using data from the head to head trial for vedolizumab and conventional non-biologic therapy but outcomes from the NMA for adalimumab and infliximab; results from these analyses are not directly comparable. According to the original company submission,[1] infliximab was dominated by vedolizumab (see Table 7.7.6.1 in CS[1] pg. 350). In response to clarification (see clarification response[2], question B1), the company acknowledged that the results based upon the NMA for all therapies should be presented to allow a fair comparison with infliximab and adalimumab. Vedolizumab is now dominated by infliximab (see Table 61).
- results are presented for moderate and severe patients at baseline separately; no details on these analyses was included within the CS[1] and the model submitted alongside the submission did not allow assessment of these subgroups. In response to clarification (see clarification response[2], question B2), the company provided an updated model with the functionality to assess these subgroups; little detail is provided by the company on how these analyses were conducted.
- results for the anti-TNF-α failure subgroup for the combined moderate to severe group are not reported in the CS[1] but were included in response to clarification (see clarification response[2] question B1) .
- the headline cost-effectiveness results presented by the company are based on the deterministic version of the model (using point estimates of parameters) rather than the expectation of the mean. Whilst PSA was undertaken by the company, probabilistic ICERs were not presented within the CS[1].

Furthermore, updated results are presented in response to clarification,[2] but these are incomplete.

Consequently, health gains and costs presented are taken directly from the updated company's model.[15] To be consistent with the company's base case assumption, results are presented at 10 years.

It should be noted that the company's model (as it stands) calculates ICERs for pairwise comparisons. In adherence with the NICE Reference Case,[64] results are presented by the ERG in a fully incrementally analysis based on the health gain and cost extracted from the company's model.

Mixed-ITT population

Table 60 summarises the estimated health gains and costs for each strategy for the mixed-ITT population for (a) the combined group of patients with moderate to severe disease at baseline, (b) patients with moderate CD only (CDAI 220-330) and (b) severe CD only (CDAI>330). Analyses are based on direct data from the GEMINI trials[11,12]

Table 60 Central estimates (based on point estimates of parameters) of cost-effectiveness for the mixed-ITT population (extracted from the company's model[15] – 10 year time horizon and price reduction to reflect the PAS)

	Costs	**QALYs**	**Inc Costs**	**Inc QALYs**	**ICER**
Moderate to severe disease at baseline[a]					
vedolizumab	£54,195	4.9802			
Conventional non-biologic therapy	£45,807	4.8469	£8,388	0.1334	**£62,903**
Subgroup: Moderate disease at baseline[b]					
vedolizumab	£50,141	5.2536			
Conventional non-biologic therapy	£43,693	4.9475	£6,447	0.3061	£21,064
Subgroup: Severe at baseline*					
vedolizumab	£53,652	4.9148			
Conventional non-biologic therapy	£45,813	4.8134	£7,840	0.1013	£77,382

[a] Presented by the company in the response to clarification (see clarification response,[2] question B1)

[b] Taken from the updated company's model[15]

Assuming a 10-year time horizon, in patients with moderate to severe CD (CDAI>220) vedolizumab is estimated to provide a greater number of QALYs compared with conventional non-biologic therapy (incremental gain = 0.13 QALYs) but at a greater cost (incremental cost = £8,388) resulting in an ICER for vedolizumab against conventional non-biologic therapy of £62,903 per QALY gained. The ICERs for the moderate and severe subgroups are estimated to be £21,064 per QALY gained and £77,382 per QALY gained respectively.

Anti-TNF-α naïve subgroup

Table 61 summarises the estimated health gains and costs for each strategy for the anti-TNF-α naive population in patients with moderate to severe CD (CDAI > 220). It should be noted that in the CS[1] some analyses were presented for the subgroup of patients with moderate or severe CD at baseline for the anti-TNF-α naive population. In the updated company's model submitted in response to clarification, no data appear to be available for patients treated with infliximab and adalimumab. Given the lack of detail and the uncertainty about these analyses, only results in patients with moderate to severe CD (CDAI > 220) are presented. Analyses are primarily based on data from the NMA (except for infliximab where the company uses data from Rutgeerts et al.[66] for the induction phase as described in Section 5.2.6).

Table 61 Central estimates (based on point estimates of parameters) of cost-effectiveness for the anti-TNF-α naive population (extracted from the company's model[15] – 10 year time horizon and price reduction to reflect the PAS)

	Costs	QALYs	Inc Costs	Inc QALYs	ICER
moderate to severe at baseline					
Conventional non-biologic therapy	£44,347	4.9300			
adalimumab	£48,493	5.1404	£4,146	0.2104	£19,705
vedolizumab	£51,990	5.1450	Extendedly dominated		
infliximab	£52,907	5.1795	£4,414	0.0391	£112,882

Assuming a 10-year time horizon and under the company's base-case assumptions, the company's model predicts that patients with moderate to severe CD at baseline on conventional non-biologic therapy gain the fewest number of QALYs (4.93) followed by adalimumab (5.14), vedolizumab (5.1450) with infliximab providing the greatest number of QALYs (5.1795). Adalimumab provided 0.2104 additional QALYs when compared with conventional non-biologic therapy for an additional cost of £4,146, resulting in an ICER of £19,705 per QALY gained. Vedolizumab provided an additional 0.0046 QALYs when compared with adalimumab for an additional cost of £3,497, leading to an ICER of £758,344 for vedolizumab versus adalimumab. In contrast, infliximab provided 0.0345 additional QALYs when compared with vedolizumab for an additional cost of £917, leading to an ICER of £26,580 for infliximab versus vedolizumab. As the ICER for infliximab versus vedolizumab is smaller than the ICER for vedolizumab compared to adalimumab (£26,528 vs. £758,344); vedolizumab is extendedly dominated. Infliximab when compared to adalimumab provided 0.0391 additional QALYs for an additional cost of £4,414 resulting in an ICER of £112,882 per QALY gained.

Anti-TNF-α failure subgroup

Table 62 summarises the estimated health gains and costs for each strategy for the anti-TNF-α failure subgroup. Analyses are based on direct data from the GEMINI trials[11,12]

In patients with moderate to severe CD at baseline (CDAI>220), vedolizumab is estimated to provide greater QALYs compared with conventional non-biologic therapy (additional 0.09 QALYs) but at a greater cost (additional £8,615) resulting in an ICER for vedolizumab against conventional non-biologic therapy of £98,452 per QALY gained. The ICERs for the moderate and severe subgroups are estimated to be £55,201 and £134,330 per QALY gained respectively.

Table 62 Central estimates (based on point estimates of parameters) of cost-effectiveness for the anti-TNF-α failure subgroup (extracted from the company's model[15] – 10 year time horizon and price reduction to reflect the PAS)

	Costs	**QALYs**	**Inc Costs**	**Inc QALYs**	**ICER**
Anti-TNF-*α* failure – moderate to severe at baseline					
vedolizumab	£54,429	4.9232			
Conventional non-biologic therapy	£45,814	4.8357	£8,615	0.0875	£98,452
Anti-TNF-*α* failure – moderate at baseline					
vedolizumab	£53,388	4.9767			
Conventional non-biologic therapy	£45,480	4.8335	£7,909	0.1433	£55,201
Anti-TNF-*α* failure –severe at baseline					
vedolizumab	£54,030	4.8485			
Conventional non-biologic therapy	£46,104	4.7895	£7,926	0.0590	£134,330

5.2.11 Sensitivity analyses

The company conducted a range of uncertainty analyses including probabilistic sensitivity analysis (PSA), deterministic univariate sensitivity analyses (SA) and scenario analyses. It should be noted that results from these analyses are not presented by the company in the clarification letter[2] using the updated company's model.[15] Consequently, results presented hereafter are taken directly from the updated company's model when possible (analyses re-run by the ERG). For brevity, only results for the moderate to severe population are reported; results for patients with moderate or severe disease at baseline are not reported given concerns expressed by the ERG in Section 5.2.9. Table 63 summarises the ranges used for the SA and distribution used in the PSA.

Table 63 Range and distribution used in SA and PSA

	Range used in SA	**Distribution assumed in PSA**
Health state costs	Upper and lower range of the calculated 95% CI (based on assumed distribution)	Gamma– assuming a 20% variance
Non-governmental costs		
Age		
Weight		
RR excess mortality		
AE decrement in utility	Upper and lower range of the calculated 95% CI (based on assumed distribution)	Beta. N assumed to equal to 100
Health states utility scores		
Probabilities of response and remission to the induction phase	Upper and lower range of the calculated 95% CI (based on assumed distribution)	Beta
AE incidence		
Percentages of responder with moderate to severe disease		
Discontinuation due to AEs		
Percent Male		
Transition probabilities during the maintenance phase	Upper and lower rang of the calculated 95% CI (based on assumed distribution) + additional constraints	Dirichlet
Probabilities and response during the maintenance phase	Not varied – but used these values are not used directly in the model	
Rate surgery complication	Not varied	
Administration cost		
Mix conventional therapy		
Cost surgery complication		
General mortality		

Deterministic one-way sensitivity analysis

Ranges used for the deterministic SA are presented in Table 63. In the company's model, the 15 variables that had the greatest impact on the ICER for vedolizumab versus each comparator (pairwise comparison) are reported in the form of tornado diagrams (see CS[1] Figures 7.7.7.1 to 7.7.7.15 pg. 353 to 365).

Due to time constraints and to limit the number of analyses, the ERG does not report results using the updated company's model; it is believed that the parameters that had the largest impact on the ICER would not change between the two versions of the model submitted.

In summary, according to the tornado diagrams present in the CS[1] the parameters that had the largest impact on the ICER included;

- vedolizumab, conventional non-biologic therapy, infliximab transition probabilities (notably for the remission health state) and adalimumab (mild and remission health state),
- conventional non-biologic therapy AE incidence,
- vedolizumab, CT, infliximab response/remission to the induction phase,
- health state costs
- health state utility values,
- surgery transition probabilities

ERG comments

As shown in Table 63, the ranges used for the deterministic SA are somewhat arbitrary for most input parameters. It should be noted, that in theory, transition probabilities need to be recalibrated when the discontinuation rates or the probabilities of response/remission to the induction phase are changed; this has not however been done.

According to Figure 7.7.7.1 (in CS pg.353), Figure 7.7.7.2 (in CS pg. 354), Figure 7.7.7.14 (in CS pg. 364) and Figure 7.7.7.15 (in CS pg. 364) the incidence of AEs in patients with conventional non-biologic therapy is a key driver of cost-effectiveness.[1] The ERG attempted to replicate this SA (varying the incidence of AEs for conventional non-biologic therapy); but contrary to the SA presented in the CS the results did not appear to be sensitive to the incidence of AEs.

Probabilistic sensitivity analysis

PSA was conducted in all three populations (mixed-ITT, anti-TNF-α naïve and anti-TNF-α failure). The distributions used in the PSA are summarised in Table 63. The CS[1] presents the results of the PSA as pairwise cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs, see 7.7.8.1 in CS[1] pg. 367-381) only.

For transparency, the ERG reports the probabilistic ICERs (Table 64) using the updated version of the company's model (analyses run by the ERG before amendment to model input parameters). It should be noted that for the anti-TNF-α naïve subgroup, an amendment to the model was necessary in order to report the mean costs and QALYs for all comparators as the submitted model only reports outcomes for pairwise comparison. In adherence with the NICE Reference Case,[64] results are presented as fully incremental comparisons.

Table 64 PSA results (moderate to severe at baseline) - 10 year time horizon and price reduction to reflect the PAS

						Probability most cost-effective	
	Costs	**QALYs**	**Inc Costs**	**Inc QALYs**	**ICER**	**at 20K**	**at 30K**
Mixed-ITT – moderate to severe at baseline							
CT	£45,707	4.8432					
vedolizumab	54,002	4.9774	£8,295	0.13	£61,825	0.13%	1.43%
Anti-TNF-α naïve – moderate to severe at baseline							
CT	£44,221	4.9247				47.20%	17.27%
adalimumab	£48,221	5.1390	£4,000	0.2143	£18,665	51.93%	78.47%
vedolizumab	£51,749	5.1431	Extendedly dominated			0.30%	1.73%
infliximab	£52,641	5.1772	£4,420	0.0383	£115,527	0.57%	2.53%
Anti-TNF-α failure – moderate to severe at baseline							
CT	£45,814	4.8402					
vedolizumab	£54,311	4.9289	£8,497	0.09	£95,852	0.13%	1.43%
CT = conventional non-biologic therapy							

It should be noted that the probabilistic ICERs are similar to the deterministic ICERs. Within the mixed-ITT population, assuming a cost-effectiveness threshold of £20,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is approximately 0.13%. Assuming a threshold of £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is approximately 1.43%.

Within the anti-TNF-α naïve subgroup, assuming a cost-effectiveness threshold of £20,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment, infliximab and adalimumab is 0.30%. Assuming a cost-effectiveness threshold of £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment, infliximab and adalimumab is 1.73%.

Within the anti-TNF-α failure subgroup, assuming a cost-effectiveness threshold of £20,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is 0.10%. Assuming a cost-effectiveness threshold of £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is 0.43%.

ERG comments

The ERG has concerns with the PSA conducted by the company. Notably, the majority of distributions appear to be arbitrary as shown in Table 63. In particular, as the calibration is conditional on the model structure, in theory, the model needs to be calibrated for each run of the PSA; this is not the case.

For the anti-TNF-α naïve subgroup, the probabilities of response and remission to the induction phase are sampled from beta distributions rather than the CI from the NMA. This is not correct. Furthermore, remission and response are sampled independently, but these two outcomes are correlated.

Health state costs are varied from a gamma distribution assuming an arbitrary variance. Clarification was sought from the company (see clarification response,[2] question B31) and states that "*The original decision was based on having a standard deviation reported from Bodger et al., 2009, without sample size. Having re-reviewed the paper by Bodger et al. we acknowledge that both descriptive statistics are available which would allow for using the published estimates to inform the probabilistic sensitivity analysis. We have tested the distributional assumptions using the standard errors (calculated from the standard deviation and sample size reported) from Bodger et al. 2009. A constant coefficient of variation was assumed from the original model to generate distributional parameters for 2013 costs. The variability of the probabilistic cost distribution was determined to be*

similar, independent of method used. Using this information from the model by Bodger et al., we would estimate the true variability of the ICER to be very similar to the initial variability." The ERG found the justification provided by the company to be confusing and believes that using the appropriate distribution is straightforward.

Similarly, the cost of surgery was derived from NHS Reference Costs and was sampled from a gamma distribution assuming an arbitrary variance. Clarification was sought on why the uncertainty was not captured using the range reported in the NHS Reference Costs (see clarification response,[2] question B32). The company states that "*This was an over-sight and the range was not considered for use in the sensitivity analyses. It is anticipated that use of the range of reference costs, rather than the current assumption would not greatly alter the CEAC."*

Utility values are sampled from a beta distribution, assuming N=100. Clarification was sought from the company on why the confidence intervals were not used in the SA and PSA (see clarification response,[2] question B42). In response, the company stated that "*Confidence intervals were not calculated for the utility values. In the absence of the values, a sample size of 100 was assumed*." The ERG believes that the justification provided by the company is unsatisfactory. The company has access to the patient-level data and therefore could calculate the CI. Correlation between health state utility values is also not included. Consequently, utility values from mild disease may be in some occasion better than the utility values for patients in remission.

Transition probabilities are varied using a Dirichlet distribution based on the predicted probability and the number of patients entering maintenance. This is arbitrary. An alternative option would have been to sample the probabilities of response and remission and calibrate the model for each sample.

It is also unclear why the confidence interval for the excess risk of mortality from Lichtenstein et al was not used; the company assumed a gamma distribution with an arbitrary variance.

Scenario analysis

The company reports cost-effectiveness results across five groups of scenarios (see CS[1] Section 7.6.9); these involved altering the model time horizon (1-year and lifetime), using an alternative source of utility values (Buxton et al 2007[72]), assuming vedolizumab assessment at week 10 and 14, using a different definition of response (CDAI drop of 100 point or more) and extending the maximum duration of biologic treatment from 1 year to 3 years.

We report in Table 65 updated results from the scenario analyses presented in the original CS[1] using the updated company's model[15] (analyses are conducted by the ERG given that results are not available in the CS[1] or clarification letter[2]). It can be seen that results are sensitive to all the scenarios considered, notably the time horizon and health state utility values.

Additional scenario analyses were conducted by the company following the clarification process. This included, assuming the same cost for conventional non-biologic therapy whilst on biologic (see clarification response,[2] question B30), assuming the same discontinuation rate per year for all biologics (see clarification response,[2] question B38) and assuming a utility value of 0.728 for surgery (see clarification response,[2] question B43). The impact on the ICER for these analyses was reported to be minimal by the company.

Table 65 Summary results of company's scenario analyses (10 year time horizon and price reduction to reflect the PAS)

Scenario	Conventional therapy	adalimumab	vedolizumab	infliximab
Mixed-ITT				
Base case	-	not evaluated	**£62,903**	not evaluated
1-year time horizon			**£192,787**	
Lifetime horizon			**£37,611**	
Utilities from Buxton et al (2007)[72]			**£39,039**	
10-week vedolizumab response assessment			**£69,204**	
14-week vedolizumab response assessment			**£77,471**	
Response (drop of 100 points or more			**£79,412**	
Maximum time on treatment =3 years			**£57,116**	
Anti-TNF-α naïve*				
Base case	-	**£19,705**	**Extendedly dominated**	**£112,882**
1-year time horizon		**£103,751**		**£249,332**
Lifetime horizon		**£9,823**		**£39,961**
Utilities from Buxton et al[72]		**£12,254**		**£67,339**
10-week vedolizumab response assessment		**£19,735**	**Dominated by ADA**	**£67,879**
14-week vedolizumab response assessment	Efficacy data not available for all comparators			
Response (drop of 100 points or more				
Maximum time on treatment =3 years	-	**£22,849**	**Dominated by ADA**	**Dominated by ADA**
Anti-TNF-α failure				
Base case	-	Not evaluated	**£98,452**	Not evaluated
1-year time horizon			**£295,901**	
Lifetime horizon			**£57,360**	
Utilities from Buxton et al[72]			**£60,961**	
10-week vedolizumab response assessment			**£98,889**	
14-week vedolizumab response assessment			**£122,700**	
Response (drop of 100 points or more			**£114,460**	
Maximum time on treatment =3 years			**£83,225**	
* incremental analysis calculated by the ERG				

5.2.12 Model validation

The company (in CS[1] pg. 392) states that the following measures were taken to validate the model structure and verify the calculations within the economic model:

- the model structure and key structural assumptions were validated by 2 clinical experts,
- the model was reviewed by two, independent health economics to ensure face validity,
- an independent modeller (not involved in the project) performed a quality assurance of the model (internal validity) which involved a detailed review of inputs and calculations
- predictions at one year were compared with trial data. Results are also compared with previous economic analyses as part of the external validity.

ERG comments

The ERG considers that the quality of the model submitted by the company to be generally poor. The implementation of the model is unnecessarily complex for a Markov model. Tracing cells to their original hardcoded source within the model is burdensome. This is made more complicated as little to no details are included within the CS[1] on the source of inputs or how the transition matrices can be derived in the economic model.[1]

The ERG did not identify any major programming errors in the company's model

A number of minor errors and inconsistencies in reporting between the CS[1]and the economic model[15] were identified;

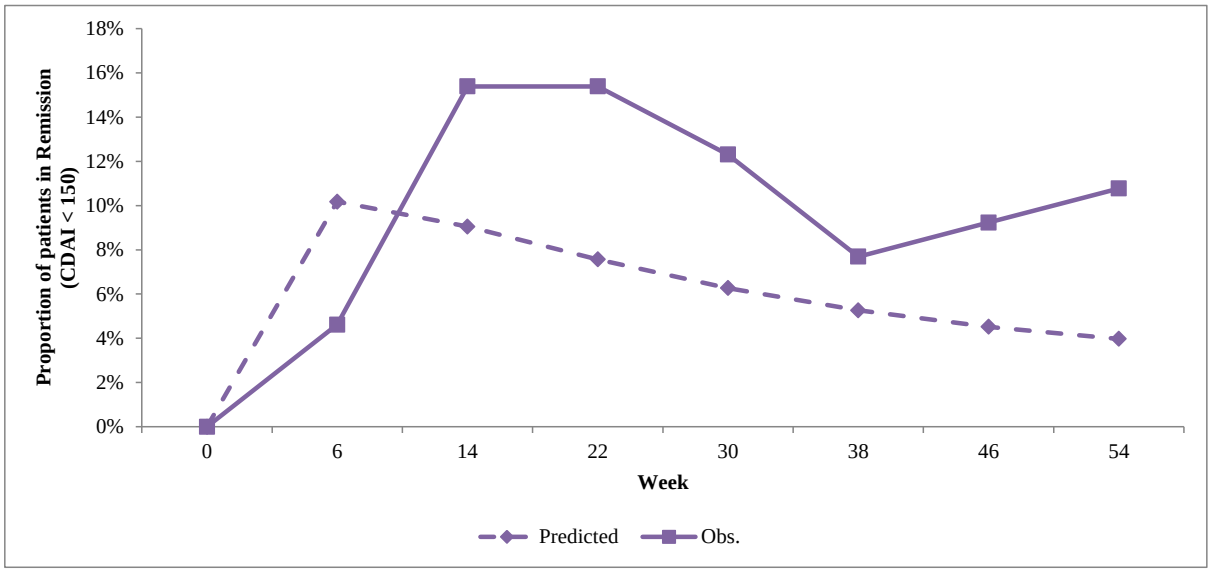
- no change in the cycle length for the decision tree for the induction tree when assessing response at 10 and 14 weeks
- use of initial induction vector on CT (using 6 week probabilities) after failure of biologic (which uses a 8 week cycle length)
- for the scenario analysis using assessment at week 10 or 14, the company assumes 3 doses of vedolizumab but only 2 administrations
- for the scenario analysis using assessment at week 10 or 14, there is no adjustment in costs for adalimumab and infliximab.

In the CS[1] (see CS[1] Table 7.7.1.1 pg.309), the company presents a comparison of the model prediction with the proportion of patients with response and remission from the GEMINI trials.[11,12] The ERG believes this comparison to be of limited value.

It order to validate the model predictions, the ERG requested (see additional clarification response[2]) data on the proportion of patients treated with placebo in remission (and other health states) at different time points (weeks 0, 6, 14, 22, 30, 38, 46 and 54) from the GEMINI II trial[11] for the anti-TNF-α failure subgroup. In the GEMINI II trial[11] patients on placebo continued to receive placebo for the full duration of the trial, irrespective of response to the induction phase. Consequently, it is possible to compare the proportion of patients in remission predicted by the model and the proportion of patients in remission from the placebo arm of the trial.

It should be noted that in the economic model, the company uses data from both GEMINI II[11] and III[12] at week 6, and that the proportion of patients in remission was higher in GEMINI III compared with GEMINI II (13% vs. 4%). Therefore, one would expect the model to predict a higher proportion of patients in remission compared with data from the GEMINI II trial[11] from week 6 and onward. As shown in Figure 15, this is not the case; the model under-predicts the proportion of patients in remission in the placebo arm, despite using pooled data from the GEMINI trials at week 6.

Figure 15 Comparison of the proportion of patients in remission predicted by the model and observed in GEMINI II in patients treated with placebo for the anti-TNF-α failure subgroup



Similarly, the ERG requested (see additional clarification response[2]) data on the proportion of patients treated with vedolizumab Q8W in remission, mild and moderate to severe CD at different time points (weeks 0, 6, 14, 22, 30, 38, 46 and 54) from the GEMINI II trial for the anti-TNF-α failure subgroup. It should be noted that the interpretation of the data is challenging due to discrepancies (possibly due to missing data at some time points), consequently Figure 16 to 19 present a crude comparison of the proportion of responders with vedolizmab in the different health state predicted by the model and in the GEMINI II trial.[11] Whilst prediction for the remission and mild CD appear reasonable (Figure 16 and 17), it can be seen that the model over-predicts by a large amount the proportion of responders to the induction phase remaining on treatment with moderate to severe CD (Figure 18). On the other side, the model under-predict by a large amount the proportion of responders to the induction phase discontinuing therapy (Figure 19).

Figure 16 Comparison of the proportion of responders to the induction phase remaining on treatment in remission predicted by the model and observed in GEMINI II in patients treated with vedolizumab (responders at week 6) for the anti-TNF-α failure subgroup

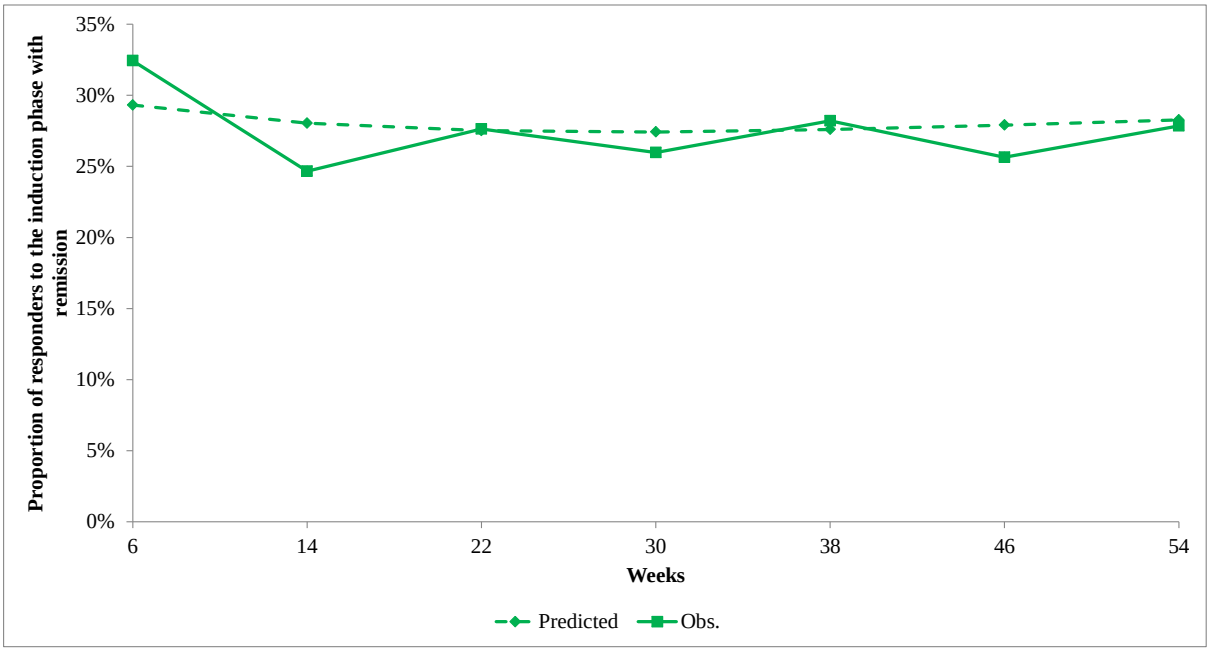


Figure 17 Comparison of the proportion of responders to the induction phase remaining on treatment with mild CD predicted by the model and observed in GEMINI II in patients treated with vedolizumab (responders at week 6) for the anti-TNF-α failure subgroup

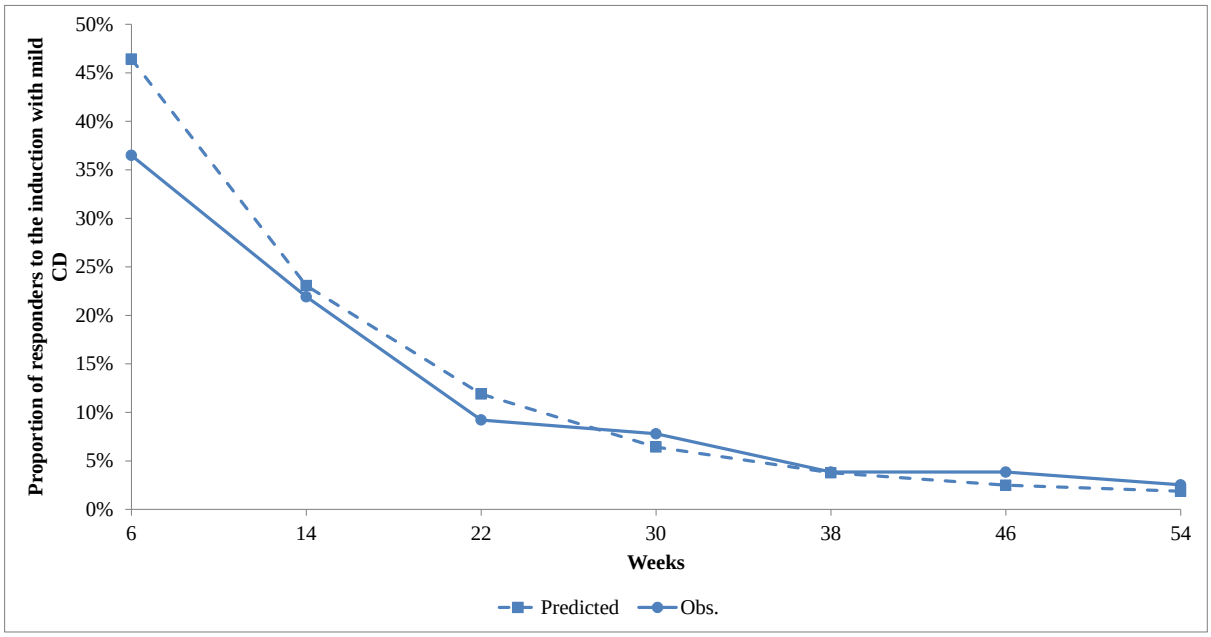


Figure 18 Comparison of the proportion of responders to the induction phase remaining on treatment with moderate to severe CD predicted by the model and observed in GEMINI II in patients treated with vedolizumab (responders at week 6) for the anti-TNF-α failure subgroup

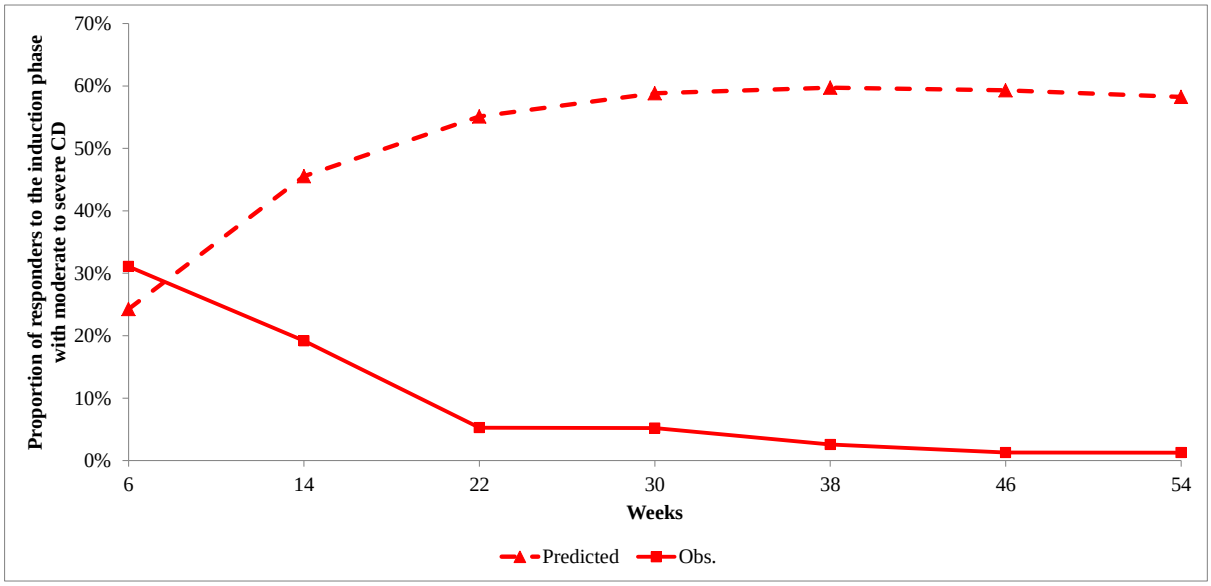
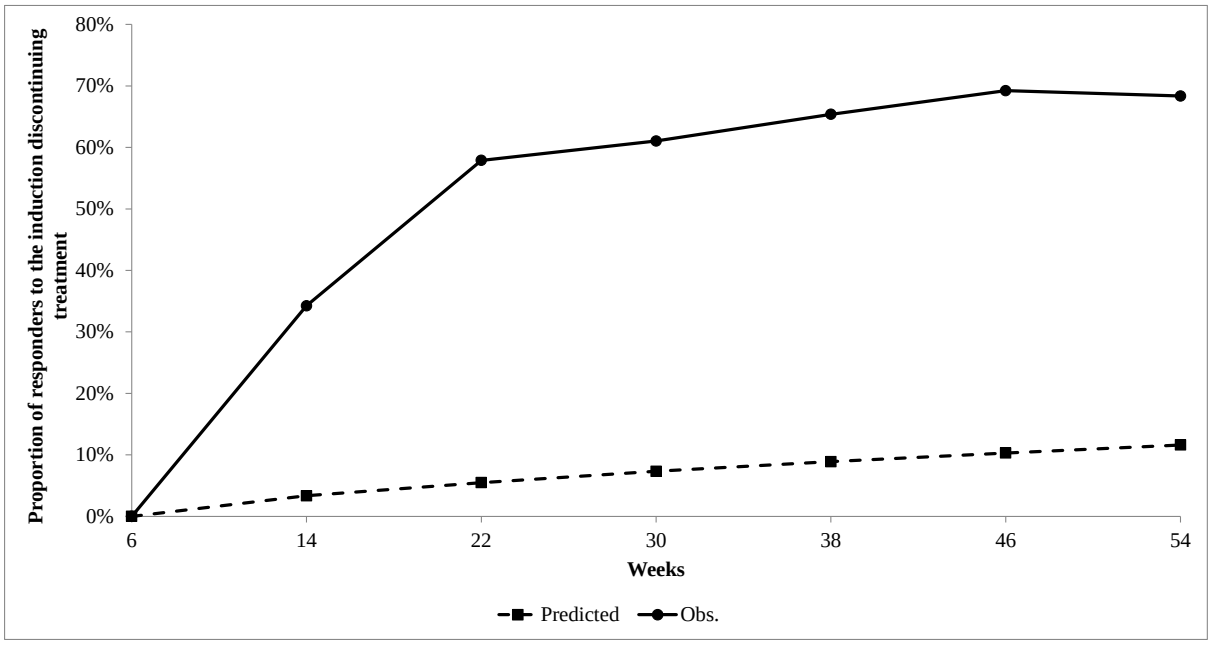


Figure 19 Comparison of the proportion of responders to the induction phase discontinuing treatment (any reason) predicted by the model and observed in GEMINI II in patients treated with vedolizumab (responders at week 6) for the anti-TNF-α failure subgroup



Furthermore, the CS,[1] responses to clarification,[2] and company's model[15] includes data from several analyses of the GEMINI trials, with no details. Therefore, it was not possible for the ERG to check whether all the values presented by the company are accurate. However, the ERG is concerned as discrepancies were found for the following inputs:

- Percentages of responders with moderate to severe disease. The company states (see clarification response[2] question B14) that the values are taken from a pooled analysis of the GEMINI trials (all arms). However, for the anti-TNF-α failure subgroup, the company reports in the economic model that amongst 136 responders, 36 patients had moderate to severe disease. This is inconsistent with the number of responders used to calculate the probability of response (116 for vedolizumab and 70 for placebo). The same issue applies to all populations evaluated.
- There are also inconsistencies between the number of patients in remission (CDAI ≤ 150), response (defined as a drop of 70 points or more in CDAI) and total number of patients (denominator) used to calculate the probabilities of response and remission for patients with moderate and severe CD at baseline for the anti-TNF-α failure and mixed ITT subgroups (Table 58 and 59). The ERG is concerned that the number of patients with moderate to severe disease does not equate to the number of patients with moderate disease plus the number of patients with severe disease.

- There also appear to be some discrepancies in the denominator for conventional non-biologic therapy for the mixed ITT population used to calculate the probability of response and remission (345 vs. 410 – Table 59) for the maintenance phase

The ERG also has concerns regarding the validity of the calibrated transition probabilities as these appear to be pasted into the model as values and it is not possible to know without refitting them whether the best solution was found.

Finally, the company uses results from the NMA (using a fixed-effect model) assuming assessment with vedolizumab to occur at week 6 or 10 (2 scenario analyses). The odd ratios are presented in Table 66. The odd ratios for adalimumab and infliximab differ within these 2 analyses. No explanation for the changes in odd ratios for adalimumab and infliximab were provided in the CS.[1]

Table 66 Odd ratios used in the economic model

	Response		**Remission**	
	Week 6	**Week 10**	**Week 6**	**Week 10**
Adalimumab 80 mg/40 mg	2.46	2.51	2.31	2.35
Infliximab 5 mg (1 dose)	25.38	25.38	25.71	24.89
Vedolizumab	1.82	1.93	2.89	2.70

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Table 67 summarises the main concerns expressed by the ERG following the critical appraisal of the CS[1] and submitted company's model.[15]

Table 67 Summary of key concerns identified by the ERG

	Section
Concerns relating to the model structure/key structural assumption	
• Potential omission of key aspects of the condition (relapsing remitting nature of CD, importance of maintaining CFR)	See Section 5.2.2
• Simplifying and debatable assumption regarding surgery	See Section 5.2.2
• Derivation of the initial induction vectors	See Section 5.2.2 and Section 5.2.6
• Derivation of the transition matrices (approach and assumptions)	See Section 5.2.2 and Section 5.2.6
• Lack of distinction between responders and non-responders with moderate to severe CD	See Section 5.2.2
• Assumption that non-responders have moderate to severe CD	See Section 5.2.2
• Same induction phase duration assumed for all therapies	See Section 5.2.2
• Relevance to clinical practice of drop of 70 or more in CDAI score to identify patients going onto receive maintenance treatment	See Section 5.2.2
• End of scheduled maintenance at approximately 1-year	See Section 5.2.2
• Potentially optimistic assumption following discontinuation whilst on biologics	See Section 5.2.2
• Omission of discontinuation due to lack of efficacy	See Section 5.2.2
Concerns relating to the population	
• Potential lack of representativeness to the UK population of patients recruited in the GEMINI trials (recruited from a large number of centres worldwide)	See Section 5.2.3
• Potential over-representation of patients with active inflammation in the GEMINI trials	See Section 5.2.3
• Difficulties in interpreting results from the mixed ITT population	See Section 5.2.3
Concerns relating to the comparators	
• Potential lack of representativeness of conventional non-biologic therapy to the UK population used in the GEMINI trials (recruited from a large number of centres worldwide)	See Section 5.2.4
• Exclusion of anti-TNF-α as a comparator for the anti-TNF-α failure	See Section 5.2.4

subgroup (although the ERG recognises the lack of data to inform such comparison)	
Concerns relating to the treatment regimen assumed / drug acquisition costs	
• The induction phase for patients on vedolizumab is assumed to be 6 week and patients are assumed to receive 2 doses before response assessment (instead of 3 doses as per SmPC)	See Section 5.2.4
• The induction phase for adalimumab is assumed to be 8 weeks and patients on adalimumab are assumed to receive a dose at week 0, 2, 4, and 6 (not in adherence with the licensing or efficacy data used).	See Section 5.2.4
Concerns relating to the time horizon	
• 10-year time horizon	See Section 5.2.5
Concerns relating to effectiveness data used to derive the initial induction vectors/transition matrices	
• Partial use of NMA for the induction – use of ACCENT-I for infliximab	See Section 5.2.6
• Lack of clarity on how the percentages of responders with moderate to severe disease were derived	See Section 5.2.6
• Potential lack of comparability between the trials included in the NMA for the maintenance phase	See Section 5.2.6
• Lack of clarity of the estimation of the discontinuation rate due to AEs	See Section 5.2.6
• Lack of clarity on the derivation of the transition probabilities from the surgery health state	See Section 5.2.6
• Debatable assumptions regarding mortality	See Section 5.2.6
• Inappropriate inclusion of adverse events (and impact on costs and HRQoL)	See Section 5.2.6 and Section 5.2.7
Concerns relating to HRQoL	
• Utility value for patients undergoing surgery	See Section 5.2.7
Concerns relating to model implementation/calculation/presentation of results	
• Pairwise comparison	See Section 5.2.10
• Arbitrary distributions used in the PSA and SA	See Section 5.2.11
• Discrepancies with data used for moderate and severe population and the subgroups of patients with moderate CD and severe CD	See Section 5.2.9

As shown in Table 67, the ERG expressed a number of concerns regarding the model structure and parameterisation of the company's model. Notably, a key concern is the derivation of the transition matrices following induction treatment. The ERG was unable to replicate the approach used by the company and therefore cannot amend the transition matrices. This is a concern as the transition matrices are a key input parameter and are conditional on the model structure and other input parameters.

The ERG also expresses concerns that non-responders at the induction phase on conventional non-biologic treatment are assumed to remain with moderate to severe CD and only discontinuation due to AEs is included for biologics. Discussion with clinical experts indicated that CD is relapsing-remitting. Some patients may improve spontaneously. Furthermore, relapse following biologics is a common and recognised effect.

Similarly, the ERG expressed some concerns with efficacy data that are used, notably the comparability of data for the different biologics at the maintenance phase, and efficacy data used for conventional non-biologic treatment.

The combination of all these issues lead to some discrepancies between the model prediction and observed data from the GEMINI II trial[11] as shown in Section 5.2.12 (figure 15 to 19).

Unfortunately, these issues cannot be addressed by the ERG without major restructuring of the economic model. It should be noted that changes to the model are challenging given the structure of the model and lack of transparency. A further concern expressed by the ERG was the assumption of the same induction duration for all biologics. Unfortunately, the ERG is not able to amend this easily within the company's existing model structure.

Consequently, results from the company's model need to be interpreted with caution. The ERG is unclear whether the ICER would improve or deteriorate following amendment of the identified structural issues.

For the sake of transparency and completeness, the ERG conducted additional scenarios analyses. The number of scenarios was limited given challenges arising from making changes to the model structure.

Additional analysis 1: Removal of AEs An analysis was undertaken whereby the impact of AEs on HRQoL and costs was removed. The ERG expressed concerns with the approach used by the company to include AE. For simplicity, the ERG removed the impact of AEs; this equate to assuming equivalent safety profile between treatments.

Additional analysis 2: Utility value for surgery An analysis was undertaken whereby the utility value for surgery is equal to the utility value for moderate to severe CD for 2 weeks and remission for the remaining 6 weeks.

Additional analysis 3: Cost for the induction phase for adalimumab. An analysis was undertaken whereby the cost for the induction phase for adalimumab is reduced to reflect the efficacy data used for the induction phase (i.e. dose of 80 mg at week 0 and 40 mg at week 2) and the additional dose before week 6 (additional 40 mg at week 4 for the proportion of patients who respond at week 4).

Additional analysis 4: Transition matrices for the different biologics. It is unclear from the available trial whether vedolizumab, infliximab and adalimumab have different efficacy in the maintenance phase. Consequently, an analysis was undertaken assuming the transition matrices for the maintenance phase for infliximab and adalimumab to be the same as the transition matrices for vedolizumab.

Additional analysis 5: Inclusion of lack of efficacy. In GEMINI II,[22] the probability of treatment failure (defined as disease worsening, need for rescue medications or surgical intervention for treatment of CD, or study drug-related AE leading to discontinuation from the study) at 1 year was 39% in the vedolizumab Q8W arm. An analysis was undertaken whereby the annual discontinuation rate was increased from 8.54% to 39% to include discontinuation due to lack of efficacy. In this analysis, the same discontinuation rate was assumed for all biologics. It should be noted that this analysis is subject to uncertainty as the transition matrices should be recalibrated but could not be done by the ERG.

Additional analysis 6: Same excess mortality rate for CD health state An analysis was undertaken assuming the same excess mortality risk rate (SMR of 1.7) for each CD health state based on Card et al.[85]

Table 68 Summary of exploratory analyses conducted by the ERG

	Mixed-ITT population	Anti-TNF-α naïve subgroup	Anti-TNF-α failure subgroup
Company's base case	**£62,903**	**Extendedly dominated**	**£98,452**
Additional analysis 1: Removal of AEs	£63,079	Extendedly dominated	£98,763
Additional analysis 2: Utility value for surgery	£63,255	Extendedly dominated	£98,798
Additional analysis 3: Cost for the induction phase for adalimumab	Not applicable	Extendedly dominated	Not applicable
Additional analysis 4: Transition matrices for the different biologics	Not applicable	Extendedly dominated	Not applicable
Additional analysis 5: Inclusion of lack of efficacy	£61,283	Extendedly dominated	£94,641
Additional analysis 6: Same excess mortality rate for CD health state	£63,765	Extendedly dominated	£99,880

Overall, the additional exploratory analyses conducted by the ERG had a limited impact on the ICER in isolation.

Unfortunately, it was not possible for the ERG to test explicitly the impact of using the Targan study[19] given concern with the derivation of the induction vector. However, using results from the Targan study[19] instead of ACCENT-1[66] would lead to an increase in the probabilities of remission and response in patients treated with infliximab at the induction phase. Vedolizumab is likely to remain extendedly dominated.

5.4 Conclusions of the cost effectiveness section

The company submitted a model-based health economic analysis as part of their submission to NICE. The analysis was undertaken from the perspective of the NHS over a 10-year time horizon. The company's analysis is presented for three populations: (1) the mixed-ITT population, which is comprised of patients who have previously received anti-TNF-α therapy and those who are anti-TNF-α naïve; (2) people who have not previously received an anti-TNF-α, and; (3) people for whom an anti-TNF-α has failed. Within all three analyses, comparators include conventional non-biologic therapies (a combination of 5-ASAs, immunomodulators and corticosteroids). Other anti-TNF-α agents (infliximab, adalimumab) are included only in the analysis of the anti-TNF-α naïve subgroup; these are excluded from the analyses of the mixed-ITT and anti-TNF-α failure subgroups.

Within the anti-TNF-α failure subgroup, the company's model[15] estimates the ICER for vedolizumab against conventional non-biologic therapy to be £62,903 per QALY gained within the mixed ITT population in patients with moderate to severe disease. The ICER for patients with moderate and severe CD at baseline were £21,064 and £77,382 per QALY gained respectively in the mixed ITT population.

Within the anti-TNF-α naïve subgroup, the CS[1] estimates that vedolizumab dominates infliximab and the ICER for vedolizumab against adalimumab is £2.602 per QALY gained. However, following a request for clarification, the company reports the ICER for vedolizumab versus adalimumab to be £758,344 and infliximab versus vedolizumab to be £26,580. Based on a fully incremental analysis (constructed by the ERG), vedolizumab is subject to extended dominance. No ICER is calculated in the model for the subgroup of patients with moderate and severe disease at baseline.

Within the anti-TNF-α failure population, the company's model[15] estimates that the ICER for vedolizumab against conventional non-biological therapy is £98,452 per QALY gained. The ICER for patients with moderate and severe CD at baseline were reported to be £55,201 and £134,330 per QALY gained respectively in this population.

The company presented a series of scenario analyses (see Section 5.2.11). Using a lifetime horizon lead to a more favourable ICER for the mixed-ITT (£37,611 per QALY gained) and anti-TNF-α failure subgroup (£57,360 per QALY gained). In contrast, assuming assessment to occur later than week 6 lead to a less favourable ICER for the mixed-ITT population (£69,204 and £77,471 per QALY gained assuming assessment at week 10 and 14 respectively) and anti-TNF-α failure subgroup (£98,889 and £122,700 per QALY gained assuming assessment at week 10 and 14 respectively).

The ERG critically appraised the company's health economic analysis and the model upon which this analysis is based. The ERG identified a number of concerns which are summarised in Table 67. Importantly, the ERG expressed a number of concerns regarding the model structure and parameterisation of the company's model. The health economic model submitted by the company is subject to a number of issues which limit the credibility of the company's results. These include (a) potential omission of key aspects of the condition such as the relapsing-remitting nature of CD, (b) simplifying and debatable assumptions regarding surgery, (c) the difficultly associated with parameterising the company's chosen structure notably the derivation of the transition matrices, and (d) debatable key structural assumptions such as assuming the same induction duration, end of scheduled maintenance at one year irrespective of achievement of remission, omission of discontinuation due to lack of efficacy and the assumptions that non-responders at the induction phase on conventional non-biologic treatment remain with moderate to severe CD (and are not able to improve). The combination of all these issues lead to some discrepancies between the model prediction and observed data from the GEMINI II trial[11] as shown in Section 5.2.12 (Figures 15 to 19).

The ERG is unclear whether the ICER would become more or less favourable following amendments of the identified issues. For the sake of transparency and completeness, the ERG conducted additional scenarios analyses. The number of scenarios was limited given challenges arising from making changes to the model structure: in isolation, these had little impact on the ICER.

Based on the company's model, vedolizumab does not appear to have an ICER below £30,000 per QALY gained in all analyses presented by the company, with the exception of patients with moderate disease at baseline for the mixed ITT population (£21,064 per QALY gained). However, the ERG is unable to confirm results from this analysis due to discrepancies in the data used and the lack of transparency regarding the derivation of model parameters. Furthermore, this analysis is compared with conventional therapy alone and no indication of the ICER for vedolizumab compared with adalimumab or with infliximab is reported.

6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

For the sake of transparency and completeness, the ERG conducted additional scenarios analyses. The number of scenarios was limited given challenges arising from making changes to the model structure: on isolation, these had little impact on the ICER.

The ERG is unclear whether the ICER would become more or less favourable following amendments of the identified structural issues.

7 END OF LIFE

NICE end of life supplementary advice should be applied in the following circumstances and when all the criteria referred to below are satisfied:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment, and;
- The treatment is licensed or otherwise indicated, for small patient populations.

The company makes no claim that vedolizumab should be appraised under the supplementary 'end of life' advice. The ERG agrees that the ends of life considerations are not applicable within this appraisal, as the first criterion is not met.

8 OVERALL CONCLUSIONS

8.1 Conclusions on the clinical effectiveness and cost-effectiveness of vedolizumab

Compared with placebo, the addition of vedolizumab to standard care in patients with moderately to severely active CD who had an inadequate response to, loss of response to, or intolerance of conventional therapy or anti-TNF-α was significantly more effective in terms of remission (defined as CDAI ≤150) at week 6 in the induction phase of GEMINI II.[11] There was no significant difference between the vedolizumab and placebo groups for the second primary outcome of enhanced clinical response (the number of patients achieving a reduction in the CDAI score of 100 or more) at week 6. In the maintenance phase of GEMINI II[11] patients treated with vedolizumab every 8 weeks (Q8W) and every 4 weeks (Q4W), had significantly higher rates of clinical remission at week 52 (defined as CDAI score of ≤ 150 points) compared with placebo. In GEMINI III[12] there was no statistically significant difference between vedolizumab and placebo in the primary endpoint of the proportion of patients achieving clinical remission at week 6 (CDAI score ≤150 points) in the anti-TNF-α failure population.

The ERG is satisfied that all relevant (published and unpublished) studies of vedolizumab were included in the CS.[1] However, there are a number of limitations and uncertainties in the evidence base which warrant caution in its interpretation. A key issue that may limit the robustness of the efficacy and safety data reported in the CS[1] relates to the high attrition rates in the maintenance phase of the GEMINI II[11] trial. As the GEMINI II study terminated at 52 weeks there are uncertainties in the evidence base regarding the efficacy and safety of the treatment for longer durations, the duration of optimal therapy, and how and when withdrawal should be introduced.
The primary endpoint was not achieved in GEMINI III; therefore, statistical evaluation of the secondary endpoints is acknowledged as exploratory by the company.

The ERG considered that the results of the NMA may underestimate the uncertainty in treatment effects since fixed effects models were used. There were also problems with the generalizability of findings to patients with strictures, patients with severe disease (CDAI >450) and to maintenance in patients who take longer to respond to induction therapy. Any generalisations to UK practice should be done with due consideration for the limitations of the evidence base.

Based on the company model, the ICER for vedolizumab against conventional non-biologic therapy is £62,903 per QALY gained within the mixed ITT population in patients with moderate to severe disease. The ICER for patients with moderate and severe CD at baseline were £21,064 and £77,382 per QALY gained respectively in the mixed ITT population.

Based on a fully incremental analysis (constructed by the ERG), within the anti-TNF-α naive subgroup, the company's model suggests that vedolizumab is extendedly dominated in the combined group of patients with moderate to severe disease (no analysis by moderate and severe possible in the company's model).Within the anti-TNF-α failure population, the company's model suggests that the ICER for vedolizumab against conventional non-biological therapy is £98,452 per QALY gained. The ICER for patients with moderate and severe CD at baseline were reported to be £55,201 and £134,330 per QALY gained respectively in this population.

Based on the company's model, vedolizumab does not appear to have an ICER below £30,000 per QALY gained in all analyses presented by the company, with the exception of patients with moderate disease at baseline for the mixed ITT population (£21,064 per QALY gained). However, the ERG is unable to confirm results from this analysis due to discrepancies in the data used and the lack of transparency regarding the derivation of model parameters. Furthermore, this analysis is compared with conventional therapy alone and no indication of the ICER for vedolizumab compared with adalimumab or with infliximab is reported.

8.2 Implications for research

- Long-term head-to-head RCTs comparing the efficacy and safety of vedolizumab with other biologics, namely infliximab and adalimumab and conventional non-biologic therapies in the treatment of patients with moderately to severely active CD.
- More evidence collected from a UK perspective.
- Further long term safety data to be collected.

9. REFERENCES

1. Takeda Pharmaceuticals. Vedolizumab for the for the treatment of adult patients with moderately to severely active Crohn's disease. Manufacturer's submission to the National Institute for Health and Care Excellence. 2014.

2. Takeda Pharmaceuticals. Response to ERG clarification questions. 2014.

3. Baumgart D.C., Sandborn W.J. Crohn's disease. Lancet 2012; 380(9853):1590-1605.

4. Mowat C., Cole A., Windsor A., Ahmad T., Arnott I., Driscoll R. et al. Guidelines for the management of inflammatory bowel disease in adults. Gut 2011; 60(5):571-607.

5. Yoshida E.M. The Crohn's Disease Activity Index, its derivatives and the Inflammatory Bowel Disease Questionnaire: a review of instruments to assess Crohn's disease. Can J Gastroenterol 1999; 13(1):65-73.

6. NICE. TA187: Infliximab (review) and Adalimumab for the treatment of Crohn's disease. 2010.

7. National Institute for Health and Care Excellence. Crohn's disease: Management in adults, children and young people (CG152). Https://Www Nice Org Uk/Guidance/Cg152 2012.

8. Vedolizumab for treating moderately to severely active Crohn's disease after prior therapy . Final scope. 2014.

9. European Medicines Agency (EMA). European Public Assessment Report - Entyvio . 2014.

10. Takeda Pharma A/S, 2014. Entyvio EMA Summary of Product Characteristics. 2014.

11. Sandborn W.J., Feagan B.G., Rutgeerts P., Hanauer S., Colombel J.F., Sands B.E. et al. Vedolizumab as induction and maintenance therapy for Crohn's disease. N Engl J Med 2013; 369(8):711-721.

12. Sands B.E., Feagan B.G., Rutgeerts P., Colombel J.F., Sandborn W.J., Sy R. et al. Effects of vedolizumab induction therapy for patients with Crohn's disease in whom tumor necrosis factor antagonist treatment failed. Gastroenterology 2014; 147(3):618-627.

13. Best W.R., Becktel J.M., Singleton J.W., Kern F., Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976; 70(3):439-444.

14. CEEU of Royal College of Physicians. UK Inflammatory Bowel Disease (IBD) audit (2013). 2013.

15. Takeda Pharmaceuticals. Cost-effectiveness for vedolizumab in crohn's disease. HE model submitted to the National Institute for Health and Care Excellence. 2014.

16. Takeda Pharmaceuticals. An Updated Systematic Literature Review and Network Meta-analysis in Crohn's Disease (Data on file). 2014.

17. Kawalec P., Mikrut A., Wisniewska N., Pilc A. Tumor necrosis factor-alpha antibodies (infliximab, adalimumab and certolizumab) in Crohn's disease: systematic review and meta-analysis. Archives of Medical Science 2013; 9(5):765-779.

18. Rutgeerts P., D'Haens G., Targan S., Vasiliauskas E., Hanauer S.B., Present D.H. et al. Efficacy and safety of retreatment with anti-tumor necrosis factor antibody (infliximab) to maintain remission in Crohn's disease. Gastroenterology 1999; 117(4):761-769.

19. Targan S.R., Hanauer S.B., van Deventer S.J., Mayer L., Present D.H., Braakman T. et al. A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. N Engl J Med 1997; 337(15):1029-1035.

20. Watanabe M., Hibi T., Lomax K.G., Paulson S.K., Chao J., Alam M.S. et al. Adalimumab for the induction and maintenance of clinical remission in Japanese patients with Crohn's disease. J Crohns Colitis 2012; 6(2):160-173.

21. National Institute for Health and Care Excellence. Single Technology Appraisal (STA) - Specification for manufacturer/sponsor submission of evidence. 2009.

22. Takeda Pharmaceuticals. Takeda Data on File. (2012a). Final clinical study report C13007 Randomized, placebo-controlled, blinded, multicenter evaluation of induction and maintenance therapy in patients with moderate to severe CD. 2012.

23. Takeda Pharmaceuticals. Takeda Data on File. (2012b). Final clinical study report C13011. A Phase 3, Randomized, Placebo-Controlled, Blinded, Multicenter Study of the Induction of Clinical Response and Remission by Vedolizumab in Patients with Moderate to Severe Crohn's Disease. 2012.

24. Feagan B.G., Greenberg G.R., Wild G., Fedorak R.N., Pare P., McDonald J.W.D. et al. Treatment of Active Crohn's Disease With MLN0002, a Humanized Antibody to the alpha 4 beta 7 Integrin. Clinical Gastroenterology and Hepatology 2008; 6(12):1370-1377.

25. Hyams J.S., Crandall W., Rosh J.R., Ruemmele F., Escher J.C., Lazar A. et al. Efficacy and Safety of Standard vs Low Dose Adalimumab Maintenance Therapy As a Function of Disease Severity in Pediatric Patients With Crohn's Disease: Subanalysis of Imagine 1. Gastroenterology 2013; 144(5):S887.

26. Veereman G., Escher J.C., Ruemmele F., Crandall W., Lazar A., Skup M. et al. Adalimumab Treatment Is Associated With Improved Quality of Life in Pediatric Crohn's Disease. Gastroenterology 2013; 144(5):S887.

27. Present D.H., Rutgeerts P., Targan S., Hanauer S.B., Mayer L., Van Hogezand R.A. et al. Infliximab for the treatment of fistulas in patients with Crohn's disease. New England Journal of Medicine 1999; 340(18):1398-1405.

28. Van Assche G., Vermeire S., Ballet V., Gabriels F., Noman M., D'Haens G. et al. Switch to adalimumab in patients with Crohn's disease controlled by maintenance infliximab: prospective randomised SWITCH trial. Gut 2012; 61(2):229-234.

29. Sands B.E., Anderson F.H., Bernstein C.N., Chey W.Y., Feagan B.G., Fedorak R.N. et al. Infliximab maintenance therapy for fistulizing Crohn's disease. New England Journal of Medicine 2004; 350(9):876-885.

30. Regueiro M., Schraut W., Baidoo L., Kip K.E., Sepulveda A.R., Pesci M. et al. Infliximab Prevents Crohn's Disease Recurrence After Ileal Resection. Gastroenterology 2009; 136(2):441-450.

31. Mazzuoli S, Regano N, Guglielmi FW. Safety of long-term infliximab and adalimumab treatments in crohn's disease. 2013.

32. Lichtenstein G.R., Thomsen O.O., Schreiber S., Lawrance I.C., Hanauer S.B., Bloomfield R. et al. Continuous Therapy With Certolizumab Pegol Maintains Remission of Patients With Crohn's Disease for up to 18 Months. Clinical Gastroenterology and Hepatology 2010; 8(7):600-609.

33. Colombel J.F., Sandborn W.J., Reinisch W., Mantzaris G.J., Kornbluth A., Rachmilewitz D. et al. Infliximab, Azathioprine, or Combination Therapy for Crohn's Disease. New England Journal of Medicine 2010; 362(15):1383-1395.

34. Bhatia J.K., Korelitz B.I., Panagopoulos G., Lobel E., Mirsky F., Sultan K. et al. A prospective open-label trial of Remicade((R)) in patients with severe exacerbation of Crohn's disease requiring hospitalization - A comparison with outcomes previously observed in patients receiving intravenous hydrocortisone. Journal of Clinical Gastroenterology 2007; 41(7):677-681.

35. Duan Z, Luo J, Li W. Efficacy of infliximab combined with azathioprine for moderate to severe Crohn's disease. 2013.

36. Travis S.P., Stange E.F., Lemann M., Oresland T., Chowers Y., Forbes A. et al. European evidence based consensus on the diagnosis and management of Crohn's disease: current management. Gut 2006; 55 Suppl 1:i16-i35.

37. Travis S.P., Stange E.F., Lemann M., Oresland T., Bemelman W.A., Chowers Y. et al. European evidence-based Consensus on the management of ulcerative colitis: Current management. J Crohns Colitis 2008; 2(1):24-62.

38. Schreiber S., Khaliq-Kareemi M., Lawrance I.C., Thomsen O.O., Hanauer S.B., McColm J. et al. Maintenance therapy with certolizumab pegol for Crohn's disease. N Engl J Med 2007; 357(3):239-250.

39. Sands B.E., Kozarek R., Spainhour J., Barish C.F., Becker S., Goldberg L. et al. Safety and tolerability of concurrent natalizumab treatment for patients with Crohn's disease not in remission while receiving infliximab. Inflamm Bowel Dis 2007; 13(1):2-11.

40. Panaccione R., Colombel J.F., Sandborn W.J., D'Haens G., Zhou Q., Pollack P.F. et al. Adalimumab maintains remission of Crohn's disease after up to 4 years of treatment: data from CHARM and ADHERE. Alimentary Pharmacology & Therapeutics 2013; 38(10):1236-1247.

41. Sandborn W.J., Hanauer S.B., Rutgeerts P., Fedorak R.N., Lukas M., MacIntosh D.G. et al. Adalimumab for maintenance treatment of Crohn's disease: results of the CLASSIC II trial. Gut 2007; 56(9):1232-1239.

42. Millennium Pharmaceuticals I. An Open-label Study of Vedolizumab (MLN0002) in Patients With Ulcerative Colitis and Crohn's Disease (GEMINI LTS). 2014.

43. Schulz K.F., Grimes D.A. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet 2002; 359(9308):781-785.

44. Hanauer S.B., Feagan B.G., MacIntosh D.G., Xu J., Milch C., Fox I. et al. Efficacy of Vedolizumab in Crohn's Disease by Prior Treatment Failure in Gemini II, a Randomized, Placebo-Controlled, Double-Blind, Multicenter Study. Gastroenterology 2013; 144(5):S772.

45. Colombel J.F., Sands B., Hanauer S., Rutgeerts P., Sandborn W., Danese S. et al. Long-term Safety of Vedolizumab for the Treatment of Ulcerative Colitis or Crohn's Disease. American Journal of Gastroenterology 2013; 108:S502-S503.

46. Colombel J.-F., Sands B.E., Feagan B.G., Loftus E.V., Sankoh S., Fox I. et al. Integrated safety analysis of vedolizumab for the treatment of ulcerative colitis or crohn's disease. Gastroenterology 2013; 144(5 SUPPL. 1):S113.

47. Feagan B.G., Rutgeerts P., Sands B.E., Hanauer S., Colombel J.F., Sandborn W.J. et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med 2013; 369(8):699-710.

48. Parikh A., Leach T., Wyant T., Scholz C., Sankoh S., Mould D.R. et al. Vedolizumab for the treatment of active ulcerative colitis: A randomized controlled phase 2 dose-ranging study. Inflammatory Bowel Diseases 2012; 18(8):1470-1479.

49. Millennium Pharmaceuticals I. Phase 2, multiple dose, open-label study to determine the long term safety of MLN002 in patients with ulcerative colitis and Crohn's disease. 2014.

50. FDA. FDA Briefing Document for the Joint Meeting of the Gastrointestinal Drugs Advisory Committee (GIDAC) and the Drug Safety and Risk Management Advisory Committee (DSaRMAC). 2014.

51. Hanauer S.B., Sandborn W.J., Rutgeerts P., Fedorak R.N., Lukas M., MacIntosh D. et al. Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn's disease: the CLASSIC-I trial. Gastroenterology 2006; 130(2):323-333.

52. Sandborn W.J., Rutgeerts P., Enns R., Hanauer S.B., Colombel J.F., Panaccione R. et al. Adalimumab induction therapy for Crohn disease previously treated with infliximab: a randomized trial. Ann Intern Med 2007; 146(12):829-838.

53. Rutgeerts P., Van A.G., Sandborn W.J., Wolf D.C., Geboes K., Colombel J.F. et al. Adalimumab induces and maintains mucosal healing in patients with Crohn's disease: data from the EXTEND trial. Gastroenterology 2012; 142(5):1102-1111.

54. Hanauer S.B., Feagan B.G., Lichtenstein G.R., Mayer L.F., Schreiber S., Colombel J.F. et al. Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. Lancet 2002; 359(9317):1541-1549.

55. Colombel J.F., Sandborn W.J., Rutgeerts P., Enns R., Hanauer S.B., Panaccione R. et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: the CHARM trial. Gastroenterology 2007; 132(1):52-65.

56. Sandborn W.J., Schreiber S., Feagan B.G., Rutgeerts P., Younes Z.H., Bloomfield R. et al. Certolizumab pegol for active Crohn's disease: a placebo-controlled, randomized trial. Clin Gastroenterol Hepatol 2011; 9(8):670-678.

57. European Medicines Agency (EMA). Summary of Product Characteristics - Infliximab. 2014.

58. European Medicines Agency (EMA). Summary of Product Characteristics - Adalimumab. 2014;1-265.

59. Clark W, Raftery J, Song F, Barton P, Cummins C F.-S.A., Burls A. Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn's disease. Health Technology Assessment 2003; 7(3):1-+.

60. Bodger K., Kikuchi T., Hughes D. Cost-effectiveness of biological therapy for Crohn's disease: Markov cohort analyses incorporating United Kingdom patient-level cost data. Aliment Pharmacol Ther 2009; 30(3):265-274.

61. Loftus E.V., Johnson S.J., Yu A.P., Wu E.Q., Chao J.D., Mulani P.M. Cost-effectiveness of adalimumab for the maintenance of remission in patients with Crohn's disease. European Journal of Gastroenterology & Hepatology 2009; 21(11):1302-1309.

62. Lindsay J., Punekar Y.S., Morris J., Chung-Faye G. Health-economic analysis: cost-effectiveness of scheduled maintenance treatment with infliximab for Crohn's disease - modelling outcomes in active luminal and fistulizing disease in adults. Alimentary Pharmacology & Therapeutics 2008; 28(1):76-87.

63. Dretzke J., Edlin R., Round J., Connock M., Hulme C., Czeczot J. et al. A systematic review and economic evaluation of the use of tumour necrosis factor-alpha (TNF-alpha) inhibitors, adalimumab and infliximab, for Crohn's disease. Health Technol Assess 2011; 15(6):1-244.

64. Guide to the methods of technology appraisal. 2013.

65. Wailoo A, Tosh J, Hemingway P. Use of tumour necrosis factor alpha (TNF a) inhibitors (adalimumab and infliximab) for Crohn's disease (2009). Report by the Decision Support Unit NICE 2009.

66. Rutgeerts P., Feagan B.G., Lichtenstein G.R., Mayer L.F., Schreiber S., Colombel J.F. et al. Comparison of scheduled and episodic treatment strategies of infliximab in Crohn's disease. Gastroenterology 2004; 126(2):402-413.

67. Blackmore L., Harris A. A prospective study of infliximab withdrawal after 12 months of treatment in patients with Crohn's disease: how will NICE guidance affect patient care? Clinical Medicine 2012; 12(3):235-238.

68. Sandborn W.J., Colombel J.F., Enns R., Feagan B.G., Hanauer S.B., Lawrance I.C. et al. Natalizumab induction and maintenance therapy for Crohn's disease. N Engl J Med 2005; 353(18):1912-1925.

69. Frolkis A.D., Dykeman J., Negron M.E., Debruyn J., Jette N., Fiest K.M. et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. Gastroenterology 2013; 145(5):996-1006.

70. Office for National Statistics. Death registrations summary tables, England and Wales 2010. 2011.

71. Lichtenstein G.R., Feagan B.G., Cohen R.D., Salzberg B.A., Diamond R.H., Chen D.M. et al. Serious infections and mortality in association with therapies for Crohn's disease: TREAT registry. Clin Gastroenterol Hepatol 2006; 4(5):621-630.

72. Buxton M.J., Lacey L.A., Feagan B.G., Niecko T., Miller D.W., Townsend R.J. Mapping from disease-specific measures to utility: an analysis of the relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. Value Health 2007; 10(3):214-220.

73. Brown R.E., Hutton J., Burrell A. Cost effectiveness of treatment options in advanced breast cancer in the UK. Pharmacoeconomics 2001; 19(11):1091-1102.

74. Porco T.C., Lewis B., Marseille E., Grinsdale J., Flood J.M., Royce S.E. Cost-effectiveness of tuberculosis evaluation and treatment of newly-arrived immigrants. BMC Public Health 2006; 6:157.

75. Hornberger J., Reyes C., Lubeck D., Valente N. Economic evaluation of rituximab plus cyclophosphamide, vincristine and prednisolone for advanced follicular lymphoma. Leuk Lymphoma 2008; 49(2):227-236.

76. Beusterien K.M., Davies J., Leach M., Meiklejohn D., Grinspan J.L., O'Toole A. et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. Health Qual Life Outcomes 2010; 8:50.

77. Beusterien K.M., Szabo S.M., Kotapati S., Mukherjee J., Hoos A., Hersey P. et al. Societal preference values for advanced melanoma health states in the United Kingdom and Australia. Br J Cancer 2009; 101(3):387-389.

78. BMJ Group, RCPCH Publications Ltd, and the Royal Pharmaceutical Society of Great Britain. British National Formulary. 2013.

79. Department of Health. PbR tariff 2012-2013. 2013.

80. Curtis L. PSSRU Unit Costs of Health and Social Care 2012. 2012.

81. Department of Health. NHS Reference Costs 2011-12. 2014.

82. NICE. TA65: Rituximab for aggressive non-Hodgkin's lymphoma. 2003.

83. NICE. TA243: Rituximab for the first-line treatment of stage III-IV follicular lymphoma: review of NICE technology appraisal guidance 110. 2012.

84. NICE. TA226: Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma. 2011.

85. Card T., Hubbard R., Logan R.F. Mortality in inflammatory bowel disease: a population-based cohort study. Gastroenterology 2003; 125(6):1583-1590.

86. Targan S.R., Feagan B.G., Fedorak R.N., Lashner B.A., Panaccione R., Present D.H. et al. Natalizumab for the treatment of active Crohn's disease: results of the ENCORE Trial. Gastroenterology 2007; 132(5):1672-1683.

87. Sandborn W.J., Feagan B.G., Stoinov S., Honiball P.J., Rutgeerts P., Mason D. et al. Certolizumab pegol for the treatment of Crohn's disease. N Engl J Med 2007; 357(3):228-238.

88. Winter T.A., Wright J., Ghosh S., Jahnsen J., Innes A., Round P. Intravenous CDP870, a PEGylated Fab' fragment of a humanized antitumour necrosis factor antibody, in patients with moderate-to-severe Crohn's disease: an exploratory study. Aliment Pharmacol Ther 2004; 20(11-12):1337-1346.

89. Schreiber S., Rutgeerts P., Fedorak R.N., Khaliq-Kareemi M., Kamm M.A., Boivin M. et al. A randomized, placebo-controlled trial of certolizumab pegol (CDP870) for treatment of Crohn's disease. Gastroenterology 2005; 129(3):807-818.

90. Ghosh S., Goldin E., Gordon F.H., Malchow H.A., Rask-Madsen J., Rutgeerts P. et al. Natalizumab for active Crohn's disease. N Engl J Med 2003; 348(1):24-32.

10. APPENDICES

Induction studies: key patient characteristics

Study	Population	Treatment	% TNF naïve	% TNF failure	Mean baseline CDAI (Intervention: control)	Fistulising	Stricturing
Treatments not licensed in the UK but which formed part of the network							
ENACT-1 Sandborn et al., 2005[68]	Moderate to severe	Natalizumab	60; 62%	27; 25	302; 303	Excluded draining fistula.	Excluded patients with a stricture or with obstructive symptoms
ENCORE Targan et al., 2007[86]	Moderate to severe	Natalizumab	50; 55	50; 45	299.5; 303.9	Excluded draining fistula.	Excluded patients with a stricture or with obstructive symptoms
PRECISE I Sandborn et al., 2007a[87]	Moderate to severe: excluded loss of response/reaction to anti-TNF-α other than infliximab	Certolizumab	70; 74	NR	300; 297	NR	Excluded patients with a stricture or with obstructive symptoms
Winter et al., 2004 [[p. 1337]][88]	Moderate to severe. No exclusion for prior anti-	Certolizumab	76%	NR	310	Excluded if had fistula abscess	

Study	Population	Treatment	% TNF naïve	% TNF failure	Mean baseline CDAI (Intervention: control)	Fistulising	Stricturing
	TNF						
Sandborn et al., 2011 [[p. 670]][56]	Moderate to severe: Anti-TNF-naïve only	Certolizumab	100%	0%	262.; 292.7	Excluded bowel perforation in last 6 months, actively draining perianal or enterocutaneous fistulae, other nonenterocutaneous fistulae	Excluded symptomatic obstructive strictures
Schreiber et al., 2005 [[p. 807]][89]	Moderate to severe; excluded primary non-responders and those with intolerance.	Certolizumab	78%	Primary non-responders: 0%	302.1	NR	Excluded non-inflammatory obstruction and abscess
Ghosh et al., 2003[90]	Moderate to severe: Anti-TNF-naïve only	Natalizumab	100%	0%	288 to 300	NR	Excluded those with symptomatic fibrotic strictures

Induction studies: treatment regimens and outcome analyses available

Study	Analysis methods	Interventions (n=randomised)	UK licenced?	Comparator	Population	Outcome Time point (week)		
						CR	ECR	CRem
Treatments not licensed in the UK but which formed part of the network								
ENACT-1 Sandborn et al., 2005[68]	ITT Missing data counted as failures	Natalizumab (IV) 300 mg at weeks 0, 4, and 8 (n = 724)	Not licenced in UK	Placebo (n=181)	Mixed: Naïve: Experienced: Failure:	2,4,6,8,10 (P),12 Calculable 10 NR	NR	2,4,6,8,10,12 Calculable 10 NR
ENCORE Targan et al., 2007[86]	NR if ITT Missing data counted as failures	Natalizumab (IV) 300 mg infusion at weeks 0, 4, and 8 (n = 259)	Not licenced in UK	Placebo (n=250)	Mixed: Naïve: Experienced: Failure:	4, 8, 12 NR NR NR	4, 8, 12 NR NR NR	4, 8, 12 NR NR NR
PRECISE I Sandborn et al., 2007a[87]	ITT Missing data counted as failures	Certolizumab pegol (IV) 400 mg at weeks 0, 2, and 4 and then every 4 weeks (n = 331)	Not licenced in UK	Placebo n = 329	Mixed: Naïve: Experienced (infliximab): Failure:	2,4,6,8,12 NR NR NR	2,4,6 (P), 8,12 NR 6 NR	2,4,6,8,12 NR NR NR
Winter et al., 2004 [[p.	ITT Imputation	Certolizumab (IV) at week 4	Not licenced in UK	Placebo n = 25	Mixed		NR)	2,4,8,12 (estimate from graph)

Study	Analysis methods	Interventions (n=randomised)	UK licenced?	Comparator	Population	Outcome Time point (week)		
						CR	ECR	CRem
1337]][88]	NR	20 mg/kg (n = 23) 10 mg/kg (n = 17) 5 mg/kg (n = 25) 1.25 mg/kg (n = 2)	Single dose study					
Sandborn et al., 2011 [[p. 670]][56]	ITT Missing data counted as failures	Certolizumab (SB) 400 mg at weeks 0, 2, and 4 (n = 223)	Not licenced in UK	Placebo n = 216	Naïve	NR	2,4,6	2,4,6 (P)
Schreiber et al., 2005 [[p. 807]][89]	ITT Missing data "advanced to end-of-study visit"	Certolizumab (IV) at weeks 0, 4, and 8 400 mg (n = 73) 200 mg (n = 72) 100 mg (n = 74)	Not licenced in UK	Placebo n = 73	Mixed (excluded primary non-responders and intolerant)	NR	NR	2,4,6,8,10,12
Ghosh et al., 2003[90]	ITT LOCF	Natalizumab (IV) 6 mg/kg, 2 infusions given 4 weeks apart (n = 51) 3 mg/kg, 2 infusions given 4 weeks apart	Not licenced in UK	Placebo n = 63	Naïve	2,4,6,8,12	NR	2,4,6 (P),8,12

Study	Analysis methods	Interventions (n=randomised)	UK licenced?	Comparator	Population	Outcome Time point (week)		
						CR	ECR	CRem
		(n = 66) 3 mg/kg, 1 infusions followed by placebo 4 weeks later (n = 68)						