Vedolizumab for the treatment of adults with moderately to severely active ulcerative colitis: A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Munira Essat, Research Associate, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Shijie Ren, Research Associate, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Alice Bessey, Research Associate, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Rachel Archer, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Sami Hoque, Consultant Gastroenterologist, Barts Health NHS Trust, Whipps Cross University Hospital, Whipps Cross Road, London, E11 1NR Alan Lobo, Consultant Gastroenterologist, Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF
Correspondence to	Munira Essat, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

| **Date completed** | September 2014 |

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 13/110/01.

Declared competing interests of the authors

Professor Alan Lobo is in preliminary discussions about participation in a company-sponsored study of vedolizumab, for which the Sheffield Teaching Hospitals NHS Foundation Trust would be reimbursed and he would not benefit personally.

Acknowledgements

We would like to thank Matt Stevenson and Rachid Rafia, ScHARR, for providing comments on the draft report and Andrea Shippam, 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:

Essat M, Tappenden P, Ren S, Bessey A, Archer R, Wong R, Hoque S, Lobo A. Vedolizumab for the treatment of adults with moderately to severely active ulcerative colitis: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2014.

Contributions of authors

Munira Essat and Rachel Archer summarised and critiqued the clinical effectiveness data reported by the manufacturer. Shijie Ren critiqued the statistical analyses undertaken by the manufacturer. Ruth Wong undertook the literature searches run by the ERG. Paul Tappenden and Alice Bessey critiqued the health economic analysis submitted by the manufacturer. Sami Hoque and Alan Lobo provided clinical advice to the ERG throughout the project. All authors were involved in drafting and commenting on the final document.

CONTENTS

	Abbreviations	vi
1	SUMMARY	1
1.1	Critique of the decision problem in the manufacturer's submission	1
1.2	Summary of clinical effectiveness evidence submitted by the manufacturer	1
1.3	Summary of the ERG's critique of clinical effectiveness evidence submitted	4
1.4	Summary of cost effectiveness submitted evidence by the manufacturer	6
1.5	Summary of the ERG's critique of cost effectiveness evidence submitted	6
1.6	ERG commentary on the robustness of evidence submitted by the manufacturer	8
1.7	Summary of exploratory and sensitivity analyses undertaken by the ERG	8
2	BACKGROUND	10
2.1	Critique of manufacturer's description of underlying health problem	10
2.2	Critique of manufacturer's overview of current service provision	10
3	CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM	12
3.1	Population	14
3.2	Intervention	16
3.3	Comparators	17
3.4	Outcomes	18
3.5	Other relevant factors	18
4	CLINICAL EFFECTIVENES	19
4.1	Critique of the methods of review(s)	19
4.2	Critique of trials of the technology of interest, their analysis and interpretation	23
4.3	Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison	49
4.4	Critique of the indirect comparison and/or multiple treatment comparison	58
4.5	Additional clinical exploratory analyses undertaken by the ERG	66
4.6	Conclusions of the clinical effectiveness section	66
5	COST-EFFECTIVENESS	68
5.1	ERG comment on manufacturer's review of cost-effectiveness evidence	68
5.2	Description of the manufacturer's model	71
5.3	Critical appraisal of the manufacturer's model	96
5.4	Exploratory and sensitivity analyses undertaken by the ERG	110
5.5	Discussion of cost-effectiveness evidence and additional analyses undertaken by the ERG	118
6	END OF LIFE CONSIDERATIONS	121

7	CONCLUSIONS	122
7.1	Conclusions on the clinical effectiveness and cost-effectiveness of vedolizumab	122
7.2	Implications for research	123
8	APPENDICES	124
Appendix 1	ERG supplementary searches in the United European Gastroenterology (UEG) website	124
Appendix 2	ERG supplementary adverse events search strategies	125
Appendix 3	Meta-analysis contrasts: anti-TNF–naïve induction	127
9	REFERENCES	138

TABLES

Table 1	Summary of additional analyses undertaken by the ERG	9
Table 2	Decision problem as outlined in the final scope issued by NICE and addressed in the MS	12
Table 3	Inclusion/exclusion criteria used to select studies of vedolizumab in the MS	21
Table 4	Characteristics of GEMINI1 study	26
Table 5	List of ongoing studies as identified by the ERG in searches of ClinicalTrials.gov and WHO ICTRP	28
Table 6	Manufacturer's quality assessment results for included RCTs	29
Table 7	GEMINI1 efficacy endpoints at week 6 in induction study	32
Table 8	GEMINI1 efficacy endpoints in maintenance study	34
Table 9	Changes in HRQoL from baseline at week 6 of UC induction therapy in GEMINI1	37
Table 10	Patient reported outcomes in UC in GEMINI1: Changes from baseline by study visit in the maintenance phase	38
Table 11	GEMINI1 patient disposition (induction phase)	40
Table 12	GEMINI1 patient disposition (maintenance phase)	42
Table 13	Key safety results: Induction trial data derived from Feagan *et al.*	44
Table 14	Adverse events occurring in ≥5% and ≥10% of patients receiving vedolizumab in maintenance phase - GEMINI1 safety population	45
Table 15	GEMINI Long-term study - Interim safety results as of July 2012, reproduced from MS	48
Table 16	Summary of trials included in the manufacturer's NMA: Induction study characteristics	51
Table 17	Summary of trials included in the manufacturer's NMA: Maintenance study characteristics	55

Table 18	Summary of data used in the network meta-analysis provided by the MS	59
Table 19	Summary of data used in the network meta-analysis	62
Table 20	Summary of studies included in the manufacturer's cost-effectiveness review	69
Table 21	Populations, comparators and sources of efficacy evidence used in manufacturer's health economic analysis	71
Table 22	Description of interventions assessed in the manufacturer's model	72
Table 23	Health states included in the manufacturer's model (induction and maintenance treatment)	74
Table 24	Summary of evidence sources used to inform the model parameters	78
Table 25	Probabilities of clinical response and clinical remission to induction treatment used in the manufacturer's model	80
Table 26	Transition vectors for induction therapy	81
Table 27	Cells manipulated within the calibration process	82
Table 28	Target data used in the calibration approach	82
Table 29	Initial starting vectors	83
Table 30	Fitted maintenance phase pre-colectomy transition probabilities	84
Table 31	Surgery and post-surgery transition probabilities	85
Table 32	Adverse event probabilities assumed within the manufacturer's model	86
Table 33	Summary of health state utility values used in the manufacturer's model	86
Table 34	Adverse event-related disutility values used in the manufacturer's model	87
Table 35	Probability of biologic discontinuation during each maintenance cycle	88
Table 36	Acquisition costs assumed within the manufacturer's model	89
Table 37	UC health state resource costs	90
Table 38	Unit costs associated with managing adverse events	90
Table 39	Central estimates of cost-effectiveness presented by the manufacturer (based on point estimates of parameters)	91
Table 40	Estimated probability vedolizumab produces the greatest net benefit at willingness to pay thresholds of £20,000 and £30,000 per QALY gained (pairwise comparisons – vedolizumab versus comparator, read from manufacturer's CEACs by ERG)	92
Table 41	Summary results of manufacturer's scenario analyses	94
Table 42	Adherence of the manufacturer's economic analysis to the NICE Reference Case	97
Table 43	Comparison of manufacturer's pairwise ICERs over different time horizons	99
Table 44	Fully incremental analysis using the manufacturer's base case model	99
Table 45	Fully incremental analysis assuming 3-year discontinuation	101
Table 46	HRQoL values	106

Table 47	Costs used in the manufacturer's model and budget impact analysis	107
Table 48	Scenario 1 – correction of conventional management maintenance transition matrix cell referencing in anti-TNF-α population (10-year time horizon)	112
Table 49	Scenario 2 - health utilities based on Woehl *et al*	113
Table 50	Scenario 3 - health utilities based on Swinburn *et al*	113
Table 51	Scenario 4 - amended transition matrix for surgery and post-surgery states	114
Table 52	Scenario 5 - removal of assumption regarding maximum biologic treatment time	115
Table 53	Scenario 6 - conventional therapy use same for all groups	115
Table 54	Scenario 7 - use of 2012/13 NHS Reference Costs to value health state resource use	116
Table 55	Scenario 8 - inclusion of costs of stoma care	117
Table 56	ERG-preferred base case	117

FIGURES

Figure 1	Overview of the induction and maintenance phase in the GEMINI1 trial	25
Figure 2	Network of evidence for anti-TNF-α naïve induction clinical response and clinical remission	60
Figure 3	Network of evidence for anti-TNF-α naïve maintenance durable clinical response and clinical remission	60
Figure 4	Network of evidence anti-TNF-α experienced/failure induction clinical response, clinical remission, and mucosal healing	61
Figure 5	Network of evidence anti-TNF-α experienced/failure maintenance durable clinical response, clinical remission, and mucosal healing	61
Figure 6	Induction treatment - decision tree model structure	73
Figure 7	Maintenance treatment – Markov model structure	74

BOXES

Box 1	Main issues identified through critical appraisal of the manufacturer's model	96

Abbreviations

ASAs / 5-ASAs	Aminosalicylates
MP / 6-MP	6-mercaptopurine
ACT1/2	The Active Ulcerative Colitis Trials 1 and 2
AE	Adverse event
CD	Crohn's disease
c.i.	Confidence interval
CEAC	Cost-effectiveness acceptability curve
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
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
MS	Manufacturer's submission
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
ONS	Office for National Statistics
PbR	Payment by Results
PML	Progressive multifocal leukoencephalopathy

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSS	Personal Social Services
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 manufacturer's submission

The population considered by the manufacturer in this assessment (adult patients with moderately to severely active ulcerative colitis [UC] who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy or a TNF-α antagonist) matches that defined in the final NICE scope. The intervention considered in the manufacturer's submission (MS), i.e. 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 then every eight weeks thereafter. The MS also includes efficacy outcomes for 4-weekly vedolizumab as maintenance therapy. The final NICE scope defines appropriate comparators to be established clinical management without vedolizumab, which may include a combination of aminosalicylates (ASAs - sulfasalazine, mesalazine, balsalazide or olsalazine), corticosteroids (beclometasone, budesonide, hydrocortisone or prednisolone), thiopurines (mercaptopurine [6-MP] or azathioprine), calcineurin inhibitors (tacrolimus or ciclosporin), anti-TNF-α agents (infliximab, adalimumab or golimumab) and surgical intervention. Surgery and calcineurin inhibitors were not included as comparators in the manufacturer's systematic review of clinical effectiveness evidence. The comparators considered within the systematic review, network meta-analysis (NMA) and health economic analysis are not consistent. Outcomes data on relapse rates were not presented in the MS. Colectomy and hospitalisation outcome data were not reported within the MS but were provided by the manufacturer following a request for clarification.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

The MS includes a systematic review and NMA of the clinical effectiveness literature. The GEMINI1 trial, which forms the main supporting evidence for the intervention, was a Phase III, multicentre (34 countries including 2 sites in the UK), randomised, double-blind, placebo-controlled trial designed to evaluate the efficacy and safety of vedolizumab as an induction treatment (Weeks 0 to 6) and maintenance treatment (Weeks 7 to 52) in patients with moderately to severely active UC who had an inadequate response to, loss of response to, or intolerance of conventional therapy or anti-TNF-α).

During the 6-week induction phase, 374 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 521 patients were enrolled in an open-label group (Cohort 2), which received the same active induction regimen given in the blinded study (Cohort 1). During the maintenance phase, patients from both cohorts (Cohort 1 and Cohort 2) who had a clinical response (defined as a reduction in the Mayo Clinic score of at least 3 points and a decrease of at least 30% from the baseline score, with a decrease of at least 1 point on the rectal bleeding subscale or an absolute rectal bleeding score of ≤1) to vedolizumab at Week 6 (n=373) 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 MS, 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. 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 general, all efficacy analyses in the GEMINI1 trial were conducted according to the intention-to-treat (ITT) principle whereby patients who withdrew prematurely were considered as treatment failures. In the induction phase, 6% [57/895] of the total population prematurely discontinued from the study. In contrast, a larger proportion of patients discontinued during the maintenance phase (44% [164/373] of the total population i.e. responders to vedolizumab during the induction phase that were re-randomised to maintenance therapy at Week 6). The main reasons for discontinuation in the vedolizumab and placebo groups were lack of efficacy or disease-related adverse events (AEs).

In the induction phase, clinical response at Week 6 was 47.1% (106/225) in the vedolizumab group compared with 25.5% (38/149) in the placebo group (difference after adjustment for stratification factors, 21.7 percentage points; 95% c.i. 11.6 to 31.7; $p<0.001$). Clinical remission occurred in 16.9% (38/225) of patients in the vedolizumab group compared with 5.4% (8/149) in the placebo group ($p=0.001$). Rates of mucosal healing were 40.9% (92/225) in the vedolizumab group and 24.8% (37/149) in the placebo group ($p=0.001$). Additional subgroup analyses showed that, compared with placebo, treatment with vedolizumab improved clinical response and remission rates at 6-weeks in patients with no prior anti-TNF-α exposure and to a lesser extent in those with prior anti-TNF-α failure (*p*-values were not provided as the manufacturer stated that 'multiple testing adjustments were not made'). A *post hoc* 'delayed responder' exploratory analysis in patients who failed to demonstrate clinical response at Week 6 in the induction phase found that the percentage of patients achieving clinical response (using partial Mayo scores) at Week 10 and Week 14 in vedolizumab-treated patients was 32% (102/322) and 39% (126/322), respectively, compared with placebo (15% [12/82] and 21% [17/82], respectively).

In the maintenance phase, higher rates of efficacy were observed in the vedolizumab (300mg i.v.) groups compared with the placebo group. At Week 52, clinical remission occurred in 41.8% (51/122) of patients who continued to receive vedolizumab every 8 weeks and in 44.8% (56/125) of patients who continued to receive vedolizumab every 4 weeks, but in only 15.9% (20/126) of patients who were re-randomised to placebo (difference after adjustment for stratification factors, 26.1 percentage

points for vedolizumab every 8 weeks versus placebo; 95% c.i.: 14.9 to 37.2; $p<0.001$ and 29.1 percentage points for vedolizumab every 4 weeks versus placebo; 95% c.i.: 17.9 to 40.4; $p<0.001$). Durable clinical response occurred in 56.6% (69/122) of patients in the vedolizumab 8-weekly group, 52% (65/125) in the vedolizumab 4-weekly group, and 23.8% (30/126) in the placebo group; $p<0.001$ in both groups versus placebo. Durable clinical remission occurred in 20.5% (25/122) of patients in the vedolizumab 8-weekly group, 24% (30/125) in the vedolizumab 4-weekly group, and 8.7% (11/126) in the placebo group; $p=0.008$ and $p<0.001$ respectively. Vedolizumab was also associated with higher mucosal healing rates (51.6% [63/122] in the vedolizumab 8-weekly group, 56% [70/125] in the vedolizumab 4-weekly group, and 19.8% [25/126] in the placebo group; $p<0.001$ in both groups versus placebo). The proportion of patients who were glucocorticoid-free at 52 weeks was significantly higher in those treated with vedolizumab compared with those who received placebo (31.4% [22/70] of the vedolizumab 8-weekly group, 45.2% [33/73] of the vedolizumab 4-weekly group, and 13.9% [10/72] in placebo group; $p=0.01$ and $p<0.001$, respectively). No clear differences in efficacy were observed between the two vedolizumab regimens. Clinical response and remission rates were generally favourable for vedolizumab compared with placebo in both the anti-TNF-α naïve and anti-TNF-α failure subgroups. However, efficacy was greater in anti-TNF-α naïve group compared with the anti-TNF-α failure group. Generally, a greater health-related quality of life (HRQoL) improvement was observed in patients treated with vedolizumab in both the induction and maintenance phase compared with the placebo group.

In the absence of any direct head-to-head randomised controlled trials (RCTs) comparing vedolizumab and other relevant biologic therapies for the treatment of moderate to severe UC, the manufacturer conducted an NMA. The NMA compared vedolizumab, adalimumab, golimumab, infliximab and placebo for the outcomes clinical response, durable clinical response, clinical remission, mucosal healing, discontinuation due to AEs, serious adverse events (SAEs) and corticosteroid-free (CSF) remission using data from the trials: GEMINI1, ULTRA1, ULTRA2, ACT1, ACT2, PURSUIT-SC, PURSUIT-M and Suzuki (2014). The size of the network for each outcome varied depending on the availability of the data in each study.

The manufacturer undertook separate NMAs of the anti-TNF-α naïve, anti-TNF-α experienced/failure subgroups and the mixed ITT population. Induction phase and maintenance phase data were synthesised separately. For the trials without re-randomisation at the end of the induction phase, the manufacturer's NMA assumes that patients who responded at the end of maintenance also all responded at end of induction. Both Bayesian fixed and random effects models were used but only the fixed effects model results were presented. All outcome measures were modelled using a binomial likelihood and a logit link function.

The fixed effects NMA suggested that in the induction phase for anti-TNF-α naïve patients, infliximab provided the largest treatment effect on clinical response, remission and mucosal healing compared with placebo, and vedolizumab has the lowest rate of discontinuations due to AEs compared with placebo. In the induction phase for anti-TNF-α experienced/failure patients, only the treatment effects of adalimumab and vedolizumab were analysed relative to placebo. Each had positive effects in term of clinical response, remission and mucosal healing, but only the effect of vedolizumab compared with placebo for the outcome of response was statistically significant. For the maintenance phase, vedolizumab was associated with the largest treatment effect compared with placebo in both the anti-TNF-α naïve and anti-TNF-α experienced/failure patient subgroups. However, patients in the GEMINI1 maintenance phase were all vedolizumab induction-responders. No data are available for the efficacy of vedolizumab for vedolizumab responders relative to placebo for placebo responders during the maintenance phase.

The frequency of AEs was similar between the vedolizumab and placebo groups in the GEMINI1 trial. The most commonly occurring AEs during the maintenance phase in the combined vedolizumab group compared with the combined placebo group were nasopharyngitis (12.9% versus 9.5%), headache (12.9% versus 10.2%), arthralgia (9.0% versus 9.1%) and upper respiratory tract infections (8.4% versus 7.6%), respectively. The majority of infusion-related reactions in the induction and maintenance phases were mild to moderate in severity with only 3 cases resulting in drug discontinuation. Although no cases of anaphylaxis, serum sickness or progressive multifocal leukoencephalopathy (PML) were observed, one patient died during the GEMINI1 trial; this was considered by the study investigators to be non-treatment related. Supplementary safety evidence from an ongoing GEMINI Long Term Safety (LTS) trial and two separate pooled safety analyses (not meta-analysed) were also provided by the manufacturer. In general, the overall safety profile of vedolizumab appeared to be similar between patients with 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. In addition 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 (TB) 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). None of the UC deaths were considered by the study investigators to be treatment-related.

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

The systematic review process followed by the manufacturer was comprehensive. Despite minor limitations in the manufacturer's search strategy, the Evidence Review Group (ERG) is confident that

all relevant studies of vedolizumab were included in the MS. The specified inclusion and exclusion criteria appear generally appropriate and reflect the information given in the decision problem. The validity assessment tool used to appraise the included studies, as suggested by NICE, was based on the quality assessment criteria for RCTs and was considered appropriate by the ERG.

Although the efficacy and safety of vedolizumab was positively demonstrated (compared with placebo) in the GEMINI1 study, there are a number of limitations and uncertainties which warrant caution in the interpretation of the available evidence. Owing to the high discontinuation rates in the maintenance phase of the GEMINI1 trial, estimates of treatment effects (including magnitude) may be confounded. The subgroup analyses undertaken to determine the efficacy of vedolizumab in patients with prior anti-TNF-α failure and in patients who were anti-TNF-α naïve were exploratory and the study was not powered for these assessments. In addition, the trial of maintenance therapy was not of sufficient size or duration to estimate the risk of uncommon AEs.

In the manufacturer's NMA, the ERG considered that the results presented may have underestimated the uncertainty in treatment effects since fixed effects models were used, despite clear evidence of heterogeneity amongst the trials included in the network. The results presented for maintenance phase clinical remission and durable clinical response may not be correct since incorrect data were used. The adjustments made by the manufacturer in the maintenance phase to the trials without re-randomisation at the end of the induction phase inflate estimates of treatment effects in both the placebo and experimental treatment groups. The impact of this adjustment on the relative treatment effect in these trials is not clear. It is also unclear if the large relative treatment effect observed for vedolizumab compared with placebo in the maintenance phase in GEMINI1 was due to the low event rates for placebo-treated vedolizumab induction-responders in the control group of the trial. Because the patient population in the maintenance phase was different between GEMINI1 trial and ULTRA2 (GEMINI1 included prior vedolizumab induction-responders only), it was not clear if the placebo groups in these two trials are comparable in the NMA for the anti-TNF-α experienced/failure subgroup. The anti-TNF-α naïve subgroup also has this comparability issue in the maintenance phase. The results of the NMA for clinical response and remission should be interpreted with further caution because these were estimated without considering the dependence/correlation between response and remission. Use of these results in the economic model ignores this dependence and may generate inappropriate samples for probabilistic sensitivity analysis (PSA).

The main uncertainties in the clinical evidence relate to the duration of treatment and generalisability of the evidence to the UK population.

1.4 Summary of cost effectiveness submitted evidence by the manufacturer

The manufacturer submitted a model-based health economic analysis as part of their submission. The analysis was undertaken from the perspective of the NHS over a 10-year time horizon. The manufacturer'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) and surgery as separate options. Other anti-TNF-α agents (infliximab, adalimumab and golimumab) are included only in the analysis of the anti-TNF-α naïve population; these therapies are excluded from the analyses of the mixed ITT and anti-TNF-α failure populations. Calcineurin inhibitors are not included in the economic analysis. All analyses include price reductions to reflect the proposed Patient Access Scheme (PAS) for vedolizumab.

The manufacturer's results were presented only as pairwise comparisons of vedolizumab versus each comparator and are thus difficult to interpret appropriately. Based on a fully incremental analysis (re-analysed by the ERG), within the mixed ITT population, the manufacturer's model suggests that surgery is dominated as it produces fewer health gains and is more costly than both conventional therapy and vedolizumab. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.15 quality adjusted life years (QALYs) at an incremental cost of £5,131; the ICER for vedolizumab versus conventional therapy is estimated to be £33,297 per QALY gained. Within the anti-TNF-α naïve population, the manufacturer's model suggests that surgery is expected to be dominated by medical therapies. Vedolizumab is expected to be the most effective option. Infliximab and golimumab are expected to be dominated by vedolizumab and are ruled out of the analysis. The ICER for adalimumab versus conventional therapy is estimated to be £3,664 per QALY gained, whilst the ICER for vedolizumab versus adalimumab is estimated to be £6,634 per QALY gained. Within the anti-TNF-α failure population, the manufacturer's model suggests that surgery is expected to be dominated. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.09 QALYs at an incremental cost of £5,839; the incremental cost-effectiveness ratio (ICER) for vedolizumab versus conventional therapy is estimated to be £64,999 per QALY gained.

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

The ERG critically appraised the manufacturer's health economic analysis and the model upon which this analysis is based. Importantly, the manufacturer's economic analysis deviates from the NICE Reference Case and the final NICE scope due to (a) missing biologic comparators in the mixed ITT

population the anti-TNF-α failure populations, (b) the use of a 10-year time horizon and (c) the use of pairwise comparisons rather than a fully incremental analysis. These issues hinder the appropriate interpretation of the manufacturer's results.

Alongside scrutinising the manufacturer's model, the ERG re-built part of the model to check for technical programming errors. One serious programming error was found; in the anti-TNF-α naïve population, the maintenance transition matrix for conventional therapy incorrectly draws on the transition matrix for infliximab. The broader critical appraisal of the manufacturer's model highlighted a number of concerns and uncertainties. The most notable of these relate to the deviations from the NICE Reference Case and final NICE scope (as discussed above), questionable assumptions regarding continuation/discontinuation of vedolizumab and other biologic therapies and highly pessimistic assumptions regarding the use, costs and benefits of colectomy. Also of particular concern is the considerable uncertainty associated with the calibration and extrapolation of the pre-colectomy maintenance transition matrices. This latter issue may have been better addressed by using the observed transitions between moderate to severe UC, response and remission states using the patient-level data collected within the GEMINI1 trial. Despite a request for these data, the manufacturer did not provide them hence the accuracy of the maintenance matrices remains unclear.

In light of the problems identified during the critical appraisal, the ERG undertook a number of additional analyses to explore the impact of likely biases on the cost-effectiveness of vedolizumab. Nine sets of additional analyses were undertaken in each of the three modelled populations; these included correcting the mistake in the maintenance transition matrix for conventional management in the anti-TNF-α naïve population, the use of alternative sources of HRQoL values, amending the surgery and post-surgical transition probabilities to better reflect clinical reality, removing assumptions regarding biologic treatment discontinuation, removing assumptions regarding the lower use of conventional therapies whilst patients are also receiving biologics, and improving the cost estimates used in the model to better account for the costs borne by the NHS. The ERG also produced a preferred base case which combines most of these exploratory analyses. The ERG's analyses indicate these issues have the propensity to dramatically shift the ICER for vedolizumab versus other therapies in all three populations. Individually, the additional analyses do not consistently favour one particular option.

The ERG-preferred base case indicates that surgery is likely to dominate all medical treatments in all three populations analysed. However, whilst surgery appears favourable within all populations, the ERG recognises that this may not be an acceptable option for all patients. Where surgery is not an acceptable option in the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £53,084 per QALY gained. Where surgery is not an acceptable option in the

anti-TNF-α naïve population, vedolizumab is expected to be dominated by adalimumab. Where surgery is not an acceptable option in the anti-TNF-α failure population, the ICER for vedolizumab versus conventional therapy is expected to be £48,205 per QALY gained.

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

1.6.1 Strengths

The manufacturer'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 MS.

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 only two of the GEMINI1 study sites 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, and there is clear evidence of heterogeneity among the trials included. There are also other problems regarding adjustment of data to account for re-randomisation which may lead to bias in the model's results.

The health economic model submitted by the manufacturer is subject to a number of issues which limit the credibility of the manufacturer's results. These include errors in model implementation, the omission of relevant comparators, deviations from the NICE Reference Case and questionable model assumptions. Whilst the manufacturer's economic analysis suggests that the ICER for vedolizumab is below £7,000 per QALY gained within the anti-TNF-α naïve population, the ERG-preferred base case indicates that vedolizumab is expected to be dominated by surgery in all three populations.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

Nine sets of additional analyses were undertaken using the manufacturer's model (refer to Section 5.4 for further details). The ICER for vedolizumab versus the next best comparator in each analysis is summarised in Table 1.

Table 1: Summary of additional analyses undertaken by the ERG

Scenario	Incremental cost per QALY gained (vedolizumab versus next best comparator)		
	Mixed ITT population	Anti-TNF-α naïve population	Anti-TNF-α failure population
Manufacturer's base case*	£33,297	£6,634	£64,999
Correction of transition matrix cell referencing error*	£33,297	£6,469	£64,999
Utilities based on Woehl *et al*	£17,140	-	ext dom
Utilities based on Swinburn *et al*	£15,267	dominating	£33,472
Amended transition matrix for surgery and post-surgery states	£44,114	£20,449	£73,931
No maximum biologic treatment time	£34,827	£3,807,239	£32,524
Same cost for conventional therapies in all groups	£22,590	dominating	£47,087
Use of NHS Reference Costs for UC health states	£27,893	£759	£51,271
Inclusion of stoma care costs	£19,630	dominating	£43,108
ERG-preferred base case	dominated	dominated	dominated

Ext dom – extendedly dominated

* assumes 10-year time horizon; all other analyses reflect a lifetime horizon

2 BACKGROUND

This chapter presents a brief commentary on the manufacturer's interpretation of the underlying health problem and the nature of current service provision.

2.1 Critique of manufacturer's description of underlying health problem

The ERG considers that the descriptions of UC pathophysiology, clinical presentation, and assessment and diagnosis detailed in Section 2 of the MS[1] appear reasonable. The descriptions of patient burden and societal burden include a summary of identified evidence of the impact of UC on HRQoL and costs of illness. The manufacturer's discussion of the context of the assessment appeared relevant to the decision problem under consideration.

2.2 Critique of manufacturer's overview of current service provision

The manufacturer stated that the goal of drug treatment in UC is to induce and maintain remission, maintain steroid-free remission, reduce complications and minimise the requirements for hospitalisations and surgery. The management of mildly to moderately active UC was described in the MS[1] as involving the initial use of oral or topical ASAs. In the event of non-response/intolerance to ASA treatment, oral corticosteroids or oral immunosuppressants would be added on to existing therapies. The ERG concurs with this broad description of UC management and with the manufacturer's view that conventional therapy options typically vary depending on the extent and location of disease. The manufacturer noted that the use of a step-wise escalation approach for the treatment of UC was recommended in NICE Clinical Guideline 166.[2] The UC treatment guidelines from the British Society for Gastroenterology[3] were also summarised in the MS.[1] If the patient does not respond to conventional therapy options, the ERG anticipates that patients may subsequently be considered for treatment using tacrolimus, i.v. steroids or anti-tumour necrosis factor-alpha (anti-TNF-α) therapy. In response to a request for clarification from the ERG[4] **(**question A13), the manufacturer noted that tacrolimus is currently unlicensed for the treatment of UC and that high quality data supporting its use are limited.

The ERG agrees that colectomy is an appropriate treatment option for patients with inadequate control of symptoms and/or poor quality of life on conventional therapy. It is stated in the MS[1] (Section 2.1) that ileostomy or ileal pouch-anal anastomosis (IPAA) are typically reserved for patients with acute severe UC who are refractory to all medical treatments and that *"indication of colectomy and surgical therapy in UC is usually failure of medical therapy leading to chronic active disease or fulminant colitis."*[1] However, the ERG notes that, whilst surgery may be required in emergency cases (e.g. acute severe/fulminant UC), patients with moderately to severely active UC may elect to undergo surgery for a number of reasons, including i) debilitating clinical course with prior treatment failures and/or frequent UC flares and the associated impacts upon patients' HRQoL, ii) increased risk of colorectal

cancer associated with long-standing UC, and iii) identification of pre-malignant dysplasia or malignant neoplasia. However, the ERG also acknowledges that surgery is associated with postoperative morbidity and death and may not be an acceptable option for some patients due to potential complications including infertility, pouchitis, wound infections, wound dehiscence and small bowel obstruction.

As described by the manufacturer, three anti-TNF-α agents are currently licensed in the UK for the treatment of UC, infliximab, adalimumab and golimumab, of which only infliximab is currently recommended by NICE for use in acute severe UC (Technology Appraisal 163[5] - treatment of acute exacerbations of severely active UC when ciclosporin is contraindicated or inappropriate). As noted by the manufacturer, there are currently no biologics recommended by NICE for patients with moderate to severe UC who are not responding to or who are intolerant to conventional therapy or TNF-α inhibitors.

It is asserted in the MS[1] that clinicians managing UC patients who are intolerant to or lose response to anti-TNF-α agents are currently likely to consider dose escalation or cycling through alternative anti-TNF-α agents before considering surgery. The manufacturer considers the latter to be potentially flawed, since a patient failing on one anti-TNF-α therapy may subsequently fail trials of other drugs which have the same mechanism of action. Since vedolizumab has a different mechanism of action (being an integrin receptor antagonist) and is licensed for use following receipt of anti-TNF-α, the manufacturer considers that vedolizumab presents an innovative treatment option for UC patients who have failed or are intolerant to conventional therapy or anti-TNF-α agents.

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

This chapter presents a summary and critique of the decision problem addressed by the MS.[1]

A summary of the decision problem as outlined in the final scope issued by NICE[6] and addressed in the MS is presented in Table 2.

Table 2: Decision problem as outlined in the final scope issued by NICE and addressed in the MS[1]

	Decision problem outlined in final scope issued by NICE	Decision problem addressed in the MS
Population	Adults with moderately to severely active UC (excluding those with acute severe ulcerative colitis that is a medical emergency and requires inpatient treatment) who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy (immunosuppressants and/or corticosteroids) or a TNF-α inhibitor	Adult patients with moderately to severely active UC who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy or a TNF-alpha inhibitor (i.e. matches population in final NICE scope)
Intervention	Vedolizumab	Vedolizumab
Comparator(s)	Established clinical management without vedolizumab, which may include a combination of aminosalicylates (sulfasalazine, mesalazine, balsalazide or olsalazine), corticosteroids (beclometasone, budesonide, hydrocortisone or prednisolone), thiopurines (mercaptopurine or azathioprine), calcineurin inhibitors (tacrolimus or ciclosporin), TNF-α inhibitors (infliximab, adalimumab or golimumab) and surgical intervention	Conventional therapy, as defined in the GEMINI1 study and used in UK clinical practice based on the UK inflammatory bowel disease (IBD) audit; TNF-α inhibitors licensed for treatment of UC in the UK (infliximab, adalimumab and golimumab). Surgical intervention and calcineurin inhibitors were not included as comparators in the manufacturer's systematic review. Surgery is included as a comparator and as part of the pathway within the manufacturer's health economic model. Other anti-TNF-α therapies are considered as comparators in the anti-TNF-α naïve model subgroup only. Calcineurin inhibitors were not included in the model.

	Decision problem outlined in final scope issued by NICE	Decision problem addressed in the MS
Outcomes	• Mortality • Measures of disease activity • Rates of and duration of response, relapse and remission • Rates of hospitalisation • Rates of surgical intervention • Time to surgical intervention • Adverse effects of treatment (including leakage and infections following surgery) • HRQoL	Data on relapse rates were not presented in the MS. Colectomy and hospitalisation data were absent from the original MS but were provided upon a request for clarification.
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 (PSS) perspective. The availability of any patient access schemes for the comparator technologies should be taken into account.	The submission includes a model-based cost-utility analysis of vedolizumab compared against infliximab, adalimumab, golimumab, conventional non-biologic therapies and surgery. Other anti-TNF-α therapies are considered as comparators in the anti-TNF-α naïve model subgroup only. The analysis was undertaken over a 10-year time horizon from the perspective of the NHS. A proposed PAS is included for vedolizumab.
Subgroups to be considered	If evidence allows the following subgroups will be considered: People who have been previously treated with one or more TNF-alpha inhibitors and people who have not received prior TNF-alpha inhibitor therapy	It was stated in the MS that pre-specified analyses would be presented examining the following subgroups: i) anti-TNF-α naïve subgroup, ii) anti-TNF-α failure subgroup, iii) GEMINI1 ITT population (comprising both anti-TNF-α naïve and anti-TNF-α experienced patients). The health economic analysis reflects these three populations.

3.1 Population

Vedolizumab has a therapeutic indication for the treatment of adult patients with moderately to severely active UC who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a TNF-α antagonist.[7]

The population described in the final NICE scope[6] was adults with moderately to severely active UC (excluding those with acute severe UC that is a medical emergency and requires inpatient treatment) who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy (immunosuppressants and/or corticosteroids) or a TNF-α inhibitor.

The population included in the MS was *"adult patients with moderately to severely active UC who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy or a TNF-α antagonist"* (MS[1] page 63).

Patients eligible for inclusion in the GEMINI1 trial were required to be aged 18 to 80 years (Feagan *et al.*, 2013).[8] Patients had to have active UC, with a Mayo score of 6 to 12, a sigmoidoscopy subscore of at least 2 and disease extending at least 15 cm from the anal verge. Eligible subjects had documented unsuccessful previous treatment (i.e. lack of response or unacceptable AEs) with one or more glucocorticoids, immunosuppressive medications (i.e. azathioprine and 6-MP) or anti-TNF-α agents. Participants were permitted to continue receiving mesalamine, up to 30mg prednisone (or equivalent) per day or immunosuppressive drugs at stable doses. Patients were ineligible if they had received anti-TNF-α therapy within 60 days before enrolment, or ciclosporin, thalidomide or investigational drugs within 30 days of enrolment, or if they had received previous treatment with vedolizumab, natalizumab, efalizumab or rituximab. Other exclusion criteria included stoma or a history of colectomy, an increased risk of infectious complications, an anticipated need for major surgery, colonic dysplasia/adenomas and malignant neoplasms. Further details on eligibility according to previous UC therapy among GEMINI1-eligible patients were provided on pages 81-82 of the MS.[1] According to the MS,[1] patients must have demonstrated over the preceding 5-year period, an inadequate response to, loss of response to, or intolerance of at least one of the following agents:

Immunomodulators:

- Signs and symptoms of persistently active disease despite a history of at least one 8-week regimen of oral azathioprine (≥1.5mg/kg) or 6-MP (≥0.75mg/kg) OR
- History of intolerance of at least 1 immunomodulator (including but not limited to nausea/vomiting, abdominal pain, pancreatitis, liver function test abnormalities, lymphopenia, TPMT genetic mutation, infection)

TNF antagonists:

- Signs and symptoms of persistently active disease despite a history of at least one 4-week induction regimen of infliximab 5mg/kg (i.v.), 2 doses at least 2 weeks apart OR
- Recurrence of symptoms during maintenance dosing following prior clinical benefit (discontinuation despite clinical benefit does not qualify) OR
- History of intolerance of infliximab (including but not limited to infusion-related reaction, demyelination, congestive heart failure, infection)

Corticosteroids:

- Signs and symptoms of persistently active disease despite a history of at least one 4-week induction regimen that included a dose equivalent to prednisone 30mg daily orally for 2 weeks or i.v. for 1 week, OR
- Two failed attempts to taper corticosteroids to below a dose equivalent to prednisone 10mg daily orally on 2 separate occasions, OR
- History of intolerance of corticosteroids (including, but not limited to, Cushing's syndrome, osteopenia/osteoporosis, hyperglycaemia, insomnia, and infection).[1]

Demographic, baseline disease characteristics and medication history of patients in the GEMINI1 trial were reported in the publication by Feagan *et al.*[8] Patients had an overall mean age of 40.3 (standard deviation [SD]=13.1) years, were predominantly white (82.0%) and male (58.7%) as a cohort, with a mean body weight of 73.4 kg (SD=18.5). Mean duration of disease was 6.9 (SD=6.4) years and patients had a mean Mayo score of 8.6 (SD=1.8). Concomitant medications for UC included glucocorticoids only (37.1%), immunosuppressants only (17.8%), glucocorticoids and immunosuppressants (16.6%) and no glucocorticoids or immunosuppressants (28.5%). Nearly half of all patients (48.2%) had received prior anti-TNF-α treatment, 41.0% having experienced ≥1 failure of anti-TNF-α therapy, due to inadequate response (48.0%), loss of response (38.4%) (i.e. subsequent loss of initial response) or unacceptable AEs (13.6%).

Patients in GEMINI1 were allowed to take conventional UC treatments in the form of mesalamine, ≥30mg prednisone (or equivalent) daily or immunosuppressive agents at stable doses. Steroid doses were unchanged until Week 6 and then were tapered using a defined regimen for clinical responders to vedolizumab. The MS[1] states that permitted immunosuppressants were kept at stable doses throughout the induction and maintenance phases, with the exception of US study sites, where these drugs were discontinued after induction. One clinical advisor to the ERG noted this difference with respect to the generalisability of the evidence to the UK clinical population. In response to a request for clarification from the ERG[4] (question A24), the manufacturer expanded on the potential impact of the different practice at US sites, anticipating that any potential effect on maintenance phase outcomes

would be minimised by the modest expected relative contribution of US patients receiving concomitant immunosuppressants during induction and by the stratification of patients among the maintenance phase treatment groups. Clinical advisors to the ERG were satisfied with the clinical relevance of the GEMINI1 trial population but noted that only two study sites were UK-based.

The ERG considered the GEMINI1 population included in the MS[1] to reflect that in the wording of the licensed indication and the final NICE scope.[6]

3.2 Intervention

The intervention described in the MS[1] matches the intervention described in the final NICE scope.[6] Vedolizumab (Entyvio®) is a humanised IgG1 monoclonal antibody or biologic. It is described as being gut-selective, binding to the α4β7 integrin, which is preferentially expressed on gut-homing T helper lymphocytes. The gut-selective mechanism of action of vedolizumab is described in the MS as being novel, avoiding the negative effects of systemic immunosuppression associated with other biologic UC therapies (e.g. TNF-α inhibitors), such as risks of infection. However, clinical advice to the ERG suggested that this gut-selective approach may also eliminate the positive effects of systemic immunosuppression, such as benefits in terms of alleviation of extra-intestinal manifestations of disease. According to the MS,[1] the novel mechanism of action of vedolizumab presents an additional treatment option for patients with UC who have failed on conventional therapy or TNF antagonists.

Vedolizumab has a therapeutic indication for the treatment of adult patients with moderately to severely active UC who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a TNF-α antagonist.

Vedolizumab is available as a powder for concentrate for solution for infusion. Each pack contains one vial containing 300mg vedolizumab. Based on correspondence between the manufacturer and NICE (21st August 2014), the NHS list price for vedolizumab is £2,050 per 300mg vial, although at the time of writing the product was not listed on the British National Formulary[9] (BNF). The MS states the price of vedolizumab (excluding VAT) to be [redacted] per 300mg vial.[1] Whilst not stated directly within the MS, this lower price includes a proposed PAS which takes the form of a confidential simple price discount for the drug.

The recommended dose for vedolizumab is 300mg administered by i.v. infusion at zero, two and six weeks and then every eight weeks thereafter. The Summary of Product Characteristics (SmPC) for vedolizumab[10] recommends that continued therapy in patients with UC should be carefully reconsidered in the absence of therapeutic benefit by week 10. Some patients with decreased response may benefit from an increase in dosing frequency to 300mg every four weeks. The MS[1] states that

patients would typically be treated until relapse, intolerance or discontinuation due to side effects and that it is expected that vedolizumab would be added-on to existing treatments in clinical practice. However, it should be noted that this does not reflect the continuation and discontinuation rules for vedolizumab within the manufacturer's health economic model (see Chapter 5).

The SmPC[10] offers guidance on the requirement to restart vedolizumab treatment following interruption of therapy, stating that dosing at every four weeks may be considered. It was also stated that, following an interruption in treatment extending up to one year in clinical trials, efficacy was still evident upon vedolizumab re-treatment with no apparent increase in infusion-related reactions or other AEs.

It is recommended in the SmPC[10] that patients should be monitored during and after vedolizumab infusions for the occurrence of acute hypersensitivity reactions. Patients may receive pre-treatment prior to infusions (e.g. with antihistamine, hydrocortisone and/or paracetamol) to ameliorate the risks of infusion-related reactions.

Contraindications to vedolizumab include active severe infections (e.g. TB, sepsis, cytomegalovirus, listeriosis), opportunistic infections (e.g. PML) and hypersensitivity to the active substance. Patients should be screened for TB prior to initiation of vedolizumab therapy.

It is noted in the SmPC[10] that some integrin antagonists and systemic immunosuppressive agents have been associated with PML, an opportunistic infection which may be fatal. Whilst it is stated in the SmPC[10] that no cases of PML were reported in the vedolizumab clinical trials, patients receiving vedolizumab should be monitored for new onset or worsening of neurological signs and symptoms. It is noted in the MS[1] that the occurrence rate of PML and TB with long-term exposure and in patients pre-treated with anti-TNF-α therapies and/or concomitant immunosuppressants is unknown.

It is also stated in the SmPC[10] that no clinical trial data are available for patients previously treated with natalizumab or rituximab and that caution should be used in considering vedolizumab treatment in such patients. Furthermore, the concomitant use of vedolizumab with biologic immunosuppressants is not recommended.

3.3 Comparators

The final NICE scope[6] describes appropriate comparators to be established clinical management without vedolizumab, which may include a combination of 5-ASAs (sulfasalazine, mesalazine, balsalazide or olsalazine), corticosteroids (beclometasone, budesonide, hydrocortisone or

prednisolone), thiopurines (6-MP or azathioprine), calcineurin inhibitors (tacrolimus or ciclosporin), TNF-α inhibitors (infliximab, adalimumab or golimumab) and surgical intervention.

The MS[1] states that included comparators were conventional therapy (as defined in the GEMINI1 study and used in UK clinical practice based on the UK IBD audit) and TNF-α inhibitors licensed for the treatment of UC in the UK (infliximab, adalimumab and golimumab). The main comparator was described by the manufacturer as being standard care, consisting of 5-ASAs, corticosteroids and immunosuppressants, reflecting baseline UC treatments in the GEMINI1 trial. Patients in GEMINI1 received vedolizumab or placebo alongside conventional UC treatments as background therapies. The manufacturer's NMA, and some health economic subgroup analyses, also include comparisons of vedolizumab against adalimumab, infliximab and golimumab.

Surgical intervention was not included as a comparator in the manufacturer's review of clinical effectiveness; it was however included in the health economic analysis. Calcineurin inhibitors were not included in the manufacturer's review, NMA or health economic analysis. Clinical advisors to the ERG noted that surgery is not acceptable to some patients and that data for the use of calcineurin inhibitors compared against biologics are very limited.

3.4 Outcomes

The final NICE scope[6] specified outcomes for consideration as follows:

- mortality
- measures of disease activity
- rates of and duration of response, relapse and remission
- rates of hospitalisation
- rates of surgical intervention
- time to surgical intervention
- adverse effects of treatment (including leakage and infections following surgery)
- HRQoL

The MS[1] states that outcomes considered were in line with those specified in the final NICE scope.[6] Clinical advisors to the ERG were satisfied with the appropriateness of clinical outcomes. Data on relapse rates were not presented in the MS. Colectomy and hospitalisation outcome data were not reported within the MS but were provided by the manufacturer following a request for clarification.

3.5 Other relevant factors

No equity issues were highlighted within the MS.[1]

4 CLINICAL EFFECTIVENESS

This chapter presents a summary and critical appraisal of the methods and results of the clinical effectiveness review and evidence synthesis presented within the MS.[1]

4.1 Critique of the methods of review(s)

4.1.1 Searches

The original search undertaken by the manufacturer to identify all relevant pharmacological intervention studies (vedolizumab, infliximab and adalimumab) was conducted in April 2013 without date and language restrictions. Update searches were conducted in February 2014 and were limited to publications from 1st April 2013 onwards. The update search included additional intervention terms such as golimumab, surgery and ciclosporin (see MS[1] pages 67-68); however, it is not clear to the ERG whether literature published prior to 2013 was considered for golimumab as the search strategy suggests that a date limit from 2013 was applied (see MS[1] page 340, Appendix 10.2, Biologics: PubMed (Medline) Literature Search Strategy: UC). Furthermore, although tacrolimus, a calcineurin inhibitor, was considered as a comparator in the final NICE scope,[6] it was not included in the manufacturer's search strategy. In response to a request for clarification from the ERG[4] (question A13), the manufacturer noted that tacrolimus may be considered as a conventional therapy and is currently unlicensed for the treatment of UC at any stage of the disease. Despite these limitations, the search strategy utilised appropriate terms to identify the condition (UC), the interventions and the type of evidence (RCTs and prospective studies).

In the original search, several electronic bibliographic databases (MEDLINE [using the PubMed platform], EMBASE [using the Elsevier Platform], the Cochrane Library [using the Wiley platform]) and research registers (ClinicalTrials.gov) were searched. For the update search, the same sources were searched including the World Health Organisation International Clinical Trials Registry Platform Search Portal (WHO ICTRP). The ERG considers the chosen electronic databases and internet sources to be appropriate and the number of hits following a repeat of the MEDLINE search strategy show numbers which are consistent with those reported in Section 6.1 of the MS.[1] However, it is unclear why the Health Technology Assessment database, which forms part of the Cochrane Library, was not searched. In addition, the terms that were used in the research registry searches were not provided in the MS,[1] hence the adequacy of these searches is unclear. Supplementary searches such as scanning of bibliographies of existing systematic reviews and meta-analyses were also undertaken. The manufacturer reported that the United European Gastroenterology website was not accessible (see MS[1] page 68). The ERG accessed the United European Gastroenterology website on 13th August 2014 and searched within the conference archive (Appendix 1). Of the 10 records retrieved between 2010 and 2013, none of these were considered to be relevant to the review.

In the NMA search (see MS[1] pages 113-114), published RCTs of vedolizumab, infliximab, adalimumab and golimumab were identified via the original and updated searches detailed above. However, separate searches relating to the AE profiles of these interventions were not undertaken by the manufacturer (see MS[1] page 140). The ERG sought clarification with the manufacturer regarding the justification for, and limitations of, the lack of searching for vedolizumab AEs (see clarification response[4] question A15). The manufacturer noted that the studies were identified through internal Takeda databases and that they were confident that the safety data provided were as complete as possible without missing data on any relevant drug-related AEs. A supplementary safety and AE search conducted by the ERG in MEDLINE (Ovid), EMBASE (Ovid), Cochrane Library (Wiley Interscience) and Toxline (National Institutes of Health) identified a total of 181 records. The ERG was unable to review the results from the search due to time constraints. The AE search strategies for vedolizumab undertaken by the ERG can be found in Appendix 2.

Despite the noted limitations, the ERG considers the search strategies to be sufficiently comprehensive to retrieve important citations relating to all eligible studies that the ERG and its clinical advisors are aware of. No relevant published studies are likely to have been missed.

4.1.2 Inclusion criteria

The MS[1] describes an appropriate method of identifying and screening references for inclusion in the systematic review of clinical effectiveness. Two independent reviewers applied pre-specified inclusion and exclusion criteria (via a two-stage sifting process) to citations identified by the searches. Any differences in selection were resolved through discussion with a third reviewer (see MS[1] page 69). A summary of the inclusion and exclusion criteria, as reported in the MS (pages 70-72; data re-tabulated by the ERG to provide further clarity), for the systematic review of vedolizumab is summarised in Table 3.

Table 3: Inclusion/exclusion criteria used to select studies of vedolizumab in the MS[1] (pages 70-71)

	Inclusion criteria	Exclusion criteria
Population	• Patients with UC (both treatment-naïve and treatment-experienced)	• Patients who do not have UC
Intervention	• Vedolizumab	
Comparator	• Infliximab (Remicade®) • Adalimumab (Humira®) • Golimumab (Simponi®) • Surgery (of any type) • Ciclosporin	• Studies that do not investigate one of the biologics of interest in at least one of the arms
Outcomes	• Clinical response • Sustained clinical response • Durable clinical response • Clinical remission • Durable clinical remission • Mucosal healing • Surgical outcomes/ complications • Safety outcomes • HRQoL outcomes • Hospitalisations • Change in Mayo score from baseline • Mean Mayo score at baseline and each subsequent visit	• For irritable bowel disease articles, exclude if results are not reported separately for UC and CD
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 • Systematic reviews and meta-analyses[a]	• 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) • 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

[a] Systematic reviews and meta-analyses were only included for identification of primary studies

The specified inclusion and exclusion criteria were mostly appropriate and generally reflect the information given in the decision problem; however, there appear to be some irregularities in the MS.[1]

The manufacturer broadly defined the included population as patients with UC (both treatment-naïve and treatment-experienced). Whilst this is appropriate, it would have been more appropriate to define the included population in line with the wording of the marketing authorisation for vedolizumab and that of the decision problem i.e. adults (≥18 years) with moderately to severely UC (both treatment-naïve and treatment-experienced) who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy (immunosuppressants and/or corticosteroids) or a TNF-α inhibitor. The excluded population was loosely defined as patients who do not have UC. Further details would have been beneficial e.g. excluding patients with acute severe UC that is a medical emergency and requires inpatient treatment or patients with mild UC and children (aged <18 years).

The statement of the decision problem proposed that tacrolimus was to be considered as a comparator and that relapse rates, colectomy rates, hospitalisation and mortality were to be considered as relevant outcomes for the appraisal. Initially, it was unclear to the ERG why these comparators and outcomes were not included in the manufacturer's systematic review as no explicit details were provided in the MS.[1] However, appropriate justifications for the exclusion of tacrolimus and outcome data for mortality, colectomy and hospitalisation were provided in the manufacturer's response to the ERG's clarification request[4] (questions A13, A22, A31 and A33). In addition, golimumab and ciclosporin were not included as part of the initial screening in the original review, but were added for the update searches. As a result, it is possible that relevant studies may have been missed, although the ERG believes that the risk of this is minimal as additional relevant studies were also identified via other sources e.g. existing systematic reviews, web searches and conference proceedings.

4.1.3 Critique of data extraction

The data extracted and presented in the clinical section of the MS[1] (including the manufacturer's response to clarification questions[4]) appear appropriate and comprehensive. As noted in the manufacturer's response to clarification question A19,[4] data extraction was undertaken independently by two reviewers.

4.1.4 Quality assessment

The validity assessment tool used to appraise the included studies in the MS[1] (page 94) was based on the quality assessment criteria for RCTs, as suggested by the NICE guideline template for manufacturers.[11] As noted in the manufacturer's response to clarification[4] (question A19), methodological quality assessment of included studies was performed by two independent researchers. The ERG considers the validity assessment tool used in the MS[1] to be appropriate.

4.1.5 Evidence synthesis

The manufacturer undertook a narrative synthesis of the evidence for vedolizumab; however, no explicit details were provided in the MS[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.[12] Despite the lack of transparency regarding the methods adopted, the ERG acknowledges that the narrative synthesis approach undertaken by the manufacturer was acceptable.

An NMA was used to perform indirect comparisons of vedolizumab, adalimumab, golimumab, infliximab and placebo. A critique of the NMA can be found in Section 4.4. No meta-analysis was performed for surgery or calcineurin inhibitors with the following reasons given by the manufacturer:

- Variation in study design; studies were not comparable
- Lack of a common comparator to connect the network; surgery studies tended to compare one approach to another without a placebo arm
- Differing outcomes in each study
- Small sample sizes[1]

With the exception of the justification regarding small sample sizes, the ERG considers these reasons to be acceptable.

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 manufacturer presented a systematic review of the clinical effectiveness and safety of vedolizumab for the treatment of moderately to severely active UC in adults who are intolerant of, or whose disease has an inadequate response or loss of response to conversional therapy or TNF-α antagonist. The manufacturer's PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram 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 MS[1] page 73) represents the identification and selection of relevant biological therapies for the treatment of UC (i.e. for the systematic review of vedolizumab and for the systematic review/ potential NMAs incorporating infliximab, adalimumab and golimumab indicated for the treatment of moderate to severe UC 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 processes for each of the reviews. Moreover, although limited details were provided for excluding two ongoing studies,[13,14] no explicit details were provided in the MS[1] or

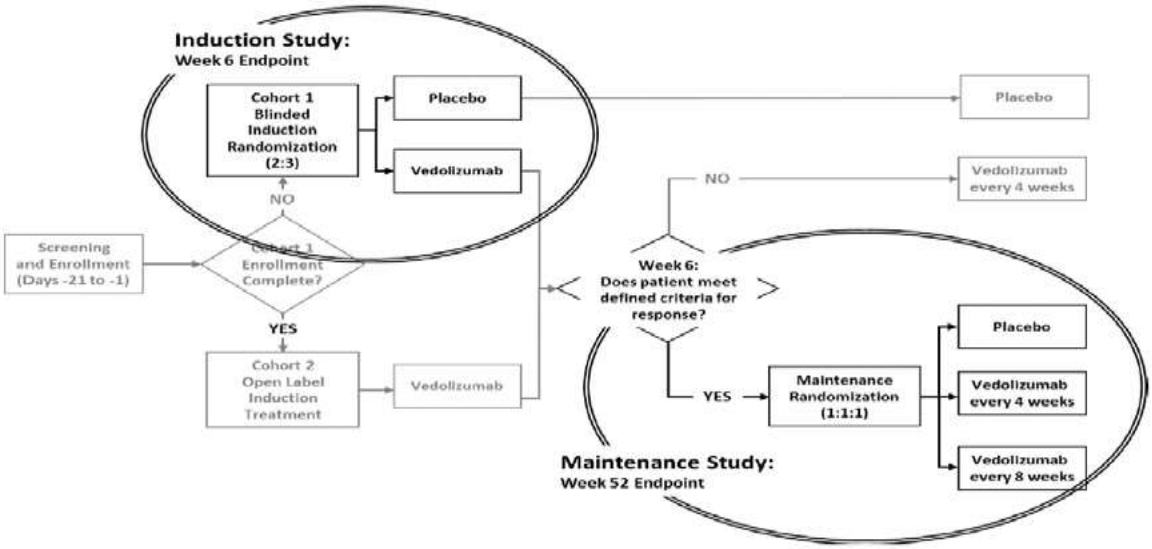
in the manufacturer's response to clarification[4] (question A20) with respect to the exclusion of other studies of vedolizumab for UC. The ERG note that one Phase I double-blind, placebo-controlled study[15] and two randomised placebo-controlled Phase II dose-ranging studies[16-18] were excluded as they used various dosing regimens based on weight. In the Phase I study, Feagan *et al*.[15] used a single dose of vedolizumab (i.v.) 0.15mg/kg, 0.5mg/kg, 3mg/kg and vedolizumab subcutaneously (s.c.) 0.15/kg in 29 patients with moderate to severe UC. Parikh *et al*.[17,18] used vedolizumab (i.v.) 2mg/kg, 6mg/kg, 10mg/kg on days 1, 15, 29 and 85 after randomisation in 47 patients with mild to severe active UC. Feagan *et al*.[16] used vedolizumab (i.v.) 0.5mg/kg or 2mg/kg on day 1 and day 29 after randomisation in 181 patients with moderate to severe active UC. As the licensed indication for vedolizumab is based on a fixed dosing schedule (300mg at zero, two and six weeks and then every eight weeks thereafter[10]), the ERG agrees that these studies were appropriately excluded from the manufacturer's review.

For the systematic review and NMA of other biological therapies, nine potential studies were excluded. As noted in the manufacturer's response to clarification[4] (question A20), three studies[20-22] were not considered to be RCTs, whereas one study[23] was an extension of the ACT1/2[24] trial and did not include any suitable time points for analysis. Two studies[17,18,25] did not report the outcome of interest. One study[25] only reported efficacy outcomes for colectomy and had no safety endpoints, whereas Parikh *et al*.[17,18] used a partial Mayo scores to report outcomes; this was not an endpoint for the NMA. Finally, two studies[26,27] had no placebo comparator arm to link to other studies in the network. The ERG agrees that the design and context of these studies were not suitable for inclusion in the NMA.

- Main evidence (pivotal study: GEMINI1 trial)[8]

The MS[1] included one Phase III, multicentre, randomised, double-blind, placebo-controlled trial designed to evaluate the efficacy and safety of vedolizumab as induction treatment (Weeks 0 to 6) and maintenance treatment (Weeks 7 to 52) in patients with moderately to severely active UC who had an inadequate response to, loss of response to, or intolerance to immunomodulators or TNF-α antagonists. An overview of the induction and maintenance phases in the GEMINI1 trial is provided in Figure 1. It is noteworthy that although the study was designed against placebo, conventional therapies (5-ASAs - corticosteroids, immunomodulators, antibiotics, probiotics, and antidiarrheals) were concomitantly administered to patients in both treatment groups. As noted in the European Medicines Agency (EMA) European Public Assessment Report[7] (EPAR), the lack of an anti-TNF-α compound comparator group represents a limitation of the study.

Figure 1: Overview of the induction and maintenance phase in the GEMINI1 trial[7]



The GEMINI1 study was conducted at 211 medical centres in 34 countries (including 2 sites in the UK) from 2008 to 2012. Of the 211 sites, enrolment at 13 sites in India was permanently discontinued at a country level due to concerns for patient safety. This arose from a parallel CD study in which SAEs led to 2 deaths; further details are provided in the supplementary appendix to Feagan *et al*.[8]

Patients eligible for inclusion in GEMINI1 were required to be adults (aged 18 to 80 years) with moderate to severe active UC as determined by a Mayo score of 6 to 12 with an endoscopic subscore ≥2 within 7 days prior to the first dose of study drug. Participants were also required to have evidence of UC extending proximal to the rectum (≥15 cm of involved colon) and an inadequate response to, loss of response to, or intolerance of at least 1 of the following: azathioprine (≥1.5mg/kg) 6-MP (≥0.75mg/kg) or anti-TNF-α (infliximab). The key exclusion criteria related to the exclusion of individuals who received anti-TNF-α therapy within 60 days prior to enrolment, or ciclosporin, thalidomide, or investigational agents within 30 days prior to enrolment, or if they had been treated previously with vedolizumab, natalizumab, efalizumab, or rituximab. Additional exclusion criteria included toxic megacolon, abdominal abscess, symptomatic colonic stricture, stoma, a history of colectomy, an increased risk of infectious complications (e.g., recent pyogenic infection, enteric pathogens detected on stool analysis, active or latent TB, immunodeficiency, hepatitis B or C, or recent live vaccination), clinically meaningful laboratory abnormalities, pregnancy or lactation, unstable or uncontrolled medical disorders, anticipated need for major surgery, colonic dysplasia or adenomas, and malignant neoplasms. A summary of the study design and population characteristics is presented in Table 4.

Table 4: Characteristics of GEMINI1 study (see MS[1] pages 75-76 and Feagan *et al.*[8])

Study	Location (sites)	Design	Population	Interventions (n=randomised)	Comparator	Primary outcome measures	Duration
GEMINI1 (Study C13006; NCT 00783718)[8]	211 medical centres in 34 countries (including 2 sites in the UK)	Phase III randomised, double-blind, placebo-controlled, induction and maintenance trial	Patients aged 18 to 80 years with moderate to severe active UC (defined as Mayo score ≥6 and an endoscopic subscore of ≥2 despite treatment with one or more of: glucocorticoids, immunosuppressive medications or TNF-α antagonists	Induction phase Vedolizumab (i.v.) 300mg at Week 0 and 2 (Cohort 1, n=225) Maintenance phase Vedolizumab (i.v.) 300mg every 8 weeks (n=122) Vedolizumab (i.v.) 300mg every 4 weeks (n=125)	Induction phase Placebo (i.v.) at Week 0 and 2 (Cohort 1, n=149) Maintenance phase Placebo (i.v.) every 4 weeks (n=126)	Induction Phase Clinical response[a] at Week 6 Maintenance Phase Clinical remission[b] at Week 52[c]	Induction phase 6 Weeks Maintenance phase 46 Weeks

[a] Defined as a reduction in complete Mayo score of ≥3 points and ≥30% from baseline, along with a decrease in rectal bleeding subscore of ≥1 points or an absolute rectal bleeding subscore of ≤1 point.
[b] Defined as a complete Mayo score of ≤2 points and no individual subscore >1 point) at Week 52
[c] Measurement point after original induction randomisation

In the induction study, 374 patients were randomised in a 3:2 ratio to receive 300mg vedolizumab i.v. or placebo (as saline) at Week 0 and Week 2 (Cohort 1), with two stratification factors: (1) concomitant use or non-use of glucocorticoids and (2) by concomitant use or non-use of immunosuppressive agents or prior use or non-use of anti-TNF-α agents. The proportion of patients with prior anti-TNF-α exposure was limited to 50%. In order to fulfil sample size requirements for the maintenance study, an additional 521 patients were enrolled in an open-label group (Cohort 2), which received the same active induction regimen given in the blinded study (Cohort 1).

In the maintenance study, patients from both cohorts (Cohort 1 and Cohort 2) who had a clinical response (defined as a reduction in the Mayo Clinic score of at least 3 points and a decrease of at least 30% from the baseline score, with a decrease of at least 1 point on the rectal bleeding subscale or an absolute rectal bleeding score of ≤1) to vedolizumab at Week 6 (n=373) were randomised in a 1:1:1 ratio to double-blind treatment with 300mg vedolizumab i.v. every 8 weeks (with placebo administered every other visit to preserve blinding), 300mg vedolizumab i.v. every 4 weeks or placebo every 4 weeks for up to 52 weeks. 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. Patients in the induction study who did not achieve clinical response at Week 6 continued to receive vedolizumab every 4 weeks and were followed to Week 52. Patients who received placebo in the induction phase continued to receive placebo and were followed up in a similar fashion.

The primary outcome in the induction trial phase was clinical response at Week 6, as defined above. Secondary endpoints included clinical remission (defined as complete Mayo Clinic score of ≤2 points and no individual subscore >1 point) and mucosal healing (defined as an endoscopic subscore of ≤1 point). The primary endpoint for the maintenance trial phase was clinical remission at Week 52. Secondary measures included durable clinical response (response at Weeks 6 and 52), durable clinical remission (remission at Weeks 6 and 52), mucosal healing at Week 52 and glucocorticoid-free remission at Week 52 in patients receiving glucocorticoids at baseline.

- *Ongoing studies of vedolizumab (MS[1] page 23)*

As reported in the MS[1] (page 23), 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 10 September 2014) which are planned for completion in the next 5 years is presented in Table 5.

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
NCT02039505[13] Sponsor: Takeda	Interventional, Phase III, multicentre, randomised, double-blinded, placebo-controlled, parallel-group study	To examine the efficacy, safety, and pharmacokinetics of 300mg vedolizumab i.v. infusion in induction and maintenance therapy in Japanese patients with moderately or severely active UC	Duration 60 weeks Estimated enrolment of 278 patients	Start date: March 2014 Expected end date: April 2018
GEMINI LTS NCT00790933 (C13008)[14] 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 manufacturer up to July 2012

4.2.2 Details of relevant studies not included in the submission

The ERG and their clinical advisors were satisfied that all relevant studies were included in the MS.[1] Repeat searches using the manufacturer's search terms were undertaken, although the ERG was not able to sift through the search results due to time constraints.

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

The validity assessment tool used to appraise the GEMINI1 trial in the MS[1] (page 95) is based on the quality assessment criteria suggested by NICE.[11] In response to a request for clarification[4] (question A19), the manufacturer confirmed that two reviewers carried out the quality assessment of the study. The completed validity assessment tool for the GEMINI1 trial, as reported in the MS,[1] is reproduced (with minor changes made by the ERG) in Table 6.

Table 6: Manufacturer's quality assessment results for included RCTs (page 95, MS)[1]

Quality assessment criteria	Trial
	GEMINI1
Was randomisation carried out appropriately?	Yes
Was the concealment of treatment allocation adequate?	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes
Were there any unexpected imbalances in drop-outs between groups?	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	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. No data were imputed for missing values in the vedolizumab pharmacokinetic or pharmacodynamics datasets.

The MS[1] (pages 79-80) states that randomisation was performed using a computer generated randomisation schedule; allocation concealment was done centrally (no further details were provided in the MS[1]) and participants and investigators were blinded to treatment allocation (double-blind). The ERG acknowledges that adequate methods of randomisation, allocation concealment and blinding were used in the conduct of GEMINI1.

In the GEMINI1 trial, patients were predominantly white (82.0%) with a mean age of 40.3 years, mean body weight of 73.4kg and male (58.7%). Mean duration of disease was 6.9 years and patients had a mean Mayo score of 8.6. Concomitant medications for UC included glucocorticoids only (37.1%), immunosuppressants only (17.8%), glucocorticoids and immunosuppressants (16.6%) and no glucocorticoids or immunosuppressants (28.5%). Approximately 48% of patients had received prior anti-TNF-α treatment. The primary published paper,[8] Clinical Study Report (CSR)[28] and the MS[1] (pages 84-85) suggest that no relevant differences in baseline demographic or clinical characteristics were observed between the treatment groups in the induction phase (vedolizumab and placebo) or in the maintenance phase (vedolizumab 4-weekly, vedolizumab 8-weekly and placebo) in the GEMINI1 trial (*p*-values were not provided). However, as noted in the MS[1] (page 82), the US Food and Drug Administration (FDA) briefing document for the Centre for Drug Evaluation and Research,[29] and the manufacturer's clarification response[4] (question A24), the US population (27% [238/895] of the total population)[30] varied from the non-US population, both in terms of entry criteria and with respect to the allowance of concomitant immunosuppressant use. 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. In addition, in the US patients were required to discontinue immunomodulators at Week 6, whilst those outside the US could continue concomitant immunomodulator therapy over the course of the trial. It is unclear to the ERG how this might have impacted on the study results.

Whilst all study withdrawals were adequately described and all patients were accounted for, 6% [57/895] of the total population in the induction phase prematurely discontinued from the study (vedolizumab Cohort 1, 3% [7/225], vedolizumab Cohort 2, 7% [36/521] and placebo, 9% [14/149]). The main reason for discontinuation was lack of efficacy (further details are provided in Section 4.2.4.2). It is noteworthy that Schulz and Grimes[31] suggest that a rate of less than 5% loss will lead to little bias. As such, the ERG acknowledges that attrition bias should be low in the induction phase of the GEMINI1 trial. In contrast however, during the maintenance phase, 44% [164/373] of the total ITT population (i.e. responders to vedolizumab during the induction phase that were randomised to maintenance therapy at Week 6) prematurely discontinued from the study (vedolizumab every 8 weeks, 37% [45/122], vedolizumab every 4 weeks, 33% [41/125] and placebo, 62% [78/126]). In general, the validity of a study may be threatened if attrition is more than 20%.[31] The main reasons for discontinuation in the vedolizumab and placebo groups were due to lack of efficacy or disease-related AEs. 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, the disproportionate discontinuation has the potential to impact on the maintenance study results, posing a serious threat to validity. In the GEMINI1 trial, efficacy analyses were conducted using the ITT approach whereby patients who withdrew prematurely were considered as treatment failures. Further details are provided in the MS[1] (pages 88-91).

It should also be noted that all subgroup analyses (i.e. patients with prior anti-TNF-α failure and those with no prior anti-TNF-α exposure, impact of concomitant therapy, correlation between partial and complete Mayo score; MS[1] pages 86-87) were exploratory and the study was not powered for these assessments.

4.2.4 Summary and critique of results

This section presents the results (as reported by the manufacturer) from the GEMINI1 trial, which forms the pivotal evidence in the MS[1] for the efficacy and safety of vedolizumab in the induction and maintenance treatment of patients with moderate to severe active UC. Additional information, not reported in the MS,[1] was provided by the manufacturer in their response to the clarification questions raised by the ERG.[4] Where applicable, data have been re-tabulated by the ERG to ensure clarity.

4.2.4.1 Efficacy

- *Induction phase*

As presented in Table 7, patients treated with vedolizumab had significantly greater rates of clinical response (primary outcome), clinical remission and mucosal healing at 6 weeks compared with placebo. The 6-week clinical response rate was 47.1% in the vedolizumab group compared with 25.5% in the placebo group (difference after adjustment for stratification factors, 21.7 percentage points; 95% c.i.: 11.6 to 31.7; $p<0.001$). Clinical remission occurred in 16.9% of patients in the vedolizumab group compared with 5.4% in the placebo group ($p=0.001$). Rates of mucosal healing were 40.9% with vedolizumab and 24.8% in the placebo group ($p=0.001$). The ERG notes that the FDA briefing document for the Centre for Drug Evaluation and Research[29] states that *'…to establish "mucosal healing" requires histologic data…'* which were not reported by the manufacturer. Details concerning how mucosal healing was confirmed were not reported in MS[1] or Feagan *et al*.[8]

Subgroup analyses

Exploratory subgroup analyses (Table 7) restricted to only those patients who were anti-TNF-α naïve, defined as patients with no prior exposure to TNF-antagonists, showed that treatment with vedolizumab improved clinical response (53.1% versus 26.3%), remission rates (23.1% versus 6.6%) and mucosal healing (49.2% versus 25%) compared with placebo at 6 weeks. *P*-values were not provided as the MS[1] stated that multiple testing adjustments were not made. However, patients with prior anti-TNF-α failure, defined as patients who failed, lost response to, or were intolerant of TNF-antagonists, had lower clinical response rates (39.0% versus 20.6% for vedolizumab and placebo, respectively), lower remission rates (9.8% versus 3.2% for vedolizumab and placebo, respectively) and lower mucosal healing rates (30.5% versus 20.6% for vedolizumab and placebo, respectively) at Week 6 than anti-TNF-α naïve patients (see Table 7).

An additional *post hoc* 'delayed responder' exploratory analysis was conducted in patients who failed to demonstrate clinical response at Week 6 in the induction phase. These patients continued on vedolizumab 300mg every 4 weeks for 46 weeks and were analysed at Weeks 10 and 14 against the placebo group. Clinical response using partial Mayo scores was achieved at Week 10 and Week 14 in vedolizumab-treated patients (Week 10: 32% [102/322], Week 14: 39% [126/322], respectively) compared with placebo patients (Week 10: 15% [12/82], Week 14: 21% [17/82], respectively). The ERG notes that these results must be interpreted with caution given that an increase in dosing frequency in patients who failed to achieve clinical response by Week 6 was not studied in a randomised manner within the GEMINI1 study.

Table 7: GEMINI1 efficacy endpoints at week 6 in induction study (MS[1] page 100, CSR[28] and Feagan *et al.*[8])

Endpoint	Vedolizumab 300mg i.v. at weeks 0 and 2	Placebo	Percentage difference from placebo[i]	95% c.i.	*p*-value
ITT patients[a]	*n=225*	*n=149*			
Clinical response[b], No. (%) (primary end point)	106 (47.1)	38 (25.5)	21.7	11.6 to 31.7	<0.001
Clinical remission[c], No. (%)	38 (16.9)	8 (5.4)	11.5	4.7 to 18.3	0.001
Mucosal healing[d], No. (%)	92 (40.9)	37 (24.8)	16.1	6.4 to 25.9	0.001
Non-ITT patients[e]	*n=521*				
Clinical response[b], No. (%)	231 (44.3)				NR
Anti-TNF-α naïve patients[f]	*n=130*	*n=76*			
Clinical response[b], No. (%)	69 (53.1)	20 (26.3)	26.8	13.7 to 39.9	NR[h]
Clinical remission[c], No. (%)	30 (23.1)	5 (6.6)	16.5	2.4 to 30.2	NR[h]
Mucosal healing[d], No. (%)	64 (49.2)	19 (25.0)	24.2	11.2 to 37.2	NR[h]
Prior anti-TNF-α failure patients[g]	*n=82*	*n=63*			
Clinical response[b], No. (%)	32 (39.0)	13 (20.6)	18.4	3.9 to 32.9	NR[h]
Clinical remission[c], No. (%)	8 (9.8)	2 (3.2)	6.6	-9.8 to 22.8	NR[h]
Mucosal healing[d], No. (%)	25 (30.5)	13 (20.6)	9.9	-4.3 to 24.0	NR[h]

c.i. - confidence interval; NR - not reported

[a] Patients with insufficient diary entries were imputed as not achieving clinical response

[b] Clinical response is defined as reduction in complete Mayo score of ≥ 3 points and ≥ 30% from baseline (Week 0) with an accompanying decrease in rectal bleeding subscore of ≥ 1 point or absolute rectal bleeding subscore of ≤ 1 point.

[c] Clinical remission is defined as complete Mayo score of ≤ 2 points and no individual subscore > 1 point.

[d] Mucosal healing is defined as Mayo endoscopic subscore of ≤ 1 point.

[e] Patients in Cohort 2 received open-label vedolizumab induction treatment

[f] Patients without prior exposure to TNF-antagonist treatment

[g] Patients with prior inadequate response, loss of response, or intolerance to TNF-antagonist treatment

[h] Although these endpoints were pre-specified, *p*-values are not provided because multiple testing adjustments were not made.

[i] Percentage differences were adjusted for two stratification factors: concomitant use or non-use of glucocorticoids, and concomitant use or non-use of immunosuppressive agents or prior use or non-use of TNF-antagonists.

Exploratory analysis in the maintenance phase comparing partial and complete Mayo scores (MS[1] page 101) showed high agreement between both Mayo scores in all 374 patients included in the ITT population, with Pearson correlation coefficients of 0.95 (95% Cl: 0.95 to 0.96) at baseline and 0.98 (95% Cl: 0.97 to 0.98) at the end of the induction phase (Week 6). Further details are provided in the MS[1] (page 101).

As noted in Section 4.2.3, all subgroup analyses were exploratory and the study was not powered for these assessments hence these should be viewed with caution.

- *Maintenance phase*

As shown in Table 8, patients receiving vedolizumab maintenance therapy either every 4 weeks or every 8 weeks were associated with significantly higher clinical remission rates at 52 weeks (primary outcome), durable clinical response (at Weeks 6 and 52), durable clinical remission (at Week 6 and Week 52), mucosal healing (at Week 52) and glucocorticoid-free remission (at Week 52) compared with placebo. At Week 52, clinical remission occurred in 41.8% of patients who continued to receive vedolizumab every 8 weeks and in 44.8% of patients who continued to receive vedolizumab every 4 weeks, but in only 15.9% of patients who were re-randomised to placebo (difference after adjustment for stratification factors, 26.1 percentage points for vedolizumab every 8-weeks vs placebo; 95% c.i.: 14.9 to 37.2; $p<0.001$ and 29.1 percentage points for vedolizumab every 4 weeks vs placebo; 95% c.i.: 17.9 to 40.4; $p<0.001$]). Durable clinical response occurred in 56.6% of patients in the vedolizumab 8-weekly group, 52% in the vedolizumab 300mg 4-weekly group, and 23.8% in placebo group; $p<0.001$ in both comparisons. Durable clinical remission occurred in 20.5% of patients in the vedolizumab 8-weekly group, 24% in vedolizumab 4-weekly, and 8.7% in placebo group; $p=0.008$ and $p<0.001$, respectively. Vedolizumab was also associated with higher mucosal healing rates (51.6% in the vedolizumab 8-weekly group, 56% in the vedolizumab 4-weekly group, and 19.8% in the placebo group; $p<0.001$ in both groups). The proportion of patients who were glucocorticoid-free at 52 weeks were significantly higher in those treated with vedolizumab than in those who received placebo (31.4% of the vedolizumab 8-weekly group, 45.2% of the vedolizumab 4-weekly group, and 13.9% of the placebo group; $p=0.01$ and $p<0.001$, respectively). However, histologic data were not reported to substantiate the claim for mucosal healing. The clinical relevance of these data is therefore unclear. In addition, the definition for CSF remission, as reported in the MS,[1] does not define a pre-specified minimum duration of time over which a patient is required to be corticosteroid-free, which is necessary to demonstrate the clinical relevance of the endpoint.[7] No clear differences in efficacy were observed between the two vedolizumab dosage groups. However, the ERG notes that the study was not powered to directly compare the 4-weekly and 8-weekly doses of vedolizumab.

Table 8: GEMINI1 efficacy endpoints in maintenance study (MS[1] pages 102-103, CSR[28] and Feagan *et al.*[8])

Study Endpoint	Vedolizumab 300mg Every 8weeks	Vedolizumab 300mg Every 4weeks	Placebo	Between group percentage difference[i]			
				Vedolizumab every 8weeks vs placebo (95% c.i.)	*p*-value	Vedolizumab every 4weeks vs placebo (95% c.i.)	*p*-value
ITT patients[a]	*n=122*	*n=125*	*n=126*				
Clinical remission[b] at Wk, 52, No. (%)	51 (41.8)	56 (44.8)	20 (15.9)	26.1 (14.9 to 37.2)	<0.001	29.1 (17.9 to 40.4)	<0.001
Durable clinical response[c], No. (%)	69 (56.6)	65 (52.0)	30 (23.8)	32.8 (20.8 to 44.7)	<0.001	28.5 (16.7 to 40.3)	<0.001
Durable clinical remission[d], No. (%)	25 (20.5)	30 (24.0)	11 (8.7)	11.8 (3.1 to 20.5)	0.008	15.3 (6.2 to 24.4)	0.001
Mucosal healing at Wk 52[e], No. (%)	63 (51.6)	70 (56.0)	25 (19.8)	32.0 (20.3 to 43.8)	<0.001	36.3 (24.4 to 48.3)	<0.001
Corticosteroid-free clinical remission at Wk 52[f], No. (%)	22/70 (31.4)	33/73 (45.2)	10/72 (13.9)	17.6 (3.9 to 31.3)	0.01	31.4 (16.6 to 46.2)	<0.001
Anti-TNF-α naïve patients[g]	*n=72*	*n=73*	*n=79*				
Clinical remission[b], No. (%)	33 (45.8)	35 (47.9)	15 (19.0)	26.8 (12.4 to 41.2)	NR[j]	29.0 (14.6 to 43.3)	NR[j]
Durable clinical response[c], No. (%)	47 (65.3)	41 (56.2)	21 (26.6)	38.7 (24.0 to 53.4)	NR[j]	29.6 (14.6 to 44.6)	NR[j]
Durable clinical remission[d], No. (%)	16 (22.2)	21 (28.8)	10 (12.7)	9.6 (-2.5 to 21.6)	NR[j]	16.1 (3.4 to 28.8)	NR[j]
Mucosal healing[e], No. (%)	43 (59.7)	44 (60.3)	19 (24.1)	35.7 (20.9 to 50.4)	NR[j]	36.2 (21.6 to 50.9)	NR[j]
Corticosteroid-free clinical remission[f] in patients with prior anti-TNF-α failure and using corticosteroids at baseline, n	*n=39*	*n=44*	*n=43*				
Achieving corticosteroid free clinical remission, No. (%)	14 (35.9)	23 (52.3)	8 (18.6)	17.3 (-1.7 to 36.3)	NR[j]	33.7 (14.9 to 52.5)	NR[j]
Prior Anti-TNF-α failure patients[h]	*n=43*	*n=40*	*n=38*				
Clinical remission[b], No. (%)	16 (37.2)	14 (35.0)	2 (5.3)	31.9 (10.3 to 51.4)	NR[j]	29.7 (7.4 to 49.4)	NR[j]
Durable clinical response[c], No. (%)	20 (46.5)	17 (42.5)	6 (15.8)	30.7 (11.8 to 49.6)	NR[j]	26.7 (7.5 to 45.9)	NR[j]
Durable clinical remission[d], No. (%)	9 (20.9)	5 (12.5)	1 (2.6)	18.3 (-3.8 to 38.9)	NR[j]	9.9 (-13.0 to 31.5)	NR[j]
Mucosal healing[e], No. (%)	18 (41.9)	19 (47.5)	3 (7.9)	34.0 (12.6 to 53.2)	NR[j]	39.6 (18.1 to 58.5)	NR[j]
Corticosteroid-free clinical remission[f] in patients with prior anti-TNF-α failure and using corticosteroids at baseline, n	*n=26*	*n=19*	*n=23*				
Achieving corticosteroid free clinical remission, No. (%)	6 (23.1)	6 (31.6)	1 (4.3)	18.7 (-9.4 to 45.2)	NR[j]	27.2 (-3.6 to 53.8)	NR[j]

NR - not reported; Wk - week

[a] Patients with insufficient diary entries were imputed as not achieving clinical response
[b] Clinical remission is defined as complete Mayo score of ≤ 2 points and no individual subscore >1 point at Week 52
[c] Durable clinical response is defined as reduction in complete Mayo score of ≥3 points and ≥ 30% from baseline (Week 0) with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point at both Weeks 6 and 52.
[d] Durable clinical remission is defined as complete Mayo score of ≤ 2 points and no individual subscore >1 point at both Weeks 6 and 52.
[e] Mucosal healing is defined as Mayo endoscopic subscore of ≤ 1 point.
[f] Corticosteroid-free clinical remission is defined as patients using oral corticosteroids at baseline (Week 0) who have discontinued corticosteroids and are in clinical remission at Week 52.
[g] Patients without prior exposure to TNF-α antagonist treatment
[h] Patients with prior inadequate response, loss of response, or intolerance to TNF-α antagonist treatment. A small number of patients (9 placebo, 7 vedolizumab every 8-weeks, and 12 vedolizumab every 4-weeks) had prior anti-TNF-α exposure without documented evidence of TNF-α antagonist failure; these patients are not included in this table.
[i] Between-group differences in percentage points were adjusted for three stratification factors: cohort, concomitant use or non-use of glucocorticoids, and concomitant use or non-use of immunosuppressive agents or prior use or non-use of TNF-α antagonists.
[j] Although these endpoints were pre-specified, *p*-values are not provided because multiple testing adjustments were not made.

Data on colectomy, surgery and rate of hospitalisation and disease activity were not reported in the MS;[1] however, these data were provided after a clarification request[4] (questions A30 and A31). Overall, in the maintenance phase, surgery (which in all cases was colectomy) was rare in all groups. In the placebo group 2/126 (1.6%, 95% c.i.: 0.2, 5.6), in the vedolizumab 8-weekly group 1/122 (0.8, 95% c.i.: 0.0, 4.5) and in the vedolizumab 4-weekly group 0/125 (0.0%, 95% c.i.: 0.0, 2.9) underwent colectomy, respectively (*p*-values not reported). Similarly rates of hospitalisation were low, with more patients in the placebo group being hospitalised 10/126 (7.9%, 95% c.i. 3.9, 14.1) compared with the vedolizumab 8-weekly group 3/122 (2.5%, 95% c.i.:0.5, 7.0) and the vedolizumab 4-weekly group 4/125 (3.2%, 95% c.i.: 0.9, 8.0), respectively (*p*-values not reported). In addition, disease activity was measured by several markers including disease worsening based on Mayo score and change in partial Mayo score over time, faecal calprotectin level and prednisone dosing. In all cases, vedolizumab disease activity was reduced over the course of the maintenance phase.

Subgroup analyses

A pre-specified subgroup analysis was performed in patients with prior anti-TNF-α failure and in patients who were anti-TNF-α naïve. Regardless of prior anti-TNF-α treatment status, the outcomes of clinical remission, durable clinical response, mucosal healing, durable clinical remission and CSF remission were greater in 8-weekly and 4-weekly vedolizumab-treated patients than placebo-treated patients (*p*-values were not provided as the manufacturer stated that 'multiple testing adjustments were not made, see Table 8). The ERG notes that the numbers of patients included in the analyses of durable clinical remission and CSF remission were small, therefore the interpretation of these subgroup results is limited.

It was reported in the MS[1] that the efficacy of vedolizumab maintenance treatment was not substantively affected by concomitant use of glucocorticoids or immunosuppressants. Further details are provided in the manufacturer response to clarification[4] (question A25). In addition, vedolizumab-treated patients had greater improvements in faecal calprotectin concentration compared with placebo-treated patients. The proportion of patients with faecal calprotectin concentrations >500mcg/g at Week 52 was 36% for placebo, 15% for vedolizumab every 8 weeks, and 21% for vedolizumab every 4 weeks (*p*-values were not reported). Furthermore, the manufacturer's clarification response[4] (question A22) stated that data on time to disease worsening and time to treatment failure were not estimable for any of the treatment groups as events were censored at Week 52 (the end of the GEMINI1 study) and patients would not be receiving the same treatment subsequently.[4] Overall, the number of patients censored at Week 52 were 67% in placebo, 83% in the vedolizumab 8-weekly group and 85% in the vedolizumab 4-weekly group.

As noted in the manufacturers clarification response[4] (question A29), subgroup analyses for clinical remission at Week 6 (induction ITT population) and at Week 52 (maintenance ITT population) showed no statistically significant effects of age, gender, race, duration from UC diagnosis to first dose, geographical region, baseline disease activity, baseline faecal calprotectin and disease localisation. The treatment benefit of vedolizumab was observed in all subgroups and for both dosing regimen groups in the maintenance phase. Further details are provided in response to clarification question A29.[4] The ERG recognises that all subgroup analyses were exploratory and the studies were not powered for these assessments, hence these should be interpreted with caution.

Patient-Reported Outcomes in GEMINI1

In the induction phase, 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) showed statistically significantly greater improvements with vedolizumab compared with placebo. A summary of these results are provided in Table 9 (further details are provided in the MS,[1] page 107).

Table 9: Changes in HRQoL from baseline at week 6 of UC induction therapy in GEMINI1 (MS,[1] page 107)

HRQoL measures	Mean difference in adjusted change from baseline vs placebo[b]	95% c.i.	*p*-value
IBDQ Total Score[a]	18.0[c]	11.0 to 24.9	NR
SF-36 Physical Component Summary[a]	2.7[c]	1.3 to 4.1	NR
SF-36 Mental Component Summary[a]	4.4[c]	2.5 to 6.4	NR
EQ-5D Score[a]	–0.5[c]	–0.7 to –0.2	NR
EQ-5D VAS Score[a]	9.6[c]	5.8 to 13.5	NR

NR- not reported in the MS[1]
[a] Higher IBDQ, SF-36, and EQ-5D VAS scores indicate improvements in HRQOL; lower EQ-5D scores indicate improvements in HRQOL.
[b] Difference = adjusted mean change for vedolizumab – adjusted mean change for placebo.
[c] Statistically significant result.

In the maintenance phase, a generally greater HRQoL improvement was observed in patients treated with vedolizumab in both the 4-weekly and 8-weekly groups compared with placebo. HRQoL, measured by the difference in mean adjusted change from baseline versus placebo (defined as adjusted mean change for vedolizumab – adjusted mean change in placebo), was statistically significant in both vedolizumab treatment groups when assessed by IBDQ score, SF-36 physical and mental component, and EQ-5D VAS at 52 weeks. Similarly in a *post hoc* analysis evaluating the Week 52 last observation carried forward (LOCF) data, patients in both vedolizumab treatment groups achieved statistically significant improvements in the total IBDQ score, SF-36 physical and

mental component, and EQ-5D VAS compared with the placebo group. Although improvements were observed in the HRQoL instruments at 30 weeks, these were not statistically significant. A summary of these results are provided in Table 10 (further details are provided in MS[1] pages 109-111).

Table 10: Patient reported outcomes in UC in GEMINI1: Changes from baseline by study visit in the maintenance phase (MS,[1] page 109)

	Vedolizumab every 8weeks			Vedolizumab every 4weeks		
HRQoL measure	Mean difference in adjusted change from baseline vs placebo[a]	95% c.i.	*p*-value	Mean difference in adjusted change from baseline vs placebo[a]	95% c.i.	*p*-value
Week 30 IBDQ	17.1	(6.6 to 27.6)	NR	11.0	(0.8 to 21.2)	NR
Week 52 IBDQ	26.1[b]	(15.2 to 36.9)	NR	25.7[b]	(15.1 to 36.3)	NR
LOCF Week 52 IBDQ[c]	21.1[b]	(11.8 to 30.4)	NR	21.6[b]	(12.4 to 30.9)	NR
Week 30 SF-36 Physical Component Summary	1.0	(–1.0 to 3.0)	NR	1.3	(–0.6 to 3.2)	NR
Week 52 SF-36 Physical Component Summary	4.7[b]	(2.3 to 7.2)	NR	3.7[b]	(1.3 to 6.1)	NR
Week 52 LOCF SF-36 Physical Component Summary[c]	3.3[b]	(1.5 to 5.2)	NR	2.8[b]	(1.0 to 4.6)	NR
Week 30 Mental Component Summary	4.4	(1.5 to 7.3)	NR	2.5	(–0.3 to 5.4)	NR
Week 52 Mental Component Summary	6.6[b]	(3.4 to 9.8)	NR	6.0[b]	(2.9 to 9.2)	NR
Week 52 LOCF Mental Component Summary[c]	4.7[b]	(2.3 to 7.2)	NR	4.8[b]	(2.3 to 7.2)	NR
Week 30 EQ-5D Score	–0.3	(–0.7 to 0.1)	NR	–0.0	(–0.4 to 0.4)	NR
Week 52 EQ-5D Score	–0.6	(–1.1 to –0.1)	NR	–0.6	(–1.1 to –0.1)	NR
Week 52 LOCF EQ-5D Score[c]	–0.4	(–0.8 to –0.1)	NR	–0.5	(–0.8 to –0.1)	NR
Week 30 EQ-5D VAS Score	6.3	(1.1 to 11.5)	NR	5.4	(0.3 to 10.4)	NR
Week 52 EQ-5D VAS Score	12.5[b]	(6.7 to 18.4)	NR	11.0[b]	(5.2 to 16.7)	NR
Week 52 LOCF EQ-5D VAS Score[c]	9.3[b]	(4.6 to 14.0)	NR	9.7[b]	(5.0 to 14.4)	NR

c.i. - confidence interval; NR- not reported in the MS[1]
[a] Difference = adjusted mean change for vedolizumab – adjusted mean change for placebo.
[b] Statistically significant result.
[c] *Post hoc* analysis.

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 UC available from the GEMINI1 trial. The manufacturer also provided supplementary supporting evidence on the use safety of vedolizumab in patients with CD.[1] Whilst separate AE searches for vedolizumab were not undertaken by the manufacturer, the manufacturer's clarification response[4] (question A15) suggests that all safety data (identified through internal databases) from all available Phase III trials in UC and CD are presented in the MS,[1] including interim results from a long-term safety study.[14]

The rates of discontinuation (including reasons for premature termination) for all participants in the induction phase of the GEMINI1 trial are presented in Table 11. Overall, 6% (57/895) of patients discontinued at the end of induction phase with no notable difference between the combined vedolizumab and placebo groups. In the ITT population, discontinuation due to AEs was reported in 3% (4/149) of placebo patients and none in the vedolizumab-treated patients. The ERG notes that the low numbers of discontinuation during this phase is reflected in the short duration of the 6-week induction phase.

Table 11: GEMINI1 patient disposition (induction phase) (MS,[1] page 97)

	Induction Phase ITT[a]		Non-ITT		
	Vedolizumab Cohort 1	Placebo	Vedolizumab Cohort 2[b]	Vedolizumab Combined	Total
Randomised/assigned, n	225	149	521	746	895
Study Populations, n (%)					
• Safety[c]	225 (100)	149 (100)	521 (100)	746 (100)	895 (100)
• Intention-to-Treat[d]	225 (100)	149 (100)	NR	225 (30)	374 (42)
• Per-Protocol[e]	215 (96)	138 (93)	NR	215 (29)	353 (39)
Completed Induction Phase, n (%)[f]	218 (97)	135 (91)	485 (93)	703 (94)	838 (94)
Discontinued	7 (3)	14 (9)	36 (7)	43 (6)	57 (6)
Adverse event	0	4 (3)	7 (1)	7 (<1)	11 (1)
Protocol violation(s)	1 (<1)	1 (<1)	6 (1)	7 (<1)	8 (<1)
Lack of efficacy	2 (<1)	5 (3)	14 (3)	16 (2)	21 (2)
Study terminated by sponsor	0	0	0	0	0
Withdrawal of consent	4 (2)	3 (2)	8 (2)	12 (2)	15 (2)
Lost to follow-up	0	1 (<1)	1 (<1)	1 (<1)	2 (<1)
Other	0	0	0	0	0

[a] All patients enrolled in Cohort 1 who were randomised to blinded induction treatment with vedolizumab or placebo.
[b] All patients enrolled in Cohort 2 who received open-label vedolizumab induction treatment.
[c] Safety population consisted of all patients who received any amount of study drug during the induction phase based on what they actually received.
[d] The ITT population consisted of all randomised patients who received any amount of blinded study drug during the induction phase based on what they were randomised to receive.
[e] The Per-Protocol population consisted of all patients without any major protocol deviations.
[f] These patients completed dosing at Weeks 0 and 2 and completed the pre-dose assessments at Week 6.

A summary of the rates of discontinuation (including reasons for premature termination) for all participants in the maintenance phase of the GEMINI1 trial are presented in Table 12. In general, a high rate of discontinuation was observed across all treatment groups (combined vedolizumab, 52% [324/620] versus combined placebo, 72% [197/275]). A statistical analysis comparing the rates of study discontinuation between the treatment groups is not reported in the MS.[1]

In the ITT population, a greater proportion of placebo-treated patients discontinued treatment compared against vedolizumab-treated patients (placebo, 62% [78/126] versus vedolizumab every 8 weeks, 37% [45/122]; vedolizumab every 4 weeks, 33% [41/125]). The most frequent reason for discontinuation across all of the ITT population treatment groups was lack of efficacy, which occurred in 48% [61/126] patients in the placebo group and less frequently in the vedolizumab groups (25% [31/122 in the 8-weekly group and 26% [33/125] in the 4-weekly group, respectively). Discontinuations due to AEs were twice as frequent in the placebo group (12% [15/126]) compared with the vedolizumab groups (6% [7/122] vedolizumab every 8 weeks and 5% [6/125] vedolizumab every 4 weeks, respectively). Similarly, drug-related AEs leading to discontinuation of treatment, as noted in the manufacturers clarification response[4] (question A27), were also more common in the placebo group (4% [5/126]) compared with the two vedolizumab groups (<1% [1/122] vedolizumab 8-weekly group and 2% [2/125] vedolizumab 4-weekly group, respectively). It is also noteworthy that the non-ITT placebo patients who were not exposed to vedolizumab in the induction phase had higher rates of discontinuation (80% [119/149], see MS,[1] pages 97-98) compared with the ITT placebo patients who originally received vedolizumab in the induction phase but were re-randomised to receive placebo in the maintenance phase (62% [62/126]). This may indicate that a beneficial effect of vedolizumab is maintained even after its use has been stopped.

In the GEMINI1 trial, adherence to study treatment, defined as the number of complete infusions (at least 75% of the infusion by volume) administered between the start (first dose of vedolizumab or placebo) and end dates (last known dose of vedolizumab or placebo) of study therapy, was high in all groups in the induction (100%) and maintenance phase (>99%). One patient in the non-ITT vedolizumab group experienced an infusion-related reaction (see manufacturer's clarification response,[4] question A26).

Table 12: GEMINI1 patient disposition (maintenance phase) (MS,[1] page 98, Feagan *et al*.)[8]

	Maintenance study ITT[a] (Responders to vedolizumab induction, randomised to maintenance at Week 6)			Non-ITT		Combined	
	Vedolizumab every 8 weeks n=122	Vedolizumab every 4 weeks n=125	Placebo n=126	Vedolizumab every 4 weeks (Week 6 non-responders) n=373	Placebo[b] (from Week 0) n=149	Vedolizumab n=620	Placebo n=275
Completed induction, n (%)	122 (100)	125 (100)	126 (100)	330 (88)	135 (91)	577 (93)	261 (95)
Randomised, n (%)	122 (100)	125 (100)	126 (100)	373 (100)	149 (100)	620 (100)	275 (100)
Randomised but not dosed	0	0	0	0	0	0	0
Safety population[d], n (%)	122 (100)	125 (100)	126 (100)	373 (100)	149 (100	620 (100)	275 (100)
Maintenance phase ITT population[a], n (%)	122 (100)	125 (100)	126 (100	NR	NR	247 (40)	126 (46)
Maintenance Phase Per-Protocol population[e]	117 (96)	121 (97)	121 (96)	NR	NR	238 (38)	121 (44)
Completed Maintenance[f]	77 (63)	84 (67)	48 (38)	135 (36)	30 (20)	296 (48)	78 (28)
Discontinued[g]	45 (37)	41 (33)	78 (62)	238 (64)	119 (80)	36 (6)	197 (72)
Adverse event	7 (6)	6 (5)	15 (12)	23 (6)	16 (11)	324 (52)	31 (11)
Protocol violations(s)	0	0	0	9 (2)	2 (1)	9 (1)	2 (<1)
Lack of efficacy	31 (25)	33 (26)	61 (48)	171 (46)	88 (59)	235 (38)	149 (54)
Study terminated by sponsor	0	0	0	0	0	0	0
Withdrawal of consent	5 (4)	2 (2)	2 (2)	32 (9)	9 (6)	39 (6)	11 (4)
Lost to follow-up	2 (2)	0	0	3 (<1)	4 (3)	5 (<1)	4 (1)
Other	0	0	0	0	0	0	0
Discontinued treatment, n (%)	1 (<1)	2 (2)	5 (4)	NR	NR	NR	NR
Enrolled into C13008 (GEMINI LTS)	108 (89)	112 (90)	113 (90)	230 (62)	112 (75)	450 (73)	225 (82)

NR, not reported.
[a] The maintenance phase ITT population consisted of all patients randomised at Week 6 (i.e. patients who received vedolizumab during the induction phase and were classified as responders at Week 6) who received any amount of blinded study drug during the maintenance phase, based on what they were randomised to receive.
[b] Patients who received placebo during the induction phase and continued to receive placebo during the maintenance phase.
[c] Patients who received vedolizumab in the induction phase but did not achieve clinical response at Week 6 and continued to receive vedolizumab every 4 weeks during the maintenance phase.
[d] The safety population consisted of all patients who received any amount of study drug at any time in the study (i.e. Week 0 through Week 50), based on what they actually received.
[e] The maintenance phase Per-Protocol population consisted of all maintenance phase ITT patients without any major protocol deviations.
[f] Completed study was defined as patients who completed the Week 52 analyses.
[g] Included patients who discontinued at any time during the study, even before Week 6

A summary of the key safety results from the GEMINI1 induction phase is presented in Table 13.[8] In general, the rates of AEs were similar between all treatment groups (Cohort 1, Cohort 2 and placebo); however, the proportion of patients experiencing SAEs in the vedolizumab group was lower compared with the placebo group (Cohort 1, 2%; Cohort 2, 4%; and placebo, 7%). However, these differences were not statistically significant (vedolizumab Cohort 1 versus placebo, *p*=0.06). Similarly, the number of serious infections in both the placebo and vedolizumab groups was low and explicit details on the nature of these infections were not provided in the MS[1] or Feagan *et al*.[8] No further details on types of SAEs were provided in the MS.[1]

Table 13: Key safety results: Induction trial data derived from Feagan *et al.*[8]

Adverse events, n (%)	Vedolizumab (300mg, i.v.)		Combined vedolizumab n=746	Placebo n=149	*p*-value[a]
	Cohort 1 (ITT) n=225	Cohort 2 (non-ITT) n=521			
Any adverse event	90 (40)	247 (47)	337 (45)	69 (46)	0.23
Serious adverse event	5 (2)	20 (4)	25 (3)	10 (7)	0.06
Common adverse event (≥5%)					
Headache	15 (7)	42 (8)	57 (8)	7 (5)	0.43
UC exacerbation	6 (3)	14 (3)	20 (3)	8 (5)	0.18
Infections					
Any infections	31 (14)	71 (14)	104 (14)	22 (15)	0.79
Serious[b] infection	1 (<1)	3 (<1)	4 (<1)	3 (2)	0.31
Infusion-related reaction	2 (<1)	1 (<1)	3 (<1)	1 (<1)	1.0
Malignant neoplasm	0	0	0	0	1.0

[a] *p*-values compare cohort 1 with the placebo group using the chi-square test (or Fisher exact test if any cell in the 2-by-2 table is ≤5).
[b] A serious infection was defined as an SAE of infection according to Medical Dictionary for Regulatory Activities (version 15) criteria. No further details were provided by the manufacturer

A summary of the main safety results from the maintenance phase of GEMINI1 trial is provided in Table 14. The most commonly occurring AEs in the combined vedolizumab group compared with the combined placebo group were nasopharyngitis (12.9% versus 9.5%), headache (12.9% versus 10.2%), arthralgia (9.0% versus 9.1%) and upper respiratory tract infection (8.4% versus 7.6%), respectively. However, *p*-values were not available for the comparisons between the combined vedolizumab groups and the combined placebo group. It should be noted that the combined placebo group included patients who were initially randomised to receive vedolizumab in the induction phase and were then re-randomised to receive placebo during the maintenance phase and also patients who were not exposed to vedolizumab.

The majority of infusion-related reactions in the induction and maintenance phase were mild to moderate in severity with only 3 cases (2 with detectable vedolizumab antibodies) resulting in drug discontinuation. No cases of anaphylaxis or serum sickness were observed.[8]

Table 14: Adverse events occurring in ≥5% and ≥10% of patients receiving vedolizumab in maintenance phase - GEMINI1 safety population* (MS[1] page 142, Feagan *et al.*[8])

Adverse events, n (%)	Vedolizumab every 8wks n=122	Vedolizumab every 4wks n=125	Placebo (ITT) (n=126)	Placebo induction (from Week 0) n=149	Vedolizumab every 4wks (Week 6 non-responders) n=373	Combined vedolizumab n=620	Combined placebo n=275	*p*-values[f]
Any adverse event	100 (82)	101 (81)	106 (84)	114 (77)	296 (79)	497 (80)	220 (80)	0.65, 0.49
Serious adverse event	10 (8)	11 (9)	20 (16)	17 (11)	56 (15)	77 (12)[a]	37 (13)[a]	0.06, 0.09
Adverse events in ≥5%								
Headache	NR	NR	NR	NR	NR	80 (12.9)	28 (10.2)	NR
Ulcerative colitis	NR	NR	NR	NR	NR	97 (15.6)	58 (21.1)	NR
Nasopharyngitis	NR	NR	NR	NR	NR	80 (12.9)	26 (9.5)	NR
Upper respiratory tract infection	NR	NR	NR	NR	NR	52 (8.4)	21 (7.6)	NR
Arthralgia	NR	NR	NR	NR	NR	56 (9.0)	25 (9.1)	NR
Nausea	NR	NR	NR	NR	NR	38 (6.1)	19 (6.9)	NR
Abdominal pain	NR	NR	NR	NR	NR	35 (5.6)	10 (3.6)	NR
Anaemia	NR	NR	NR	NR	NR	35 (5.6)	16 (5.8)	NR
Fatigue	NR	NR	NR	NR	NR	33 (5.3)	10 (3.6)	NR
Cough	NR	NR	NR	NR	NR	36 (5.8)	13 (4.7)	NR
Any serious adverse event	NR	NR	NR	NR	NR	77 (12.4)	37 (13.5)	NR
Any serious infection[b]	NR	NR	NR	NR	NR	12 (1.9)	8 (2.9)	NR
Any cancer	NR	NR	NR	NR	NR	1 (0.2)[e]	3 (1.1)[c]	NR
Common adverse event ≥10%								
Nasopharyngitis	19 (16)	18 (14)	15 (12)	11 (7)	43 (12)	80 (13)	26 (9)	0.40, 0.56
Upper respiratory tract infection	12 (10)	12 (10)	13 (10)	8 (5)	28 (8)	52 (8)	21 (8)	0.90, 0.85
Influenza	8 (7)	2 (2)	3 (2)	3 (2)	20 (5)	30 (5)	6 (2)	0.13, 1.0
Bronchitis	7 (6)	6 (5)	7 (6)	5 (3)	11 (3)	24 (4)	12 (4)	0.95, 0.79
Gastroenteritis	3 (2)	5 (4)	5 (4)	0	11 (3)	19 (3)	5 (2)	0.72, 1.0
Sinusitis	2 (2)	3 (2)	6 (5)	2 (1)	10 (3)	15 (2)	8 (3)	0.28, 0.50

Adverse events, n (%)	Vedolizumab every 8wks n=122	Vedolizumab every 4wks n=125	Placebo (ITT) (n=126)	Placebo induction (from Week 0) n=149	Vedolizumab every 4wks (Week 6 non-responders) n=373	Combined vedolizumab n=620	Combined placebo n=275	*p*-values[f]
Urinary tract infection	5 (4)	1 (<1)	6 (5)	5 (3)	8 (2)	14 (2)	11 (4)	1.0, 0.12
Infections								
Any infection	87 (71)	90 (72)	89 (71)	66 (44)	214 (57)	371 (60)	155 (56)	0.91, 0.81
Serious[b]	3 (2)	2 (2)	4 (3)	4 (3)	7 (2)	12 (2)[a]	8 (3)[a]	1.0, 0.68
Infusion-related reaction	7 (6)	10 (11)	2 (2)	1 (<1)	28 (8)	49 (8)	3 (1)	0.10, 0.02
Malignant neoplasm	1 (<1)[e]	NR	2 (2)[c]	1 (<1)[d]	NR	1(<1)[a]	3 (1)[a]	1.0, 0.50

NR - not reported

[a] The exposure-adjusted relative risk for patients receiving vedolizumab versus placebo group was 0.71 (95% c.i.: 0.45, 1.10) for SAEs, 0.56 (95% c.i. 0.22, 1.44) for serious infections, and 0.09 (95% c.i. 0.01, 0.89) for malignancies; values consist of events per person per year of exposure, using patient data from both Cohort 1 and Cohort 2 of the induction and maintenance trials. Exposure was calculated as days from first dose to last dose inclusive for patients who completed or were rescued to open-label vedolizumab in a separate study; exposure was calculated as first dose to last dose date plus up to 113 days, depending on length of follow-up, to account for pharmacologically relevant exposure for patients who permanently discontinued therapy. Days were converted into years. Exposure-adjusted incidence rates were calculated as total number of events/total patient-years.

[b] A serious infection was defined as an SAE of infection according to the Medical Dictionary for Regulatory Activities (version 15) criteria.

[c] Colon cancer (n=1), transitional cell carcinoma (n=1).

[d] Squamous cell carcinoma of the skin (n=1).

[e] Colon cancer (n=1) in vedolizumab group.

[f] The first *p*-value is derived from the comparison of vedolizumab every 8 weeks with placebo, and the second is derived from the comparison of vedolizumab every 4 weeks versus placebo. The test is derived from chi-square test (or Fisher exact test if any cell in the 2-by-2 table is ≤5).

* AEs were classified according to the MedDRA SOC categorisation and preferred terms. Patients with >1 event in a category were counted only once if the start and stop dates of the multiple events overlapped or if the start and stop dates were the same; if the start and stop dates of the multiple events did not overlap, they were counted as separate events.
The safety population was defined as all patients who received at least one dose of the study drug.
The vedolizumab group includes patients who received maintenance therapy with vedolizumab (patients who had a response to vedolizumab as induction therapy and who were assigned to vedolizumab every 4 weeks or every 8 weeks during the trial of maintenance therapy and patients who did not have a response to vedolizumab as induction therapy).
The placebo group includes patients who did not receive maintenance therapy with vedolizumab (patients assigned to placebo during the trial of induction therapy and patients who had a response to vedolizumab during that trial and who were assigned to placebo in the trial of maintenance therapy).

Supplementary safety evidence

The manufacturer provided supplementary safety evidence from the following studies:

- The GEMINI Long-Term Safety study (GEMINI LTS C13008)[14;32] (MS[1] pages 143-144).
- Pooled safety analyses of randomised placebo-controlled trials of vedolizumab in UC and CD.[33,34]

The GEMINI Long-Term Safety study (GEMINI LTS C13008)[14] – Interim results (MS[1] pages 143-144).

This is an ongoing Phase III, open-label, multicentre, long-term safety study (up to 7 years) evaluating vedolizumab in patients with UC and CD. As noted in the manufacturer's response to clarification[4] (question A32), patients were enrolled from the following studies: Study C13004 (Phase II long-term follow-up in patients with CD and UC),[18] Study C13006 (GEMINI I in patients with UC),[8] Study C13007 (GEMINI II in patients with CD)[35] and Study 13011 (GEMINI III in patients with CD).[36] The objective of this study is to collect and characterise important clinical safety events resulting from chronic vedolizumab administration (300mg vedolizumab i.v. every 4 weeks). The primary outcome measures were safety parameters such as AEs, SAEs, results of standard laboratory tests and ECGs, time to major IBD-related events (hospitalisations, surgeries or procedures), and improvements in quality of life. Limited interim results, as reported in the MS,[1] are summarised in Table 15. The mean age of patients with UC was 41.3 years (SD ±13.30) and 37.7 years (SD ±12.52) for those with CD. Vedolizumab exposure was ≥6, ≥12, and ≥24 months for 1,534 patients, 1,149 patients, and 502 patients, respectively. As reported in the MS[1] (pages 143-144) the safety profile of vedolizumab was similar between UC and CD 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, except for anal abscess, which occurred in 2% of CD patients. No cases of systemic candidiasis, disseminated herpes zoster, cytomegalovirus hepatitis or encephalitis, pneumocystis pneumonia or PML were reported. 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 MS.[1]

Table 15: GEMINI Long-term study - Interim safety results as of July 2012, (reproduced from MS page[1] 144)

Adverse events, 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)*	3 (<1)†

* Respiratory failure, acute stroke, pulmonary embolism
† Septicaemia, traumatic intracranial haemorrhage, suicide

Pooled safety analyses

The manufacturer undertook two separate pooled safety analyses. The first was a pooled analysis (not meta-analysis) of two Phase III, randomised, placebo-controlled, double-blind studies in adults with moderately to severely active UC (GEMINI1)[8] and CD (GEMINI2)[35] despite previous anti-TNF-α and/or other therapy. In general, the results of this analysis found that patients receiving vedolizumab (300mg vedolizumab i.v. every 4 weeks or every 8 weeks) had higher rates of overall AEs and SAEs (including gastrointestinal disorders and infections) compared with placebo; however, the overall incidence of AEs, adjusted for patient-years, was higher for the placebo groups than the vedolizumab group. Further details are provided on page 145 of the MS.[1]

A second pooled safety analysis (not meta-analysis) included 6 studies including two Phase II trials,[17,18] three Phase III trials (GEMINI I,[8] GEMINI II,[35] GEMINI III)[36] and one open-label long-term safety study (GEMINI LTS).[14] In general, as noted in the MS[1] (page 145), the baseline characteristics of the safety population were similar 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=1,107) and CD patients (n=1,723) with the most common AEs 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 146-149 of the MS.[1]

SAEs were low with vedolizumab treatment (see MS[1] pages 146-149) with the most common SAEs being exacerbation of CD, exacerbation of UC, abdominal pain and 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. However, the EPAR for vedolizumab[7] notes

that although no cases of PML have been described in the clinical programme to date, there is an absence of long-term safety data. As a result, the risk of PML is being monitored in the post-approval safety studies. In addition, a total of 26 vedolizumab-treated patients in the integrated safety population had been diagnosed with malignancy; 18 of these met SAE criteria. Of these, skin cancer (n=5) and colon cancer (n=4) were most common. 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 and no extra pulmonary manifestations or dissemination were reported.

Limited data on deaths were provided in the MS.[1] As noted in the FDA briefing document,[29] a total of 13 deaths (as of June 2013) occurred across all controlled and uncontrolled studies in UC and CD: GEMINI1 (UC patient, n=1 [vedolizumab Cohort 2 group]), GEMINI2 (CD patients, n=5 [1 in placebo group and 4 in vedolizumab group]) and 7 in the GEMINI LTS study (UC patients, n=3; CD patients, n=4). As noted in the manufacturer's clarification response[4] (question A33), none of the deaths in the UC patients were considered by the study investigators to be related to vedolizumab. Moreover, as noted in the EPAR,[7] a total of 9 post-study deaths have occurred as of March 2013 in the vedolizumab clinical program: GEMINI1,[8] n=2; GEMINI2,[35] n=1 and GEMINI LTS,[14] n=5 and one in a Phase II study.[18] Of these 9 deaths, sepsis was reported in a total of 3 subjects, malignancies were the cause of 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 EPAR[7] 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 indirect comparison and/or multiple treatment comparison

In the absence of any direct head-to-head RCTs comparing vedolizumab and other biologic therapies (infliximab, adalimumab, golimumab), surgery or calcineurin inhibitors for the treatment of moderate to severe UC, the manufacturer conducted an NMA. This is an extension of the conventional pairwise meta-analysis, combining 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 manufacturer conducted a systematic review to collate the published RCTs which assess the efficacy and safety of biological therapies prescribed for the treatment of UC.[1] The inclusion criteria for the NMA systematic review were as follows (see MS[1] pages 113-114): the population of interest

was people with moderate to severe UC; the interventions of interest were vedolizumab, other available biologics licensed for UC (infliximab, adalimumab, golimumab), surgery or ciclosporin; the relevant study design related to RCTs and the outcomes included efficacy and safety. The systematic review methods undertaken for the NMA (e.g. literature searching, study selection, data extraction and quality assessment) were the same as those undertaken for the vedolizumab systematic review. As noted in Section 4.1.1 adequate systematic searches were also undertaken to identify all relevant RCT studies assessing the efficacy and safety of infliximab, adalimumab, golimumab, surgery and ciclosporin for the treatment of UC.

The manufacturer's systematic review identified eight RCTs of varying methodological quality, that compared either vedolizumab,[8] infliximab,[24] adalimumab,[37-39] or golimumab[40,41] with placebo in patients with moderate to severe UC. A summary of the design and study characteristics as reported in the MS,[1] is provided in Table 16 for the induction phase and Table 17 for the maintenance phase (further data are reported in the MS,[1] pages 116-122). As noted on page 115 of the MS,[1] although studies were identified for surgery (n=6) and ciclosporin (n=5), these were not deemed suitable for inclusion in the NMA due to variation in study design, lack of common comparator to connect the network, differing outcomes in each study and small sample sizes.

Table 16: Summary of trials included in the manufacturer's NMA: Induction study characteristics (MS[1] pages 116-119)

Study	Design	Patient characteristics[a]	Treatment, dose, and sample size (ITT)	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
Vedolizumab versus placebo						
GEMINI1 Feagan *et al*, 2013[8]	Randomised, double-blind, placebo controlled trial (n=374)	Adults with moderate to severely active UC with inadequate response to, loss of response to, or intolerance of ≥1 of IM or TNF-antagonist Anti-TNF-α naïve: 51-58% Mean age: 40.1-41.2 years Male: 59-62%	Vedolizumab (i.v.) 300mg at Weeks 0 and 2 (n=225) Placebo (n=149)	6	6	Clinical response (primary endpoint) Clinical remission Mucosal healing Serious AEs Discontinuation due to AEs
Adalimumab versus placebo						
ULTRA1 Reinisch *et al*, 2011[37]	Randomised, double-blind, placebo-controlled trial (n=390)	Ambulatory adults with moderate to severely active UC despite concurrent and stable treatment with oral CSs and/or IMs Anti-TNF-α naïve: 100% Mean age: NR (median, 36.5-40 years) Male: 60.0-63.8%	Adalimumab (s.c.) 160/80: 160mg at Week 0, 80mg at Week 2, 40mg at Weeks 4 and 6 (n=130) Adalimumab (s.c.) 80/40: 80mg at Week 0, 40mg at Weeks 2, 4 and 6 (n=130) Placebo (n=130)	8	8	Clinical remission (primary endpoint) Clinical response Mucosal healing Serious AEs Discontinuation due to AEs

Study	Design	Patient characteristics[a]	Treatment, dose, and sample size (ITT)	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
ULTRA2 Sandborn *et al*, 2012[38]	Randomised, double-blind, placebo-controlled trial (n=518)	Adults with moderate to severely active UC for ≥3 months despite concurrent therapy with steroids and/or AZA or 6-MP Anti-TNF-α naïve: 58.9-60.9% Mean age: 39.6-41.3 years Male: 57.3-61. %8	Adalimumab (s.c.) 160mg at Week 0, 80mg at Week 2 and then 40mg EOW beginning at Week 4 (n=258) Placebo (n=260)	8	52	Clinical remission (primary endpoint) Clinical response Mucosal healing
Suzuki *et al*, 2014[39]	Randomised, double-blind, placebo controlled, trial (n=273)	Japanese patients age ≥15 years with biopsy-confirmed, moderately to severely active UC despite concurrent treatment with stable doses of oral CSs and/or IMs Anti-TNF-α naïve: 100% Mean age: 41.3-44.4 years Male: 57.5-67.8%	Adalimumab (s.c.) 160mg at Week 0, 80mg at Week 2, and then 40mg EOW beginning at Week 4 (n=90) Adalimumab (s.c.) 80mg at Week 0, 40mg at Week 2, and then 40mg EOW beginning at Week 4 (n=87) Placebo (n=96)	8	52	Clinical response Clinical remission Mucosal healing Serious AEs Discontinuation due to AEs

Study	Design	Patient characteristics[a]	Treatment, dose, and sample size (ITT)	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
Infliximab versus placebo						
ACT1 Rutgeerts *et al*, 2005[24]	Randomised, double-blind, placebo-controlled trial (n=364)	Adults with moderate to severely active UC despite concurrent treatment with CS ± AZA or 6-MP[b] Anti-TNF-α naïve: 100% Mean age: 41.4-42.4 years Male: 59-64.5%	Infliximab (i.v.), 5mg/kg at Weeks 0, 2, and 6 (n=121) Infliximab (i.v.), 10mg/kg at Weeks 0, 2, and 6 (n=122) Placebo n = 121	8	54	Clinical response (primary endpoint) Clinical remission
ACT2 Rutgeerts *et al*., 2005[24]	Randomised, double-blind, placebo-controlled trial (n=364)	Adults with moderate to severely active UC despite concurrent treatment with CS ± AZA or 6-MP and 5-ASA-containing medications[c] Anti-TNF-α naïve: 100% Mean age: 39.3-40.5 years Male: 56.7-62.8%	Infliximab (i.v.) 5mg/kg at Weeks 0, 2, and 6 n = 121 Infliximab (i.v.) 10mg/kg at Weeks 0, 2, and 6 (n=120) Placebo (n=123)	8	30	Clinical response (primary endpoint) Clinical remission Mucosal healing

Study	Design	Patient characteristics[a]	Treatment, dose, and sample size (ITT)	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
Golimumab versus placebo						
PURSUIT-SC Sandborn *et al*., 2014[40]	Randomised, double-blind, placebo-controlled trial (n=1,065)	Adults with moderate to severely active UC; no minimum disease duration; and inadequate response to, or intolerance of ≥1 of conventional therapies;[d] or were CS-dependent Anti-TNF-α naïve: 100% Mean age: 39-40.9 years Male: 52.9-60.7%	Golimumab (s.c.) 400mg at Week 0 and 200mg at week 2 (n=331) Golimumab (s.c.) 200mg at Week 0 and 100mg at Week 2 (n=331) Golimumab (s.c.) 100mg at Week 0 and 50mg at Week 2 (n=72) Placebo (n=331)	6	6	Clinical response (primary endpoint) Clinical remission Mucosal healing Serious AEs Discontinuation due to AEs

AZA - azathioprine; CS - corticosteroid; EOW - every other week; IM - immunomodulator; NR - not reported
Notes: Clinical remission = a Mayo score of 2 points or lower and no individual subscore above 1.
Clinical response = A decrease from baseline in the total Mayo score by at least 3 points and at least 30% with an accompanying decrease in rectal bleeding subscore of at least 1 point or an absolute rectal bleeding subscore of 0 or 1.
Mucosal healing was defined as a Mayo endoscopy subscore of 0 or 1.
[a] A range of values indicates across treatment groups (e.g., mean age).
[b] In ACT1, concurrent therapy was not required of patients who had no response to corticosteroids within 18 months prior to enrolment or no response to azathioprine or MP within 5 years prior to enrolment, or patients who could not tolerate corticosteroids, azathioprine, or MP. Rutgeerts *et al*, 2005[24]
[c] In ACT2, concurrent therapy was not required of patients who had no response to corticosteroids or 5-ASA-containing medications within 18 months prior to enrolment or no response to azathioprine or MP within 5 years prior to enrolment, or patients who could not tolerate corticosteroids, azathioprine, MP, or 5-ASA-containing medications. Rutgeerts *et al*, 2005[24]
[d] Conventional therapies are oral mesalamine, oral CSs, AZA, and 6-MP.

Table 17: Summary of trials included in the manufacturer's NMA: Maintenance study characteristics (MS[1] pages 120-122)

Trial, Study	Design	Patient characteristics[a]	Treatment, dose, and sample size	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
Vedolizumab						
GEMINI1 Feagan *et al*, 2013[8]	Randomised, double-blind trial (n=373, ITT) Central randomisation; stratified by 1) concomitant oral CS use and 2) prior anti-TNF-α or concomitant IM Responders to 6 weeks of vedolizumab induction were randomised to maintenance therapy	Adults with moderate to severely active UC with inadequate response to, loss of response to, or intolerance of ≥1 of IM or anti-TNF Anti-TNF-α naïve: 58-63% (ITT) Mean age: 38.6-41 years (ITT) Male: 54-57% (ITT)	Vedolizumab (i.v.) 300mg every 4 Weeks from week 6 to week 50 (n=125) Vedolizumab (i.v.) 300mg every 8 Weeks from Week 6 to Week 50 (n=122) Placebo (n=126)	52	46 (excluding induction phase)[b]	Clinical remission (primary endpoint) Durable clinical response (clinical response at both 6 and 52 Weeks) Durable clinical remission at 52 Weeks CSF remission Mucosal healing Serious AEs Discontinuation due to AEs
Adalimumab (s.c.)						
ULTRA2 Sandborn *et al*, 2012[38]	Randomised, double-blind trial (n=494 treated) Central randomisation and stratification by prior infliximab or other anti-TNF-α exposure Patients were randomised to an induction plus maintenance regimen at baseline	Adults with moderate to severely active UC for ≥3 months despite concurrent therapy with steroids and/or AZA or 6-MP Anti-TNF-α naïve: 58.9-60.5% Mean age: NR Male: NR	Adalimumab (s.c.) 160mg at Week 0, 80mg at week 2 and then 40mg EOW beginning at week 4 to through Week 52 (n=NR [248 treated]) Placebo (n=NR [246 treated])	52	52	Clinical remission (primary endpoint) Durable clinical response (clinical response at both Weeks 8 and 52) Clinical response at Week 52 Mucosal healing Discontinuations due to AEs

Trial, Study	Design	Patient characteristics[a]	Treatment, dose, and sample size	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
Suzuki *et al*, 2014[39]	Randomised, double-blind trial (n=273 treated) Randomisation was based on a centrally designed randomisation table. Patients were randomised to an induction plus maintenance regimen at baseline	Japanese patients aged ≥15 years with biopsy-confirmed, moderately to severely active UC despite concurrent treatment with stable doses of oral CSs and/or IMs Anti-TNF-α naïve: 100% Mean age: NR male: NR	Adalimumab (s.c.) 160mg at Week 0, 80mg at week 2, or 80mg at Week 0, 40mg at week 2; and then 40mg EOW beginning at Week 4 (n = NR [177 treated]) Placebo (n=NR [96 treated])	52	52	Clinical response at Week 52 Clinical remission Mucosal healing
Infliximab						
ACT-1 Rutgeerts *et al*, 2005[24]	Randomised, double-blind trial (n=364) Central randomisation; stratified by investigational site and CS-refractory UC. Patients were randomised to an induction plus maintenance regimen at baseline	Adults with moderate to severely active UC despite concurrent treatment with CS ± AZA or 6-MP Anti-TNF-α naïve: 100% Mean age: 41.4-42.4 years Male: 59-64. %5	Infliximab (i.v.), 5mg/kg every 8 weeks through to Week 46 (n=121) Infliximab (i.v), 10mg/kg every 8 weeks through to Week 46 (n=122) Placebo (n=121)	8	54	Clinical remission Durable clinical response (clinical response at both Weeks 8 and 30) Clinical response at Week 54 Discontinuation due to AEs CSF remission Mucosal healing

Trial, Study	Design	Patient characteristics[a]	Treatment, dose, and sample size	Primary endpoint time (weeks)	Study duration (weeks)	Key outcomes measured
Golimumab						
PURSUIT-M Sandborn *et al*, 2014[41]	Randomised, double-blind trial (n=564) Adaptive randomisation based on investigational site, clinical remission status, and CS use at PURSUIT-M baseline, and induction therapy. Responders to 6-weeks of induction golimumab were randomised at maintenance baseline visit	Adults with moderate to severely active UC; no minimum disease duration; and inadequate response to, or intolerance of ≥1 of conventional therapies; or were CS-dependent Anti-TNF-α naïve: 100% Mean age: 39.1-41.4 years Male: 48.1-57.8%	Golimumab (s.c.) 50mg every 4 weeks through to 52 Weeks (n=154) Golimumab (s.c.) 100mg every 4 weeks through to 52 Weeks (n=154) Placebo (n=156)	54	54	Durable clinical response (maintained from induction response to Week 54) (primary endpoint) Clinical response at Week 54 Clinical remission Serious AEs Discontinuation due to AEs

AZA - azathioprine; CS - corticosteroid; EOW - every other week; IM - immunomodulator; NR - not reported
Notes: Clinical remission = A Mayo score of 2 points or lower and no individual subscore above 1.
Clinical response = A decrease from baseline in the total Mayo score by at least 3 points and at least 30% with an accompanying decrease in rectal bleeding subscore of at least 1 point or an absolute rectal bleeding subscore of 0 or 1.
Mucosal healing was defined as a Mayo endoscopy subscore of 0 or 1.
[a] A range of values indicates across treatment groups (e.g., mean age).
[b] The manufacturer reported the study duration as 66 weeks

The main differences noted between the studies in both the induction phase and maintenance phase relate to patient characteristics, study design (randomisation at baseline or re-randomisation of biologic induction-responders) and study duration. GEMINI1[8] and ULTRA2[38] both included patients with prior anti-TNF-α exposure and anti-TNF-α naïve patients, whilst ACT1,[24] ACT2,[24] PURSUIT-SC,[40] Suzuki 2014,[39] and ULTRA1[37] included only patients who were anti-TNF-α naïve. Within PURSUIT-M,[41] all recruited patients were golimumab induction-responders.[40] Patients with prior anti-TNF-α exposure may be a more difficult to treat population than those who are anti-TNF-α naïve. In two of the maintenance trials (GEMINI1[8] and PURSUIT-M[41]), only patients who responded to biologic induction therapy were included in the maintenance phase analysis; these patients were re-randomised to either active treatment or placebo at the start of the maintenance phase. In contrast, in ULTRA2,[38] ACT1/2,[24] and Suzuki 2014,[39] patients were randomised to induction and maintenance regimens at baseline. As noted in the MS[1] (page 124), these differences would have implications for the efficacy results. In addition, the duration of studies varied both in the induction phase (between 6 to 8 weeks) and the maintenance phase (between 52 to 54 weeks, further details are provided in Table 16 and Table 17). The MS[1] (page 125) notes that the difference in study duration in the maintenance phase would not have a great impact on the results; the ERG agrees with this statement.

Data for the study quality (validity) assessment of the RCT studies included in the NMA (see MS,[1] pages 116-122) appear to be derived from the published trial reports. Although a detailed evaluation of each of the included studies was not undertaken by the ERG, the studies appear to be reasonably well conducted (MS[1] pages 353-355). With the exception of GEMINI1, these trials have previously been reviewed as part of the multiple technology appraisal of infliximab, adalimumab and golimumab for the treatment of moderately to severely active UC after failure of conventional therapy.[42]

For the statistical analysis (MS,[1] pages 126-129), the manufacturer 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. For the trials without re-randomisation at the end of the induction phase, the manufacturer's NMA assumes that patients that responded at the end of maintenance also all responded at end of induction. All outcome measures were modelled separately using a binomial likelihood and a logit link function. The models are reported on page 127 of the MS.[1]

4.4 Critique of the indirect comparison and/or multiple treatment comparison

An NMA was performed to compare treatment effects between vedolizumab, adalimumab, golimumab, infliximab and placebo for the outcomes of clinical response, clinical remission, discontinuation due to AEs and SAEs (Table 18) using data from the trials: GEMINI1,[8] ULTRA1,[37]

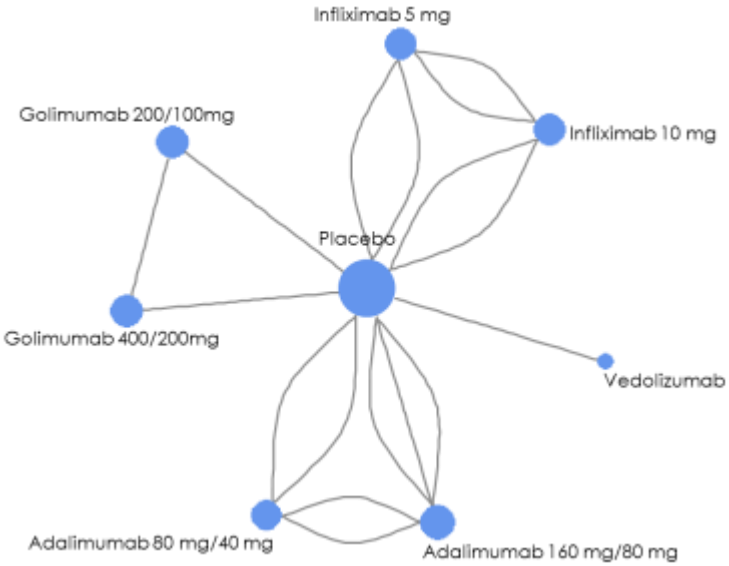
ULTRA2,[38] ACT1,[24] ACT2,[24] PURSUIT-SC,[40] PURSUIT-M[41] and Suzuki 2014.[39] The size of the network for each outcome varies depending on the availability of the data in each study.

Table 18: Summary of data used in the network meta-analysis provided by the MS[1]

Study Population (Study phase)	Clinical response	Clinical remission	Discontinuation due to AEs	SAEs
ITT (induction)	√	√	√	√
ITT (maintenance)		√	√	√
Anti-TNF-α naïve (induction)	√	√	√	
Anti-TNF-α naïve (maintenance)		√	√	
Anti-TNF-α experienced/failure (induction)	√	√		
Anti-TNF-α experienced/failure (maintenance)		√		

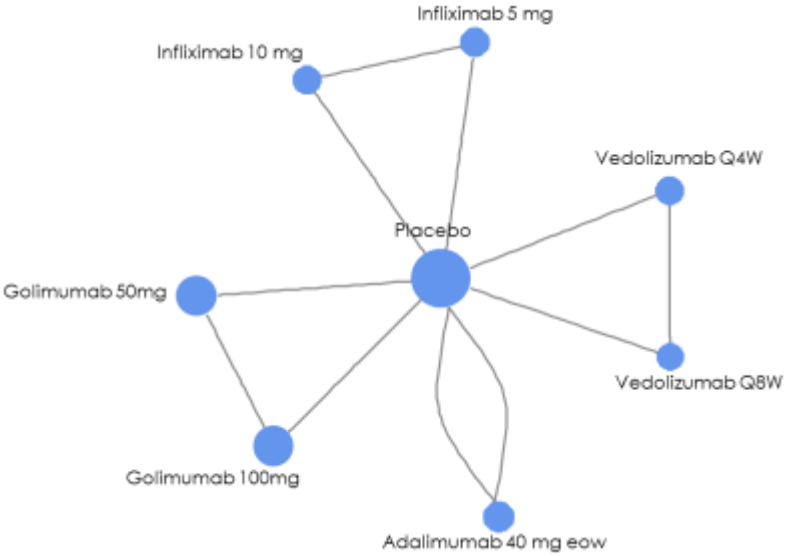
Whilst network diagrams were not reported within the MS,[1] these were provided by the manufacturer in response to a request for clarification from the ERG (see Figures 2-5). The ERG believes that there are mistakes in the diagrams provided by the manufacturer. The PURSUIT trial included in Figure 3 should be PURSUIT-M[41] rather than PURSUIT-SC.[40] The trials included in Figures 4 and 5 should be GEMINI1[8] and ULTRA2.[38] The outcomes analysed have not been reported consistently in the MS[1] or in the manufacturer's response to clarification.[4] The ERG has summarised the outcomes analysed in Table 19. It is not clear why some of the outcomes which have been measured in the trials have not been synthesised, e.g. mucosal healing for the ITT population. The ERG considers that data for durable clinical response in the maintenance phase should be not be synthesised because not all trials measured this outcome and the definition of durable clinical response may differ in those trials which do report this outcome.

Figure 2: Network of evidence for anti-TNF-α naïve induction clinical response and clinical remission



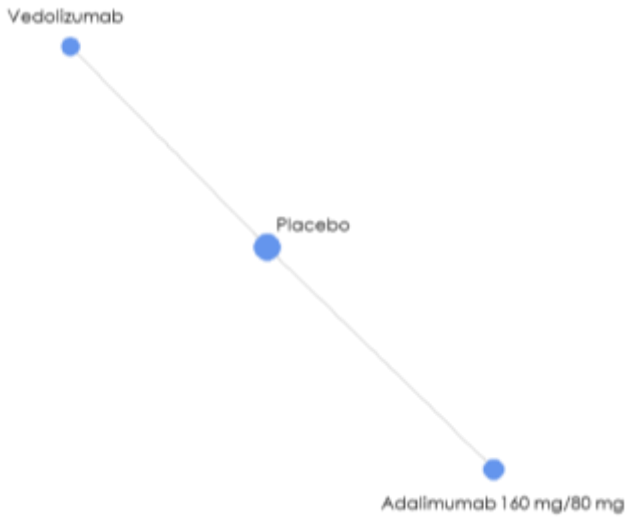
Trials included: GEMINI1;[8] ULTRA1;[37] ULTRA2;[38] ACT1;[24] ACT2;[24] PURSUIT-SC;[40] Suzuki 2014

Figure 3: Network of evidence for anti-TNF-α naïve maintenance durable clinical response and clinical remission



Trials included: GEMINI1;[8] ULTRA2;[38] ACT1;[24] PURSUIT-SC;[40] Suzuki 2014[39]

Figure 4: Network of evidence anti-TNF-α experienced/failure induction clinical response, clinical remission, and mucosal healing



Trials included: GEMINI1;[8] ULTRA1[37]

Figure 5: Network of evidence anti-TNF-α experienced/failure maintenance durable clinical response, clinical remission, and mucosal healing

Trials included: GEMINI1;[8] Suzuki 2014[39]

Table 19: Summary of data used in the network meta-analysis

Study Population (Study Phase)	Clinical response	Durable clinical response	Clinical remission	Mucosal healing	Discontinuation due to AEs	SAEs	CSF remission
ITT (induction)	√		√		√	√	
ITT (maintenance)			√		√	√	
Anti-TNF-α naïve (induction)	√		√	√	√		
Anti-TNF-α naïve (maintenance)		√	√	√	√		√
Anti-TNF-α experienced/failure (induction)	√		√	√			
Anti-TNF-α experienced/failure (maintenance)		√	√	√			

The results of the manufacturer's NMA are presented in Tables 31 to 39 of the MS[1] in terms of odds ratios for vedolizumab versus each treatment (although there is some confusion within the MS and the clarification response regarding what these odds ratios represent). Following a request for clarification, the manufacturer also provided fixed effects results for each contrast for the anti-TNF-α naïve and anti-TNF-α failure populations (see Appendix 3); the equivalent results for the mixed ITT population were not provided by the manufacturer.

These analyses suggested that in the induction phase for anti-TNF-α naïve patients, infliximab provided the largest treatment effect on clinical response, remission and mucosal healing compared with placebo, and vedolizumab has the lowest rate of discontinuations due to AEs compared with placebo. In the induction phase for anti-TNF-α experienced/failure patients, only the treatment effect of adalimumab and vedolizumab were analysed relative to placebo. Each had positive effects in term of clinical response, remission and mucosal healing, but only the effect of vedolizumab compared with placebo in response was statistically significant. For the maintenance phase, vedolizumab was associated with the largest treatment effect compared with placebo in both the anti-TNF-α naïve and experienced/failure patient subgroups. However, patients in the GEMINI1 maintenance phase were all vedolizumab induction-responders. No efficacy data were available for vedolizumab induction-responders relative to placebo induction-responders during the maintenance phase.

The ERG considers that all of the manufacturer's NMA results should be interpreted with caution since they were based on an assumption of no between-study variance yet the patient populations and trial designs were different between included studies. In addition, there are several mistakes in the data presented by the manufacturer. The ULTRA2[38] trial data for the anti-TNF-α experienced/failure population maintenance phase clinical remission reported in Table 142 in the MS[1] report were

incorrect as the N for placebo and adalimumab should be 29 and 36, respectively. The Suzuki 2014[39] trial data for the maintenance phase clinical remission reported in Table 132 and 138 in the MS[1] report were incorrect as the N for placebo and adalimumab 40mg EOW should be 8 and 41, respectively. The PURSUIT-M[41] trial data for maintenance phase durable clinical response in the clarification response[4] page 59 were incorrect as the N for placebo, golimumab 50mg and golimumab 100mg should be 154, 151 and 151, respectively; the N for golimumab 50mg and golimumab 100mg should be 71 and 76, respectively. The ULTRA2[38] trial data for maintenance phase durable clinical response in the clarification response[4] page 59 were also incorrect as the N for adalimumab 40mg EOW should be 59. The ERG has not checked all the data presented by the manufacturer; hence it is unclear if data used for other outcome measures were all correct.

The manufacturer undertook separate NMAs of anti-TNF-α naïve and anti-TNF-α experienced/failure subgroups. However, the manufacturer did not provide a rationale for conducting such analysis on subgroups separately. The ERG considers that the disadvantage of conducting separate analyses is that the possibility of an interaction between treatment and subgroup cannot be explored. The ERG asked the manufacturer to conduct an additional meta-regression including type of population as a covariate to assess if there is an interaction. The manufacturer's response stated that *"when such a model is fitted to a small network, the model may pick up on variation which could be caused by any number of study differences (known or unknown) causing the result to be spuriously significant or not significant, e.g. due to a lack of data. At least 20 studies would be needed."*[4] and that because the maximum number of studies in any of the network was 7, no such analysis was performed. The ERG considers this point to be reasonable for conducting meta-regression in general. However, whether it is possible to undertake meta-regression analysis also depends on the number of treatments included and the assumption of the model coefficients. If conducting a meta-regression is indeed not possible, then the predictive distribution of treatment effects which incorporates extra variability should be presented.

Induction phase and maintenance phase data were synthesised separately by the manufacturer. The ERG considers this to be appropriate. The MS[1] acknowledges that the study designs of ULTRA2,[38] Suzuki 2014[39] and ACT1[24] are different from the designs employed within the GEMINI1 and PURSUIT-M[41] trials. In order to allow for comparison with adalimumab and infliximab, the manufacturer made the following adjustment to the trials without re-randomisation after the induction phase. When conducting the NMA for the maintenance phase, the manufacturer assumed that the responders at the end of induction were the same as the responders at the end of maintenance in calculating the probability of durable clinical response, clinical remission, mucosal healing, and CSF remission. However, this approach ignores the fact that non-responders at the end of induction could become responders at the end of the maintenance phase, and the number of events at the end of

maintenance could be contributed to by both responders and non-responders at the end of the induction phase. Event rates in both the placebo arms and experimental treatment arms were inflated using this approach. The magnitudes of the inflation in both arms of all trials in which this adjustment was made were not the same depending on the actual observed data, hence it is difficult to predict the impact of this adjustment on the relative treatment effect.

The ERG considers that there is no empirical evidence available to estimate relative treatment effects in the maintenance phase of vedolizumab compared with placebo for placebo-treated patients (patients treated with placebo in both induction and maintenance phase). It is not clear if the maintenance phase results in GEMINI1[8] (or PURSUIT-M[41]) overestimate or underestimate this relative treatment effect. The ERG believes that the adjustment applied to the trials without re-randomisation at the end of the induction phase by the manufacturer did not adjust the bias sufficiently, rather, it is possible that their adjustment method actually introduced more bias into the analysis. Consequently, all the maintenance phase results produced from the manufacturer's NMA should be interpreted with caution.

The manufacturer stated that *"Where there were closed loops in the network, consistency analyses were performed and studies were found to be consistent unless otherwise stated"* (MS,[1] page 139). This was contradicted by a later statement made by the manufacturer *"no tests could be performed to look at consistency/inconsistency in the network"* (MS,[1] page 139). The MS[1] also stated that *"Heterogeneity checks on placebo response rates were also performed to investigate the similarity of patient populations between trials, unless otherwise stated, patient populations were found to be consistent"* (MS,[1] page 139). The precise checks undertaken by the manufacturer are not clear.

The manufacturer reported that both fixed effects and random effects models were used. However, it was not clear in the MS[1] which of these models the results presented were based on. The ERG asked the manufacturer to clarify which model was used and for the manufacturer to justify the model choice (see clarification response,[4] question A1). The manufacturer responded by stating that both models were used and most of the results are based on Bayesian fixed effect models because of a lack of robust closed loops. The manufacturer also stated that *"The use of random effects MTCs was restricted to instances when closed loops existed in the network."*[4] The ERG does not consider that this justification is valid. Random effects MTCs can be used for non-closed loop networks. The existence of closed loops could be used to check inconsistency. The manufacturer also justified the model choice by stating that, *"The random effects MTCs did not have good convergence as observed through iteration plots and Gelman–Rubin diagnostic tests and in many cases appeared to have greatly inflated errors compared to the equivalent fixed effects and frequentist models."* The ERG does not consider this as a sufficient justification for the use of fixed effects models. The ERG

considers that heterogeneity in treatment effect in different studies is to be expected. The existence of heterogeneity between trials was also supported by the manufacturer, as the MS[1] (page 124) noted that *"the clinical trials varied in terms of study design and patient populations; (i.e., heterogeneity between trials)."* The use of a random effects model would explicitly model heterogeneity and capture uncertainty in the true treatment effect. A fixed effects model would underestimate this uncertainty. When data are not very informative, careful consideration of the prior distribution is required.

The manufacturer modelled clinical response rate and remission rate separately using a binomial likelihood (see MS[1] page 127). The ERG considers that this approach is partially appropriate. The results for clinical response and remission should be interpreted with caution, because these results were estimated without considering the dependence/correlation between response and remission. Ideally, the NMA should take account of the nature of the data i.e. ordered categorical. Use of these results in the economic model ignores this dependence and would potentially generate inappropriate samples for PSA. The complementary log-log model was also used to take into account length of time for discontinuation due to AEs. The ERG considers this to be appropriate.

The results presented in the MS[1] were generated using a total of 60,000 iterations with burn-in of 20,000 iterations and thinning by 50 from 3 chains. Despite a request for clarification[4] (question A10), the methods used by the manufacturer used for assessing convergence and the number of simulations to retain remain unclear (the manufacturer stated that *"The burn-in number of iterations and total number of iterations were chosen to give adequate time for models to converge"*[4]). The ERG considers that it was highly likely that convergence had occurred in most analyses but this is unclear in the case when the number of patients experiencing outcomes was very low (e.g., discontinuing due to AEs and experiencing SAEs).

The MS purports to present results using odds ratios for each comparator vs. vedolizumab for the anti-TNF-α naïve and anti-TNF-α experienced/failure subgroup using a fixed effects model.[1] The ERG noted that the reported odds ratios seem to suggest that vedolizumab could be worse than placebo which contradicts the GEMINI1 trial findings.[8] The ERG requested clarification of the results presented in Tables 31 to 42 of the MS.[1] The manufacturer stated that *"The MTC results are presented as odds ratios (ORs) for each treatment included in the MTC* ***relative to placebo****. We acknowledge that there was an error in the submission, with the ORs stated to be versus vedolizumab."* However the ERG believes that this is also incorrect. The results presented in Tables 31 to 42 in the MS[1] report should be the odds ratio for ***vedolizumab relative to*** each treatment included in the NMA.

4.5 Additional clinical exploratory analyses undertaken by the ERG

As the manufacturer undertook a comprehensive systematic review (no major limitations were noted) of vedolizumab of treatment of adults with moderate to severe active UC who are intolerant of, or whose disease has had an inadequate response or loss of response to conventional therapy or a TNF-α antagonist, no additional work was undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

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

The clinical evidence in the MS[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 UC. The ERG is content that all relevant (published and unpublished) studies of vedolizumab were included in the MS.[1] Although the ERG acknowledges the exclusion of ciclosporin and surgery from the NMA due to incomparable study design, lack of common comparator and differing outcomes, they were included in the final NICE scope and therefore should have been considered as relevant comparators.

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

A key issue that may limit the robustness of the efficacy and safety data reported in the MS[1] relates to the high dropout rates in the maintenance phase of the GEMINI1[8] trial. High rates of discontinuation were observed across all treatment groups (combined vedolizumab, 52% [324/620] versus combined placebo, 72% [197/275]). In general, the validity of a study may be threatened if attrition is more than 20%.[31] Another issue that may limit the robustness of the efficacy evidence in the anti-TNF-α failure and anti-TNF-α naïve patients and other subgroup analyses in participants from the GEMINI1[8] trial is the exploratory approach used. These trials were not powered for these exploratory subgroup analyses.

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 generalisability to the UK population. Further details are provided below.

Duration of treatment

The duration of treatment of vedolizumab in the GEMINI1[8] trial was 52 weeks, followed by enrolment in the ongoing GEMINI LTS study.[14] 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 therapy or with respect to how to predict instances in which this can be successfully achieved. The SmPC[10] for vedolizumab advises for

monitoring and reporting of any suspected adverse reactions after authorisation especially for new onset or worsening of neurological signs and symptoms.

Generalisability to the population of England and Wales

The total population in the GEMINI1[8] trial was predominantly white (82.0%) with a mean age of 40.3 years, mean body weight of 73.4kg and male (58.7%). Mean duration of disease was 6.9 years and patients had a mean Mayo score of 8.6. Approximately 48% of patients had received prior anti-TNF-α treatment. In addition, of the 211 study sites from which patients were recruited for the GEMINI1 trial,[8] only two were UK-based and 63 were US-based. In contrast to the other study sites, in the US, permitted immunosuppressants were discontinued after the induction phase. As such there is some uncertainty regarding the generalisability 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. In addition, as noted in the SmPC,[10] no clinical data are available for patients previously treated with rituximab and thus caution should be used in considering vedolizumab treatment in such patients.

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. The results presented for clinical remission and durable clinical response in the maintenance phase may not be correct since incorrect data were used. The adjustment made by the manufacturer in the maintenance phase to the trials without re-randomisation at the end of induction phase inflates absolute treatment effects in both the placebo and experimental treatment arm of each trial. The impact of this adjustment on the relative treatment effect in these trials is not clear. It is also noteworthy that the maintenance phase of GEMINI1[8] only recruited vedolizumab induction-responders, and PURSUIT-M[41] only recruited golimumab induction-responders. Therefore, the placebo group in GEMINI1[8] and PURSUIT-M,[41] and in other trials without re-randomisation at the end of the induction phase (ULTRA2,[38] ACT1,[24] and Suzuki 2014[39]) may not be comparable. It is unclear if the large relative treatment effect observed for vedolizumab compared with placebo in the maintenance phase was due to the low event rates for placebo-treated vedolizumab-responders.

5 COST-EFFECTIVENESS

This chapter presents a summary and critical appraisal of the methods and results of the manufacturer's review of published economic evaluations and the *de novo* health economic analysis presented within the MS.[1] Additional analyses undertaken by the ERG are also presented within this chapter.

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

5.1.1 Description of manufacturer's review

The MS[1] presents a systematic review of evidence relating to the cost-effectiveness of vedolizumab and other treatments for patients with UC. A systematic literature review was initially performed by the manufacturer in April 2013 and updated in March 2014. Search terms for databases included combinations of free text and MeSH headings including 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 MS states that searches were restricted to studies published after 2002 as prior to that date, biologic drugs used in the treatment of UC had not been approved for use in the UK, and resource use and cost studies would be out of date.[1] The manufacturer's search strategy was comprised of searches of the following databases:

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

The manufacturer'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 manufacturer's selection of studies for inclusion in the review was guided by inclusion and exclusion criteria (see MS[1] Table 49). 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.

Two full UK economic evaluations[43,44] were included in the manufacturer's systematic review (see Table 20). The study reported by Tsai *et al*[44] assessed the cost-utility of infliximab versus conventional non-biologic therapies (5-ASAs, immunomodulators and corticosteroids) in patients with moderate to severe UC from the perspective of the NHS over a 10-year time horizon. The study reported by Punekar and Hawkins[43] assessed the cost-utility of infliximab, ciclosporin and conventional therapies (i.v. steroids in addition to existing immunomodulators) in hospitalised patients with acute severe UC who were not responding to 72 h of i.v. steroid therapy from the perspective of the NHS over a 1-year time horizon. The MS presents a quality assessment summary of the two studies (see MS[1] Table 51) which suggests that, in the manufacturer's view, both studies are applicable to the decision problem for the appraisal but that each study is subject to minor limitations.

Table 20: Summary of studies included in the manufacturer's cost-effectiveness review

Study	Tsai *et al* 2008[44]	Punekar and Hawkins 2010[43]
Analysis type	Cost-utility analysis	Cost-utility analysis
Population	Patients with moderate to severe UC	Acute UC patients
Economic comparisons included	• infliximab • conventional non-biologic treatments Separate analyses were conducted for continuation of infliximab in patients achieving (1) response and remission; (2) remission only.	• infliximab • ciclosporin • conventional non-biologic treatments
Perspective	NHS	NHS
Time horizon	10 years	1 year
Headline findings	*Responder continuation rule* ICER for infliximab versus conventional treatment = £27,424 per QALY gained *Remission only continuation rule* ICER for infliximab versus conventional treatment = £19,696 per QALY gained	ICER for infliximab versus ciclosporin = £19,545 per QALY gained ICER for ciclosporin versus standard care = dominating

5.1.2 ERG comment on the review of cost-effectiveness evidence

The ERG consider that the search methods detailed in Section 7.1.2 and Appendix 10.10 of the MS[1] were clearly reported and the sources searched were largely appropriate for the review. However, the ERG found that the terms differed in the original search and the update search in that further terms for conventional treatments (ASAs, corticosteroids and immunomodulators) were added in the update search; it is unclear why these terms were omitted from the original search. The ERG requested clarification for this discrepancy (see clarification response,[4] question A18). In their response, the manufacturer stated *"Regarding the comparator searches, the update was in response to the NICE scoping advice given to Takeda."*[4] The ERG does not believe that this presents a full justification for the discrepancy. The ERG also notes that the use of a publication cut-off date may fail to identify relevant evidence relating to the cost-effectiveness of conventional non-biologic UC therapies.

Searches for evidence on the cost-effectiveness of surgery and ciclosporin were not included in the MS.[1] In response to a request for clarification[4] (question A17), the manufacturer stated that ciclosporin was beyond the scope of the appraisal. The ERG notes that calcineurin inhibitors (tacrolimus and ciclosporin) were specified in the final NICE scope. The manufacturer also noted that searches for surgery had been undertaken but omitted, in error, from the MS. Further details of these searches are presented in Appendix 2 of the manufacturer's response to clarification.[4]

With respect to study selection, the ERG notes that only the study reported by Tsai *et al*[44] relates to the moderate to severe UC population and is directly applicable to this appraisal. Importantly however, this study does not include all relevant comparators (surgery and other biologic therapies were not considered) and the study adopts only a 10-year rather than lifetime horizon. The ERG believes that the inclusion of the study reported by Punekar and Hawkins[43] within the manufacturer's review is inappropriate as this study includes a population of patients that have been hospitalised for acute severe UC; this population is specifically excluded from the NICE scope.[6] In response to a request for clarification on this issue from the ERG (question B3), the manufacturer stated that the Punekar study was included as the original search did not include exclusion criteria related to terms for chronic or acute UC.[4] However, even if the search identified the study, the application of appropriate study selection criteria should have resulted in its exclusion from the review. Further, whilst the MS[1] (Table 51) suggests that the patient population included in Punekar *et al*[43] is relevant to the appraisal, the manufacturer's subsequent response to clarification[4] also states that the patient population considered in the paper is different from the license for vedolizumab. The ERG also believes that other non-UK economic analyses (for example Xie *et al,*[45] mentioned elsewhere in the MS[1]) may have provided useful information for the appraisal, hence these should also have been included in the manufacturer's review.

5.2 Description of the manufacturer's model

5.2.1 Health economic evaluation scope

The health economic analysis presented by the manufacturer uses a model-based approach to compare vedolizumab versus other medical therapies and surgery from the perspective of the UK NHS (see Table 21). Costs borne by the PSS are excluded from the economic analysis; the manufacturer states that these are expected to be minimal (see MS[1] Table 53). The manufacturer's health economic 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) and surgery. Other anti-TNF-α agents (infliximab, adalimumab and golimumab) are included only in the analysis of the anti-TNF-α naïve population; these are excluded from the analyses of the mixed ITT and anti-TNF-α failure populations. Calcineurin inhibitors are not included in the analysis. The efficacy data and the methods used to synthesise these, differ between the populations included in the analyses. Within the economic analyses in all three populations, cost-effectiveness results are presented as pairwise comparisons in terms of the incremental cost per QALY gained for vedolizumab versus each individual comparator. The manufacturer's base case analysis adopts a 10-year time horizon; a lifetime horizon is considered in the sensitivity analysis. All costs and health outcomes are discounted at a rate of 3.5% per annum.

Table 21: Populations, comparators and sources of efficacy evidence used in manufacturer's health economic analysis

Population	Interventions compared	Source of efficacy data
(1) Mixed ITT population	• Vedolizumab • Conventional non-biologic therapies (5-ASAs, immunomodulators, corticosteroids) • Surgery	Induction: Observed outcomes from GEMINI1[8,28] used to inform probabilities of response/remission Maintenance: Model fitted to probability of achieving response/remission observed in the GEMINI1 trial[8,28]
(2) Anti-TNF-α naïve population	• Vedolizumab • Infliximab • Adalimumab • Golimumab • Conventional non-biologic therapies (5-ASAs, immunomodulators, corticosteroids) • Surgery	Induction: Manufacturer's NMA[1] used to inform probabilities of response/remission for each option Maintenance: Model maintenance transition matrix fitted against 1-year probabilities of response/remission predicted by manufacturer's NMA[1]
(3) Anti-TNF-α failure population	• Vedolizumab • Conventional non-biologic therapies (5-ASAs, immunomodulators, corticosteroids) • Surgery	Induction: Observed outcomes from GEMINI1[8,28] used to inform probabilities of response/remission Maintenance: Model fitted to probability of response/remission observed in the GEMINI1 trial[8,28]

Table 22 summarises the treatment regimens included within the manufacturer's model. Vedolizumab induction therapy is assumed to be given as an i.v. infusion at a dose of 300mg at Weeks 0 and 2. It should be noted that this reflects the GEMINI1 trial rather than the EMA marketing authorisation[10] (the latter recommends three induction doses). Subsequent maintenance therapy is assumed for those patients who respond to treatment at a dose of 300mg every 8 weeks thereafter. Adalimumab induction therapy is assumed to be given as a self-administered s.c. injection at a dose of 160mg at Week 0, 80mg at Week 2 and 40mg at Weeks 4 and 6. Subsequent maintenance therapy is assumed for those patients who respond to treatment at a dose of 40mg every other week (EOW). Every week (EW) dosing[46] is not included for adalimumab. Golimumab induction therapy is assumed to be given as an s.c. injection at a dose of 200mg at Week 0 and 100mg at Week 2. Subsequent maintenance therapy is assumed for those patients who respond to treatment at a dose of 50mg every 4 weeks. The 100mg dose for patients with body mass >80kg[47] is not included in the model. Infliximab is assumed to be given at a dose of 5mg/kg at Weeks 0, 2 and 6, with subsequent maintenance therapy for those patients who respond to treatment at a dose of 5mg/kg every 8 weeks. Conventional non-biologic treatments are assumed to be taken daily indefinitely and include a mix of 5-ASAs (balsalazide, mesalazine, olsalazine, sulfasalazine and budesonide), corticosteroids (prednisolone) and immunomodulators (azathioprine, 6-MP and methotrexate); the specific products assumed are not specified by the manufacturer in either their model or submission.[1]

The manufacturer's model assumes that treatment using biologic therapy is discontinued if the patient fails to achieve response during induction or if the patient experiences AEs which warrant discontinuation of therapy. All patients are assumed to discontinue biologic therapy at 1-year irrespective of whether they have maintained clinical response or remission up to that point. Prior to 1-year, the model assumes that patients receiving biologic maintenance therapy cannot lose response; that is, they remain on biologic therapy even if they have moderate to severe disease. Subsequent to the discontinuation of biologic treatment, patients are assumed to receive conventional non-biologic therapies. Surgery is included in the model as a subsequent part of the pathway.

Table 22: Description of interventions assessed in the manufacturer's model

Treatment	Induction regimen	Maintenance regimen	Administration
Vedolizumab	300mg at Week 0 and 2	300mg every 8 weeks	i.v. infusion
Infliximab	5mg/kg at Week 0, 2 and 6	5mg/kg every 8 weeks	i.v. infusion
Adalimumab	160mg at Week 0, 80mg at Week 2, 40mg at Weeks 4 and 6	40mg every 2 weeks	self-administered s.c. injection
Golimumab	200mg at Week 0, 100mg at Week 2	50mg every 4 weeks	self-administered s.c. injection
Conventional non-biologic treatments	Various – all treatments appear to be assumed to be given daily indefinitely		
Surgery	n/a	n/a	n/a

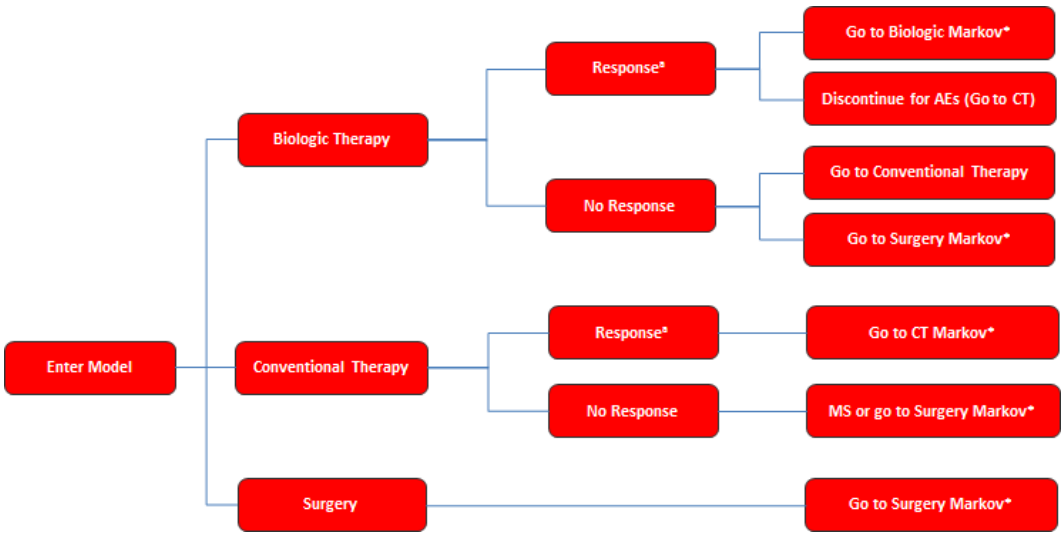
n/a – not applicable

5.2.2 Model structure and logic

5.2.2.1 Model methodology and definition of health states

The manufacturer's model adopts a hybrid approach whereby a decision tree is used to evaluate outcomes at the end of initial induction therapy and a Markov structure is used to evaluate subsequent outcomes during maintenance treatment (including subsequent induction treatment using conventional therapies for patients who discontinue biologic treatments). The model initially adopts a 6-week cycle length to reflect outcomes at the end of induction therapy. Subsequently, the model adopts an 8-week cycle length during maintenance treatment. The manufacturer's diagrammatic representations of the model structure for induction treatment and maintenance treatment are shown in Figures 6 and 7, respectively. The model includes a total of 16 mutually exclusive health states, as shown in Table 23. These are divided into two sets of 8 health states, which reflect (a) whether the patient is receiving, or (b) has most recently previously received, a biologic treatment or other conventional non-biologic treatments.

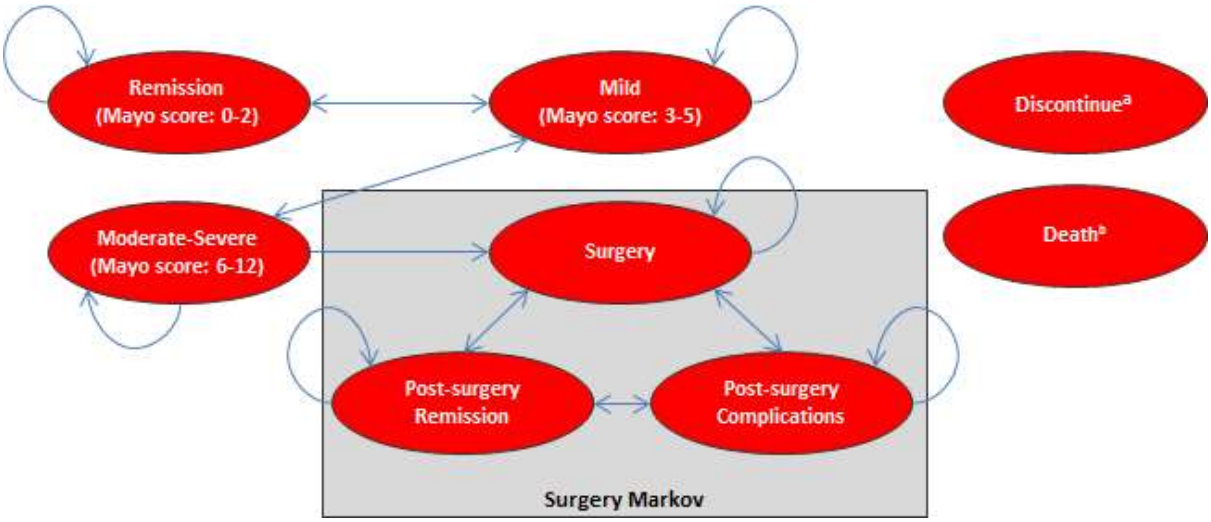
Figure 6: Induction treatment - decision tree model structure[1]



AE, adverse event; CT, conventional therapy; MS, moderate-severe.

[a] *Response is defined as a drop in Mayo score of 3 points or more. This includes patients who also achieve remission, as remission is a subset of response. Remission is defined as a Mayo score less than 3.*

Figure 7: Maintenance treatment – Markov model structure[1]



a Reasons for discontinuation include lack of response and AEs. Discontinuation due to AEs is applicable only to responders on biologic treatments, because nonresponders on biologics switch to conventional therapy and continue receiving such until the end of the model's time horizon or until the patients require surgery.

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

Table 23: Health states included in the manufacturer's model (induction and maintenance treatment)

	(a) States for patients who are currently receiving or who have most recently received a biologic therapy	*(b) States for patients who are currently receiving or who have most recently received conventional non-biologic therapy*
Pre-colectomy health states	Remission	Remission
	Mild	Mild
	Moderate-severe (response)	Moderate-severe (response)
	Moderate-severe (no response)	Moderate-severe (no response)
Surgery and post-surgical health states	Surgery	Surgery
	Post-surgery remission	Post-surgery remission
	Post-surgery complications	Post-surgery complications
	Dead	Dead

It should be noted that the description of model states within the manufacturer's diagram does not directly reflect the actual health states included in the Markov component of the model as it does not account for patients who are responders with moderate to severe disease and those who are non-responders with moderate to severe disease.

The general model structure is the same for the biologic treatments, conventional non-biologic treatments and surgery options; the treatment group under consideration influences whether patients enter the model in (a) the biologic states, or (b) the non-biologic states.

The logic of the manufacturer's model is described below.

5.2.2.2 Summary of model logic

Biologic treatment groups (vedolizumab, infliximab, adalimumab and golimumab) – health state transitions

Patients enter the model in the moderate to severe UC (no response) state. At the end of the first cycle, patients are redistributed across the model health states according to probabilities of having mild disease or remission based on the manufacturer's NMA of induction studies[1] (anti-TNF-naïve-α population only) or according to the observed outcomes within the GEMINI1 trial[8] (mixed ITT and anti-TNF-failure-α populations). During the induction cycle, a proportion of patients are assumed to undergo surgery and transit to the surgery and post-surgery states. A proportion of patients who achieve response are assumed to have moderate to severe UC and are thus classed as moderate to severe responders (i.e. whilst achieving the response criteria, their Mayo score is greater than 6). Patients who achieve response or remission during induction therapy using biologic treatments are assumed to continue to receive the same biologic as maintenance therapy; patients who do not achieve response or remission during induction are assumed to discontinue and subsequently receive conventional non-biologic therapy. The induction transition vector is applied only during the first model cycle. During the induction phase, 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 UC.

During the maintenance phase of the model, patients may remain on biologic treatment provided they do not experience AEs sufficient to warrant discontinuation and provided they have not received biologic treatment for more than 1-year. Patients are assumed to continue biologic maintenance therapy even if they have moderate to severe disease. Probabilities of transiting between the pre-colectomy, colectomy and post-colectomy health states are determined by a 6*x*6 matrix of transition probabilities which has been derived by calibrating the model-predicted probabilities of response/remission at 1-year against the predicted estimates from the manufacturer's NMA of induction treatments and NMA[21] of maintenance treatments (anti-TNF-naïve-α population only) or against the observed results of the GEMINI1 trial[8] (mixed ITT and anti-TNF-failure-α populations). The transition probabilities between surgery and post-surgical states were derived from other published literature. A proportion of patients who achieve response are assumed to be in the moderate to severe UC state and remain on biologic treatment. A proportion of patients are assumed to discontinue therapy during each maintenance cycle due to AEs. The 8-week transition probabilities used to reflect maintenance treatment effects are applied indefinitely until the maximum biologic treatment duration has been reached. During maintenance, 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 UC.

At approximately 1-year, a forced treatment switch is applied to all patients receiving biologic treatment; any patients who are currently receiving biologic therapy at this point are assumed to discontinue and subsequently receive conventional non-biologic treatments, irrespective of their current level of response to treatment.

Conventional non-biologic treatment – health state transitions

Transitions between the conventional non-biologic states are similar to those for the biologic portion of the model, except that once patients enter these states they remain in them for the remainder of the model time horizon (they cannot subsequently receive biologic therapy). After discontinuation of biologic treatment, patients enter the conventional non-biologic portion of the model in their current UC state and a vector of probabilities of induction response is applied based on the results for the placebo arm of the GEMINI1 trial[8,28] (mixed ITT population and anti-TNF-α failure population) or the equivalent placebo estimates from the manufacturer's NMA (anti-TNF-α naïve population). Subsequent probabilities of transiting between the pre-colectomy, colectomy and post-colectomy states are driven by a 6*x*6 matrix of transition probabilities which has been derived by calibrating the model-predicted probabilities of response/remission at 1-year against the predicted estimates from the manufacturer's NMAs of induction and maintenance treatments[1] (anti-TNF-α naïve population) or against the observed results of the placebo arm of the GEMINI1 trial[8] (mixed ITT and anti-TNF-α failure populations). The transition probabilities between surgery and post-surgical states were derived from other published literature. The 8-week transition probabilities used to reflect maintenance treatment effects are applied indefinitely for the remainder of the model time horizon. The risk of mortality is applied as a baseline other-cause mortality rate, with state-specific relative risks to reflect an excess risk of death due to UC.

Surgery and post-surgery states

Patients may transit to surgery from the moderate to severe UC state within either the biologic treatment portion of the model or the non-biologic treatment portion of the model during induction, but only from the non-biologic portion of the model during maintenance (the ERG notes that this may be an unintended programming error). During each maintenance cycle, patients in the surgery state may stay in the surgery state, transit to post-surgery remission or post-surgery complications, or die. Patients in the post-surgical remission state and post-surgical complications states can also subsequently transit back to the surgery state. Mortality is applied as a baseline other-cause mortality rate, with state-specific relative risks to reflect an excess risk of death due to UC. A peri-operative mortality risk is not included in the model.

Health-related quality of life

Different levels of HRQoL are applied to each of the model health states. HRQoL is also reduced according to the incidence and impact of AEs associated with individual biologic and non-biologic treatments. Total QALYs in each treatment group are driven by health state sojourn time, the incidence and health impact of treatment-related AEs and state-specific relative risks of mortality.

Resource costs

The model includes costs associated with drug acquisition, drug administration (vedolizumab and infliximab), surgery, consultant visits, hospitalisation episodes, blood tests, elective/emergency endoscopy and AEs.

5.2.2.3 Key structural assumptions employed within the manufacturer's model

The manufacturer's model employs the following key structural assumptions:

- All patients receiving biologic treatment will discontinue that treatment after 1-year irrespective of their current level of response
- Prior to 1-year, patients who commence biologic maintenance therapy are assumed to remain on biologic maintenance therapy even if they have moderate to severe disease
- Probabilities of induction response for conventional treatment are applied to patients who have previously discontinued biologic therapy; this assumes that response to non-biologic treatment is independent of previous biologic use
- Patients may transit to surgery immediately after receiving either conventional non-biologic therapy or biologic therapy during induction, but only after receiving conventional non-biologic therapy during maintenance
- Maintenance transition probabilities apply indefinitely over the model time horizon
- HRQoL is dependent on the severity of UC, medical treatments received and whether the patient is undergoing or has previously undergone surgery. Disutilities associated with the incidence of treatment-related AEs are assumed to apply for the entire cycle in which the patient receives that treatment
- Mortality is dependent on the severity of UC and whether the patient is undergoing or has previously undergone surgery.

5.2.3 Evidence sources used to inform the model

5.2.3.1 Summary of evidence sources used to inform the model parameters

Table 24 summarises the evidence sources used to inform the manufacturer's model parameters. The derivation of the manufacturer's model parameter values using these sources is described in further detail in the following sections.

Table 24: Summary of evidence sources used to inform the model parameters

Parameter group	Sources
Induction treatment – probabilities of transition between remission, mild, moderate to severe UC (response), moderate to severe UC (no response)	**(1) Mixed ITT population** Vedolizumab and conventional treatment – estimated directly from GEMINI1 trial[8] Surgery – n/a **(2) Anti-TNF-α naïve population** Vedolizumab, infliximab, adalimumab, golimumab and conventional treatment – transition vector estimated using odds ratios from the manufacturer's NMA[1] Surgery – n/a **(3) Anti-TNF-α failure population** Vedolizumab and conventional treatment – estimated directly from GEMINI1 trial[28] Surgery – n/a Probability patient achieving response is in moderate to severe UC state based on proportion observed in GEMINI1 trial populations (applied to all biologic therapies)
Maintenance treatment - probabilities of transition between remission, mild and moderate to severe UC (response or no response)	**(1) Mixed ITT population** Vedolizumab and conventional treatment – estimated by calibrating model predictions of probabilities of response and remission at 1-year against observed probabilities of response and remission during maintenance conditional on observed probabilities of response and remission during induction from GEMINI1 trial[8] Surgery – n/a **(2) Anti-TNF-α naïve population** Conventional treatment, vedolizumab, infliximab, adalimumab and golimumab – transition matrices derived by calibrating model predictions of probabilities of response and remission at 1-year against the expected probabilities of response/remission at 1-year using the manufacturer's NMAs of induction and maintenance treatments[1] Surgery – n/a **(3) Anti-TNF-α failure population** Vedolizumab and conventional treatment – estimated by calibrating model predictions of the probabilities of response and remission at 1-year against observed probabilities of response and remission during maintenance conditional on observed probabilities of response and remission during induction within GEMINI1 trial[28] Surgery – n/a
Colectomy rate	1-year estimates of probability of undergoing colectomy taken from Frolkis *et al*[48] and converted to reflect duration of induction and maintenance cycles within the model
Transition probabilities between surgery, post-surgical remission and post-surgery complications	Estimates taken from Loftus *et al*,[49] Mahadevan *et al*[50] and Xie *et al*[45] and converted to reflect model cycle length

Parameter group	Sources
Other cause mortality	Exponential model fitted to ONS life tables[51]
Relative risks of excess UC mortality	No excess risk assumed for remission or mild states. Relative risk of death due to UC in moderate to severe state based on Button *et al.*[52] Relative risk for surgery and post-surgery states based on Jess *et al.*[53]
Health-related quality of life	HRQoL for remission, mild and moderate to severe states taken from GEMINI1.[1] HRQoL for surgery and post-surgery states based on Punekar and Hawkins[43] which in turn are reported to be taken from Woehl *et al.*[54]
Probability of discontinuation of biologic therapy due to AEs during maintenance therapy	**(1) Mixed ITT population** Vedolizumab - estimate taken from GEMINI1 CSR[28] Infliximab, adalimumab and golimumab – not evaluated in this population **(2) Anti-TNF-α naïve population** Vedolizumab - estimate taken from GEMINI1 CSR[28] Infliximab – estimate taken from ACT1[24] Adalimumab – estimate taken from Suzuki *et al*[39] Golimumab – estimate taken from PURSUIT-Maintenance[41] **(3) Anti-TNF-α failure population** Vedolizumab - estimate taken from GEMINI1 CSR[28]
Incidence of AEs due to medical treatments	Naïve pooling of data from clinical trials[24,28,37,38,41]
AE disutilities	Disutilities based on Brown *et al*,[55] Porco *et al*,[56] Hornberger *et al*,[57] Beusterien *et al*[58] and Beusterien *et al*[59]
Drug acquisition costs (biologic and non-biologic therapies)	Price of vedolizumab sourced from manufacturer, including proposed Patient Access Scheme. Costs of other products taken from British National Formulary (BNF) 2013[9]
Infusion costs (vedolizumab and infliximab only)	Infusion cost taken from the Payment by Results (PbR) mandatory tariff 2013/14[60] (code FZ37F)
Usage of conventional non-biologic treatments	Interviews with two consultant gastroenterologists
Health state resource use and costs associated with endoscopy, consultant visits, blood tests and hospitalisations	Resource use and cost estimates taken from Tsai *et al*[44]
Costs of surgery	Taken from Buchanan *et al*[61]

n/a - not applicable

5.2.3.2 Induction treatment - transition probabilities between remission, mild, moderate to severe UC (response), moderate to severe UC (no response)

Within the mixed ITT and anti-TNF-α failure populations, the probabilities of remission, response (excluding remission), and no response for vedolizumab and conventional treatment were estimated directly using the GEMINI1 trial data[8,28] (see Table 25). Within the anti-TNF-α naïve population, the probabilities of remission, response (excluding remission), and no response for medical treatments were estimated using the manufacturer's NMA.[1]

Table 25: Probabilities of clinical response and clinical remission to induction treatment used in the manufacturer's model

Treatment option	Clinical response r/N (%)	Clinical remission r/N (%)	Source
Mixed ITT population			
Vedolizumab	106/225 (47.1%)	38/225 (16.9%)	Feagan *et al*[8]
Conventional treatment	38/149 (25.5%)	8/149 (5.4%)	
Infliximab	Not evaluated within this population		
Adalimumab			
Golimumab			
Surgery	Not applicable		
*Anti-TNF-α naïve population**			
Vedolizumab	62.35%	30.25%	Manufacturer's NMA[1]
Conventional treatment	34.29%	8.93%	
Infliximab	68.18%	33.41%	
Adalimumab	49.60%	15.14%	
Golimumab	57.05%	25.78%	
Surgery	Not applicable		
Anti-TNF-α failure population			
Vedolizumab	32/82 (39.0%)	8/82 (9.8%)	GEMINI1 CSR[28]
Conventional treatment	13/63 (20.6%)	2/63 (3.2%)	
Infliximab	Not evaluated within this population		
Adalimumab			
Golimumab			
Surgery	Not applicable		

* Number of patients not reported

Within the GEMINI1 ITT population,[8] 38 of 149 (25.5%) patients randomised to placebo and 106 of 225 (47.1%) patients randomised to vedolizumab achieved clinical response. Eight (5.4%) patients randomised to placebo and 38 (16.9%) patients randomised to vedolizumab achieved clinical remission. These values are used in the manufacturer's mixed ITT population analysis to inform estimates of the probability of achieving response/remission for vedolizumab and conventional therapy. Transition probabilities for response and remission are not applicable to the comparator of surgery.

Within the GEMINI1 anti-TNF-α failure population,[8] 13 of 63 (20.6%) patients randomised to placebo and 32 of 82 (39.0%) patients randomised to vedolizumab achieved clinical response. Within this subgroup, 2 (3.2%) patients randomised to placebo and 8 of 82 (9.8%) patients randomised to vedolizumab achieved clinical remission.[28] These values are used in the manufacturer's anti-TNF-α failure population analysis to inform estimates of the probability of achieving response/remission for vedolizumab and conventional therapy. Transition probabilities for response and remission are not applicable to the comparator of surgery.

Within the analysis of the anti-TNF-α naïve population, the probabilities of remission, response (excluding remission), and no response for each medical treatment were estimated using odds ratios for response and remission estimated using the manufacturer's NMA.[1] Transition probabilities for response and remission are not applicable to the comparator of surgery.

The manufacturer's model uses these estimates of proportions of patients achieving response and remission for each therapy, together with an estimate of the proportion of patients responding to treatment who have moderate to severe disease (13.2%, 10.1% and 20.9% in the mixed ITT, anti-TNF-α naïve and anti-TNF-α failure populations in GEMINI1, respectively) and the proportion of patients expected to undergo surgery, to estimate the initial transition vector from the moderate to severe (no response) state to remission, mild, moderate-severe UC (responders), moderate-severe UC (non-responders) and surgery for the induction phase. The probability of achieving response is adjusted by subtracting the proportion of patients who achieve remission (a subset of response). A fixed proportion of patients (0.58%) are assumed to undergo surgery during the first induction cycle, based on Frolkis *et al*.[48] Table 26 shows the transition vectors applied during the induction phase of the model.

Table 26: Transition vectors for induction therapy

	Remission	Mild	Moderate to severe (responders)	Moderate to severe (non-responders)	Surgery
Mixed ITT population					
Vedolizumab	0.169	0.240	0.062	0.523	0.006
Conventional treatment	0.054	0.168	0.034	0.739	0.006
Infliximab	Not evaluated within this population				
Adalimumab					
Golimumab					
Surgery	Not applicable				
Anti-TNF-α naïve population					
Vedolizumab	0.302	0.258	0.063	0.371	0.006
Conventional treatment	0.089	0.219	0.035	0.651	0.006
Infliximab	0.334	0.279	0.069	0.312	0.006
Adalimumab	0.151	0.294	0.050	0.498	0.006
Golimumab	0.258	0.255	0.058	0.424	0.006
Surgery	Not applicable				
Anti-TNF-α failure population					
Vedolizumab	0.098	0.211	0.082	0.604	0.006
Conventional treatment	0.032	0.131	0.043	0.788	0.006
Infliximab	Not evaluated within this population				
Adalimumab					
Golimumab					
Surgery	Not applicable				

5.2.3.3 Maintenance phase – transition probabilities

In all three populations, the probabilities of maintaining response to biological and non-biologic treatments were estimated using a process of model calibration. A linear programming approach was used to fit the 1-year model-predicted estimates of the proportion of patients in remission and response to those observed within from the GEMINI1 trial (mixed ITT and anti-TNF-α failure populations, vedolizumab and conventional treatment only) or estimated using the manufacturer's NMA (anti-TNF-α naïve population only, all medical treatments). The manufacturer's calibration method uses the Microsoft Excel Solver add-in to minimise the sum squared error of the "observed" and predicted estimates by manipulating seven of nine pre-colectomy transition probabilities (quantities x_1 to x_7 in Table 27) conditional on the model structure, the initial starting matrix for calibration and a series of constraints defined by the manufacturer (see MS[1] Appendix 10.15).

Table 27: 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

Table 28 shows the target data and the sources used in the manufacturer's calibration.

Table 28: Target data used in the calibration approach

Treatment option	Probability response at 1-year*	Probability remission at 1-year*	Source
Mixed ITT population			
Vedolizumab	0.197	0.070	Feagan *et al*[8]
Conventional treatment	0.040	0.020	
Infliximab	Not evaluated within this population		
Adalimumab			
Golimumab			
Surgery	Not applicable		
*Anti-TNF-α naïve population**			
Vedolizumab	0.358	0.144	Manufacturer's NMA[1]
Conventional treatment	0.093	0.058	
Infliximab	0.216	0.171	
Adalimumab	0.210	0.043	
Golimumab	0.222	0.123	
Surgery	Not applicable		
Anti-TNF-α failure population			
Vedolizumab	0.145	0.036	GEMINI1 CSR[28]
Conventional treatment	0.011	0.022	
Infliximab	Not evaluated within this population		
Adalimumab			
Golimumab			
Surgery	Not applicable		

* Estimates conditional on probability of response at the end of induction

As stated in the MS[1] (Appendix 10.15, page 395), the manufacturer's calibration method makes the following assumptions:

- *"No more than 99.5% of patients remain in remission over each 8-week cycle. Given the opportunity for the optimisation problem to have many optimal solutions, this constraint avoids the solution of all patients in remission remaining in remission.*
- *No more than 20% of patients with mild disease may transition into remission. This constraint is intended to depict the progressive nature of the disease.*
- *The probability of staying in mild disease is greater than the probability of going from mild disease to moderate-severe disease. In other words, we assumed that patients are more likely to remain in their current health state.*
- *The probability of staying in moderate-severe disease is greater than moving from moderate-severe disease to mild disease. In other words, we assumed that patients are more likely to remain in their current health state.*
- *The probability of moving from remission to moderate-severe (and vice versa) is zero. This constraint is based on the assumption that the disease progression/improvement rate is not fast enough to justify a transition between the two extreme states. All transition probabilities must be non-negative.*
- *The sum of probabilities from one state to all other states is constrained to equal 1. This constraint preserves the Markovian assumption."*[1]

For each biologic treatment option, the calibration process used the same initial transition matrix, as shown in Table 29. A different initial transition matrix was used for conventional treatment. Justification for using different initial matrices for different treatments is not reported within the MS.[1]

Table 29: 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	0.05	0.942
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	0.05	0.942

Table 30 shows the fitted pre-colectomy transition matrices estimated by the manufacturer using the calibration process; these values are directly used within the manufacturer's model.

Table 30: Fitted maintenance phase pre-colectomy transition probabilities

Mixed ITT population			
Vedolizumab			
From state\ To state	Remission	Mild	No response
Remission	0.97	0.03	0.00
Mild	0.09	0.60	0.32
No response	0.00	0.12	0.87
Conventional therapy			
From state\ To state	Remission	Mild	No response
Remission	0.91	0.09	0.00
Mild	0.03	0.55	0.42
No response	0.00	0.02	0.97
Anti-TNF-α naïve population			
Vedolizumab			
From state\ To state	Remission	Mild	No response
Remission	0.93	0.07	0.00
Mild	0.20	0.62	0.18
No response	0.00	0.28	0.71
*Conventional therapy**			
From state\ To state	Remission	Mild	No response
Remission	0.93	0.07	0.00
Mild	0.04	0.56	0.41
No response	0.00	0.03	0.96
Infliximab			
From state\ To state	Remission	Mild	No response
Remission	0.92	0.08	0.00
Mild	0.03	0.68	0.30
No response	0.00	0.16	0.83
Adalimumab			
From state\ To state	Remission	Mild	No response
Remission	0.98	0.02	0.00
Mild	0.15	0.56	0.29
No response	0.00	0.08	0.91
Golimumab			
From state\ To state	Remission	Mild	No response
Remission	0.95	0.05	0.00
Mild	0.05	0.62	0.32
No response	0.00	0.18	0.81
Anti-TNF-α failure population			
Vedolizumab			
From state\ To state	Remission	Mild	No response
Remission	0.99	0.01	0.00
Mild	0.12	0.57	0.31
No response	0.00	0.07	0.92
Conventional therapy			
From state\ To state	Remission	Mild	No response
Remission	0.84	0.16	0.00
Mild	0.00	0.59	0.41
No response	0.00	0.03	0.96

* The manufacturer's model includes a cell-referencing error which results in the conventional therapy matrix drawing in transition probabilities for infliximab (see Section 5.3). The corrected values are shown in this table.

5.2.3.4 Colectomy rate

The probability of undergoing colectomy was based on a systematic review and meta-analysis of population-based studies reported by Froklis *et al*.[48] This study reports the probability of undergoing surgery after a diagnosis of UC to be 4.9%, 11.6% and 15.6% at 1-, 5- and 10-years, respectively. The manufacturer's model uses the 1-year estimate (4.9%) and adjusts this to reflect the durations of the induction phase and the maintenance phase assuming a constant rate (induction probability=0.58%, maintenance probability = 0.77%).

5.2.3.5 Surgery and post-surgery transition probabilities

The probabilities of transiting between the surgery and post-surgery states within the manufacturer's model were estimated from the literature[45,49,50] (see Table 31). Loftus *et al* report that within 180 days post-colectomy, 15.3% patients underwent further unplanned surgeries. The manufacturer converted this 6-month probability to an 8-week probability assuming a constant rate; the manufacturer's model then applies this 8-week probability to all surgery-related states during each cycle of the model (probability=0.05). The probability of transiting from the surgery state to the post-surgery complications state was based on a study by Mahadevan *et al*,[50] which reported that an estimated 31% patients experience early complications of colectomy within 30 days; this estimate was converted to an 8-week probability assuming a constant rate (probability=0.50). The probability of experiencing late complications (transiting from post-surgery remission to post-surgery complications) was based on estimates of complications within 6-months of surgery reported by Loftus *et al*;[49] this estimate was converted to an 8-week probability assuming a constant rate (probability=0.17). The probability of transiting from post-surgery complications to post-surgery remission was based on a previous economic modelling study reported by Xie *et al*;[45] an estimate of 0.84 (time interval not specified in the paper, assumed by the manufacturer to reflect 1-year) was converted to an 8-week probability assuming a constant rate (probability=0.245). This matrix of probabilities is applied to each cycle within the maintenance phase.

Table 31: Surgery and post-surgery transition probabilities

Health State	Surgery	Post-surgery remission	Post-surgery complications
Surgery	0.050	0.450	0.500
Post-surgery remission	0.050	0.777	0.173
Post-surgery complications	0.050	0.245	0.705

5.2.3.6 UC-related and other-cause mortality

The model includes other-cause mortality and relative risk multipliers for moderate to severe, surgery and post-surgery UC states. The probability of dying from other causes was modelled by fitting an exponential curve to ONS life tables.[51] Relative risks for moderate to severe UC (relative risk=1.90), surgery (relative risk=1.30) and post-surgery UC states (relative risk=1.30) were taken from Button *et*

al[52] and Jess *et al*.[53] Each relative risk is applied to the baseline other-cause mortality rate during each cycle.

5.2.3.7 Incidence of adverse events

Estimates of the incidence of AEs were derived through a simple (unadjusted) pooling of AE data reported in the publications of the pivotal clinical trials of the biologics. The estimates used in the model are summarised in Table 32. AE probabilities were assumed to be the same across all three populations.

Table 32: Adverse event probabilities assumed within the manufacturer's model

Treatment option	Serious infection	Skin reaction	Acute hypersensitivity reaction	Source
Vedolizumab	0.002	0.001	0.000	GEMINI1 CSR[28]
Conventional treatment	0.003	0.006	0.000	
Infliximab	0.004	0.021	0.003	ACT1/2[24]
Adalimumab	0.001	0.000	0.000	ULTRA1/2[37,38]
Golimumab	0.002	0.009	0.001	PURSUIT[41]
Surgery	Not applicable			

5.2.3.8 Health-related quality of life

Table 33 summarises the health utility values assumed within the manufacturer's model. Utility scores for the pre-surgical states were derived from the GEMINI1 EQ-5D values for each state (all valuations at all study visits combined). No difference in HRQoL is assumed for moderate to severe responders and moderate to severe non-responders. Values for the post-surgery state reported within the submission were drawn from the previous economic evaluation reported by Punekar and Hawkins.[43] The value of 0.42 for post-surgery complications reflects the value for moderate to severe disease within Woehl *et al*.[54] The value of 0.60 for post-surgical remission does not actually reflect any of the values reported by Woehl *et al*.[54]

Table 33: Summary of health state utility values used in the manufacturer's model

Health state	Utility value	Source
Remission	0.86	GEMINI1 EQ-5D value for all study visits combined.[1] No difference assumed for moderate to severe responders and non-responders.
Mild	0.80	
Moderate to severe (responder)	0.68	
Moderate to severe (non-responder)	0.68	
Surgery	0.42	Reported in MS to be based on Punekar and Hawkins[43] but appears to be originally sourced from Woehl *et al.*[54] Woehl *et al* EQ-5D study misreferenced by Punekar as an epidemiology and resource use study.[62] Actual values used by Punekar do not coincide with estimates in the EQ-5D study reported in Woehl *et al.*[54]
Post-surgery complications	0.42	
Post-surgery remission	0.60	

Table 34 summarises the disutilities assumed within the model.

Table 34: Adverse event-related disutility values used in the manufacturer's model

Adverse event	Disutility value used in model	Source	Elicitation methods
Serious infection	-0.520	Brown *et al*[55]	Proxy utility values derived from 180 nurses using SG methods. Reported utility of 0.48 converted to disutility assuming baseline of perfect health.
TB	-0.550	Porco *et al*[56]	Elicitation method unclear. Reported utility of 0.45 converted to disutility assuming baseline of perfect health.
Malignancy (including lymphoma)	-0.195	Hornberger *et al*[57]	Elicitation method unclear. Reported utility of 0.805 for follicular lymphoma (pre-progression) converted to disutility assuming baseline of perfect health.
Acute hypersensitivity reactions	-0.110	Beusterien *et al*[58]	Cross-sectional SG using members of the general public. Disutility directly estimated as part of analysis.
Skin site reactions	-0.030	Beusterien *et al*[59]	Cross-sectional SG using members of the general public. Disutility directly estimated as part of analysis.

The disutility for serious infection was estimated using a published economic evaluation of treatments for advanced breast cancer.[55] Within this study, standard gamble (SG) methods were used to elicit utility values for a variety of health states using 180 nurses as proxy. The disutility estimated by the manufacturer assumes a baseline utility of 1.0 (perfect health).

The disutility for TB was estimated using a published economic evaluation of TB evaluation and treatment of newly-arrived immigrants.[56] The elicitation methods within this study are unclear; estimates appear to be based on other literature and assumptions. The disutility estimated by the manufacturer assumes a baseline utility of 1.0 (perfect health).

The disutility for malignancy was estimated using a published economic evaluation of rituximab plus cyclophosphamide, vincristine and prednisolone for advanced follicular lymphoma.[57] The elicitation methods within this study are unclear, as reported estimates appear to be based on other literature. The disutility estimated by the manufacturer assumes a baseline utility of 1.0 (perfect health).

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.[58] A disutility for grade 3/4 pyrexia was reported; this value was used directly in the manufacturer'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.[59] A disutility of 0.03 was reported by UK responders; this value was used directly in the manufacturer's model.

It should be noted that the disutilities associated with treating TB and lymphoma are not actually used in the manufacturer's model as the incidence rate for these events is zero for all treatment options in all three populations.

5.3.2.9 Biologic discontinuation rate due to adverse events (maintenance therapy)

The model assumes that a proportion of patients receiving biologic treatment will discontinue therapy due to AEs; the probabilities of discontinuation of each biologic treatment during each maintenance cycle within the manufacturer's model are summarised in Table 35. Within the mixed ITT population and the anti-TNF-α failure populations, the proportions of patients discontinuing biologic treatment were estimated using observed discontinuation rates from the GEMINI1 trial.[28] Within the anti-TNF-α naïve population, discontinuation rates for each treatment group were taken from individual clinical trials of each biologic treatment.[24,28,39,41]

Table 35: Probability of biologic discontinuation during each maintenance cycle

Treatment option	Probability of discontinuation	Source
Mixed ITT population		
Vedolizumab	0.0088	GEMINI1 trial[1]
Conventional treatment	Not applicable	
Infliximab	Not evaluated within this population	
Adalimumab		
Golimumab		
Surgery	Not applicable	
Anti-TNF-α naïve population		
Vedolizumab	0.0064	GEMINI1 trial[1]
Conventional treatment	Not applicable	
Infliximab	0.0127	ACT1[24]
Adalimumab	0.0191	Suzuki *et al*[39]
Golimumab	0.0080	Pursuit-Maintenance[41]
Surgery	Not applicable	
Anti-TNF-α failure population		
Vedolizumab	0.0143	GEMINI1 trial[1]
Conventional treatment	Not applicable	
Infliximab	Not evaluated within this population	
Adalimumab		
Golimumab		
Surgery	Not applicable	

5.3.2.10 Drug acquisition and administration costs

The acquisition costs of biologic and non-biologic therapies included in the manufacturer's model are summarised in Table 36.

Table 36: Acquisition costs assumed within the manufacturer's model

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	12	4	£5,035.44	£1,678.48
Adalimumab (40mg prefilled pen/syringe)	£352.14	8	4	£2,817.12	£1,408.56
Golimumab (50mg prefilled pen/syringe)	£762.97	6	2	£4,577.82	£1,525.94
Conventional treatment	£3.66	Mix of various products		£153.60	£204.80*

* Assumed to be £102.40 for patients whilst receiving biologic treatment

The basic NHS list price of vedolizumab is £2,050 per 300mg vial. The manufacturer's model includes a lower drug acquisition cost to reflect the proposed PAS for vedolizumab; the price used in the model is [redacted] per 300mg vial. The proposed PAS takes the form of a simple price discount for the NHS. The acquisition costs of infliximab, adalimumab and golimumab were based on drug prices reported within the BNF 2013.[9]

A number of conventional non-biologic treatments (balsalazide, mesalazine, olsalazine, sulfasalazine and budesonide, prednisolone, azathioprine, 6-MP and methotrexate) are assumed within the model; these appear to be based on a daily cost which is applied indefinitely. Usage of these products was based on expert opinion from two gastroenterologists; the prices of each product are reported in Table 79 of the MS.[1] The model assumes that whilst patients are receiving biologic therapy, the costs of conventional therapies are halved (cost=£102.40).

The costs associated with the administration of infusional biologics (infliximab and vedolizumab) were taken from the PbR tariff 2012/13[60] and were assumed to be £308 per administration visit.

5.3.2.11 UC health state resource costs

Resource use associated with consultant visits, hospitalisations, blood tests and endoscopy were taken from Tsai *et al*[44] (see Table 37). The cost of surgery was based on Buchanan *et al.*[61] The MS states that unit costs associated with UC health state costs were based on NHS Reference Costs 2011-2012, however the values used within the manufacturer's model (and presented in Table 80 of the MS[1]) are actually taken directly from Tsai *et al*[44] and uplifted to current prices.

Table 37: UC health state resource costs

Resource component	Unit cost	Units per cycle					
		Remission	Mild	Moderate to severe	Surgery	Post-surgery remission	Post-surgery complications
Consultant visit	£105.73	0.31	0.69	1	-	0.23	0.27
Hospitalisation	£3,399.36	0.46	0.05	0.05	-	-	0.50
Surgery	£13,577.27	-	-	-	1	-	-
Blood tests	£3.35	0.50	0.6	1	-	0.23	0.50
Elective endoscopy	£1,497.12	0.03	0.08	0.13	-	0.18	0.10
Emergency endoscopy	£2,026.09	-	0.04	0.12	-	0.08	0.02
Per-cycle cost	-	£236.52	£424.02	£957.77	£13,577.27	£467.65	£1,913.24

5.3.2.12 Costs of managing adverse events

Unit costs associated with the management of AEs associated with biologic and non-biologic treatments were taken from NHS Reference Costs 2011-2012[63] and three previous NICE Technology Appraisals (see Table 38).

Table 38: Unit costs associated with managing adverse events

Adverse Event	Total Cost	Source
Serious infection	£1,470.00	NHS Reference Costs 2011/12. Average of 5 different types of serious infections: sepsis, pneumonia, urinary tract infection, respiratory infection, and bronchitis
TB	£2,272.00	NHS Reference Costs 2011/12. Average of non-elective short-stay and long-stay tuberculosis
Lymphoma	£14,975.00	NICE (2003), NICE (2012), and NICE (2011). Average of lymphoma costs from three technological appraisals for rituximab (TA65, TA243, and TA226)
Hypersensitivity	£3,188.00	NHS Reference Costs 2011/12. Average of non-elective short-stay and long-stay pyrexia
Injection site reactions	£1,363.28	NHS Reference Costs 2011/12. Average of procedures associated with skin disorders

It should be noted that the costs associated with treating TB and lymphoma are not actually used in the manufacturer's model as the incidence rate for these events is zero for all treatment options in all three populations.

5.3.3 Cost-effectiveness results presented by the manufacturer

5.3.3.1 Central estimates of cost-effectiveness presented by the manufacturer

Table 39 summarises the estimated health gains and costs for each treatment option within each of the three populations considered within the manufacturer's model. It should be noted that the manufacturer did not undertake a fully incremental analysis hence all ICERs presented in the table are pairwise comparisons of vedolizumab versus each individual comparator. It should also be noted that

the headline cost-effectiveness results presented by the manufacturer 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 manufacturer, probabilistic ICERs are not presented within the MS.[1]

Table 39: Central estimates of cost-effectiveness presented by the manufacturer (based on point estimates of parameters)

Option	QALYs	Costs	Incremental QALYs	Incremental costs	Pairwise ICER (vedolizumab versus comparator)
Mixed ITT population					
Vedolizumab	5.55	£77,056	-	-	-
Conventional therapy	5.40	£71,925	0.15	£5,131	£33,297
Surgery	4.28	£107,831	1.27	-£30,775	dominating
Anti-TNF-α naïve population					
Vedolizumab	5.90	£69,075	-	-	-
Infliximab	5.82	£73,952	0.08	-£4,877	dominating
Golimumab	5.79	£70,387	0.11	-£1,312	dominating
Adalimumab	5.76	£68,157	0.14	£918	£6,634
Conventional therapy	5.56	£67,406	0.34	£1,669	£4,862
Surgery	4.28	£107,831	1.67	-£38,756	dominating
Anti-TNF-α failure population					
Vedolizumab	5.46	£78,409	-	-	-
Conventional therapy	5.37	£72,570	0.09	£5,839	£64,999
Surgery	4.28	£107,831	1.182	-£29,422	dominating

Within the mixed ITT population, the manufacturer's base case analysis suggests that vedolizumab dominates surgery. Within this population, the ICER for vedolizumab versus conventional therapy is estimated to be £33,297 per QALY gained.

Within the anti-TNF-α naïve population, the manufacturer's base case analysis indicates that vedolizumab dominates infliximab, golimumab and surgery. The pairwise ICER for vedolizumab versus adalimumab is estimated to be £6,634 per QALY gained. The pairwise ICER for vedolizumab versus conventional therapy is estimated to be £4,862 per QALY gained.

Within the anti-TNF-α failure population, the manufacturer's base case analysis indicates that vedolizumab dominates surgery. The pairwise ICER for vedolizumab versus conventional therapy is estimated to be £64,999 per QALY gained.

5.3.3.2 Uncertainty analysis conducted by the manufacturer

The manufacturer conducted a range of uncertainty analyses including PSA, deterministic one-way sensitivity analyses and scenario analyses for all comparators in all three populations; the results of these analyses are summarised below.

Probabilistic sensitivity analysis

PSA was conducted in all three populations. The MS presents the results of the PSA as pairwise cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs, see MS[1] Figures 42-50). The probability that vedolizumab produces the greatest amount of net benefit at cost-effectiveness thresholds of £20,000 per QALY gained and £30,000 per QALY gained are summarised within Table 40. Incremental CEACs comparing all options in each population were not included within the MS.[1]

Table 40: Estimated probability vedolizumab produces the greatest net benefit at willingness to pay thresholds of £20,000 and £30,000 per QALY gained (pairwise comparisons – vedolizumab versus comparator, read from manufacturer's CEACs by ERG)

Population	Conventional therapy	Infliximab	Adalimumab	Golimumab	Surgery
Probability vedolizumab produces greatest net benefit at λ=£20,000 per QALY gained					
Mixed ITT population	0.20	not evaluated	not evaluated	not evaluated	1.00
Anti-TNF-α naïve population	0.91	0.97	0.82	0.97	1.00
Anti-TNF-α failure population	0.05	not evaluated	not evaluated	not evaluated	1.00
Probability vedolizumab produces greatest net benefit at λ=£30,000 per QALY gained					
Mixed ITT population	0.40	not evaluated	not evaluated	not evaluated	1.00
Anti-TNF-α naïve population	0.96	0.97	0.91	0.99	1.00
Anti-TNF-α failure population	0.10	not evaluated	not evaluated	not evaluated	1.00

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 0.20. Assuming a threshold of £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is approximately 0.40. Assuming a cost-effectiveness threshold of £20,000 per QALY gained or £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than surgery is approximately 1.0.

Within the anti-TNF-α naïve population, assuming a cost-effectiveness threshold of £20,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment, infliximab, adalimumab, golimumab and surgery is greater than 0.82. Assuming a cost-effectiveness threshold of £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment, infliximab, adalimumab, golimumab and surgery is greater than 0.91.

Within the anti-TNF-α failure population, assuming a cost-effectiveness threshold of £20,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is 0.05. Assuming a cost-effectiveness threshold of £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than conventional treatment is 0.10. Assuming a cost-effectiveness threshold of £20,000 per QALY gained or £30,000 per QALY gained, the probability that vedolizumab produces more net benefit than surgery is approximately 1.0.

Deterministic one-way sensitivity analysis

A number of one-way sensitivity analyses were conducted whereby each variable was replaced with higher and lower values than those used in the base case analysis (see Table 68 of the MS[1]). Starting age and average weight were varied by +/-5% whilst relative risks of all-cause mortality, health state utilities, AE disutilities, health state costs, and AE costs were varied by +/-20%. Drug costs, including those associated with conventional therapy, were not varied in the one-way sensitivity analysis. For all other parameters, the manufacturer based the upper and lower values on the 95% confidence intervals. The 15 variables that had the greatest impact on the ICER for vedolizumab versus each comparator in the three populations were presented as tornado diagrams (see MS[1] Figures 33-41). For comparisons of vedolizumab versus surgery in all three populations, the surgery and post-surgical transition probabilities, health state costs, and health state utilities had the largest impacts on the ICER for vedolizumab. For the other comparators, the parameters that had the largest impact on the ICER included remission transition probabilities for conventional therapy, vedolizumab, and infliximab, the efficacy of vedolizumab, infliximab, and conventional therapy in the initial response period and health state costs.

Scenario analysis

The manufacturer reports cost-effectiveness results across five groups of scenarios (see MS Section 7.6.9, summarised in Table 41); these involved altering the model time horizon, using alternative sources of utility values (Punekar *et al*,[43] Tsai *et al*[44] and Arseneau *et al*[64] as separate scenarios), excluding the excess mortality risk for UC, using 10-week response data rather than 6-week response data and extending the maximum duration of biologic treatment from 1 year to 3 years. These scenario analyses are presented in the MS for all pairwise comparisons of vedolizumab in all three populations except for the comparison of vedolizumab versus surgery in the anti-TNF-α naïve population; this analysis was not presented within the MS.[1]

Table 41: Summary results of manufacturer's scenario analyses

Scenario	Incremental cost per QALY gained (pairwise - vedolizumab versus comparator)				
	Conventional therapy	Infliximab	Adalimumab	Golimumab	Surgery
Mixed ITT population					
Base case	£33,297	not evaluated	not evaluated	not evaluated	dominating
1-year time horizon	£188,640	not evaluated	not evaluated	not evaluated	dominating
Lifetime horizon	£20,599	not evaluated	not evaluated	not evaluated	dominating
Utilities from Punekar *et al*[43]	£17,857	not evaluated	not evaluated	not evaluated	£117,134*
Utilities from Arseneau *et al*[64]	£18,008	not evaluated	not evaluated	not evaluated	£26,438*
Utilities from Tsai *et al*[44]	£18,627	not evaluated	not evaluated	not evaluated	£46,733*
Excluding excess mortality risk	£33,675	not evaluated	not evaluated	not evaluated	dominating
10-week vedolizumab response assessment	£31,414	not evaluated	not evaluated	not evaluated	dominating
Maximum time on treatment =3 years	£39,575	not evaluated	not evaluated	not evaluated	dominating
Anti-TNF-α naïve population					
Base case	£4,862	dominating	£6,634	dominating	dominating
1-year time horizon	£139,885	dominating	£135,406	£51,918	not reported
Lifetime horizon	dominating	dominating	dominating	dominating	not reported
Utilities from Punekar *et al*[43]	£2,469	dominating	£3,342	dominating	not reported
Utilities from Arseneau *et al*[64]	£2,375	dominating	£3,190	dominating	not reported
Utilities from Tsai *et al*[44]	£2,375	dominating	£3,459	dominating	not reported
Excluding excess mortality risk	£4,647	dominating	£6,452	dominating	not reported
10-week vedolizumab response assessment	£12,726	dominating	£21,006	£6,916	not reported
Maximum time on treatment =3 years	£26,152	dominating	£50,607	£15,548	not reported
Anti-TNF-α failure population					
Base case	£64,999	not evaluated	not evaluated	not evaluated	dominating
1-year time horizon	£230,671	not evaluated	not evaluated	not evaluated	dominating
Lifetime horizon	£44,132	not evaluated	not evaluated	not evaluated	dominating
Utilities from Punekar *et al*[43]	£35,830	not evaluated	not evaluated	not evaluated	£67,866*
Utilities from Arseneau *et al*[64]	£35,355	not evaluated	not evaluated	not evaluated	£22,164*
Utilities from Tsai *et al*[44]	£37,589	not evaluated	not evaluated	not evaluated	£35,732*
Excluding excess mortality risk	£66,025	not evaluated	not evaluated	not evaluated	dominating
10-week vedolizumab response assessment	£55,763	not evaluated	not evaluated	not evaluated	dominating
Maximum time on treatment =3 years	£55,149	not evaluated	not evaluated	not evaluated	dominating

* Results are in the South West quadrant of the cost-effectiveness plane: the ICER represents the cost-effectiveness of surgery compared with vedolizumab

Model time horizon of 1 year and lifetime (63 years)

Compared with the base case ICERs, truncating the model time horizon to 1-year increases the ICER for vedolizumab compared against all comparators in all populations substantially. Assuming a 1-year horizon, the pairwise ICERs for vedolizumab range from £51,918 per QALY gained in the anti-TNF-α naïve population (vedolizumab versus golimumab) to £230,671 per QALY gained in the anti-TNF-α failure population (vedolizumab versus conventional therapy). Extending the model time horizon to the patients' remaining lifetime decreases the ICERs in all populations and leads to vedolizumab dominating all comparators evaluated in the anti-TNF-α naïve population. Surgery remains dominated in both scenarios in the mixed ITT population and the anti-TNF-α failure population.

Alternative sources of utility values

Using alternative values for the health utility parameters reduces the ICER for vedolizumab versus all comparators except surgery. In the mixed ITT population, the ICER for vedolizumab versus conventional therapy is reduced from £33,297 per QALY gained to below £19,000 per QALY gained. For vedolizumab versus surgery in the mixed ITT and anti-TNF-α failure populations, the use of alternative utility values results in a situation whereby surgery produces more QALYs than vedolizumab. As vedolizumab has a lower estimated cost than surgery, this means that the ICER moves to the South West quadrant of the cost-effectiveness plane (vedolizumab is less expensive and less effective than surgery). The ICERs presented by the manufacturer in these cases represent the cost-effectiveness of surgery compared with vedolizumab and range from £22,164 per QALY gained to £67,866 per QALY gained in the anti-TNF-α failure population and £26,438 per QALY gained to £117,134 per QALY gained in the mixed ITT population.[1] These scenarios were not evaluated for comparisons of vedolizumab versus surgery in the anti-TNF-α naïve population. Infliximab and golimumab remain dominated in all scenarios in the anti-TNF-α naïve population.

No additional UC mortality risk

The exclusion of UC specific mortality has only a very minor impact on the ICERs for vedolizumab versus all other comparators.

10-week continuation rule

The use of Week 10 response data rather than Week 6 response data increases the ICER for vedolizumab compared against conventional therapy, adalimumab and golimumab in the anti-TNF-α naïve population. Within this population, the ICER for vedolizumab versus adalimumab is increased from £6,634 per QALY gained to £21,006 per QALY gained, whilst for the comparison against golimumab, vedolizumab moves from a position of dominance to an ICER of £6,916 per QALY gained. The ICER for vedolizumab versus conventional therapy is increased from £4,862 per QALY gained to £12,726 per QALY gained. In the mixed ITT population and anti-TNF-α failure

populations, the ICERs for vedolizumab versus conventional therapy also decrease. Surgery and infliximab remain dominated in all scenarios evaluated.

Duration of treatment increased from one year to three years

Changing the maximum treatment duration to 3 years increases the ICER for vedolizumab within the mixed ITT and anti-TNF-α failure populations (except for comparisons against infliximab and surgery as these remain dominated by vedolizumab). For vedolizumab versus adalimumab, the assumption of a 3 year maximum treatment duration increases the ICER from £6,634 per QALY gained to £50,607 per QALY gained. In the anti-TNF-α failure population, the ICER for vedolizumab versus conventional therapy is decreased from £64,999 per QALY gained to £55,149 per QALY gained.

5.3 Critical appraisal of the manufacturer's model

5.3.1 Summary of main issues identified through critical appraisal of the manufacturer's model

This section presents a critical appraisal of the manufacturer's health economic analysis and the model upon which this analysis is based. This process was undertaken using published checklists[65,66] for the critical appraisal of economic evaluations and models together with scrutiny and a partial re-build of the manufacturer's model by the ERG.

The manufacturer's economic analysis is subject to a number of issues, as summarised in Box 1.

Box 1: Main issues identified through critical appraisal of the manufacturer's model

1. Deviations from the NICE Reference Case[67] and final NICE scope[6]
2. Use of continuation and discontinuation rules for biologic treatments
3. Questionable methods for estimating maintenance transition probabilities
4. Partial use of the NMA within the health economic model
5. Implausible transition probabilities between surgery and post-surgical states
6. Questionable assumptions regarding health-related quality of life
7. Concerns relating to resource use and unit costs within the manufacturer's model
8. Inappropriate assumptions concerning adverse events of biologic and non-biologic therapies
9. Use of incremental cost-effectiveness ratios based on point estimates of parameters (deterministic)
10. Issues concerning model implementation

These issues are discussed in the following sections.

(1) Deviations from the NICE Reference Case[67] and final NICE scope[6]

Table 42 demonstrates the extent to which the manufacturer's economic analysis adheres to the NICE Reference Case.[67]

Table 42: Adherence of the manufacturer's economic analysis to the NICE Reference Case

Element of HTA	Reference Case	ERG comments
Defining the decision problem	The scope developed by NICE	The scope of the manufacturer's health economic analysis is partly in line with that developed by NICE.[6]
Comparator(s)	As listed in the scope developed by NICE	Surgery and conventional non-biologic treatments are considered in all three population analyses. Other biologics (infliximab, adalimumab and golimumab) are evaluated only for the anti-TNF-α naïve subgroup. Other biologic comparators are not considered in the mixed ITT or anti-TNF-α failure populations. Further anti-TNF-α agents may be used in patients after failure of prior anti-TNF-α therapy (although the effectiveness of golimumab and infliximab is unclear); this is not considered within the manufacturer's health economic analysis.
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Health benefits for patients are measured and valued over a 10-year horizon.
Perspective on costs	NHS and PSS	An NHS perspective was adopted. PSS costs were not considered to be relevant.
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The economic analysis takes the form of a cost-utility analysis. A fully incremental analysis is not presented within the MS.[1] Vedolizumab is compared against each comparator in a pairwise fashion.
Time horizon	Long enough to reflect all important differences between the technologies being compared	A 10-year time horizon is used in the manufacturer's base case analysis. A lifetime horizon considered in a sensitivity analysis.
Synthesis of evidence on health effects	Based on systematic review	The manufacturer's NMAs of the effects of biologic and conventional treatments are based on a systematic review, however only the induction NMA is directly used in the model (maintenance transition probabilities were calibrated). Transition probabilities relating to surgery and post-surgical states were drawn from a targeted review of the literature (details not provided by the manufacturer)
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults	Health outcomes are valued using QALYs, derived from patients with UC using the EQ-5D questionnaire.
Source of data for measurement of HRQoL	Reported directly by patients and/or carers	
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	HRQoL estimates valued using the preferences of the general public.
Equity weighting	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.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Costs relate to NHS resource use and are valued using prices relevant to the NHS.
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	Costs and health outcomes are discounted at 3.5%.

The manufacturer's model has been implemented partly in line with NICE's Reference Case (see Table 42) and the economic analysis is generally in line with the final NICE scope.[6] Three deviations from the NICE Reference Case warrant more detailed discussion: these relate to (i) relevant comparators excluded from analyses of mixed ITT and anti-TNF-α naïve populations; (ii) use of a 10-year time horizon, and (iii) failure to undertake a fully incremental analysis.

(i) Relevant comparators missing from analyses of mixed ITT and anti-TNF-α naïve populations

In the mixed ITT and anti-TNF-α failure populations, vedolizumab is compared against conventional non-biologic therapy and surgery. Anti-TNF-α agents (infliximab, adalimumab and golimumab) are not included in the health economic comparisons within these two populations, but are included in the analysis within the anti-TNF-α naïve population. Since 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, yet these therapies are considered within the manufacturer's analysis of the anti-TNF-α naïve population but not the broader mixed ITT population, it is unclear how one should interpret the results of the analysis. Furthermore, the manufacturer's analysis within the anti-TNF-α failure population excludes all other biologic therapies. The use of a second anti-TNF-α agent following the failure of a first anti-TNF-α agent may be possible, however there is only limited evidence available to estimate efficacy.[8,38] Within the MS,[1] the manufacturer argues that the GEMINI1 anti-TNF-α failure subgroup and the ULTRA2[38] anti-TNF-α experienced subgroups may not be comparable. The ERG agrees with this statement, however adalimumab was included in the manufacturer's NMA of the anti-TNF-α failure subgroup but excluded from the health economic analysis of this subgroup. Further, whilst the effectiveness of golimumab and infliximab is not clear within this population, the exclusion of these therapies altogether is questionable. Calcineurin inhibitors are not evaluated in any of the three populations.

(ii) Use of a 10-year time horizon

The manufacturer's model adopts a 10-year time horizon in the base case analysis. The NICE methods guide stipulates that the time horizon of the analysis should be long enough to reflect all important differences in costs or outcomes between the technologies being compared.[67] The MS[1] (Table 53) states that the choice of time horizon is *"in line with model by Tsai et al. (2008) and with previous models submitted to NICE."* However, this represents an insufficient justification as it is not clear whether all relevant differences in health gains and costs would be captured within this 10-year period. The manufacturer does present pairwise ICERs for vedolizumab versus each comparator within all three populations over a lifetime horizon (see MS[1] Tables 107 to 114); in some instances these suggest very different (more favourable) results from the manufacturer's base case analysis (see Table 43). 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 evidence over a lifetime horizon is subject to considerable uncertainty.

Table 43: Comparison of manufacturer's pairwise ICERs over different time horizons

	Pairwise ICER (vedolizumab versus comparator)		
	10-years (base case)	1-year	Lifetime
Mixed ITT population			
Vedolizumab	-	-	-
Conventional therapy	£33,297	£188,640	£20,599
Surgery	dominating	dominating	dominating
Anti-TNF-α naïve population			
Vedolizumab	-	-	-
Infliximab	dominating	dominating	dominating
Golimumab	dominating	£51,918	dominating
Adalimumab	£6,634	£135,406	dominating
Conventional therapy	£4,862	£139,885	dominating
Surgery	dominating	not evaluated	not evaluated
Anti-TNF-α failure population			
Vedolizumab	-	-	-
Conventional therapy	£64,999	£230,671	£44,132
Surgery	dominating	dominating	dominating

(iii) Use of pairwise rather than fully incremental comparisons

The results of the manufacturer's health economic analyses are not presented as fully incremental comparisons. Within each of the three populations, vedolizumab was compared against each comparator in a pairwise fashion. For example, in the mixed ITT population, vedolizumab was compared against conventional therapy and vedolizumab was compared against surgery, but conventional therapy was not compared against surgery in the same analysis. In the absence of a fully incremental comparison of all relevant treatment options, the correct interpretation of the manufacturer's health economic results is problematic. Table 44 presents a fully incremental re-analysis of the manufacturer's base case undertaken by the ERG.

Table 44: Fully incremental analysis using the manufacturer's base case model

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER (vedolizumab versus comparator)
Mixed ITT population					
Vedolizumab	5.55	£77,056	0.15	£5,131	£33,297
Conventional therapy	5.40	£71,925	-	-	-
Surgery	4.28	£107,831	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	5.90	£69,075	0.14	£918	£6,634
Infliximab	5.82	£73,952	-	-	dominated
Golimumab	5.79	£70,387	-	-	dominated
Adalimumab	5.76	£68,157	0.21	£751	£3,664
Conventional therapy	5.56	£67,406	-	-	-
Surgery	4.28	£107,831	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	5.46	£78,409	0.09	£5,839	£64,999
Conventional therapy	5.37	£72,570	-	-	-
Surgery	4.28	£107,831	-	-	dominated

Within the mixed ITT population, the manufacturer's model suggests that surgery is dominated as it produces fewer health gains and is more costly than both conventional therapy and vedolizumab. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.15 QALYs at an incremental cost of £5,131; the ICER for vedolizumab versus conventional therapy is estimated to be £33,297 per QALY gained.

Within the anti-TNF-α naïve population, the manufacturer's model suggests that surgery is expected to be dominated by medical therapies. Vedolizumab is expected to be the most effective option. Infliximab and golimumab are expected to be dominated by vedolizumab and are ruled out of the analysis. The ICER for adalimumab versus conventional therapy is estimated to be £3,664 per QALY gained, whilst the ICER for vedolizumab versus adalimumab is estimated to be £6,634 per QALY gained.

Within the anti-TNF-α failure population, the manufacturer's model suggests that surgery is expected to be dominated by conventional therapy. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.09 QALYs at an incremental cost of £5,839; the ICER for vedolizumab versus conventional therapy is estimated to be £64,999 per QALY gained.

(2) Use of continuation and discontinuation rules for biologic treatments

As noted in Section 5.2.1, the manufacturer's model assumes that all patients who are still receiving anti-TNF-α therapy at 1-year will discontinue and subsequently receive non-biologic therapies, irrespective of whether they are currently responding to treatment. Page 188 of the MS states:

"Within the model discontinuation of treatment can be due to a lack of response by the end of the induction phase or due to adverse events. In addition, it is assumed in the model that treatment with a biologic (VEDO, infliximab, adalimumab or golimumab) is limited to one year and all patients on therapy at week 54 of the model switch to conventional therapy."[1]

This proposed discontinuation rule is not discussed elsewhere in the MS. In response to a request for clarification from the ERG[4] regarding the rationale for this assumption (question B2), the manufacturer stated:

"In the absence of a stopping rule, it is uncertain what the average duration of treatment is with vedolizumab, golimumab, adalimumab and infliximab. A treatment duration of 1-year in responding patients was chosen to reflect the follow-up within clinical trials, particularly the GEMINI I trial

upon which the model is mostly based. The impact on the ICER of patients receiving vedolizumab, golimumab, adalimumab or infliximab for 3 years was presented in the submission."[4]

Whilst there is uncertainty with respect to the long-term efficacy of vedolizumab, infliximab, adalimumab and golimumab as the randomised phases of trials of these therapies adopted a maximum follow-up of 54 weeks, the wording of the marketing authorisations for the biologics does not stipulate if or when responding patients should discontinue therapy.[10,46,47,68] Furthermore, it is not clear whether the discontinuation rule for treatment with vedolizumab and other biologic therapies adopted in the manufacturer's model would be adhered to in routine practice as it may not be preferable to patients and clinicians to withdraw biologic therapy when a patient is still obtaining clinical benefit from it.

The manufacturer undertook a scenario analysis (see MS[1] Tables 107 to 114) to explore the impact of adopting a longer period of time on biologic treatment (all patients discontinue at 3-years). Table 45 presents the results of this scenario as a fully incremental analysis. The ICER increases by around £6,000 for vedolizumab compared with conventional therapy in the mixed ITT population and decreases by around £10,000 for vedolizumab compared against conventional therapy in the anti-TNF-α failure population. In the anti-TNF-α naïve population, the ICER for vedolizumab increases substantially from £6,634 per QALY gained to £50,607 per QALY gained as the next most effective comparator changes from infliximab to adalimumab. It should be noted however that there is no obvious rationale for assuming a 3-year maximum treatment duration either; the ERG believes that such a discontinuation rule should not have been included in the manufacturer's base case analysis.

Table 45: Fully incremental analysis assuming 3-year discontinuation

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	5.603	£80,073	0.206	£8,148	£39,575
Conventional therapy	5.397	£71,925	-	-	-
Surgery	4.281	£107,831	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	5.924	£77,052	0.130	£6,556	£50,607
Adalimumab	5.794	£70,496	0.239	£3062	£12,812
Golimumab	5.772	£74,693	-	-	ext dom
Infliximab	5.765	£80,378	-	-	dominated
Conventional therapy	5.555	£67,407	-	-	-
Surgery	4.281	£107,831	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	5.514	£80,326	0.141	£7,757	£55,149
Conventional therapy	5.373	£72,570	-	-	-
Surgery	4.281	£107,831	-	-	dominated

Ext dom – extendedly dominated

It should also be noted that the model assumes that patients will continue to receive biologic maintenance therapy up to 1-year even if response was lost after induction. The use of such a continuation rule is unlikely to be clinically realistic.

(3) Questionable methods for estimating maintenance transition probabilities

The manufacturer's model estimates transition probabilities for the maintenance phase using a method of linear programming optimisation (see MS[1] Section 7.3.2 and Appendix 10.15). As noted in Section 5.2.3.3, this approach involves using the Microsoft Excel Solver add-in to determine seven non-zero transition probabilities by comparing the model-predicted proportion of patients in remission or response at 1-year against the observed proportion of patients in remission or response at 1-year in GEMINI1 or against the predicted proportion based on the manufacturer's NMAs of induction and maintenance therapies (note - the target datapoints and their derivation depend on the population considered in the analysis). The ERG believes that there are five problems associated with this approach:

i) The manufacturer's use of calibration methods discards their empirical trial data.
ii) The initial starting matrix of transitions used in the optimisation approach appears to be largely arbitrary
iii) The constraints imposed in the optimisation approach appear to be largely arbitrary
iv) Fitting seven unknown parameters to two known datapoints is likely to result in overfitting. Many possible combinations of transition probabilities could fit the two 1-year datapoints on response and remission.
v) The fitting process ignores those patients who achieved response but had moderate to severe disease.

The manufacturer has access to patient-level GEMINI1 trial data on the observed transitions between remission/mild/moderate-to-severe within the maintenance phase; these could have been used to directly calculate the observed probability of transiting between pre-colectomy health states (remission, mild and moderate to severe). This approach could have been adopted both for the vedolizumab group and the conventional management group. Instead however, the manufacturer has adopted an approach which "guesses" seven unknown parameters by fitting these to two datapoints (probabilities of response and remission at 1-year) 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. These constraints and starting matrices (see Section 10.15 of the MS[1]) are based on assumptions made by the manufacturer which are not adequately justified using evidence. Arbitrarily, a different starting matrix is used for biologic therapies and for conventional therapies. The ERG sought further clarification on the constraints and initial matrices used within the calibration process (question B17); the manufacturer did not provide

justification regarding the values used as constraints and stated that *"The initial values were chosen to be reasonable estimates."*[4] Overall, the ERG believes that this approach represents a poor use of the available trial data for vedolizumab and conventional therapy. This issue is complicated however by the existence of the other biologic therapies for which the manufacturer would not have access to patient-level trial data; within the anti-TNF-α naïve subgroup, an appropriate method for deriving maintenance transition matrices is not immediately obvious.

It should also be noted that the target datapoints used in the fitting process relate to the probability of achieving response and the probability of achieving remission at 1-year. These datapoints are derived using the GEMINI1 trial results for the mixed ITT population and the anti-TNF-α failure population, whilst the results of the manufacturer's NMAs are used to estimate the target datapoints within the anti-TNF-α naïve population. The calibration process attempts to fit the proportion of patients in remission and mild health states to these target datapoints. However, the manufacturer's model structure also attempts to account for those patients who achieved response but had moderate to severe disease; these patients are not accounted for in the calibration process. The ERG believes that these patients should have been included in the manufacturer's calibration.

(4) Partial use of the NMA within the health economic model

The MS includes a description of NMAs conducted to estimate the relative treatment effects of biologic therapies and conventional non-biologic therapies for inducing and maintaining response and remission. However, these analyses are not used (or even considered) in either the mixed ITT population analysis or the anti-TNF-α failure population analysis. Furthermore, in the anti-TNF-α naïve population analysis, only the induction NMA is used directly within the health economic model. For the maintenance phase, the NMA is used as a basis for predicting the probability that a patient is in remission or has mild disease at 1-year; i.e. it is used to inform the target datapoints for calibration rather than to directly inform the health economic model parameters themselves. Conceptually, this means that the calibration is attempting to fit the health economic model predictions to the NMA model predictions rather than empirical evidence.

(5) Implausible transition probabilities between surgery and post-surgical states

The ERG has a number of concerns regarding the assumed transition probabilities between the surgery and post-surgery states. The transition probabilities assumed in each treatment group are presented in Table 31. The ERG believes that the methods used by the manufacturer to estimate and apply these transition probabilities on a repeated basis is highly likely to overestimate the probability of undergoing surgical procedures and the time spent in the post-surgical complications state, thereby substantially inflating the cost and reducing the health gains associated with this treatment option.

(a) Probability of repeated surgery (transitions from surgery to surgery, post-surgical remission, or post-surgical complications)

The probability of returning to surgery from the surgery, post-surgical remission, or post-surgical complications health states is based on the probability of having a further unplanned surgery in the 6 months following colectomy, based on a retrospective analysis of claims data of privately insured UC patients in the US reported by Loftus *et al*.[49] Within this study, the authors reported a 6-month probability of unplanned surgery of 15.3%.[49] The manufacturer's model converts this estimate to an 8-week probability using standard methods[69] (probability=0.0503). However, this probability is applied during each cycle over the entire model time horizon. The manufacturer does not present any evidence to suggest that this probability should be applied indefinitely. In reality, further operations are more likely within the first 12-months of the initial colectomy, but the same rate would not apply on a repeated basis indefinitely. The consequence of this assumption is that the manufacturer's model is highly likely to overestimate the number of surgical procedures undergone by any patient. Ignoring death, if all patients are assumed to enter the model in the surgery health state within the model, the application of the manufacturer's post-colectomy transition matrix suggests that all patients undergo a further 3.3 surgical procedures over a 10 year time horizon (total=4.3 surgeries). This estimate would be higher still given a longer time horizon. Furthermore, the costs of additional planned surgeries, for example the second or third stage of an IPAA, are already included in the cost estimates reported by Buchanan *et al*[61] and used by the manufacturer; these costs are applied every time a patient transits to the surgery state. This assumption therefore substantially overestimates the cost for patients undergoing surgery in all treatment groups but most notably biases against the colectomy group in favour of medical therapies.

(b) Probability of transiting from surgery to post-surgical complications (early complications)

The probability of transiting from surgery to the post-surgical complications state is based on a study which reported a 31% probability of experiencing complications in the first 30 days following colectomy.[50] The manufacturer's model converts this estimate to an 8-week probability using standard methods[69] (probability=0.50). This probability is applied during each cycle over the remaining model time horizon. However, this is likely to substantially overestimate the complication rate as evidence suggests that colectomy-related complications are more common in the first year following surgery.[70,71] The repeated application of this complication rate over each model cycle, combined with the over-estimate of the probability of undergoing repeated surgery, is likely to substantially overestimate the overall complication rate, thus inflating the costs and reducing the health gains associated with patients undergoing surgery. This assumption overestimates costs and reduces health gains for patients undergoing surgery in all treatment groups but most notably biases against the colectomy group in favour of medical therapies.

(c) Probability of transiting from post-surgical remission to post-surgical complications (late complications)

The probability of transiting from post-surgical remission to post-surgical complications is based on the study reported by Loftus *et al.*[49] The probability of experiencing late complications presented in the period 31 days to 6 months post-surgery was reported to be 0.457. The manufacturer converted this estimate to an 8-week probability using standard methods.[69] This probability is then applied during each cycle for the duration of the model time horizon. However, as noted above, evidence suggests that complications are more likely to occur in the first year following surgery and the risk of complications decreases substantially after this time.[70,71] The repeated use of the initial 6-month complication rate for the duration of the model time horizon will likely overestimate the number of patients experiencing post-surgical complications thereby increasing costs and decreasing health gains for patients undergoing surgery. Again, this assumption overestimates costs and reduces health gains for patients undergoing surgery in all treatment groups but most notably biases against the colectomy group in favour of medical therapies.

The ERG also believes that the manufacturer has been inconsistent in their use of evidence to estimate post-surgical complication rates. The manufacturer's model uses Mahadevan *et al*[50] to calculate the probability of transiting from surgery to post-surgical complications (early complications, up to 30 days) and the probability of transiting from post-surgical remission to post-surgical complications (late complications, 30 days to 6 months) from Loftus *et al.*[49] Both studies report early and late complication rates. The ERG believes that it would have been more consistent to use a single source of evidence (i.e. the most relevant) to inform the probabilities of early and late complications.

(d) Probability of transiting from post-surgical complications to post-surgical remission

The probability of transiting from post-surgical complications to post-surgical remission within the manufacturer's model is based on an "annual probability" of 0.84 based on a previous health economic model reported by Xie *et al.*[45] The manufacturer's model converts this annual probability to an 8-week probability using standard methods[69] (probability=0.245). This probability is applied during each cycle over the model time horizon. In the original source of this estimate (Raval *et al*[70]) the probability of 0.84 does not clearly relate to 1-year and relates only to pouch leaks. Converting this estimate to an 8-week probability and repeatedly applying this value is likely to underestimate the probability of recovering from complications and overestimate the amount of time spent in the post-surgical complications health state, thereby inflating costs and reducing health gains for patients undergoing surgery. Again, this biases against the colectomy group in favour of medical therapies.

(6) Questionable assumptions regarding health-related quality of life

The manufacturer's model uses HRQoL estimates from the GEMINI1 trial[8] for pre-colectomy states, and estimates based on Punekar *et al*[43] for the surgery and post-surgery states. The ERG is satisfied

that the use of the GEMINI trial EQ-5D estimates is reasonable. However, the use of the estimates reported by Punekar *et al* to inform surgery and post-surgery HRQoL values is dubious. The HRQoL estimates for the surgery state (utility=0.60) and the post-surgery health states (utility=0.42) reported in the Punekar paper cite the source of the values to be a study of the epidemiology and costs of CD.[62] A health utility study reported by Woehl *et al* does exist and reports HRQoL for patients with UC in various health states (remission, response, moderate to severe UC and post-surgery). However, the utility values reported by Punekar *et al*[43] do not coincide with those reported by Woehl *et al.*[54] As shown in Table 46, the values presented in Woehl *et al* for patients who have undergone surgery are substantially higher than those for the surgical remission health state presented in Punekar *et al.*[43] The values for the pre-surgical states are also slightly different. It appears that these transcription errors have also been applied in the manufacturer's model; these will downweight health gains accrued by patients undergoing surgery and will bias against the surgery group in favour of medical therapies.

Table 46: HRQoL values

Health state	Utilities reported by Woehl *et al*[54]	Utilities reported by Punekar *et al*[43]	Utilities used in manufacturer's model
Remission	0.87	0.88	0.86
Mild	0.76	-	0.80
Active UC or moderate/severe disease	0.41	0.42	0.68
Surgery	IPAA=0.71, ileostomy=0.72	-	0.42
Post-surgical remission		0.60	0.60
Post-surgical complications		0.42	0.42

It is also noteworthy that within the manufacturer's model, the utility value for patients in post-surgical remission (utility=0.60) is lower than that for moderate/severe disease (utility=0.68). This appears to be inconsistent - if patients can expect a lower HRQoL after surgery compared to before surgery, it is unclear why any patient would ever elect to undergo such procedures.

As noted by Punekar *et al*,[43] utility values for post-surgical complications were not explicitly included in the Woehl *et al* study.[54] Punekar *et al* instead assumed that the HRQoL value for patients with post-surgical complications is equivalent to that for patients with active UC. The manufacturer's model also assumes this HRQoL value of 0.42 despite the fact that the utility value for moderate to severe UC in their model, derived from the GEMINI1 trial, is substantially higher (utility=0.68). A more consistent approach would have involved using the value of 0.68 to represent the utility score for patients experiencing surgical complications, although given the manufacturer's assumptions regarding the utility score for the surgery state (transcribed incorrectly), this would have led to a higher utility score being assigned to surgical complications (0.68) than to surgical remission (0.60).

(7) Concerns relating to resource use and unit costs within the manufacturer's model

In the manufacturer's model, the costs of health states, excluding those for the surgery state, appear to be higher than current estimates from NHS Reference Costs and are therefore unlikely to reflect the current costs borne by the NHS. Whilst the MS[1] and the manufacturer's response to clarification[4] (question B22) state that the unit costs for endoscopy, consultant visits, blood tests and hospitalisations have been taken from 2011/12 NHS Reference Costs, the manufacturer's model actually uses the unit costs reported in Tsai *et al* (the source of which is cited as 2006/07 NHS Reference Costs) uplifted to current prices using inflation indices. However, in the budget impact analysis, (see MS[1] Table 123), the manufacturer has used 2012/13 NHS Reference Costs to value resource use presented in Tsai *et al*[44] for the post-surgical health states. As shown in Table 47, excluding the cost of consultant visits, the estimates from 2012/13 NHS Reference Costs are substantially lower than those reported Tsai *et al.*[44] The consequence of this error is that the costs of the post-surgical states are overestimated; applying the current 2012/13 NHS Reference Cost estimates reduces the cost of the post-surgical complication state from £12,470 to £9,109 per year and reduces the cost of the post-surgical remission health state from £3,048 to £1,447 per year.

Table 47: Costs used in the manufacturer's model and budget impact analysis

Resource item	Inflated cost from Tsai *et al*[44] used in the manufacturer's economic model	NHS Reference Costs used in the manufacturer's budget impact analysis
Consultant visit	£105.73	£115.48
Hospitalisation episode	£3,399.36	£2,574.02
Blood tests	£3.35	£2.95
Elective endoscopy	£1,497.12	£635.68
Emergency endoscopy	£2,026.09	£950.00

It is also noteworthy that the costs of the post-surgery states, based on Tsai *et al*,[44] may not include all relevant resource items. In particular, it is not clear whether the costs of stoma care (nurse visits and consumables) are included in the resource use estimates reported within this study. This issue was recognised in the appraisal of infliximab, adalimumab and golimumab for moderately to severely active UC.[42] Additional work undertaken by the Assessment Group (some of whom are authors of this ERG report) included an estimate of approximately £466 per year, based on Buchanan *et al.*[61] The omission of these costs will underestimate the costs of patients undergoing surgery and will bias against the medical therapies.

The cost of treating AEs used within the manufacturer's model may also represent overestimates. The Health Resource Group (HRG) codes used by the manufacturer to calculate the AE costs all included an inpatient stay of either short or long stay duration. The MS does not justify assumptions regarding

treatment setting for managing AEs. Whilst some of the AEs will require inpatient admission, it is unlikely that this is true for all AEs (e.g. skin reactions).

The manufacturer's model also assumes that whilst patients are receiving biologic therapies, the costs associated with conventional non-biologic therapies will be half of those incurred by patients who are receiving conventional therapies only. This is not justified in the MS. In response to a request for clarification on this issue,[4] the manufacturer stated:

"A detailed assessment of the use of conventional therapy alongside vedolizumab would be complex. The use of conventional therapy within the GEMINI-1 trial 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. 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. £204.80 per cycle). In this scenario, the ICER would be £35,893 per QALY compared with conventional therapy for the mixed patient population, compared to an ICER of £33,295 per QALY in the base-case in the main submission."[4]

Finally, the manufacturer's model includes the cost of topical rather than oral prednisolone. Replacing the cost of topical prednisolone with that for oral prednisolone reduces the overall cost of conventional therapy; the ERG notes that this does not however have a material impact upon the ICER for vedolizumab.

(8) Inappropriate assumptions concerning adverse events of biologic and non-biologic therapies

The AE probabilities for conventional therapy were calculated from rates of AEs in the placebo arms of the included trials for vedolizumab, infliximab, adalimumab, golimumab. As part of the trials, placebo-treated patients received a placebo transfusion or injection. In the calculation of the probability of 'rash' for conventional therapy, infusion site rash was included for patients in the conventional therapy group of the model. In normal practice patients on conventional therapy would not be receiving infusions. Whilst no patients experienced an infusion site rash the arm of the vedolizumab trial included in the calculation of AEs for conventional therapy, it is not clear that the infusion site rash is excluded from the general category of rash used for other comparators.

The manufacturer applied disutility values to patients who experienced AEs. For serious infection, TB and malignancy, these disutilities were calculated by subtracting the utility of patients with experiencing that AE from a baseline value of 1. This may overestimate the disutility, firstly, as it assumes that those patients who were not experiencing the event have perfect quality of life, and secondly, because the disutility is assumed to apply indefinitely whilst the patient is receiving the given treatment. It should also be noted that the disutility of the AEs may already, to some degree, be reflected in the pre-colectomy utility values derived from the GEMINI1 trial, although as HRQoL was estimated by state rather than by treatment, the differential impacts of treatment-related AEs would not be captured.

(9) Use of incremental cost-effectiveness ratios based on point estimates of parameters (deterministic)
The cost-effectiveness results presented in the MS are based on point estimates of parameters rather than the expectation of the mean. Whilst the manufacturer has undertaken PSA, this is not used as the basis for the estimates of cost-effectiveness reported within the MS. This is inappropriate as there may be non-linearity within the model which could lead to different estimates of incremental cost-effectiveness. It is also noteworthy that the manufacturer's model only compares two options simultaneously and their PSA routine does not use a common set of random numbers hence it is not straightforward to generate probabilistic ICERs using consistent random numbers for each stochastic iteration.

(10) Issues concerning model implementation
The ERG re-built the vedolizumab and conventional therapy arms of the manufacturer's model in order to assess its robustness. The ERG identified only one serious programming error in the model's implementation – in the anti-TNF-α naïve population, the maintenance transition matrix for conventional therapy incorrectly uses the maintenance transition matrix for infliximab (this applies to all patients in the conventional therapy group and to all patients in the biologic therapy groups following discontinuation of biologic therapy). In addition, the ERG notes the following:

- The model assumes that patients can transit from biologic therapy to surgery during induction but subsequently, patients must first receive conventional therapies. This structural aspect of model implementation is unlikely to affect the model results but seems counterintuitive (for example patients receiving biologic therapies may transition to surgery during a symptomatic flare). Removing this inconsistent assumption is unlikely to have a material influence on the estimated ICERs for vedolizumab.
- Upon discontinuation of biologic therapy, the induction transition vector for conventional therapy is applied; this assumes that patients who have failed on a biologic (or two biologics in the anti-TNF-α failure group) have the same probability of achieving response and remission as those who have not.

- As noted above, the model structure is limited in that only two options are compared simultaneously in the results worksheet. Combined with the absence of a common random number set, it is not straightforward to produce an accurate fully incremental analysis using the probabilistic model given its current form.
- The implementation of the model is unnecessarily complex for a Markov model. Tracing cells to their original hardcoded source within the model is burdensome.

5.4 Exploratory and sensitivity analyses undertaken by the ERG

5.4.1 Description of additional analyses undertaken by the ERG

This section presents additional analyses undertaken using the manufacturer's model; this includes the development of an ERG-preferred base case analysis. Based on the issues identified within the critical appraisal of the manufacturer's model (see Section 5.3), the following analyses were undertaken:

Additional analysis 1: Correction of cell-referencing error for conventional therapy. In the anti-TNF-α naïve population, the cell-referencing error for the maintenance transition matrix for conventional therapy has been corrected.

Additional analysis 2: Health utilities based on Woehl et al.[54] An analysis was undertaken whereby health utilities were based on the estimates reported within the study reported by Woehl *et al:*[54] remission utility=0.87; response utility=0.76; moderate to severe utility=0.41; post-surgery remission=0.71. The utility score for the surgery state was assumed to be the same as that for the moderate to severe state (utility=0.41). The utility score for the post-surgical complications state was assumed to equal that for post-surgical remission less a disutility for complications sourced from Arseneau *et al*[64] (disutility=0.17).

Additional analysis 3: Health utilities based on Swinburn et al.[72] An analysis was undertaken whereby health utilities were based on the estimates reported within the study reported by Swinburn *et al:*[72] remission utility=0.91; response utility=0.80; moderate to severe utility=0.55; post-surgery remission=0.59. The utility score for the surgery state was assumed to be the same as that for the moderate to severe state (utility=0.55). The utility score for the post-surgical complications state was assumed to be equal to that for post-surgical remission less a disutility for complications sourced from Arseneau *et al*[64] (disutility=0.17).

Additional analysis 4: Amended transition matrix for surgery and post-surgery states. A further analysis was undertaken whereby the transition matrix for surgery and post-surgical states was amended to better reflect clinical practice. Within this analysis, the probability of repeat surgery was set equal to zero (as the costs of these are already reflected in the Buchanan surgery cost[61]). The

probability of experiencing late complications was based on the probability of chronic pouchitis reported within Arai *et al.*[71] Upon leaving the surgery state, patients either remain in remission or remain in the post-surgery complications state for the remainder of the modelled time horizon.

Additional analysis 5: Removal of assumption regarding maximum biologic treatment time. An analysis was undertaken whereby the manufacturer's assumption of a 1-year maximum treatment duration was removed. Within this analysis, patients may continue to receive biologic therapies beyond 1-year provided they achieve and maintain response or remission from those therapies.

Additional analysis 6: Removal of assumption regarding lower use of conventional therapies whilst patients are receiving biologic therapies. An analysis was undertaken whereby the cost of conventional therapies per cycle is the same for all medical treatment groups irrespective of whether the patient is currently receiving biologic therapy.

Additional analysis 7: Use of 2012/13 NHS Reference Costs to value health state resource use. An analysis was undertaken whereby the costs of health state resource use were based on 2012/13 NHS Reference Costs rather than the estimates reported in Tsai *et al*.[44]

Additional analysis 8: Inclusion of the costs of stoma care. An analysis was undertaken whereby the costs of stoma care were included in the post-surgery states of the model. The 6-monthly cost of stoma consumables for patients undergoing surgery was estimated to be £178.09 based on Buchanan *et al*;[61] this was uplifted to current prices and applied to 40% of patients who are assumed to have an ileostomy. The cost of 1.5 nurse visits per 6-months was estimated to be £136.88. The combined cost of stoma consumables and nurse visits per 8-week cycle was added to the other costs for the post-surgery states within the manufacturer's model.

Additional analysis 9: ERG-preferred base case. This analysis incorporates all additional analyses except for analysis 3 (utilities reported by Swinburn *et al*[72]).

All additional analyses in the anti-TNF-α population (additional analysis 1) include the correction of the cell-referencing error for the conventional management maintenance phase transition matrix (see Section 5.3). With the exception of additional analysis 1, each analysis is presented for the mixed ITT population, the anti-TNF-α naïve population and the anti-TNF-α failure population over a lifetime horizon. Due to the absence of a common random number set and the ability of the model to consider only two options simultaneously, the results presented here are based on point estimates of parameters rather than the expectation of the mean. All analyses are presented as fully incremental comparisons of all treatment options.

It should be noted that as part of the request for clarification,[4] the ERG requested summary tables of the numbers of patients in the GEMINI1 trial who transited between moderate to severe UC, response and remission; access to these data would have allowed for the calculation of the observed probabilities of transiting between pre-colectomy health states in the vedolizumab and conventional therapy groups. However, these data were not made available to the ERG.

5.4.2 Results of additional analyses undertaken by the ERG

The results of additional analyses 1-9 are presented in Tables 48 to 56.

Table 48: Scenario 1 – correction of conventional management maintenance transition matrix cell referencing in anti-TNF-α population (10-year time horizon)

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER (vedolizumab versus comparator)
Mixed ITT population					
Vedolizumab	5.55	£77,056	0.15	£5,131	£33,297
Conventional therapy	5.40	£71,925	-	-	-
Surgery	4.28	£107,831	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	5.90	£69,400	0.14	£909	£6,469
Infliximab	5.81	£74,427	-	-	dominated
Golimumab	5.78	£70,798	-	-	dominated
Adalimumab	5.75	£68,492	0.21	£609	£2,868
Conventional therapy	5.54	£67,883	-	-	-
Surgery	4.28	£107,831	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	5.46	£78,409	0.09	£5,839	£64,999
Conventional therapy	5.37	£72,570	-	-	-
Surgery	4.28	£107,831	-	-	dominated

The correction of the cell referencing error in the maintenance transition matrix for conventional therapy within the anti-TNF-α population analysis impacts upon the estimated costs and health gains for all medical treatment options. Within the corrected analysis, surgery, golimumab and infliximab are ruled out due to dominance. Vedolizumab is expected to be the most effective option. The ICER for adalimumab versus conventional therapy is estimated to be £2,868 per QALY gained. The ICER for vedolizumab versus adalimumab is estimated to be £6,469 per QALY gained.

Table 49: Scenario 2 - health utilities based on Woehl *et al*[54]

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Surgery	13.05	£248,631	2.21	£48,285	£21,881
Vedolizumab	10.84	£200,346	0.24	£4,144	£17,140
Conventional therapy	10.60	£196,202	-	-	-
Anti-TNF-α naïve population					
Surgery	13.05	£248,631	1.15	£66,163	£57,725
Vedolizumab	11.90	£182,468	-	-	-
Infliximab	11.71	£188,902	-	-	dominated
Golimumab	11.64	£185,861	-	-	dominated
Adalimumab	11.57	£184,190	-	-	dominated
Conventional therapy	11.07	£187,392	-	-	dominated
Anti-TNF-α failure population					
Surgery	13.05	£248,631	2.49	£51,599	£20,714
Vedolizumab	10.70	£202,259	-	-	ext dom
Conventional therapy	10.56	£197,032	-	-	-

Ext dom – extendedly dominated

The use of utilities reported by Woehl *et al*[54] has a substantial impact upon the model results. Within the mixed ITT population, surgery becomes the most effective option. The ICER for vedolizumab versus conventional therapy is estimated to be £17,140 per QALY gained. The ICER for surgery versus vedolizumab is estimated to be £21,881 per QALY gained. Within the anti-TNF-α naïve population, surgery becomes the most effective option. All other medical options are expected to be dominated by vedolizumab. The ICER for surgery versus vedolizumab is estimated to be £57,725 per QALY gained. Within the anti-TNF-α failure population, surgery becomes the most effective option. Vedolizumab is expected to be ruled out due to extended dominance. The ICER for surgery versus conventional therapy is estimated to be £20,714 per QALY gained.

Table 50: Scenario 3 - health utilities based on Swinburn *et al*[72]

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	11.36	£200,346	0.27	£4,144	£15,267
Conventional therapy	11.09	£196,202	-	-	-
Surgery	10.86	£248,631	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	12.50	£182,468	-	-	dominating
Infliximab	12.28	£188,902	-	-	dominated
Golimumab	12.21	£185,861	-	-	dominated
Adalimumab	12.13	£184,190	-	-	dominated
Conventional therapy	11.59	£187,392	-	-	dominated
Surgery	10.86	£248,631	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	11.20	£202,259	0.16	£5,227	£33,472
Conventional therapy	11.04	£197,032	-	-	-
Surgery	10.86	£248,631	-	-	dominated

Compared to Scenario 2, the use of utilities reported by Swinburn *et al*[72] results in considerably more favourable estimates of the clinical effectiveness and cost-effectiveness for vedolizumab versus all comparators in all three populations. Within the mixed ITT population, surgery is expected to be dominated. The ICER for vedolizumab versus conventional therapy is estimated to be £15,267 per QALY gained. Within the anti-TNF-α naïve population, vedolizumab is expected to dominate all other options. Within the anti-TNF-α failure population, surgery is expected to be dominated. The ICER for vedolizumab versus conventional therapy is estimated to be £33,472 per QALY gained.

Table 51: Scenario 4 - amended transition matrix for surgery and post-surgery states

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	13.29	£131,111	0.17	£7,478	£44,114
Conventional therapy	13.12	£123,634	0.81	£32,989	£40,839
Surgery	12.31	£90,645	-	-	-
Anti-TNF-α naïve population					
Vedolizumab	14.01	£125,340	1.70	£34,696	£20,449
Infliximab	13.88	£129,552	-	-	dominated
Golimumab	13.83	£125,659	-	-	dominated
Adalimumab	13.78	£123,078	-	-	ext dom
Conventional therapy	13.44	£120,285	-	-	ext dom
Surgery	12.31	£90,645	-	-	-
Anti-TNF-α failure population					
Vedolizumab	13.19	£131,271	0.10	£7,300	£73,931
Conventional therapy	13.09	£123,971	0.78	£33,326	£42,769
Surgery	12.31	£90,645	-	-	-

Ext dom – extendedly dominated

Using alternative assumptions regarding the transition probabilities for the surgery and post-surgical states in the model also has a considerable impact upon the results of the manufacturer's model as surgery is no longer dominated in any of the three analysis populations. Vedolizumab is expected to be the most effective option in all three populations. Within the mixed ITT population, the ICER for conventional therapy versus surgery is estimated to be £40,839 per QALY gained. The ICER for vedolizumab versus conventional therapy is estimated to be £44,114 per QALY gained. Within the anti- TNF-α naïve population, all options except vedolizumab and surgery are expected to be ruled out of the analysis due to simple or extended dominance. The ICER for vedolizumab versus surgery is estimated to be £20,449 per QALY gained. Within the anti-TNF-α failure population, the ICER for conventional therapy is estimated to be £42,769 per QALY gained. The ICER for vedolizumab versus conventional therapy is estimated to be £73,931 per QALY gained.

Table 52: Scenario 5 - removal of assumption regarding maximum biologic treatment time

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	12.85	£210,883	0.42	£14,681	£34,827
Conventional therapy	12.43	£196,202	-	-	-
Surgery	10.81	£248,631	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	13.39	£207,168	0.00	£11,711	£3,807,239
Adalimumab	13.39	£195,457	0.58	£8,065	£13,908
Golimumab	13.10	£197,159	-	-	dominated
Infliximab	13.07	£202,159	-	-	dominated
Conventional therapy	12.81	£187,392	-	-	-
Surgery	10.81	£248,631	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	12.89	£212,963	0.49	£15,931	£32,524
Conventional therapy	12.40	£197,032	-	-	-
Surgery	10.81	£248,631	-	-	dominated

Removing the assumption that patients will discontinue biologic therapy at 1-year irrespective of disease control substantially impacts upon the model results in the anti-TNF-α naïve and anti-TNF-α failure populations. Vedolizumab is expected to be the most effective option in all three populations. Within the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £34,827 per QALY gained. Surgery is expected to be dominated. Within the anti- TNF-α naïve population, surgery, infliximab and golimumab are expected to be dominated. The ICER for adalimumab versus conventional therapy is estimated to be £13,908 per QALY gained. The ICER for vedolizumab versus adalimumab is estimated to be in excess of £3.8million per QALY gained. Within the anti-TNF-α failure population, surgery is expected to be dominated. The ICER for vedolizumab versus conventional therapy is estimated to be £32,524 per QALY gained.

Table 53: Scenario 6 - conventional therapy use same for all groups

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	12.63	£200,747	0.20	£4,544	£22,590
Conventional therapy	12.43	£196,202	-	-	-
Surgery	10.81	£248,631	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	13.47	£182,961	-	-	dominating
Infliximab	13.31	£189,421	-	-	dominated
Golimumab	13.26	£186,319	-	-	dominated
Adalimumab	13.20	£184,584	-	-	dominated
Conventional therapy	12.81	£187,392	-	-	dominated
Surgery	10.81	£248,631	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	12.52	£202,608	0.12	£5,577	£47,087
Conventional therapy	12.40	£197,032	-	-	-
Surgery	10.81	£248,631	-	-	dominated

Removing the assumption that patients receiving biologic therapies receive half as much conventional therapy as those who are not receiving biologic therapies has a marked impact upon the cost-effectiveness of vedolizumab within the mixed ITT population and the anti-TNF-α failure population. Within the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £22,590 per QALY gained. Surgery remains dominated. Within the anti- TNF-α naïve population, vedolizumab is expected to dominate all other options. Within the anti-TNF-α failure population, surgery is expected to be dominated. The ICER for vedolizumab versus conventional therapy is estimated to be £47,087 per QALY gained.

Table 54: Scenario 7 - use of 2012/13 NHS Reference Costs to value health state resource use

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	12.63	£151,516	0.20	£5,611	£27,893
Conventional therapy	12.43	£145,905	-	-	-
Surgery	10.81	£202,284	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	13.47	£139,548	0.27	£203	£759
Infliximab	13.31	£144,864	-	-	dominated
Golimumab	13.26	£141,434	-	-	dominated
Adalimumab	13.20	£139,346	-	-	-
Conventional therapy	12.81	£139,727	-	-	dominated
Surgery	10.81	£202,284	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	12.52	£152,558	0.12	£6,072	£51,271
Conventional therapy	12.40	£146,486	-	-	-
Surgery	10.81	£202,284	-	-	dominated

The results of the analysis using 2012/13 NHS Reference Costs to value UC health states suggests some impact upon the cost-effectiveness of vedolizumab within all three populations. Within the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £27,893 per QALY gained. Surgery is expected to be dominated. Within the anti-TNF-α naïve population, surgery, conventional therapy, golimumab and infliximab are expected to be dominated. The ICER for vedolizumab versus adalimumab is estimated to be £759 per QALY gained. Within the anti-TNF-α failure population, surgery is expected to be dominated, whilst the ICER for vedolizumab versus conventional therapy is estimated to be £51,271 per QALY gained.

Table 55: Scenario 8 - inclusion of costs of stoma care

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Vedolizumab	12.63	£204,395	0.20	£3,949	£19,630
Conventional therapy	12.43	£200,447	-	-	-
Surgery	10.81	£257,874	-	-	dominated
Anti-TNF-α naïve population					
Vedolizumab	13.47	£185,809	-	-	dominating
Infliximab	13.31	£192,372	-	-	dominated
Golimumab	13.26	£189,382	-	-	dominated
Adalimumab	13.20	£187,764	-	-	dominated
Conventional therapy	12.81	£191,317	-	-	dominated
Surgery	10.81	£257,874	-	-	dominated
Anti-TNF-α failure population					
Vedolizumab	12.52	£206,410	0.12	£5,105	£43,108
Conventional therapy	12.40	£201,305	-	-	-
Surgery	10.81	£257,874	-	-	dominated

The inclusion of stoma care costs impacts upon the cost-effectiveness of vedolizumab within the mixed ITT population and the anti-TNF-α failure population. Within the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £19,630 per QALY gained. Surgery is expected to be dominated. Within the anti- TNF-α naïve population, vedolizumab is expected to dominate all other options. Within the anti-TNF-α failure population, surgery is expected to be dominated. The ICER for vedolizumab versus conventional therapy is estimated to be £43,108 per QALY gained.

Table 56: ERG-preferred base case

Option	QALYs	Costs	Incremental QALYs	Incremental costs	ICER
Mixed ITT population					
Surgery	14.60	£65,204	-	-	dominating
Vedolizumab	11.78	£107,604	-	-	dominated
Conventional therapy	11.31	£82,940	-	-	dominated
Anti-TNF-α naïve population					
Surgery	14.60	£65,204	-	-	dominating
Adalimumab	12.39	£102,666	-	-	dominated
Vedolizumab	12.37	£115,240	-	-	dominated
Golimumab	12.05	£98,594	-	-	dominated
Infliximab	12.01	£102,916	-	-	dominated
Conventional therapy	11.73	£81,501	-	-	dominated
Anti-TNF-α failure population					
Surgery	14.60	£65,204	-	-	dominating
Vedolizumab	11.84	£110,025	-	-	dominated
Conventional therapy	11.28	£83,094	-	-	dominated

Within the ERG base case, surgery is expected to dominate all other options in all three populations. Whilst surgery appears favourable within all three analysis populations, the ERG notes that this may not be an acceptable option for all patients. Where surgery is not an acceptable option in the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £53,084 per QALY gained. Where surgery is not an acceptable option in the anti-TNF-α naïve population, vedolizumab is expected to be dominated by adalimumab. Where surgery is not an acceptable option in the anti-TNF-α failure population, the ICER for vedolizumab versus conventional therapy is estimated to be £48,205 per QALY gained.

5.5 Discussion of cost-effectiveness evidence and additional analyses undertaken by the ERG

The manufacturer submitted a model-based health economic analysis as part of their submission to NICE.[1] The analysis was undertaken from the perspective of the NHS over a 10-year time horizon. The manufacturer'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) and surgery as separate options. Other anti-TNF-α agents (infliximab, adalimumab and golimumab) are included only in the analysis of the anti-TNF-α naïve population; these are excluded from the analyses of the mixed ITT and anti-TNF-α failure populations. Calcineurin inhibitors were not evaluated in any of the three populations.

Based on a fully incremental analysis (re-analysed by the ERG), within the mixed ITT population, the manufacturer's model suggests that surgery is dominated as it produces fewer health gains and is more costly than both conventional therapy and vedolizumab. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.15 QALYs at an incremental cost of £5,131; the ICER for vedolizumab versus conventional therapy is estimated to be £33,297 per QALY gained. Within the anti-TNF-α naïve population, the manufacturer's model suggests that surgery is expected to be dominated by medical therapies. Vedolizumab is expected to be the most effective option. Infliximab and golimumab are expected to be dominated by vedolizumab and are ruled out of the analysis. The ICER for adalimumab versus conventional therapy is expected to be £3,664 per QALY gained, whilst the ICER for vedolizumab versus adalimumab is expected to be £6,634 per QALY gained. Within the anti-TNF-α failure population, the manufacturer's model suggests that surgery is expected to be dominated by conventional therapy. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.09 QALYs at an

incremental cost of £5,839; the ICER for vedolizumab versus conventional therapy is estimated to be £64,999 per QALY gained.

The ERG critically appraised the manufacturer's health economic analysis and the model upon which this analysis is based. Importantly, the manufacturer's economic analysis deviates from the NICE Reference Case and the final NICE scope due to (a) missing comparators in the mixed ITT population the anti-TNF-α failure populations, (b) the use of a 10-year time horizon and (c) the use of pairwise comparisons rather than a fully incremental analysis. These issues hinder the appropriate interpretation of the manufacturer's results.

The ERG scrutinised the manufacturer's model and partially re-built the model to check for technical programming errors. The ERG identified one serious programming error – in the anti-TNF-α naïve population, the maintenance transition matrix for conventional therapy incorrectly draws in values for the maintenance transition matrix for infliximab. The broader critical appraisal of the manufacturer's model highlighted a number of other concerns and uncertainties. The most notable of these relate to the deviations from the NICE Reference Case and final NICE scope (as discussed above), assumptions regarding continuation/discontinuation of vedolizumab and other biologic therapies, and highly pessimistic assumptions regarding the use, costs and benefits of colectomy. Also of particular concern is the considerable uncertainty associated with the calibration and extrapolation of the pre-colectomy maintenance transition matrices. This latter issue could have been better addressed by using the observed transitions between moderate to severe UC, response and remission states using the patient-level trial within the GEMINI1 trial. Despite a request for these data, the manufacturer did not provide them hence the credibility and accuracy of the model's maintenance phase transition matrices remain unclear.

In light of the problems identified during the critical appraisal, the ERG undertook a number of additional analyses to explore the impact of likely biases on the cost-effectiveness of vedolizumab. Nine sets of additional analyses were undertaken in each of the three modelled populations; these included correcting the mistake in the maintenance transition matrix for conventional management in the anti-TNF-α naïve population, the use of alternative sources of HRQoL values, amending the surgery and post-surgical transition probabilities to better reflect clinical reality, removing assumptions regarding biologic treatment discontinuation, removing assumptions regarding the lower use of conventional therapies whilst patients are also receiving biologics, and improving the cost estimates used in the model to better reflect the costs borne by the NHS. The ERG also produced a preferred base case which combines most of these additional analyses. The results of these additional analyses do not consistently favour one particular option but indicate that these issues have the propensity to dramatically shift the ICER for vedolizumab versus other therapies in all three

populations. The ERG-preferred base case indicates that surgery is likely to dominate all medical treatments in all three populations analysed. However, surgery may not be an acceptable option for all patients. Where surgery is not an acceptable option in the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £53,084 per QALY gained. Where surgery is not an acceptable option in the anti-TNF-α naïve population, vedolizumab is expected to be dominated by adalimumab. Where surgery is not an acceptable option in the anti-TNF-α failure population, the ICER for vedolizumab versus conventional therapy is estimated to be £48,205 per QALY gained.

6 END OF LIFE CONSIDERATIONS

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 manufacturer makes no claim that vedolizumab should be appraised under the supplementary 'end of life' advice. The ERG agrees that the end of life considerations are not applicable within this appraisal.

7 CONCLUSIONS

7.1 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 UC who had an inadequate response to, loss of response to, or intolerance of conventional therapy or TNF-α antagonist was more effective in both the induction and maintenance phase of the GEMINI1 trial. However, there are a number of limitations and uncertainties in the evidence base which warrant caution in its interpretation. Due to the high discontinuation rates in the maintenance phase of the GEMINI1 trial, efficacy of treatment effects (including magnitude) may be confounded. The subgroup analyses to determine the efficacy of vedolizumab in patients with prior TNF-α antagonist failure and in patients who were TNF-α antagonist naïve were exploratory and the study was not powered for these assessments. In addition, the trial of maintenance therapy was not large enough or of sufficient duration to estimate the risk of uncommon AEs. In the NMA, the ERG considered that the results presented may have underestimated 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. The impact of adjustments made to the trial data to reflect differences in study design (specifically relating to re-randomisation of biologic induction-responders and randomisation at baseline only), on treatment effects is unclear. The main uncertainties in the clinical evidence relate to the duration of treatment and generalisability to the UK population.

Based on a fully incremental analysis (re-analysed by the ERG), within the mixed ITT population, the manufacturer's model suggests that surgery is dominated as it produces fewer health gains and is more costly than both conventional therapy and vedolizumab. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.15 QALYs at an incremental cost of £5,131; the ICER for vedolizumab versus conventional therapy is estimated to be £33,297 per QALY gained. Within the anti-TNF-α naïve population, the manufacturer's model suggests that surgery is expected to be dominated by medical therapies. Vedolizumab is expected to be the most effective option. Infliximab and golimumab are expected to be dominated by vedolizumab and are ruled out of the analysis. The ICER for adalimumab versus conventional therapy is expected to be £3,664 per QALY gained, whilst the ICER for vedolizumab versus adalimumab is expected to be £6,634 per QALY gained. Within the anti-TNF-α failure population, the manufacturer's model suggests that surgery is expected to be dominated by conventional therapy. Vedolizumab is expected to be the most effective option. Compared against conventional therapy, vedolizumab is expected to produce an additional 0.09 QALYs at an incremental cost of £5,839; the ICER for vedolizumab versus conventional therapy is estimated to be £64,999 per QALY gained.

The ERG undertook a number of additional analyses; some of these improve the ICER for vedolizumab, whilst others increase the ICER substantially or lead to a situation whereby vedolizumab is dominated. The ERG-preferred base case indicates that surgery is likely to dominate all medical treatments in all three populations analysed. Whilst surgery appears favourable within all three analysis populations, the ERG notes that this may not be an acceptable option for all patients. Where surgery is not an acceptable option in the mixed ITT population, the ICER for vedolizumab versus conventional therapy is estimated to be £53,084 per QALY gained. Where surgery is not an acceptable option in the anti-TNF-α naïve population, vedolizumab is expected to be dominated by adalimumab. Where surgery is not an acceptable option in the anti-TNF-α failure population, the ICER for vedolizumab versus conventional therapy is expected to be £48,205 per QALY gained.

7.2 Implications for research

- A long-term head-to-head RCT comparing the efficacy and safety of vedolizumab, other biologics and conventional non-biologic therapies in the treatment of patients with moderately to severely active UC (including EQ-5D data pre- and post-colectomy) would be of considerable value.
- Longer term epidemiological studies and clinical experience are required to fully assess the risk of AEs associated with vedolizumab (particularly the risk of PML, cancer, infections, teratogenicity and fertility).
- Further research is required to evaluate the optimal dose and treatment duration of vedolizumab in clinical practice and the advantages/disadvantages associated with longer treatment duration (including whether benefits are maintained following cessation of vedolizumab treatment).
- Research is required to investigate the interaction between vedolizumab and coexisting therapies and whether the latter can be weaned or discontinued.
- Efficacy and safety studies of vedolizumab should be assed in a “real world experience” in patients with comorbidities including special circumstances e.g. pregnancy, lactation, children and in patients with prior exposure to rituximab.

8 APPENDICES

Appendix 1: ERG supplementary searches in the United European Gastroenterology (UEG) website

Searched the UEG (https://www.ueg.eu/education/library/)
12th August 2014

3 results found for 'vedolizumab' and '2013'
4 results found for **'**vedolizumab' and '2012':
1 result found for 'vedolizumab' and '2011'
2 results found for 'vedolizumab' and '2010'

Appendix 2: ERG supplementary adverse events search strategies

Summary of results

Source	**Date searched**	**Economic**
Medline and Medline in Process	14th August 2014	36
Embase	14th August 2014	181
Cochrane Library: CDSR	14th August 2014	1
Cochrane Library: HTA	14th August 2014	2
Cochrane Library: DARE	14th August 2014	0
Toxline	14th August 2014	10
Total	-	230
Total unique records	-	183

Medline and Medline In-Process: Ovid. 1946 to Present

14th August 2014

36 records

1. vedolizumab.mp.
2. vedo.mp.
3. entyvio.tw.
4. 943609-66-3.rn.
5. 1 or 2 or 3 or 4
6. ae.fs.
7. to.fs.
8. po.fs.
9. ((side or adverse or undesirable) adj2 (event$ or effect$ or reaction$ or outcome$)).ab,ti.
10. adrs.ab,ti.
11. (safe or safety).ab,ti.
12. (treatment adj emergent).ab,ti.
13. tolerability.ab,ti.
14. toxicity.ab,ti.
15. or/6-14
16. 5 and 15

Embase: Ovid. 1974 to 2014 August 13

14th August 2014

181 records

1. vedolizumab/
2. vedolizumab.mp.

3. vedo.mp.
4. entyvio.tw.
5. 943609-66-3.rn.
6. or/1-5
7. ae.fs.
8. to.fs.
9. po.fs.
10. ((side or adverse or undesirable) adj2 (event$ or effect$ or reaction$ or outcome$)).ab,ti.
11. adrs.ab,ti.
12. (safe or safety).ab,ti.
13. (treatment adj emergent).ab,ti.
14. tolerability.ab,ti.
15. toxicity.ab,ti.
16. or/7-15
17. 6 and 16

Cochrane Library: Wiley Interscience.

Cochrane Database of Systematic Reviews (CDSR): Wiley Interscience. 1996-present

Health Technology Assessment Database (HTA): Wiley Interscience. 1995-present

NHS Economic Evaluation Database (NHS EED): Wiley Interscience. 1995-present

14th August 2014

10 records

#1	vedolizumab:ti,ab,kw
#2	vedo:ti,ab,kw
#3	entyvio:ti,ab,kw
#4	943609-66-3:ti,ab,kw
#5	{or #1-#4}

Toxline:US NIH. http://toxnet.nlm.nih.gov/

14th August 2014

10 records

Appendix 3: Meta-analysis contrasts: anti-TNF–naïve induction

All Pairwise Treatment Comparisons Anti-TNF–Naïve Induction Clinical Response

All pairwise comparisons: Bayesian MTC (BUGS)

	Placebo	Adalimumab 80 mg/40 mg	Adalimumab 160 mg/80 mg	Golimumab 200/100mg	Golimumab 400/200mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab
Placebo	1	0.72 (0.5 , 1)	0.56 (0.44 , 0.7)	0.4 (0.28 , 0.6)	0.34 (0.24 , 0.5)	0.24 (0.17 , 0.4)	0.26 (0.18 , 0.4)	0.38 (0.24 , 0.6)
Adalimumab 80 mg/40 mg	1.39 (0.98 , 2)	1	0.79 (0.55 , 1.1)	0.55 (0.33 , 0.9)	0.48 (0.29 , 0.8)	0.34 (0.2 , 0.6)	0.37 (0.22 , 0.6)	0.53 (0.3 , 0.9)
Adalimumab 160 mg/80 mg	1.78 (1.36 , 2.3)	1.27 (0.88 , 1.8)	1	0.7 (0.45 , 1.1)	0.61 (0.39 , 1)	0.43 (0.28 , 0.7)	0.47 (0.29 , 0.7)	0.68 (0.4 , 1.1)
Golimumab 200/100mg	2.52 (1.75 , 3.6)	1.83 (1.1 , 3)	1.43 (0.93 , 2.2)	1	0.88 (0.62 , 1.2)	0.62 (0.37 , 1)	0.67 (0.39 , 1.1)	0.96 (0.54 , 1.7)
Golimumab 400/200mg	2.91 (2.02 , 4.2)	2.09 (1.26 , 3.4)	1.64 (1.05 , 2.6)	1.14 (0.8 , 1.6)	1	0.71 (0.43 , 1.2)	0.76 (0.45 , 1.3)	1.1 (0.61 , 2)
Infliximab 5 mg	4.11 (2.84 , 6)	2.96 (1.76 , 4.9)	2.32 (1.45 , 3.6)	1.61 (0.98 , 2.7)	1.42 (0.85 , 2.3)	1	1.08 (0.74 , 1.6)	1.56 (0.85 , 2.7)
Infliximab 10 mg	3.78 (2.6 , 5.6)	2.71 (1.63 , 4.5)	2.14 (1.38 , 3.5)	1.5 (0.91 , 2.6)	1.31 (0.78 , 2.2)	0.92 (0.64 , 1.3)	1	1.44 (0.78 , 2.6)
Vedolizumab	2.62 (1.69 , 4.2)	1.88 (1.09 , 3.4)	1.48 (0.9 , 2.5)	1.04 (0.58 , 1.8)	0.91 (0.51 , 1.6)	0.64 (0.36 , 1.2)	0.69 (0.39 , 1.3)	1

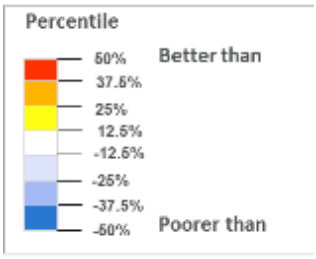
Percentile

50% Better than
37.5%
25%
12.5%
-12.5%
-25%
-37.5%
-50% Poorer than

All Pairwise Treatment Comparisons Anti-TNF–Naïve Induction Clinical Remission

All pairwise comparisons: Bayesian MTC (BUGS)

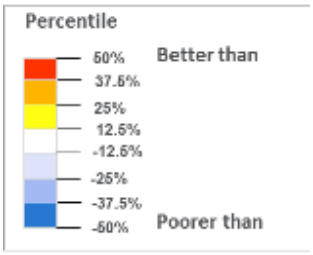
	Placebo	Adalimumab 80 mg/40 mg	Adalimumab 160 mg/80 mg	Golimumab 200/100mg	Golimumab 400/200mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab
Placebo	1	0.81 (0.48 , 1.4)	0.57 (0.39 , 0.8)	0.29 (0.15 , 0.5)	0.3 (0.16 , 0.5)	0.2 (0.12 , 0.3)	0.26 (0.16 , 0.4)	0.27 (0.11 , 0.6)
Adalimumab 80 mg/40 mg	1.24 (0.72 , 2.1)	1	0.71 (0.42 , 1.2)	0.35 (0.16 , 0.8)	0.37 (0.17 , 0.8)	0.24 (0.11 , 0.5)	0.32 (0.16 , 0.7)	0.33 (0.12 , 0.9)
Adalimumab 160 mg/80 mg	1.75 (1.18 , 2.6)	1.42 (0.84 , 2.4)	1	0.5 (0.24 , 1)	0.53 (0.25 , 1.1)	0.34 (0.19 , 0.6)	0.46 (0.25 , 0.9)	0.48 (0.18 , 1.1)
Golimumab 200/100mg	3.51 (2.02 , 6.6)	2.83 (1.32 , 6.3)	2 (1.03 , 4.1)	1	1.06 (0.69 , 1.7)	0.68 (0.32 , 1.5)	0.93 (0.44 , 2.1)	0.95 (0.32 , 2.6)
Golimumab 400/200mg	3.3 (1.85 , 6.2)	2.68 (1.23 , 6)	1.89 (0.93 , 4)	0.94 (0.6 , 1.5)	1	0.64 (0.31 , 1.4)	0.87 (0.4 , 2)	0.9 (0.33 , 2.5)
Infliximab 5 mg	5.12 (3.2 , 8.4)	4.17 (2.01 , 8.9)	2.96 (1.56 , 5.3)	1.46 (0.68 , 3.1)	1.55 (0.71 , 3.3)	1	1.35 (0.91 , 2)	1.39 (0.51 , 3.5)
Infliximab 10 mg	3.81 (2.27 , 6.3)	3.09 (1.47 , 6.4)	2.15 (1.14 , 4)	1.07 (0.48 , 2.3)	1.15 (0.51 , 2.5)	0.74 (0.51 , 1.1)	1	1.04 (0.38 , 2.6)
Vedolizumab	3.67 (1.67 , 9.1)	3 (1.15 , 8.3)	2.09 (0.88 , 5.7)	1.05 (0.39 , 3.1)	1.11 (0.4 , 3.1)	0.72 (0.29 , 1.9)	0.97 (0.39 , 2.6)	1



All Pairwise Treatment Comparisons Anti-TNF–Naïve Induction Discontinuations Due to AEs

All pairwise comparisons: Bayesian MTC

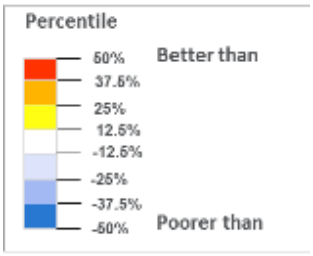
	Placebo	Adalimumab 80 mg/40 mg	Adalimumab 160 mg/80 mg	Golimumab 200/100mg	Golimumab 400/200mg	Vedolizumab
Placebo	1	1.35 (0.59, 3.44)	0.88 (0.43, 1.73)	4.08 (0.42, 95.8)	3.77 (0.4, 115)	>1000 (5.67, >1000)
Adalimumab 80 mg/40 mg	0.74 (0.29, 1.69)	1	0.65 (0.26, 1.55)	2.95 (0.27, 79.1)	2.79 (0.24, 92)	>1000 (3.6, >1000)
Adalimumab 160 mg/80 mg	1.14 (0.58, 2.31)	1.54 (0.65, 3.82)	1	4.59 (0.42, 110.3)	4.42 (0.42, 137.4)	>1000 (6.17, >1000)
Golimumab 200/100mg	0.24 (0.01, 2.38)	0.34 (0.01, 3.67)	0.22 (0.01, 2.38)	1	0.98 (0.03, 43)	838.4 (0.66, >1000)
Golimumab 400/200mg	0.26 (0.01, 2.49)	0.36 (0.01, 4.2)	0.23 (0.01, 2.39)	1.02 (0.02, 33.2)	1	>1000 (0.61, >1000)
Vedolizumab	0 (0, 0.18)	0 (0, 0.28)	0 (0, 0.16)	0 (0, 1.51)	0 (0, 1.63)	1



All Pairwise Treatment Comparisons Anti-TNF–Naïve Induction Mucosal Healing

Odds ratios for all pairwise comparisons: Bayesian FE MTC

	Placebo	Adalimumab 80 mg/40 mg	Adalimumab 160 mg/80 mg	Golimumab 200/100mg	Golimumab 400/200mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab
Placebo	1	0.89 (0.62, 1.29)	0.65 (0.48, 0.88)	0.52 (0.37, 0.75)	0.48 (0.33, 0.7)	0.29 (0.17, 0.5)	0.28 (0.16, 0.47)	0.34 (0.19, 0.63)
Adalimumab 80 mg/40 mg	1.13 (0.77, 1.62)	1	0.73 (0.51, 1.06)	0.59 (0.35, 1)	0.55 (0.33, 0.91)	0.33 (0.17, 0.63)	0.31 (0.16, 0.6)	0.38 (0.18, 0.77)
Adalimumab 160 mg/80 mg	1.53 (1.14, 2.07)	1.36 (0.94, 1.95)	1	0.81 (0.51, 1.26)	0.74 (0.46, 1.19)	0.45 (0.24, 0.82)	0.43 (0.23, 0.79)	0.52 (0.27, 1.03)
Golimumab 200/100mg	1.91 (1.33, 2.73)	1.69 (1, 2.9)	1.24 (0.79, 1.98)	1	0.92 (0.65, 1.31)	0.56 (0.29, 1.04)	0.54 (0.27, 1.01)	0.65 (0.32, 1.32)
Golimumab 400/200mg	2.08 (1.43, 2.99)	1.83 (1.1, 3.07)	1.36 (0.84, 2.16)	1.09 (0.77, 1.54)	1	0.61 (0.32, 1.16)	0.58 (0.3, 1.15)	0.71 (0.35, 1.42)
Infliximab 5 mg	3.42 (2, 5.94)	3.03 (1.58, 5.8)	2.23 (1.21, 4.14)	1.79 (0.96, 3.42)	1.65 (0.86, 3.17)	1	0.94 (0.57, 1.63)	1.17 (0.52, 2.63)
Infliximab 10 mg	3.63 (2.11, 6.14)	3.2 (1.66, 6.28)	2.34 (1.27, 4.36)	1.86 (0.99, 3.73)	1.73 (0.87, 3.37)	1.06 (0.61, 1.76)	1	1.22 (0.54, 2.8)
Vedolizumab	2.97 (1.59, 5.37)	2.61 (1.3, 5.44)	1.92 (0.97, 3.76)	1.54 (0.76, 3.09)	1.42 (0.7, 2.87)	0.86 (0.38, 1.91)	0.82 (0.36, 1.86)	1

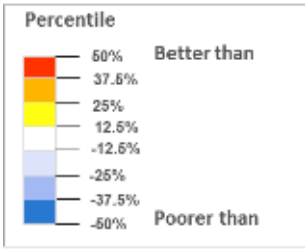


Meta-analysis contrasts: anti-TNF–naïve maintenance

All Pairwise Treatment Comparisons Anti-TNF–Naïve Maintenance Durable Clinical Response (no ACT-1 three-time-point information)

Odds ratios for all pairwise comparisons: Bayesian MTC

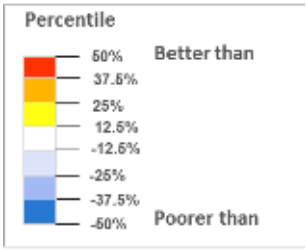
	Placebo	Adalimumab 40 mg eow	Golimumab 50mg	Golimumab 100mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.8 (0.47, 1.4)	0.52 (0.32, 0.81)	0.44 (0.28, 0.7)	0.59 (0.29, 1.29)	0.43 (0.21, 0.95)	0.28 (0.14, 0.54)	0.19 (0.09, 0.37)
Adalimumab 40 mg eow	1.26 (0.72, 2.11)	1	0.64 (0.31, 1.31)	0.56 (0.27, 1.16)	0.74 (0.3, 1.82)	0.54 (0.22, 1.37)	0.35 (0.15, 0.82)	0.24 (0.1, 0.57)
Golimumab 50mg	1.94 (1.24, 3.09)	1.55 (0.76, 3.27)	1	0.87 (0.55, 1.33)	1.16 (0.49, 2.76)	0.84 (0.36, 2.06)	0.54 (0.24, 1.23)	0.37 (0.16, 0.85)
Golimumab 100mg	2.25 (1.42, 3.56)	1.79 (0.86, 3.71)	1.15 (0.75, 1.82)	1	1.35 (0.57, 3.15)	0.97 (0.42, 2.39)	0.63 (0.28, 1.42)	0.43 (0.19, 0.97)
Infliximab 5 mg	1.69 (0.78, 3.48)	1.35 (0.55, 3.3)	0.86 (0.36, 2.04)	0.74 (0.32, 1.76)	1	0.73 (0.38, 1.41)	0.47 (0.17, 1.21)	0.32 (0.12, 0.87)
Infliximab 10 mg	2.33 (1.06, 4.8)	1.84 (0.73, 4.63)	1.19 (0.49, 2.79)	1.03 (0.42, 2.39)	1.37 (0.71, 2.62)	1	0.65 (0.23, 1.73)	0.44 (0.15, 1.21)
Vedolizumab Q4W	3.55 (1.85, 7.22)	2.85 (1.23, 6.87)	1.84 (0.82, 4.15)	1.59 (0.71, 3.56)	2.13 (0.83, 5.88)	1.55 (0.58, 4.34)	1	0.68 (0.35, 1.27)
Vedolizumab Q8W	5.27 (2.7, 11)	4.21 (1.75, 10)	2.71 (1.18, 6.42)	2.33 (1.03, 5.37)	3.09 (1.15, 8.6)	2.28 (0.83, 6.52)	1.47 (0.79, 2.89)	1



All Pairwise Treatment Comparisons Anti-TNF–Naïve Maintenance Durable Clinical Response (ACT-1 three-time-point information)

Odds ratios for all pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 40 mg eow	Golimumab 50mg	Golimumab 100mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.76 (0.45, 1.27)	0.51 (0.32, 0.83)	0.45 (0.28, 0.72)	0.47 (0.22, 1.01)	0.39 (0.18, 0.83)	0.28 (0.14, 0.56)	0.19 (0.09, 0.38)
Adalimumab 40 mg eow	1.31 (0.79, 2.22)	1	0.68 (0.34, 1.37)	0.58 (0.3, 1.16)	0.62 (0.25, 1.55)	0.52 (0.21, 1.33)	0.36 (0.15, 0.89)	0.25 (0.1, 0.59)
Golimumab 50mg	1.94 (1.21, 3.08)	1.48 (0.73, 2.97)	1	0.86 (0.55, 1.37)	0.91 (0.38, 2.22)	0.77 (0.3, 1.86)	0.54 (0.23, 1.22)	0.36 (0.16, 0.83)
Golimumab 100mg	2.24 (1.39, 3.6)	1.72 (0.86, 3.36)	1.16 (0.73, 1.82)	1	1.05 (0.45, 2.6)	0.88 (0.36, 2.2)	0.62 (0.27, 1.44)	0.42 (0.19, 0.99)
Infliximab 5 mg	2.12 (0.99, 4.47)	1.62 (0.64, 3.98)	1.1 (0.45, 2.62)	0.95 (0.38, 2.22)	1	0.85 (0.44, 1.56)	0.59 (0.21, 1.57)	0.4 (0.14, 1.08)
Infliximab 10 mg	2.54 (1.2, 5.41)	1.91 (0.75, 4.8)	1.3 (0.54, 3.32)	1.13 (0.45, 2.78)	1.18 (0.64, 2.29)	1	0.7 (0.25, 1.89)	0.47 (0.17, 1.32)
Vedolizumab Q4W	3.59 (1.8, 7.08)	2.75 (1.13, 6.48)	1.87 (0.82, 4.26)	1.61 (0.69, 3.75)	1.68 (0.64, 4.68)	1.42 (0.53, 3.94)	1	0.67 (0.35, 1.3)
Vedolizumab Q8W	5.3 (2.63, 11)	4.08 (1.68, 9.58)	2.76 (1.21, 6.43)	2.39 (1.01, 5.36)	2.49 (0.92, 6.92)	2.13 (0.76, 5.9)	1.48 (0.77, 2.82)	1



All Pairwise Treatment Comparisons Anti-TNF–Naïve Maintenance Clinical Remission

Odds ratios for all pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 40 mg eow	Golimumab 50mg	Golimumab 100mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.58 (0.31, 1.08)	0.58 (0.35, 0.99)	0.55 (0.33, 0.89)	0.79 (0.37, 1.64)	0.63 (0.29, 1.29)	0.25 (0.12, 0.52)	0.27 (0.13, 0.54)
Adalimumab 40 mg eow	1.73 (0.93, 3.2)	1	1 (0.44, 2.2)	0.95 (0.43, 2.11)	1.38 (0.53, 3.61)	1.07 (0.41, 2.8)	0.44 (0.16, 1.14)	0.47 (0.17, 1.24)
Golimumab 50mg	1.73 (1.02, 2.87)	1 (0.45, 2.29)	1	0.95 (0.59, 1.51)	1.37 (0.57, 3.29)	1.08 (0.44, 2.82)	0.44 (0.18, 1.04)	0.48 (0.19, 1.12)
Golimumab 100mg	1.83 (1.13, 3.02)	1.05 (0.47, 2.34)	1.05 (0.66, 1.68)	1	1.44 (0.58, 3.44)	1.13 (0.45, 2.85)	0.46 (0.19, 1.08)	0.5 (0.2, 1.22)
Infliximab 5 mg	1.26 (0.61, 2.69)	0.72 (0.28, 1.88)	0.73 (0.3, 1.77)	0.69 (0.29, 1.73)	1	0.8 (0.41, 1.49)	0.32 (0.11, 0.96)	0.34 (0.12, 0.97)
Infliximab 10 mg	1.6 (0.78, 3.47)	0.93 (0.36, 2.47)	0.92 (0.35, 2.29)	0.88 (0.35, 2.21)	1.26 (0.67, 2.47)	1	0.41 (0.14, 1.16)	0.44 (0.15, 1.28)
Vedolizumab Q4W	3.93 (1.94, 8.48)	2.28 (0.88, 6.18)	2.26 (0.96, 5.71)	2.16 (0.92, 5.38)	3.16 (1.05, 8.88)	2.46 (0.86, 7.15)	1	1.08 (0.56, 2.1)
Vedolizumab Q8W	3.67 (1.85, 7.88)	2.14 (0.81, 5.82)	2.1 (0.9, 5.32)	2.01 (0.82, 4.97)	2.93 (1.03, 8.46)	2.3 (0.78, 6.66)	0.93 (0.48, 1.79)	1

Percentile

50%	Better than
37.5%	
25%	
12.5%	
-12.5%	
-25%	
-37.5%	
-50%	Poorer than

All Pairwise Treatment Comparisons Anti-TNF–Naïve Maintenance Discontinuation Due to AEs

All pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 40 mg eow	Golimumab 50mg	Golimumab 100mg	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	-0.8 (0.16, 1.12)	0.23 (0.48, 3.37)	-0.39 (0.28, 1.57)	0.11 (0.47, 2.73)	0.01 (0.43, 2.69)	2.54 (1.86, 298)	1.17 (0.89, 16)
Adalimumab 40 mg eow	0.8 (0.89, 6.36)	1	1.03 (0.73, 12)	0.42 (0.42, 5.5)	0.92 (0.65, 9.65)	0.82 (0.63, 9.32)	3.36 (3.31, 828)	1.97 (1.48, 49)
Golimumab 50mg	-0.23 (0.3, 2.07)	-1.03 (0.08, 1.36)	1	-0.62 (0.22, 1.34)	-0.11 (0.23, 3.37)	-0.21 (0.21, 3.09)	2.34 (1.15, 257)	0.96 (0.51, 17)
Golimumab 100mg	0.39 (0.64, 3.59)	-0.42 (0.18, 2.36)	0.62 (0.75, 4.64)	1	0.51 (0.49, 5.86)	0.42 (0.45, 5.5)	2.94 (2.32, 446)	1.55 (1.01, 30)
Infliximab 5 mg	-0.11 (0.37, 2.15)	-0.92 (0.1, 1.53)	0.11 (0.3, 4.35)	-0.51 (0.17, 2.05)	1	-0.08 (0.36, 2.14)	2.44 (1.29, 299)	1.07 (0.59, 18)
Infliximab 10 mg	-0.01 (0.37, 2.34)	-0.82 (0.11, 1.59)	0.21 (0.32, 4.66)	-0.42 (0.18, 2.21)	0.08 (0.47, 2.75)	1	2.54 (1.46, 345)	1.14 (0.68, 19)
Vedolizumab Q4W	-2.54 (0, 0.54)	-3.36 (0, 0.3)	-2.34 (0, 0.87)	-2.94 (0, 0.43)	-2.44 (0, 0.78)	-2.54 (0, 0.68)	1	-1.37 (0.01, 2.49)
Vedolizumab Q8W	-1.17 (0.06, 1.13)	-1.97 (0.02, 0.67)	-0.96 (0.06, 1.97)	-1.55 (0.03, 0.99)	-1.07 (0.05, 1.69)	-1.14 (0.05, 1.48)	1.37 (0.4, 95)	1

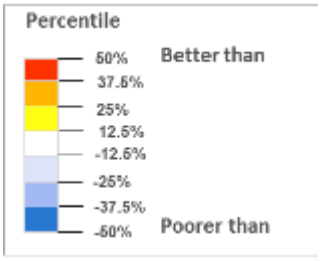
Percentile

50%	Better than
37.5%	
25%	
12.5%	
-12.5%	
-25%	
-37.5%	
-50%	Poorer than

All Pairwise Treatment Comparisons Anti-TNF–Naïve Maintenance Mucosal Healing

Odds ratios for all pairwise comparisons: Bayesian FE MTC

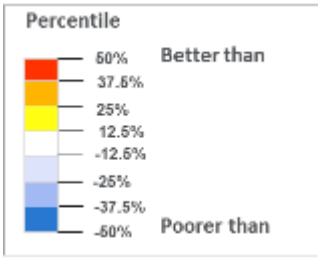
	Placebo	Adalimumab 40 mg eow	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.67 (0.42, 1.06)	0.5 (0.25, 1.05)	0.3 (0.13, 0.64)	0.2 (0.1, 0.42)	0.21 (0.1, 0.43)
Adalimumab 40 mg eow	1.49 (0.95, 2.39)	1	0.77 (0.32, 1.75)	0.45 (0.17, 1.12)	0.31 (0.13, 0.71)	0.31 (0.14, 0.75)
Infliximab 5 mg	1.98 (0.96, 4.04)	1.31 (0.57, 3.12)	1	0.6 (0.28, 1.16)	0.4 (0.14, 1.07)	0.41 (0.15, 1.14)
Infliximab 10 mg	3.33 (1.57, 7.69)	2.21 (0.89, 5.89)	1.68 (0.86, 3.55)	1	0.69 (0.24, 1.97)	0.7 (0.25, 2)
Vedolizumab Q4W	4.88 (2.39, 9.94)	3.27 (1.41, 7.46)	2.48 (0.93, 7.02)	1.45 (0.51, 4.18)	1	1.02 (0.53, 1.97)
Vedolizumab Q8W	4.79 (2.33, 9.93)	3.21 (1.33, 7.35)	2.43 (0.87, 6.66)	1.43 (0.5, 3.97)	0.98 (0.51, 1.89)	1



All Pairwise Treatment Comparisons Anti-TNF–Naïve Maintenance CSF Remission

Odds ratios for all pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 40 mg eow	Infliximab 5 mg	Infliximab 10 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.48 (0.11, 1.47)	0.66 (0.23, 1.69)	0.97 (0.32, 2.66)	0.2 (0.07, 0.52)	0.39 (0.13, 1.09)
Adalimumab 40 mg eow	2.06 (0.68, 8.7)	1	1.36 (0.3, 7.41)	1.97 (0.41, 11)	0.41 (0.09, 2.11)	0.83 (0.17, 4.4)
Infliximab 5 mg	1.51 (0.59, 4.31)	0.74 (0.13, 3.31)	1	1.45 (0.64, 3.41)	0.31 (0.07, 1.22)	0.6 (0.14, 2.57)
Infliximab 10 mg	1.04 (0.38, 3.08)	0.51 (0.09, 2.41)	0.69 (0.29, 1.56)	1	0.2 (0.05, 0.87)	0.41 (0.09, 1.94)
Vedolizumab Q4W	4.91 (1.94, 15)	2.42 (0.47, 11)	3.26 (0.82, 14)	4.9 (1.15, 20)	1	1.98 (0.84, 4.99)
Vedolizumab Q8W	2.57 (0.92, 7.57)	1.2 (0.23, 5.97)	1.66 (0.39, 6.99)	2.46 (0.52, 11)	0.51 (0.2, 1.2)	1

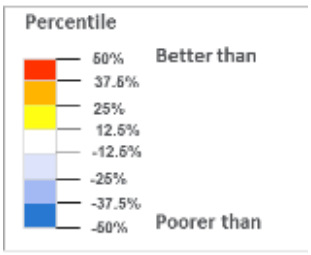


Meta-analysis contrasts: anti-TNF-experienced/failure induction

All Pairwise Treatment Comparisons Anti-TNF–Experienced/Failure Induction Clinical Response

Odds ratios for all pairwise comparisons: Bayesian MTC

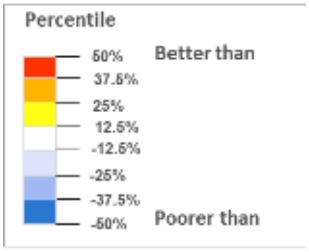
	Placebo	Adalimumab 160 mg/80 mg	Vedolizumab
Placebo	1	0.7 (0.4, 1.3)	0.4 (0.2, 0.8)
Adalimumab 160 mg/80 mg	1.4 (0.8, 2.6)	1	0.6 (0.2, 1.5)
Vedolizumab	2.5 (1.2, 5.5)	1.7 (0.7, 4.4)	1



All Pairwise Treatment Comparisons Anti-TNF–Experienced/Failure Induction Clinical Remission

Odds ratios for all pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 160 mg/80 mg	Vedolizumab
Placebo	1	0.7 (0.2, 2.1)	0.3 (0, 1.2)
Adalimumab 160 mg/80 mg	1.4 (0.5, 4)	1	0.4 (0, 2.3)
Vedolizumab	3.7 (0.9, 28)	2.7 (0.4, 24)	1



All Pairwise Treatment Comparisons Anti-TNF–Experienced/Failure Induction Mucosal Healing

Odds ratios for all pairwise comparisons: Bayesian FE MTC

	Placebo	Adalimumab 160 mg/80 mg	Vedolizumab
Placebo	1	0.9 (0.5, 1.7)	0.6 (0.3, 1.3)
Adalimumab 160 mg/80 mg	1.1 (0.6, 2.1)	1	0.6 (0.2, 1.7)
Vedolizumab	1.7 (0.8, 3.8)	1.6 (0.6, 4.2)	1

Percentile

- 50% Better than
- 37.5%
- 25%
- 12.5%
- -12.5%
- -25%
- -37.5%
- -50% Poorer than

Meta-analysis contrasts: anti-TNF-experienced/failure maintenance

All Pairwise Treatment Comparisons Anti-TNF–Experienced/Failure Maintenance Durable Clinical Response

Odds ratios for all pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 40 mg eow	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.4 (0.14, 1.11)	0.24 (0.07, 0.72)	0.2 (0.06, 0.58)
Adalimumab 40 mg eow	2.47 (0.9, 6.99)	1	0.59 (0.12, 2.6)	0.49 (0.11, 2.25)
Vedolizumab Q4W	4.15 (1.38, 14)	1.69 (0.39, 8.38)	1	0.84 (0.32, 2.06)
Vedolizumab Q8W	4.89 (1.74, 16)	2.04 (0.44, 9.01)	1.19 (0.48, 3.09)	1

Percentile

- 50% Better than
- 37.5%
- 25%
- 12.5%
- -12.5%
- -25%
- -37.5%
- -50% Poorer than

All Pairwise Treatment Comparisons Anti-TNF–Experienced/Failure Maintenance Clinical Remission

Odds ratios for all pairwise comparisons: Bayesian MTC

	Placebo	Adalimumab 40 mg eow	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.28 (0.05, 0.99)	0.09 (0.01, 0.38)	0.08 (0.01, 0.32)
Adalimumab 40 mg eow	3.6 (1.01, 18)	1	0.33 (0.03, 2.98)	0.29 (0.03, 2.53)
Vedolizumab Q4W	11 (2.62, 76)	3.06 (0.34, 31)	1	0.91 (0.37, 2.23)
Vedolizumab Q8W	12 (3.14, 78)	3.4 (0.4, 33)	1.1 (0.45, 2.73)	1

Percentile
50% Better than
37.5%
25%
12.5%
-12.5%
-25%
-37.5%
-50% Poorer than

All Pairwise Treatment Comparisons Anti-TNF–Experienced/Failure Maintenance Mucosal Healing

Odds ratios for all pairwise comparisons: Bayesian FE MTC

	Placebo	Adalimumab 160 mg/80 mg	Vedolizumab Q4W	Vedolizumab Q8W
Placebo	1	0.74 (0.26, 2.01)	0.09 (0.02, 0.3)	0.11 (0.02, 0.37)
Adalimumab 160 mg/80 mg	1.36 (0.5, 3.91)	1	0.12 (0.02, 0.58)	0.15 (0.02, 0.74)
Vedolizumab Q4W	11 (3.37, 57)	8.58 (1.72, 54)	1	1.26 (0.52, 3.07)
Vedolizumab Q8W	9.09 (2.74, 40)	6.72 (1.36, 41)	0.79 (0.33, 1.93)	1

Percentile
50% Better than
37.5%
25%
12.5%
-12.5%
-25%
-37.5%
-50% Poorer than

9. REFERENCES

1. Takeda Pharmaceuticals. Vedolizumab (Entyvio) for the treatment of adults with moderate to severe ulcerative colitis. Manufacturer's submission to the National Institute for Health and Care Excellence. 2014. High Wycombe, Takeda.

2. National Institute for Health and Care Excellence. Clinical Guideline 166. Ulcerative colitis: Management in adults, children and young people. 2013. London, NICE.

3. Mowat C, Cole A, Windsor Al, Ahmad T, Arnott I, Driscoll R et al. Guidelines for the management of inflammatory bowel disease in adults. *Gut* 2011; 60(5):571-607.

4. Takeda Pharmaceuticals. Response to ERG clarification questions. 2014. High Wycombe, Takeda.

5. National Institute for Health and Care Excellence. NICE Technology Appraisal 163: Infliximab for acute exacerbations of ulcerative colitis. 2008. London, NICE.

6. National Institute for Health and Care Excellence. Vedolizumab for the treatment of adults with moderate to severe ulcerative colitis. Single Technology Appraisal. Final scope. 2014.

7. European Medicines Agency. European public assessment report - vedolizumab. 2014; 1-166.

8. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *New England Journal of Medicine* 2013; 369(8):699-710.

9. BMJ Group, RCPCH Publications Ltd, and the Royal Pharmaceutical Society of Great Britain. British National Formulary. 2013. London, BNF.

10. European Medicines Agency. Summary of product characteristics - vedolizumab. 2014; 1-35.

11. National Institute for Health and Care Excellence. Single Technology Appraisal (STA) - Specification for manufacturer/sponsor submission of evidence. 2009. London, National Institute for Health and Care Excellence (NICE).

12. Centre for Reviews and Dissemination (CRD). Systematic review: CRD's guidance for undertaking reviews in health care. CRD, University of York; 2009. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Higgins, J, Green, S, editors. 2011.

13. Takeda Pharmaceuticals. Phase III Study of MLN0002 (300 mg) in the Treatment of Ulcerative Colitis. 2014.

14. Millennium Pharmaceuticals, I. An Open-label Study of Vedolizumab (MLN0002) in Patients With Ulcerative Colitis and Crohn's Disease (GEMINI LTS). 2014.

15. Feagan B, Macdonald J, Greenberg G, et al. An ascending dose of a humanized alpha 4 beta 7 antibody in ulcerative colitis (UC). *Gastroenterology* 2000; 118(4):A874.21.

16. Feagan BG, Greenberg GR, Wild G, et al. Treatment of ulcerative colitis with a humanized antibody to the á4â integrin. *New England Journal of Medicine* 2005; 253:2499-2507.

17. Parikh A, Leach T, Wyant T, Scholz C, Sankoh S, Mould DR 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.

18. 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.

19. Takeda Pharma A/S, 2014. Entyvio EMA Summary of Product Characteristics. 2014.

20. Reinisch W, Sandborn WJ, Panaccione R, Huang B, Pollack PF, Lazar A et al. 52-week efficacy of adalimumab in patients with moderately to severely active ulcerative colitis who failed corticosteroids and/or immunosuppressants. *Inflamm Bowel Dis* 2013; 19(8):1700-1709.

21. Sjoberg D, Larsson M, Nielsen AL, Holmqvist L, Ronnblom A. Conservative approach results in lower colectomy rates in patients with severe ulcerative colitis data from 70 patients with S3 in the ICURE study. *Journal of Crohn's and Colitis* 2013; 7(Suppl 1):S233.

22. Gustavsson A, Jarnerot G, Hertervig E, Friis-Liby I, Blomquist L, Karlen P et al. Clinical trial: colectomy after rescue therapy in ulcerative colitis - 3-year follow-up of the Swedish-Danish controlled infliximab study. *Aliment Pharmacol Ther* 2010; 32(8):984-989.

23. Reinisch W, Sandborn WJ, Rutgeerts P, Feagan BG, Rachmilewitz D, Hanauer SB et al. Long-term infliximab maintenance therapy for ulcerative colitis: the ACT- 1 and -2 extension studies. *Inflamm Bowel Dis* 2012; 18:201-211.

24. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *New England Journal of Medicine* 2005; 353:2462-2476.

25. Jarnerot G, Hertervig E, Friis-Liby I, Blomquist L, Karlen P, Granno C et al. Infliximab as rescue therapy in severe to moderately severe ulcerative colitis: a randomized, placebo-controlled study. *Gastroenterology* 2005; 128(7):1805-1811.

26. Laharie D, Bourreille A, Branche J, Allez M, Bouhnik Y, Filippi J et al. Ciclosporin versus infliximab in patients with severe ulcerative colitis refractory to intravenous steroids: a parallel, open-label randomised controlled trial. *The Lancet* 2012; 380(9857):1909-1915.

27. Probert CS, Hearing SD, Schreiber S, Kuhbacher T, Ghosh S, Arnott ID et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. *Gut* 2003; 52(7):998-1002.

28. Takeda Pharmaceuticals. Takeda Data on File, 2012. FINAL CLINICAL STUDY REPORT C13006 A Phase 3, Randomized, Placebo-Controlled, Blinded, Multicenter Study of the Induction and Maintenance of Clinical Response and Remission by Vedolizumab (MLN0002) in Patients with Moderate to Severe Ulcerative Colitis. 2012.

29. 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. Food and Drug Administration (FDA), Center for Drug Evaluation and Research.

30. Millennium Pharmaceuticals, I. Study of Vedolizumab (MLN0002) in Patients With Moderate to Severe Ulcerative Colitis (GEMINI I). 2014.

31. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *The Lancet* 2002; 359(9308):781-785.

32. Colombel, J., Sands, B., et al. OP359: Long term Safety of Vedolizumab for the treatment of Ulcerative Colitis and Crohn's disease. In United European Gastroenterology Week. Berlin Germany. Abstract OP359. 2013.

33. Long term Safety of Vedolizumab for the Treatment of Ulcerative Colitis or Crohn ' s Disease (oral presentation). United European Gastroenterology Week; Berlin, Germany: 2013.

34. Colombel JF, Sands BE, Feagan BG, Loftus EV, 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.

35. Sandborn WJ, Feagan BG, Rutgeerts P, Hanauer S, Colombel JF, Sands BE et al. Vedolizumab as induction and maintenance therapy for Crohn's disease. *New England Journal of Medicine* 2013; 369(8):711-721.

36. Sands BE, Feagan BG, , R.P., Colombel JF, Sandborn WJ, Sy R et al. Effects of Vedolizumab Induction Therapy for Patients with Crohn's Disease in Whom Tumor Necrosis Factor Antagonist Treatment Had Failed. Gastroenterology, (in press). 2014.

37. Reinisch W, Sandborn WJ, Hommes DW, D'Haens G, Hanauer S, Schreiber S et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. *Gut* 2011; 60(6):780-787.

38. Sandborn WJ, van Assche G, Reinisch W, Colombel JF, D'Haens G, Wolf DC et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2012; 142(2):257-265. e1-3.

39. Suzuki Y, Motoya S, Hanai H, Matsumoto T, Hibi T, Robinson AM et al. Efficacy and safety of adalimumab in Japanese patients with moderately to severely active ulcerative colitis. *Journal of Gastroenterology* 2014; 49(2):283-294.

40. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014; 146(1):85-95.

41. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014; 146:96-109.

42. Archer R, Tappenden P, Ren S, Martyn-St James, Harvey R, Basirir H et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy. Technology Assessment Report: Final report to the National Institute for Health and Care Excellence. 2014.

43. Punekar YS, Hawkins N. Cost-effectiveness of infliximab for the treatment of acute exacerbations of ulcerative colitis. *European Journal of Health Economics* 2010; 11:67-76.

44. Tsai HH, Punekar YS, Morris J, Fortun P. A model of the long-term cost-effectiveness of scheduled maintenance treatment with infliximab for moderate-to-severe ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 2008; 28:1230-1239.

45. Xie F, Blackhouse G, Assasi N, Gaebel K, Robertson D, Goeree R. Cost-utility analysis of infliximab and adalimumab for refractory ulcerative colitis. *Cost Effectiveness and Resource Allocation* 2009; 7(20).

46. European Medicines Agency. Summary of product characteristics - adalimumab. 2014; 1-265.

47. European Medicines Agency. Summary of product characteristics - golimumab. 2014; 1-197.

48. Frolkis AD, Dykeman J, Negron ME, Debruyn J, Jette N, Fiest KM et al. Risk of surgery for inflammatory bowel diseases has decreased over time: A systematic review and meta-analysis of population-based studies. *Gatroenterology* 2013; 145(5):996-1006.

49. Loftus EV, Delgado DJ, Friedman HS, Sandborn WJ. Colectomy and the incidence of postsurgical complications among ulcerative colitis patients with private health insurance in the United States. *American Journal of Gastroenterology* 2008; 103:1737-1745.

50. Mahadevan U, Loftus EV, Tremaine WJ, Pemberton JH, Harmsen WS, Schleck CD et al. Azathioprine or 6-mercaptopurine before colectomy for ulcerative colitis is not associated with increased postoperative complications. *Inflammatory Bowel Disease* 2002; 8(5):311-316.

51. Office for National Statistics. Death registrations summary tables, England and Wales 2010. 2011. ONS.

52. Button LA, Roberts SE, Goldacre MJ, Akbari A, Rodgers SE, Williams JG. Hospitalized prevalence and 5-year mortality for IBD: Record linkage study. *World Journal of Gastroenterology* 2010; 16(4):431-438.

53. Jess T, Gamborg M, Munkholm P, Sorensen TIA. Overall and cause-specific mortality in ulcerative colitis: Meta-analysis of population-based inception cohort studies. *American Journal of Gastroenterology* 2007; 102(3):609-617.

54. Woehl A, Hawthorne A, McEwan P. The relation between disease activity, quality of life and health utility in patients with ulcerative colitis. *Gut* 2008; 57(Suppl1):A153.

55. Brown RE, Hutton J, Burrell A. Cost effectiveness of treatment options in advanced breast cancer in the UK. *Pharmacoeconomics* 2001; 19(11):1091-1102.

56. Porco TC, Lewis B, Marseille E, Grinsdale J, Flood JM, Royce SE. Cost effectiveness of tuberculosis evaluation and treatment of newly-arrived immigrants. *BMC Public Health* 2006; 6(157).

57. Hornberger J, Reyes C, Lubeck D, Valente N. Economic evaluation of rituximab plus cyclophosphamide, vincristine and prednisolone for advanced follicular lymphoma. *Leukemia & Lymphoma* 2008; 49(2):227-236.

58. Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: A cross-sectional utility study. *Health and Quality of Life Outcomes* 2010; 8(50).

59. Beusterien KM, Szabo SM, Kotapati S, Mukherjee J, Hoos A, Hersey P et al. Societal preference values for advanced melanoma health states in the United Kingdom and Australia. *British Journal of Cancer* 2009; 101:387-389.

60. Department of Health. PbR tariff 2012-2013. 2013. London, DH.

61. Buchanan J, Wordsworth S, Ahmad T, Perrin A, Vermeire S, Sans M et al. Managing the long term care of inflammatory bowel disease patients: The cost to European health care providers. *Journal of Crohn's and Colitis* 2011; 5:301-316.

62. Woehl A, Hawthorne AB, Morgan CL, Punekar Y, McEwan P. The epidemiology and healthcare resource use in patients with ulcerative colitis: A population based UK study. *Value in Health* 2007; 10(A335).

63. Department of Health. NHS Reference Costs 2011-12. 2014. London, DH.

64. Arseneau KO, Sultan S, Provenzale DT, Onken J, Bickston SJ, Foley E et al. Do patient preference influence decisions on treatment for patients with steroid-refractory ulcerative colitis? *Clinical Gastroenterology and Hepatology* 2006; 4:1135-1142.

65. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford University Press, Oxford; 2005.

66. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C et al. Principles of good practice for decision analytic modeling in health care evaluation: Report of the ISPOR Task Force on good research practices--modeling studies. *Value in Health* 2003; 6(1):9-17.

67. Guide to the methods of technology appraisal. 2013. London, NICE.

68. European Medicines Agency. Summary of product characteristics - Infliximab. 2014; 1-56.

69. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. 1st ed. Oxford University Press, Oxford; 2006.

70. Raval MJ, Schnitzler M, O'Conner BI, Cohen Z, McLeod RS. Improved outcome due to increased experience and individualized management of leaks after ileal pouch-anal anastomosis. *Annals of Surgery* 2007; 246(5):763-770.

71. Arai K, Koganei K, Kimura H, Akatani M, Kitoh F, Sugita A et al. Incidence and outcome of complications following restorative proctocolectomy. *The American Journal of Surgery* 2005; 190:39-42.

72. Swinburn P, Elwick H, Bean K, Curry A, Patel S, Bodger K et al. The impact of surgery on health related quality of life in ulcerative colitis. *Gut* 2012; 61:A237.